Estimating many properties of quantum systems via few repetitions

THESIS

submitted in partial fulfillment of the
requirements for the degree of

BACHELOR OF SCIENCE
in
PHYSICS

Author :	Thomas Rothe
Student ID :	1930443
Supervisor :	Dr. Jordi Tura Brugués
2^{nd} corrector :	Dr. Wolfgang Löffler

Leiden, The Netherlands, July 8, 2021

Estimating many properties of quantum systems via few repetitions

Thomas Rothe

Huygens-Kamerlingh Onnes Laboratory, Leiden University
P.O. Box 9500, 2300 RA Leiden, The Netherlands

July 8, 2021

Abstract

Characterizing quantum states is a central, yet involved, task in quantum information processing. In experiments, the unknown quantum state of interest must be prepared and measured multiple times to learn its properties. Unfortunately, a full tomographic description is prohibitive by the exponential scaling of the quantum state description with the system size. In practice, only a few quantities are of interest for which protocols involving informationally incomplete measurements are preferable. After studying existing data acquisition protocols, we discuss classical shadow estimation, a particular experimentally feasible method for estimating many system properties. We extend the applicability to quantum many-body systems with higher dimensional subspaces and derive similar performance guarantees to the qubit case. Ultimately we implement the generalized protocol in a modular and economic numerical framework and demonstrate the accuracy along with the favourable scaling of classical shadow estimation in unbiased numerical experiments. In particular, we suggest and discuss the near-term application to 4-photon OAM entangled systems.

Contents

Chapter 1

Quantum state tomography: an introduction

Classical computers and networks started increasingly to transform the way we live since the beginning of the information age in the mid 20th century. Following up the technology of the classical information age, quantum technology is starting to become available. Although some milestones - like quantum supremacy in 2019 [1] - have been achieved, large-scale quantum devices are still far from accessible. To build up towards such systems, first the fundamental noisy behaviour of existing intermediate-scale quantum (NISQ) devices has to be overcome. For large-scale devices we then can apply quantum-error-correction which is not possible for such NISQ devices. [2]

The task of identifying, characterizing and reconstructing quantum states has become essential for bringing these developments forward. This process we call: Quantum State Tomography (QST). [3] Predicting which values qubits (quantum bits) take at the output of a quantum computer, is just one simple example of possible applications. The processing of qubits requires new kind of circuits with well-defined quantum gates that can manipulate qubits reliably. To check the way whole quantum circuits or individual gates work on qubits, we can send in known quantum states and measure the states that leave at the output. By comparison with the desired transformation, we can optimize our circuit designs. We thus also need QST for this so-called standard quantum process tomography.[4]. For calibrating quantum measurement devices, we can even characterize the unknown applied quantum measurements completely from the measurement outcomes.

These also relate to one of the currently most promising applications of NISQ devices: Variational Quantum Eigensolvers. Such an Eigensolver could be useful for studying and designing complex molecules in fields like material science, chemistry and pharmaceutical research. [2] [5] [6]. The role of quantum state tomography is to identify eigenstates of atoms and molecules and subsequently their corresponding energy spectrum. Modelling and engineering of bandstructures and other microscopic molecular properties, for example, can then be improved in both the efficiency in application and synthetisation of such molecules.

Beside the broad relevance for quantum computers and quantum information in general, there are also numerous more fundamental applications. The basic task of determining the output of an (unknown) source of quantum particles, for example. Similarly, in quantum optics research QST is widely used in the characterization of optical signals. [7]

There is, however, a fundamental problem associated to the task of quantum state tomography. Through the exponential scaling of possible quantum state configurations with the number of system constituents and the restriction that no direct copy of an unknown quantum system can be made [8], the amount of resources required to achieve the task scales exponentially as well if the system size is increased. In the context of quantum computers, this means that by an increase of the number of fundamental information storage and processing units, qubits, the number of required identical copies of the qubit register must increase exponentially to completely characterize it using quantum state tomography. Practically, to create each copy without cloning the register, this implies complete recalculation of the qubit register that is to be read out. With exponentially many required copies, this would take an exponential amount of time which interferes with the idea of an exponential speed-up of quantum computers over classical devices. On the other hand, calculating all copies in parallel would need an exponential amount of material resources.

The aim of this thesis is to illustrate the power of choosing, scheduling, executing and processing measurements wisely to characterize a quantum system to a maximum with minimal resources and effort. Doing this requires us to drop the requirement of describing the quantum state of the system fully and focus on the characterization based on its most relevant properties. This task has been named *shadow tomography*. [9] This offsets itself from existing solutions for this task by adapting existing protocols to specific systems to gain maximum efficiency but also flexibility. In par-

ticular the system of spatially entangled photon pairs is considered that is mostly promising in its application to information processing applications like quantum networks. [2] [10] [11]

1.1 Fundamental concepts

Before we can introduce a more exact description of the problem associated with quantum state tomography, we first need to introduce some basic concepts from quantum mechanics and mathematical preliminaries. Both the conventional notation and definitions will be given and can be found in various textbooks like [8] and lecture notes on Quantum Computation and Quantum Information Theory, e.g. [12], [13].

1.1.1 State-vectors

To understand how to measure and interpret quantum states, we need to present the notation and formalism to represent a state of any quantum system in a unique way.
A quantum state is typically represented with a so-called *state-vector* in *state-space*. It is used to determine probability distributions for present and future behaviour of a quantum system. The complex vector space of state vectors is a Hilbert space denoted $\mathcal{H}$.

Definition 1.1.1 (State-vector)**.** *For an arbitrary Hilbert space $\mathcal{H}$, any $|\psi\rangle \in \mathcal{H}$ such that $|||\psi\rangle|| = 1$ is called a* ***state-vector****. The subset of vectors satisfying this condition is denoted $\mathcal{S}(\mathcal{H})$*

Here the norm, whenever not stated explicitly, refers to the standard 2-norm. Similarly it is useful to introduce a norm $||\cdot||_\infty := \mathrm{Tr}\,(\cdot)^2$ on the operator space $\mathcal{L}(\mathcal{H})$ of state-space $\mathcal{H}$ called *Hilbert-Schmidt norm* (or simply *operator norm*). It arises specifically from the notion of an inner-product on $\mathcal{L}(\mathcal{H})$, the *Hilbert-Schmidt innerproduct*: Given $P, Q \in \mathcal{L}(\mathcal{H})$ and their corresponding *ket-vector representations* $|P\rangle, |Q\rangle$ we define $\langle P, Q\rangle = \langle P||Q\rangle := \mathrm{Tr}\, P^\dagger Q$. The ket-vector representation is defined through an isomorphism $\mathcal{L}(\mathcal{H}) \to \mathcal{H} \otimes \mathcal{H} : P \mapsto |P\rangle$. More specifically this is the map $|i\rangle\langle j| \mapsto |i\rangle|j\rangle$ for an orthonormal basis $\{|i\rangle\}_{i\in\{1,dim(\mathcal{H})\}}$ of $\mathcal{H}$, which defines the operations complete. It is a convenient way to write operators in a similar way as quantum states, especially if more operators are involved.

The vector notation that we used is bra-ket notation in which a *ket* $|\psi\rangle \in \mathcal{H}$ is represented as a column vector in state-space. A *bra* $\langle\psi| \in \mathcal{H}$ is the

complex conjugate, which is an element of the dual space of the original state-space. Short notations like the inner-product $\langle|\psi\rangle, |\phi\rangle\rangle = \langle\psi|\phi\rangle$ and the outer-product $|\psi\rangle\langle\phi|$ will be used throughout.
With regard to quantum computation and information one often works more specifically in states of objects that can only be in two distinguishable (i.e. orthogonal) states. Like the bit in classical computation, represented by digits 0 and 1, we label for quantum computation the two states with the orthonormal state-vectors $|0\rangle$ and $|1\rangle$ respectively. However, in a quantum system the state can take any superposition of these two as well. Such a 2-state quantum-system is called a *qubit*. The Hilbert space of such a simple system is then taken as $\mathcal{H} = \mathbb{C}^2$ and the orthonormal set of state vectors $\{|0\rangle, |1\rangle\}$ forms a basis for $\mathcal{H}$ called the *computational basis*. Any quantum superposition can then be written: $|\psi\rangle = a|0\rangle + b|1\rangle$. Note that for being a state-vector, the condition $|a|^2 + |b|^2 = 1$ must always be satisfied. This is equivalent to the statement for conservation of probability.
In this context, we also introduce another regularly used basis for qubits called the *Hadamard basis* (or Pauli-X-basis) which is the set

$$\{|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\}$$

While we by convention (from the theory of spinors) label the computational basis states $|0\rangle$, $|1\rangle$ as the two levels along the Z-direction in a three dimensional system, we often choose the Hadamard basis as natural basis states along the X-direction and yet another (Pauli-Y) basis along the Y-direction. This encodes the incompatibility of the distinct directions in determining the two-level state: From the perspective of Pauli-X-basis states $|+\rangle$,$|-\rangle$, we see that the norm of projections on $|0\rangle$ and $|1\rangle$ are equal. i.e. from a deterministic state in the Pauli-X-basis, the superposition has no preferred state in the Z direction. The same holds for the Pauli-Y-basis and vice versa.
These collections of bases are called *mutually unbiased bases* (MUBs), that is a set of orthogonal bases with constant mutual inner product (up to a phase). Let p be a prime and $0 < \alpha \in \mathbb{Z}^+$. For $d = p^\alpha$ dimensional state-spaces, $d + 1$ MUBs do then exist (up to a unitary transformation of each basis).[14] The set of computational basis, Pauli-X basis and Pauli-Y-basis is therefore complete for single qubits ($d = 2$). For arbitrary d the number of MUBs is still an open question. [15]

A very convenient way to illustrate state vectors for single qubits is the *Bloch sphere*. This unit sphere in the complex state-space has poles labelled

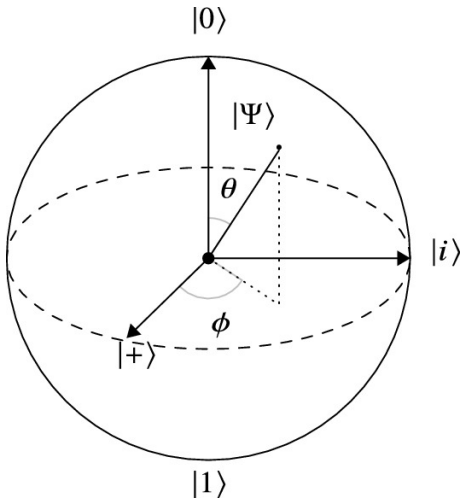

Figure 1.1: *[16] Bloch sphere for a single qubit system. On the poles are, by convention, the two computational basis vectors, while on the equator the Hadamard and Pauli-Y basis are denoted.*

by the two computational basis states and along the equator the two remaining MUBs are distributed. (see figure 1.1) This picture allows to associate a single point on the sphere to a unique superposition. For higher state-space dimensions this intuitive picture vanishes since the shape does no longer correspond to an intuitive geometric shape.

Finally, within the state-vector formalism, the equivalence of states with different phase must be mentioned. That is, a state $|\psi\rangle$ will be called *equivalent* with $e^{i\theta}|\psi\rangle$ for any real phase angle θ. This can be seen by noting that physical measurements only yield real values and expectation values involve the inner product with an Hermitian (measurement) operator $\hat{A}$, therefore: $\langle(|\psi\rangle e^{i\theta})|\hat{A}|(|\psi\rangle e^{i\theta})\rangle = e^{i\theta}e^{-i\theta}(\langle\psi|\hat{A}|\psi\rangle) = (\langle\psi|\hat{A}|\psi\rangle)$. This is exactly the expectation of $|\psi\rangle$ and thus no experimental distinction can be made between the two states. This idea is equivalent to representing the states as a point on the Bloch sphere and rotating all states by a certain phase θ (i.e. rotating the whole sphere). The rotation does not change the sphere, and the distribution of points on it, at all. Therefore the physical picture of the system is invariant under the global rotation i.e. the physical state is the same.

For multiple distinguishable qubits or quantum subsystems one can easily extend the concept of a state vector by taking the tensor product of the state vectors of each qubit or subsystem: $|\psi\rangle_{AB} = |\psi\rangle_A \otimes |\psi\rangle_B$. Vice versa,

however, not every state of a *multipartite system* can be written as a tensor product. Entangled qubits are an important example. Descriptions for individual subsystems can still be obtained but, as the term entanglement suggests, those separate descriptions do not give us very much information as long as they are not considered together. Just like a cookbook with ingredients and directions printed on different pages. Considering a single page does not make much sense.

1.1.2 Density Operator

The state-vector formalism becomes, however, inconvenient when describing *mixed states*. That are states that incorporate also classical information, or, more precisely, classical uncertainty. There is then no fundamental, unavoidable uncertainty (like for superpositions), but the uncertainty is rather an artefact of incomplete knowledge of the observer about which quantum state a system is in. Even stronger, if two or more subsystems are entangled we have no clear way anymore to write it in terms of state vectors and the formalism becomes obsolete.

For making the notation consistent for systems with both classical and quantum information, for removing confusion of equivalent states and for handling composite and entangled quantum systems easily, we can alternatively describe a state with density operators:

Definition 1.1.2 (Density operator/matrix)**.** *A density operator/matrix is an operator $\rho \in \mathcal{L}(\mathcal{H})$ on an arbitrary Hilbert space $\mathcal{H}$ if it is semi-positive-definite ($\rho \geq 0$) and $\operatorname{Tr}\rho = 1$. The set of Density operators is denoted $\mathcal{D}(\mathcal{H})$.*

Also density matrices can be extended to multiple systems or qubits by taking the tensor product of the separate density matrices, i.e. yielding density matrices $\rho_{\mathcal{H}\mathcal{H}'} \in \mathcal{D}(\mathcal{H} \bigotimes \mathcal{H}')$. The subscript labels the involved subsystems.

Equally to compositing subsystems to a larger system, systems can be *traced out* to yield the description of the state in a smaller subsystem. For this we introduce the *partial trace*:

Definition 1.1.3 (Partial Trace)**.** *Consider Hilbert spaces $\mathcal{H}, \mathcal{H}'$. The partial trace , denoted $\operatorname{Tr}_{\mathcal{H}}(\cdot)$, with respect to subsystem $\mathcal{H}$ of a density matrix $\rho_{\mathcal{H}\mathcal{H}'} \in \mathcal{D}(\mathcal{H} \bigotimes \mathcal{H}')$ is the superoperator*

$$\operatorname{Tr}_{\mathcal{H}}(\cdot) \coloneqq \operatorname{Tr}(\cdot) \bigotimes \mathbb{I}_{\mathcal{H}'} : \mathcal{L}(\mathcal{H}) \bigotimes \mathcal{L}(\mathcal{H}') \mapsto \mathcal{L}(\mathcal{H}')$$

With $\mathbb{I}_{\mathcal{H}'} : \mathcal{L}(\mathcal{H}') \mapsto \mathcal{L}(\mathcal{H}')$ *the identity operator on the operator space of* $\mathcal{H}'$.

So the *reduced density operator* of subsystem $\mathcal{H}'$ is $\rho_{\mathcal{H}'} = \mathrm{Tr}_{\mathcal{H}} (\rho_{\mathcal{H}\mathcal{H}'})$.

If the rank $rank(\rho) = 1$ then the density operator is called *pure*. Any state-vector $|\psi\rangle \in \mathcal{S}(\mathcal{H})$ can be mapped to such a pure density operator $|\psi\rangle\langle\psi| \in \mathcal{D}(\mathcal{H})$. This formalism directly accounts for the global phase invariance, as mentioned for state vectors.

Describing now any quantum state, including classical information in form of probabilities P_i, for a pure state $|\psi_i\rangle\langle\psi_i|$ is as easy as taking the expectation values of the density operators $\sum_i P_i|\psi_i\rangle\langle\psi_i| \in \mathcal{D}(\mathcal{H})$. For the set $\{|\psi_i\rangle\}_{i\in\{1,\ldots,dim(\mathcal{H})\}}$ forming an orthonormal basis of $\mathcal{H}$, this is just the usual spectral decomposition of an density operator of a (*mixed*) state. So any state that is not necessarily pure. From the superposition we can also extract the trace condition for density operators of single qubits:

$$\mathrm{Tr}\,(\rho) = \mathrm{Tr} \sum_{i=1}^{dim(\mathcal{H})} P_i|\psi_i\rangle\langle\psi_i|) = \sum_{i=1}^{dim(\mathcal{H})} P_i \,\mathrm{Tr}\,|\psi_i\rangle\langle\psi_i|) = \sum_{i=1}^{dim(\mathcal{H})} P_i\langle\psi_i|\psi_i\rangle = \sum_{i=1}^{dim(\mathcal{H})} P_i := 1$$

. The last equality follows from $\{P_i\}_i$ being a complete set of probabilities.

In the picture of the Bloch sphere, a mixed state can be represented by a unique point *inside* the sphere. In fact, the center of the sphere corresponds to the maximally mixed state in which indeed the state is completely random. Pure states reside on the surface sphere and their convex hull generate the mixed states. This generalizes naturally to the more complex shapes for higher-dimensional state-spaces.[17] [18]

Finally, unifying the descriptions of pure states as state vectors and mixed states as density matrices, we remark that any mixed state $\rho \in \mathcal{D}(\mathcal{H})$ can be obtained from a pure state in a composite system $|\psi\rangle\langle\psi| \in \mathcal{D}(\mathcal{H} \bigotimes \mathcal{H}')$ by tracing out all subsystems but $\mathcal{H}$: $\rho_{\mathcal{H}} = \mathrm{Tr}_{\mathcal{H}'} (|\psi\rangle\langle\psi|)$. Such a pure state, known as a *purification* of the mixed state, is only unique up to unitary operation on the ancillary system $\mathcal{H}'$ to $\mathcal{H}$. Therefore many purifications can be found.

1.1.3 Measurements

Measurements are central as a tool to determine certain properties of a quantum system. It is one of the essential aims in this thesis to find out

which measurements to make and how to carry out measurements such that the quantum state (or its properties) can be recovered efficiently.
For the sake of completeness, first the definition for a general measurement will be given:

Definition 1.1.4 (Measurement)**.** *A measurement is specified by a set of operators (called measurement operators) $\{M_i\}$ such that $\sum_i M_i^\dagger M_i = \mathbb{I}$ (the state must be completely within these possible outcomes i) with $i \in I$ and I a non-empty finite set of possible measurement outcomes.*
The following results can be extracted from the measurement operators and a quantum state described by density operator ρ:

- *The probability for measuring outcome i is: $P_i = \operatorname{Tr} M_i^\dagger M_i \rho$. Where $P_i \geq 0$ is a valid probability since $M_i^\dagger M_i \geq 0$*
- *If an outcome i is obtained from a measurement with $P_i \neq 0$, the quantum state collapses to the post-measurement state: $\frac{1}{\sqrt{P_i}}(M_i \rho M_i^\dagger)$*

Note that when we say "operator" in the context of measurements on density operators, we actually refer to superoperators. So elements of $\mathcal{L}(\mathcal{L}(\mathcal{H}))$. However, this measurement framework can also be applied one-to-one in the state-vector formalism and therefore the form of the operators is the same. We will use these terms here interchangeably.

That a state collapses to a (different) post-measurement state with a given probability when a measurement is applied, is the basis of the problem in state tomography. The probabilistic nature of quantum mechanics requires multiple measurements to fully estimate a quantum state, but each measurement must be applied to an own copy of the state since, generally, the measurement changes the state.
The most relevant kind of *general* measurements is that of POVMs:

Definition 1.1.5 (POVM)**.** *A positive operator-valued measure (POVM) is a measurement with a finite set of operators E_i that statisfy $\sum_i E_i = \mathbb{I}$ and $E_i \geq 0$ for all i out of a non-empty finite set. (In definition 1.1.4 identify $M_i = \sqrt{E_i}$, for example).*
The probability of measurement outcome i is then $P_i = \operatorname{Tr} E_i \rho$

POVMs are of interest because they are able to describe most practically realistic measurements. However, one disadvantage of POVMs in some situations might be that the post-measurement state can not be determined within this measurement framework, as the operators E_i define

no unique operators M_i $(= \sqrt{E_i})$ in definition 1.1.4. Applying any unitary U_i to the M_i's yield different measurement operators with the same underlying POVM E_i: e.g. for measurement $\mathfrak{M}_i = U_i\sqrt{E_i}$ we see that its POVM consist of $\mathfrak{M}_i^\dagger\mathfrak{M}_i = \sqrt{E_i}U_i^\dagger U_i\sqrt{E_i} = \sqrt{E_i}I\sqrt{E_i} = E_i$. Therefore the corresponding measurement operators to a POVM are fixed by the exact way of implementation of the measurements.

Often it is sufficient and easier to use even more restricted conditions on the operators. Also for this thesis, we will mainly be concerned with *projective measurements*:

Definition 1.1.6 (Projective Measurements)**.** *Projective/von Neumann measurements are Hermitian operators $M \in \mathcal{L}(\mathcal{H})$ —in this context also called observables —with a spectral decomposition into orthogonal projectors Π_m onto the eigenspaces of M: $M = \sum_m m\Pi_m$, where the measurement outcomes m (i.e. eigenvalues of M) form a non-empty finite set. In fact, the set is restricted in size (as opposed to POVMs) to the dimension of the Hilbert space $\mathcal{H}$. Orthogonal projectors need to satisfy $\Pi_m^2 = \Pi_m$, $\sum_m \Pi_m = I$ and are Hermitian as well. This implies that all Π_m only have eigenvalue 0 or 1, as opposed to the most general operators in a POVM, whose eigenvalues must lie within the interval $[0, 1]$.*
With a projective measurement represented by an observable M of a system in a quantum state described by ρ, the following results can be obtained:

- *The probability for measurement outcome m: $P_m = \operatorname{Tr}\Pi_m\rho$*
  Notice the consistency of this definition since indeed $\sum_m P_m = \operatorname{Tr}(\sum_m \Pi_m)\rho = \operatorname{Tr}\rho = 1$
- *The post-measurement state: $\dfrac{1}{\sqrt{P_m}}(\Pi_m\rho\Pi_m^\dagger)$*
- *The expectation value for observable M: $\langle M\rangle = \langle M, \rho\rangle = \operatorname{Tr} M\rho = \sum_m m \operatorname{Tr}\Pi_m\rho$. Where the inner-product is the Hilbert-Schmidt inner-product.*

Note that this defines a conventional but specific kind formalism for projective measurements, since more generally the orthogonal projections do not need to be on the eigenspaces of an observable. Rather a mutually orthogonal set of orthogonal projectors $\{M_i\}_i$ on non-overlapping subspaces can define a measurement on itself and involve more abstract labels of possible measurement outcomes i. In most physical applications, however, the concept of an observable is used because of its straightforward physical interpretation. In this thesis we will refer with *"measurements"* to such projective measurements.

Projective measurements compare to POVMs like pure states to mixed states: When considering systems that consist of multiple parts (e.g. $\mathcal{A}$, $\mathcal{B}$), for any subsystem with any mixed state we can always find a state that is pure in the bigger multipartite system. One reobtains the mixed state in the single subsystem by the operation of 'tracing out' —i.e. ignoring —a subsystem in this pure state. Equivalently, from the perspective of a single subsystem, applying a POVM to the subsystem is the same as applying a projective measurement on the bigger system.
In contrast to POVMs, for projective measurements also the state after measurement and statistical quantities are very easily calculated from the decomposition of an observable. The average or expectation of observable M ,$\langle M \rangle$, for example.

1.1.4 Operators

Having seen now states and measurements being represented by operators and superoperators respectively, it is still unclear which concrete operators can be considered in general and what might be a (conventional) basis for the operator space.

We start by briefly mentioning a convenient and very generic, albeit incomplete, notion of operations on quantum systems, *quantum channels*:

Definition 1.1.7 (Quantum Channel)**.** *Let $\mathcal{H}, \mathcal{H}'$ be two Hilbert spaces. A Quantum Channel is defined to be a completely positive and trace preserving (CPTP) map $\mathscr{T} : \mathcal{L}(\mathcal{H}) \mapsto \mathcal{L}(\mathcal{H}')$.*
That is for any operator $R \geq 0$ it satisfies $\operatorname{Tr} \mathscr{T}(R) = \operatorname{Tr} R$ and $\mathscr{T}(\tilde{R}) \geq 0$ with $\tilde{R} = \mathcal{H} \bigotimes \tilde{\mathcal{H}} \geq 0$, $\tilde{R} \in \mathcal{L}(\mathcal{H} \bigotimes \tilde{\mathcal{H}})$ for $\mathbb{I}_{\tilde{\mathcal{H}}}$ the identity operator of any separate Hilbert space $\tilde{\mathcal{H}}$.

This definition ensures density operator $\rho \in \mathcal{D}(\mathcal{H})$ is mapped to another valid density operator $\rho' \in \mathcal{D}(\mathcal{H}')$. The partial trace introduced in section 1.1.2 above is an example for a CPTP map.

Although quantum channels are very useful in describing many operations in quantum information processing, it is often more convenient to consider time evolution of quantum systems by unitary operators U instead. A Hamiltonian H, for example, can be applied easily by the unitary operator $e^{-iHt/\hbar}$ to a quantum state. In analogy to purifications for quantum states, quantum channels are equivalent to unitary operations

on larger composite systems. That is, any quantum channel $\mathscr{T} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}')$ can be written as $\mathscr{T}(\cdot) = \mathrm{Tr}_{\mathcal{H}\tilde{\mathcal{H}}}\left(U\left((\cdot) \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0|\right)\right)$ with unitary U acting on $\mathcal{H} \otimes \tilde{\mathcal{H}} \otimes \mathcal{H}'$. So describing the evolution of a quantum system can be done using only unitary operations. In practice, however, both quantum channels and unitaries are convenient in different situations. [19] [8]

From a more intuitive perspective, unitaries will be mostly sufficient for our current discussion. To get more concrete on the purpose and kind of unitaries, we introduce the most conventional choice for a generating set of unitary operations:

Definition 1.1.8 (Pauli Matrices)**.** *The Pauli matrices/operators $\sigma_x, \sigma_y, \sigma_z$ are defined as:*
$\sigma_x := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$
Therefore the Pauli matrices are Hermitian and satisfy: $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = \mathbb{I}$, $\mathrm{Tr}\,\sigma_x = \mathrm{Tr}\,\sigma_y = \mathrm{Tr}\,\sigma_z = 0$

To keep notation clean we will equivalently write simply X, Y, Z.
Now it is very easy to form a basis for the real subspace of Hermitian operators in $\mathcal{L}(\mathcal{H})$ since a 2×2 Hermitian matrix takes the general form:
$\begin{pmatrix} a & c - d \cdot i \\ c + d \cdot i & b \end{pmatrix}$ for explicitly real parameters a,b,c and d.
Therefore the set $\{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}$ ($\mathbb{I}$ being the 2×2 identity) is complete and:
$\alpha \cdot \mathbb{I} + \beta \cdot \sigma_x + \gamma \cdot \sigma_y + \delta \cdot \sigma_z = \begin{pmatrix} \alpha + \delta & \beta - \gamma \cdot i \\ \beta + \gamma \cdot i & \alpha - \delta \end{pmatrix} = \mathbf{0}$
has clearly only the solution $\alpha = 0$, $\beta = 0$, $\gamma = 0$, $\delta = 0$. So the matrices are linearly independent and complete with respect to the Hermitian subspace of $\mathcal{L}(\mathcal{H})$. Therefore they form a basis of Hermitian operators over $\mathbb{R}$. Equally they form also a basis over all matrices over $\mathbb{C}$.

This can also be used to write density operators in the basis of Pauli matrices:

$$\rho = \frac{1}{2}(\mathbb{I} + \alpha\sigma_x + \beta\sigma_y + \gamma\sigma_z) \tag{1.1}$$

This is called *Bloch sphere representation* of the density operator. Note that the identity (which solely contributes to the trace) had to drop a coefficient to fix $\mathrm{Tr}\,\rho = 1$. This restriction removes thus one degree of freedom, consistently with the number of unknown density matrix elements. However, the expression is even more intuitive since the identity lets the state start in

the maximally mixed state at the center of the Bloch sphere and the Bloch vector $\overrightarrow{n} = \{\alpha, \beta, \gamma\}$ defines the deviation from the center to reach the point in or on the Bloch sphere that corresponds to the state.
It might sometimes be convenient to express the density matrix in a few operators of *any* basis, which is possible with any linearly independent set of operators that span the space of Hermitian operators. The simple form of the above representation, however, holds only with orthogonality.

Also, more importantly, basic measurement operators can be constructed with only the set $\{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}$, since these four matrices form a basis of Hermitian 2×2 matrices over $\mathbb{R}$. Taking the definition of the tensor product into account, this generalizes easily to situations with n qubits, forming also a basis of the Hermitian subspace of $\mathcal{L}(\mathcal{L}(\mathcal{H}))^{\otimes n}$. The choices of certain bases for all single-qubit measurements are called *settings* of the measurement. Considering 3 spatial dimensions, the number of settings are 3^n for n qubits if the operator is *global*. That is, if it is acting non-trivially on all qubits in the system, while a *local* observable only acts on a part of the system.

The compilation of operators from Pauli observables is a very important feature as resources for applications like quantum devices are limited. Standardization of operations that can be performed during measurements might make the process sometimes less efficient but practically intuitive and feasible. There exist also more extensive measurements that, for example, use entanglement between qubits. Often the implementation for these has been restricted a lot by decoherence, destroying the entanglement state over time. [20] [21]

Finally it will become useful to exploit that Pauli operators form also a group under inclusion of appropriate phases:

Definition 1.1.9 (Pauli Group)**.** *The n-qubit Pauli group $\mathcal{P}^n$ is generated by tensor products of Pauli operators, so*

$$\mathcal{P}^n = \{p \bigotimes_{k=1}^{n} \sigma_{j_k} | p \in \{\pm 1, \pm i\}, j_k \in \{I, x, y, z\}\} \tag{1.2}$$

under matrix multiplication.

Another closely related group of interest is the so-called (unitary) n-qubit Clifford group $\mathscr{C}_d^n -$ of which the n-qubit Pauli group is, by defini-

tion, a normal subgroup. In particular $\mathscr{C}^n$ normalizing $\mathcal{P}^n$ means

$$CPC^{-1} \in \mathcal{P}^n; \forall C \in \mathscr{C}^n, P \in \mathcal{P}^n \tag{1.3}$$

This will be a key identity when working with unitary transformation of Pauli operators later. Note further that for $n = 1$ the single qubit Clifford group $\mathscr{C}^1$ is isomorphic to the single qubit Pauli group $\mathcal{P}^1$. [22] [23] Therefore we will consider the group of n-qubit tensor products of $\mathscr{C}^1$ to be equivalent to the Pauli group $\mathcal{P}^n$.

We remark that some authors use an alternative definition of the Clifford group as the *Clifford quotient group* $\mathcal{C}^n := \mathscr{C}^n/\mathcal{U}(1)$ with the unitary group of complex scalars $\mathcal{U}(1) = \{e^{i\theta}\mathbb{I}_d^n; \theta \in \mathbb{R}\}$ [24]. Thus it is the group $\mathscr{C}^n$ under exclusion of phases, generated by equivalence classes $[X] = \{e^{i\theta}X\}_{\theta\in\mathbb{R}}$. [25] The exact difference will not be important for the discussions in this thesis and we will use the term Clifford group interchangeably in context.
The Clifford group is of main interest to efficiently describe a finite and discrete set of interesting operations on quantum systems. [26].

Quantum Circuits

The Clifford group $\mathcal{C}^n$ is generated by a few basic operations: The Hadamard gate, Phase gate and the CNOT gate. We write their matrix representation in the computational basis as:

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} \quad P = \begin{pmatrix}1 & 0\\ 0 & i\end{pmatrix} \quad CNOT = \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\end{pmatrix} \tag{1.4}$$

A *quantum gate* is just a unitary transformation in a quantum system but is used mainly to name components in a *quantum circuit*. That is a detailed instruction set constructed from a finite number of elementary gates. Quantum gates are therefore completely equivalent to reversible logic (e.g. AND, XOR) gates in classical electronic circuits. Figure 1.2 gives the relevant example of generating the maximally entangled 3-qubit Greenberger-Horne-Zeilinger (GHZ) state from the vacuum ground state $|000\rangle$ in a circuit diagram.

Since the Clifford group does not span the whole space of unitaries, there is a large variety of other gates and circuits. Although required explicitly for any circuit that is unique to quantum devices they will not be of direct reference or use in this thesis.

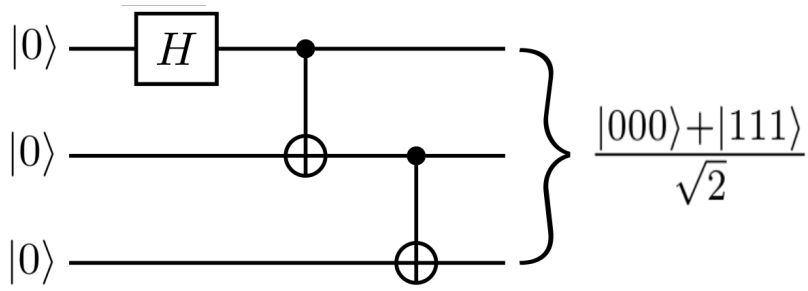

Figure 1.2: *Quantum circuit for transforming the vacuum state for 3-qubits $|0\rangle^{\otimes 3}$ into a maximally entangled Greenberger-Horne-Zeilinger (GHZ) state. A Hadamard gate is applied to the first qubit, followed by two CNOT gates with qubit 1 and 2 as control qubits and qubit 2 and 3 as target qubits respectively.*

Generalizations to multilevel systems

In our numerical experiments we consider the application to multilevel, i.e. qudit instead of qubit, systems. Therefore we will briefly discuss common extensions on the definitions of the Pauli and Clifford group above to subsystem dimensions of $d > 2$.

Although multiple generalizations of the Pauli group exist, the most natural choice turns out to be the finite and discrete Weyl-Heisenberg group $\mathcal{W}^d$ of d^2 elements that is generally generated by X_d and Z_d under matrix multiplication and inclusion of a phase $\sqrt{\omega} = e^{i\pi/d}$. [27] Here defining a dth primitive root of unity $\omega = e^{i2\pi/d}$. These generalized Pauli operators X_d, Z_d are also known as clock and shift operators of order d such that $X_d^d = \mathbb{I} = Z_d^d$ and are defined as the unitary operators satisfying $X_d|j\rangle = |j+1 \bmod d\rangle$ and $Z_d|j\rangle = \omega^j|j\rangle$ respectively for computational basis state $|j\rangle \in \{|0\rangle \dots |d-1\rangle\}$ [28] [25]

We write an element of the group as a Pauli operator P with a phase $\tilde{\omega}^\alpha$ as:

$$\tilde{\omega}^\alpha \cdot P(\overrightarrow{a}) = \tilde{\omega}^\alpha \cdot X^{a_1} Z^{a_{n+1}} \bigotimes \dots \bigotimes X^{a_n} Z^{a_{2n}} \in \mathcal{W}_d$$

with $\overrightarrow{a} \in \mathbb{Z}_d^{2n}$ and: [28]

$$\tilde{\omega} = \begin{cases} \sqrt{\omega} & \text{if } d \text{ even} \\ \omega & \text{if } d \text{ odd} \end{cases}$$

Here the modulus is understood component-wise.
Action on a n-qudit computational basis state $|\overrightarrow{x}\rangle$ with bitstring $\overrightarrow{x} \in \mathbb{Z}_d^n$

can then be summarized as

$$P(\overrightarrow{a})|x\rangle = \omega^{\overrightarrow{a}_{n+1\ldots 2n}\cdot\overrightarrow{x}}|\overrightarrow{x}+\overrightarrow{a}_{1\ldots n} \bmod d\rangle$$

Also we identify the product rule: [28]

$$P(\overrightarrow{a})P(\overrightarrow{b}) = \tilde{\omega}^{2\left(\overrightarrow{a}_{n+1\ldots 2n}\cdot\overrightarrow{b}_{n+1\ldots 2n}\right)}P(\overrightarrow{a}+\overrightarrow{b} \bmod d)$$

Distinguishing phase factors with $\tilde{\omega}$ for odd or even d is required to preserve the order d for all elements in the complete Weyl-Heisenberg group in both cases.
If d is odd $P(\overrightarrow{a})^d = \sqrt{\omega}^{d(d-1)||\overrightarrow{a}_{n+1\ldots 2n}||^2}P(\overrightarrow{0}) = e^{i(d-1)\pi}\mathbb{I} = \mathbb{I}$ as required, but $e^{i(d-1)\pi} = -1$ if d is even. [28]
We conclude that for even d the number of roots of unity doubles and the order of elements in the group becomes $2d$, while it stays d in the odd case. This is reflected in $\tilde{\omega}$ introduced above. [27] [25]. To keep an unified formalism for all d, we also track phases modulo $2d$ for odd d because it does not change the results.

Note further that although elements of this group are still unitary, these generalized Pauli operators are not Hermitian anymore and therefore can be assigned unphysical complex eigenvalues. In context of applications this, however, is no problem. Also non-commuting Pauli operators do not need to anticommute anymore, instead the commutation relation

$$P(\overrightarrow{a})P(\overrightarrow{b}) = \omega^{\overrightarrow{a}\odot\overrightarrow{b}}P(\overrightarrow{b})P(\overrightarrow{a}) \tag{1.5}$$

holds with $\odot$ the symplectic inner product in $\mathcal{F} = \mathbb{Z}_d^{2n}$. Where members of $\mathbb{Z}_d^{2n}$ are $2n \times 2n$-arrays with elements on the finite field $\mathbb{Z}_d^{2n}$. [28] [25].
In its matrix representation this bilinear form can be written as

$$\overrightarrow{a}\odot\overrightarrow{b} = \overrightarrow{a}^T\Lambda\overrightarrow{b} \text{ with } \Lambda = \begin{pmatrix} 0 & -\mathbb{I}_n \\ \mathbb{I}_n & 0 \end{pmatrix} \in \mathbb{Z}_d^{2n\times 2n} \text{ and } \overrightarrow{a},\ \overrightarrow{b} \in \mathbb{Z}_d^{2n}$$

This implies that the symplectic inner-product defines the phase factor in the commutation relation between $P(\overrightarrow{a})$ and $P(\overrightarrow{b})$. If it vanishes, the Pauli operators commute. This also holds for qubits, but becomes even more meaningful since a symplectic inner-product of 1 in that case corresponds directly to anticommuting Pauli operators.

For the (generalized) Pauli group we will regularly also refer to the quotient $\mathcal{P}_n^d := \mathcal{W}_n^d / U(1)$ as the group consisting of (tensor products of) Pauli operators $P(\overrightarrow{a})$ without phases.

The generalized Clifford group can now be introduced in analogy to the qubit case as the normalizer to the Weyl-Heisenberg group i.e.

$$\mathscr{C}_d^n = \{C \in \mathcal{U}(n) | CPC^{-1} \in \mathcal{W}_d^n; \forall P \in \mathcal{W}_d^n\}$$

for $\mathcal{U}(n)$ the group of unitary matrices of size $n \times n$. [29] [25]. Also here the quotient $\mathcal{C}_d^n = \mathscr{C}_d^n / \mathcal{U}(1)$ is considered. Because of the complicated structure of the Clifford group in higher dimensions ($d > 2$) when including phases, we will mostly refer to this *generalized Clifford quotient group* $\mathcal{C}_d^n$ as the generalized Clifford group. [25] [24]

Generalized Clifford gates do exist and generate the n-qudit Clifford group $\mathcal{C}_d^n$. While phase P and CNOT gates are still general generators, we generalize the Hadamard gate by the quantum Fourier tranform F. Their actions on computational basis states $|i\rangle, |j\rangle$ with $i, j \in \mathbb{Z}_d$ are: [28] [30] [24]

$$F|j\rangle = \frac{1}{\sqrt{d}} \sum_{i \in \mathbb{Z}_d} \omega^{j \cdot i} |i\rangle \quad P|j\rangle = \tilde{\omega}^{j(j+d)} |j\rangle \quad CNOT|i\rangle|j\rangle = |i\rangle|(i+j) \ mod \ d\,\rangle$$

As usual this choice is far from unique (in literature), we follow [30] which showed that these operators form indeed the minimal set of operators which are sufficient to generate the n-qudit Clifford group. Also the choice of generators was made to ensure the convergence to the qubit case for $d = 2$, both for the generalized CLifford group and the generalized Pauli group.

Especially the identification of the quantum Fourier transform with the Hadamard gate is conceptually interestingly since it allows to directly see the connection between Pauli operators X_d and Z_d and direct phase space representations of quantum states. As in any physical system where we can use Fourier transformations to transform between position and momentum space, we know that applying a Hadamard gate (or Fourier transform) to the Pauli-X basis transforms it to the Pauli-Z basis. This allows to interpret the operators X_d and Z_d similarly as the momentum and position displacement operators in discrete phase space. Here $\hat{p}$ and $\hat{x}$ are the quantum mechanical position and momentum operators respectively. [31]. This is the underlying notion from which the finite Weyl-Heisenberg

group $\mathcal{W}_d^n$ originally arose in describing the quantum kinematics of quantum systems. Clifford operations are then those operations that keep the system kinematics invariant.

1.2 Quantum state tomography

Within the technical framework of the previous section, we can now finally state and interpret the general problem of quantum state tomography (QST).

We consider the simplest possible system of n independent, non-interacting spins with spin 1/2. The state of this system can be described by a state-vector or a density matrix in the computational basis of $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$. Each element of this basis represents a configuration in which each spin can be 1 (up) or 0 (down).
Each spin is therefore just a qubit in state $\alpha|0\rangle + \beta|1\rangle$ ($|\alpha|^2 + |\beta|^2 = 1$)
The general state-vector $|\psi\rangle = \sum_{b\in\{0,1\}^n} \alpha_b|b\rangle$, with coefficients $\alpha_b \in \mathbb{C}$ and $\sum_{b\in\{0,1\}^n} |\alpha_b|^2 = 1$, describes each individual configuration just by a string of 0ś and 1ś. While the state-vector contains only D elements to denote in which configuration or superposition of configurations the system is in, the density matrix ρ needs $D^2 - 1$ parameters for a unique description. Here D is the Hilbert space dimension $dim(\mathcal{H}) = 2^n$, which is the number of possible distinguishable configurations, and -1 appears because of the constraint $\mathrm{Tr}\,\rho = 1$ that makes one parameter redundant. For the benefits mentioned earlier, we will keep using the density operator formalism from here on. The problem of shadow tomography, for any similar system, can then be stated as follows:

> Let ρ be a density matrix representing an (unknown) quantum state of system with state-space $\mathcal{H}$. Given the observables O_i, which form a set of measurements $\{O_i\}_{i\in\{1,\dots,m\}}$, and their outcomes $b_i = \mathrm{Tr}\,O_i\rho$ (see definition 1.1.6) with $b_i \in \mathbb{R}$. How can the density matrix ρ of the system's quantum state be fully reconstructed and under which constraints on the set $\{O_i\}_{i\in\{1,\dots,m\}}$, measurement outcomes b_i and density matrix ρ?

A straightforward mathematical solution to this problem is the vectorization of the problem and using linear inversion to solve the resulting system of linear equations.

This can be done by identifying the *Hilbert-Schmidt inner-product* on $\mathcal{L}(\mathcal{H})$ as usual:

$$\langle A, B\rangle = \mathrm{Tr}\, A^\dagger B$$

for any $A, B \in \mathcal{L}(\mathcal{H})$. Note that observables are per definition Hermitian, therefore the expectation values are just $b_i = \mathrm{Tr}\, O_i\rho = \mathrm{Tr}\, O_i^\dagger\rho = \langle O_i, \rho\rangle$. For convenience we can transform this to an inner-product $\langle\overrightarrow{O_i}|\,\overrightarrow{\rho}\rangle$ on $\mathbb{C}^{D^2}$, where D^2 is again the dimension of parameter space for ρ. The $\overrightarrow{O_i}$ and $\overrightarrow{\rho}$ are vectorizations of the operators that are obtained by stacking the column-vectors of the matrix representations into a single long vector. It is straightforward to see that this indeed corresponds to the sum of the diagonal elements of the matrix product $O_i\rho$, i.e. the Hilbert-Schmidt inner-product of O_i and ρ. To compactly write this for all $i \in 1, \ldots, m$ we note that $\overrightarrow{\rho}$ is equal for all i and therefore can be written as $\mathcal{O}\overrightarrow{\rho} = \overrightarrow{b}$ with b the vector with components b_i and $\mathcal{O}$ the $m \times D^2$-matrix with row vectors $\overrightarrow{O_i}$. Inverting $\mathcal{O}$ solves the problem under the condition that $\mathcal{O}$ is invertible, thus for $m = D^2$ and for linearly independent observables O_i.

We can do even better by avoiding a matrix inversion if we restrict the observables further to be orthogonal. Then we use that for a complete set of observables, that is for the number of observables being equal to the dimension of the operator space $dim(\mathcal{L}(\mathcal{H})) = D^2$, we have$\sum_i \overrightarrow{O_i}\overrightarrow{O_i^\dagger} = \mathbb{I}$ So then $\overrightarrow{\rho} = \mathbb{I}\,\overrightarrow{\rho} = \sum_i \overrightarrow{O_i}\overrightarrow{O_i^\dagger}\,\overrightarrow{\rho} = \sum_i \overrightarrow{O_i}\langle\overrightarrow{O_i}|\,\overrightarrow{\rho}\rangle = \sum_i \overrightarrow{O_i}\cdot b_i$. This can be expressed in operator form as $\rho = \sum_i O_i \cdot b_i$.

In both cases we can, as mentioned before, reduce the dimension of the vectorized equations by one due to the restriction that $\rho = 1$, so that a complete set of measurements consists of only $D^2 - 1$ observables.
Also we can easily interpret these solution for any number of qubits that constitute an arbitrary quantum state as no assumptions were made on dimensionality. Still it is useful to see that the above also works in the case of single qubit measurements on an arbitrary quantum state of n qubits as such measurements are often used: Let then any effective observable O_i on the whole system be separable as $O_i = \bigotimes_k O_i^{(k)}$,with $O_i^{(k)}$ the single-qubit observable on qubit k, such that $\mathrm{Tr}\left[(\bigotimes_k O_i^{(k)})\rho\right]$.
We reattain then for an orthogonal and complete set of observables O_i: $\overrightarrow{\rho} = \mathbb{I}_4^{\otimes n}\rho = \sum_i(\bigotimes_k O_i^{(k)})((\bigotimes_k O_i^{(k)})^\dagger)\,\overrightarrow{\rho} = \sum_i(\bigotimes_k O_i^{(k)})\cdot b_i$ or $\rho = \sum_i(\bigotimes_k O_i^{(k)})\cdot b_i$. Where we used that a tensor product of identities is just

the identity in the higher dimension and vectorization works the same with the Kronecker product as tensor product.

Although the above solution works mathematically and forms an absolute lower bound to obtain a unique solution *for an arbitrary quantum state* of the system, it does not guarantee the result to be physical. That is, it might give a matrix that is not a density matrix i.e. not positive-semidefinite or resulting in invalid predictions for probability values. Since there is no easy expression for physical constraints like the semi-positiveness of a matrix, there thus needs to be a more advanced way of solving the problem. This should either let us impose the physical constraints implicitly or be based on a procedure that is physically argued, rather then being of purely mathematical kind. That is, more generally, elements of the density matrix ρ should not be regarded independent at all.
This is all beside the practical issues of implementing, operating and dealing with accuracy of the measurements represented by the observables.

Establishing a procedure that gives even physical results, however, does not imply practical feasibility. While the general lower bound of required linearly independent observables is $D^2 - 1$ to fully reconstruct the density matrix of a state, D itself does grow exponentially with the number of qubits n ($D = 2^n$). Therefore, in the most ideal case, the number of measurements needed for estimating a quantum state completely scales exponentially with the number of constituents in our quantum system. [32]. As for almost all measurements the wave function of the state collapses and, thus, the measured state changes. In consequence the state of the system (which we want to measure) must be prepared anew for every measurement.
One exception to this is when required measurement operators do commute. In this case the operators have a common set of eigenvectors and can therefore be measured simultaneously in the shared eigenbasis. [33] [34] [8] Nevertheless the problem persists for many applications in which the measurement operators certainly do not commute. Especially when measures of entanglement are involved. [8] [35] [36] This makes generalized and complete quantum state tomography infeasible for any larger number of qubits. This is the main issue that we are addressing in this thesis. The (scaling in the) number of required copies of a state for a state tomographic procedure is called the *sample complexity*.

Illustrating the problem more concretely: In a quantum computing device the state preparation (i.e. calculation) could potentially be exponentially

faster, but on the other hand one needs to repeat it exponentially more often for increasing number of qubits to read out the output of a quantum algorithm, for example. This restricts the usefulness of such quantum devices heavily.

It gets even worse if we consider that a measurement with an observable has also an underlying statistical distribution, which naturally scales as $O(1(/\epsilon^2)$ in accuracy ϵ,and that a system's state might not always be prepared equally well. This increases the sample complexity even further by a certain factor in order to determine the expectation values of observables also with a suitable accuracy. In most cases this latter problem is equivalent to the effort of estimating a Bernoulli parameter for each of the elements/parameters of the density matrix ρ [17].

1.3 Protocols

In the late 1980s, after several suggestions that mostly relate to the naive linear inversion approach from the last section, one of the first ideas [37] for a *tomographic* method (which coined the term of QST) and its experimental demonstration [7] was the technique of Optical Homodyning Tomography (OHT). All proposals from before that guaranteed physical results had the problem of being nearly experimentally infeasible. They had requirements like preparation of completely pure states [38] or implementation of a lot of complicated shaped potentials [39].
Homodyning tomography relies upon sampling easy representable marginal distributions over phase space for an optical signal. Expressed more visually, this comes down to cutting (by integration) through phase space distributions at different angles and calculating a marginal distribution along each direction. For ensuring accuracy, the measurement at each angle is repeated a large number of times. From this set of distributions, the so-called Wigner distribution is calculated. This quasi-probability distribution can describe the quantum state fully and maps one-to-one to the concrete description of the state: the density matrix. [7] [40]

Lots of (more general) procedures have been suggested since OHT and the applications have become broader. for example, in characterizing entanglement states not only in quantum optics for photons, but also for atoms and even molecules [32] [41]
One particular method that is worth highlighting as well, because of its wide application in experimental procedures in past and present, is the

maximum likelihood estimation method (MLE). The essence of MLE is the goal of maximizing the likelihood to find the given measurement outcomes (b_i) for a certain density matrix. While such optimization methods always involve far more computational complexity than direct calculations, the estimated density matrices in MLE can be guaranteed, under the right constraints [42], to be physically valid states. In this case the most likely *valid* density matrix is estimated.

To perform maximum likelihood estimation one needs to define 3 essential components: a density matrix parametrization, a likelihood function$\mathcal{L}(\rho)$ that is to be maximized and an optimization algorithm to finally execute MLE under the defined conditions. [3]. While the latter is a rather technical choice, defining the shape of a density matrix and a optimization function is a very general problem for the many optimization based protocols that have been proposed.
Basic methods often use *Bloch vectors* [43] $\overrightarrow{r} \in \mathbb{R}^{4^n-1}$ to express a density matrix for n qubits: $\rho = \frac{1}{2}(\mathbb{I} + \sum_{i=1}^{4^n-1} r_i\sigma_i)$ for $\sigma_i \in \{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}^{\otimes n}$. This follows the decomposition of ρ in equation 1.1. This parametrization of density matrices is convenient since, in the case of Pauli matrices, the density matrix representation as a point in or on the Bloch sphere is then a simple and intuitive interpretation.
Alternatively, matrix elements are simply chosen as parameters for optimization and structured such that the resulting matrix will be psoitive semidefinite. (e.g. choosing a Cholesky decomposition of a density matrix $\frac{1}{\mathrm{Tr}\,(T^\dagger T)} T^\dagger T$ with T being a low-triangular matrix and the elements of T being the optimization parameters. [3])

More advanced approaches like in [42] use perturbation theoretic approaches to get a density matrix that has not necessarily full rank. An important property of maximum likelihood estimation, namely, is *asymptotic normality*. It holds generally only for (near) full-rank density matrices and states a guarantee of convergence. If it doesn't hold, one needs to be careful with results of the MLE technique. When the estimated density matrix, for example, has zero eigenvalues, it is probably not the indication of being the most probable state but of a solution that lies outside the physical (positive semidefinite) domain of matrices while maximizing the likelihood function. Physically zero eigenvalues are valid but it is rather a very unrealistic case if the value would be exactly zero rather than just very small in experimental settings, especially with a measurement error-bar. [44].

Moving the value, proportional to the error-bar, a little above zero would yield a better estimate than just keeping it zero. So the MLE estimate is certainly not optimal with finite amount of measurements if it can be improved by such a simple manual adjustment. Nevertheless this technique is applied in lots of physical experiments because, for very few qubits and a lot of measurement repetitions in each measurement basis, one can still extract some useful information with respect to those eigenvalues which are physically reasonable. Still it is very inefficient, we do not have control over which aspects can estimated well and for more Hilbert space dimensions the number of physically unreasonable eigenvalues grows fast.

The likelihood function $\mathcal{L}(\rho) = p(\mathcal{M}|\rho)$ is often taken as a Gaussian [3] or multinomial distribution [43] [42]. The state estimate can then be found by optimizing:

$$\rho_{MLE} = \underset{\rho \in \mathcal{D}(\mathcal{H})}{\arg\max}\, \mathcal{L}(\rho)$$

Beside the computational expense, the finite number of measurements and the required sampling of the search space of possible density matrices are the central problems of MLE if accurate estimates are required.

From this perspective, Bayesian Mean Estimation (BME) has been proposed which is found to be working better on the domain where MLE fails. [42] [44]. Rather than maximizing a likelihood function, the goal is to take a weighted average of all states that are compatible with the measurement outcomes and where the weight of every state is just its likelihood;

$$\rho_{BME} = \frac{\int \rho \cdot \mathcal{L}(\rho) \cdot d\rho}{\int \mathcal{L}(\rho) \cdot d\rho}$$

Here $\mathcal{L}(\rho)$ contains now the likelihood function from MLE and an additional factor called the *prior* that represents the knowledge of the tomographic observer before the measurement. This follows the philosophy of Bayesian statistics: an estimate should represent the knowledge about the system from the measurement outcomes but it also considers all other alternatives to the states that would fit the measurement outcomes best. In fact, in the case that the prior distribution is chosen such that it really and accurately represents the prior knowledge of the observer , the mean squared error of the Bayesian mean estimate is *optimal*. [45] That is, it satisfies lower bounds.

The variety of protocols proposed since then is enormous, to name just

a few : state estimation by compressed sensing [46], local asymptotic normality [47], (projected) least squares [48], direct tomography [49], neural network tomography [50], spectrum estimation [51], partial (reduced density matrix) tomography [33] ,permutationally invariant tomography [52], forced purity routine [53], (generalized) overlap tomography [54], minimax estimation [55]

1.3.1 Framework

We will now turn our focus to describing protocols in a more general way to understand the essential ingredients of existing quantum tomography protocols and identify which choices one has to make when constructing one. Although breaking down a single procedure for the large variety of protocols is not possible, the fundamental framework and ingredients of a QST protocol are often the same.

Initially a set of copies of a state with density matrix ρ is prepared, a set of quantum measurements $\{M_i\}$ are taken to obtain a classical description of the state in the form of a set $\{X_i\}$ of measurement outcomes. Using classical (or in the future quantum) post-processing of this data, an estimate of the density matrix is constructed.
The details of each of these steps are diverse and although most protocols stick to above setting, it's not essential for being a valid way to reconstruct all the information of a quantum state. The way of practical implementation might sometimes also influence how a protocol works. Setup related ways of physically measuring in certain bases and performing error mitigation, for example, require often not just additions but very different protocols.

Within each of the steps, there are a few considerations to make when constructing a protocol. For the input state simplifications should be considered. If the protocol is designed application specific, structure of the input state and its density matrix might be exploitable. For example, if the state can be assumed low-rank (i.e. nearly pure) [56] such that the spectral decomposition $\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$ reduces to $rank(\rho) \ll 4^n$ terms only. Thus the number of parameters that need to be estimated decreases and makes the estimation easier.
For states that have a so-called matrix-product-state (MPS) representation [57] [58] there exist efficient algorithms that have considerable effect on the computational complexity in the post-processing step.

A MPS is an expansion of the full state vector (interpreted as a tensor with n rank-d indices) into a contraction of many tensors with indices of much lower dimension. For full density matrices the number of indices of the contracted tensor is $2n$ instead. The upshot is that we can process many low dimensional components separately instead of one large dimensional matrix. Instead of scaling exponentially in the number of qubits, the computational expense can then scale linearly. Assuming that the MPS is a good approximation to the state, also the number of estimation parameters scale linearly in n. In fact, for many quantum many-body systems of interest this turns out to be a very good approximation [59].
Equally, any other kind of structure in the density matrix can reduce the number of estimation parameters considerably. This emphasizes the importance of examining the structure and symmetry of the quantum system in consideration for gaining a lot of efficiency.

When making measurements, we have to balance between complexity in practical implementation and performance by choosing the type of measurement. That can be POVM measurements, general projective measurements or only 2-outcome measurements. This is in descending order of complexity. Equally, *how* such a measurement is performed is often significantly different. Applying a measurement to all copies of the quantum state at once (global), to each copy individually or to each qubit within a copy separately (local). Also here the complexity in required quantum circuits decreases from global to local. The reason is simply that global circuits work only with gates that exploit entanglement between copies and their qubits. But this scales for more qubits very fast and is not robust with quantum noise, while a local measurement is even for large system easy to implement.
While the kind of measurements can be summarized that way, obtaining the exact measurement operators is where the individual diversity enters. Not less because it is closely connected to the way post-processing is done. While advanced protocols propose parametrizations of measurement operators with numerical optimization [60] to find efficient measurements before or during the experiment, others tend to perform random projective measurements, continuous/time-dependent measurements [61] [62] or fixed pre-selected measurements. Paths of experimental and computational implementation for each them are so drastically individual that one can not select an overall leader in performance for the general case.

Measurement strategies like *overlapping tomography* [63] or *adaptive tomography* [64] have also demonstrated possibilities to find and compress re-

quired measurements in as few measurement *settings* as possible. On-chip implementation is enabled by this for example, which gives estimates by iterative fixed linear transformations. [65]. Measuring with such efficiency can reduce the sample complexity for readily existing protocols significantly as well.

Note that tomographic completeness (i.e. measurement operators forming an operator basis on the Hilbert space $\mathcal{H}$) might not be necessary for every kind of protocol since, for example, highly structured states from above need less information to be described and thus less measurement (repetitions). [66]

Finally, for estimating a quantum state, the desired form of the estimation should be chosen and the accuracy of this estimation should be measured. Different norms, like the *trace norm* and *fidelity*, do exist. [12][67] Beside the obvious option of a point estimate —i.e. estimating a certain single quantum state —we can choose to get a more rough region estimate. That is a bounded region in the space of density matrices in which we expect the real state to be with a given confidence. Another useful option is a Bayesian posterior distribution or any other form of a state estimate that makes sense for the target application. Also this can simplify the problem as the required detail of information for the estimate can be reduced.

Exact inversion, treating expectation values as random variables which together give rise to a direct probability distribution over quantum states (e.g. machine learning of quantum states) and numerical optimization are the three ways to do post-processing. Because of the very limited tools for purely the first two approaches, optimization has become a central part of post-processing in most recent protocols. Might it be only the mapping of an unphysical state estimate to its nearest physical neighbour. Optimization takes some cost function and tries to minimize the cost by changing an hypothesis to the real state. This cost function could be simply the average distance between the matrix elements of density matrices, the likelihood function for the measurement outcomes (like in MLE), a variance measure of expectation values or state-specific properties (e.g. its rank).

Also how good a protocol in the end really is, depends heavily upon what is needed. While in this thesis we focus mainly on sample complexity (i.e. how many copies of the quantum state are needed), like a lot of authors already did, others take the minimization of computational resources more into consideration [68] [53]. Practical feasibility is for readily

well-established protocols also of central discussion when considering requirements on quantum circuits for quantum measurements. The depth of a quantum circuit is a factor, but also whether entanglement of quantum gates for measurements on multiple qubits or copies of a state are required. [59] Implementing and conserving entanglement of quantum gates on a large scale is still a challenge.

1.3.2 Evaluation

Evaluating and comparing protocols is also not a matter of comparing just a single measure for how good a protocol is doing, but consists of several measures that are all sensitive to different aspects, like the effective error in individual density matrix elements, the entanglement properties of the original state and the purity of the original state. [12] [8] Comparing such properties between the original state and the state estimation allows to choose, design or modify a certain state tomography protocol accordingly. This is important since in most applications the accuracy in some property is more significant than for another.

Most common measures include *fidelity*, *trace-distance*, *coincidence* and *entropy*. Since our discussion will only involve the measure of *fidelity*, we define it briefly: [12], [8]

Definition 1.3.1 (Fidelity)**.** *A measure of overlap between two states (e.g. an estimated state $\sigma \in \mathcal{D}(\mathcal{H})$ and a real state $\rho \in \mathcal{D}(\mathcal{H})$) is called the Fidelity. It is the simple inner product of the states:*

$$F := \left(\mathrm{Tr} \sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} \right)^2$$

The fidelity for pure states $|\psi\rangle, |\phi\rangle \in \mathcal{H}$ can be reduced to:

$$|\langle\psi|\phi\rangle|^2$$

Also we introduce a very specific, but widely used set of measures of entanglement [12]

Definition 1.3.2 (Rényi-entropy)**.** *Given some (possibly multipartite) quantum system with Hilbert space $\mathcal{H}_A$, we define the Rényi-α-entropy*

$$H_\alpha(\rho_A) := H_\alpha(\rho_A) = \frac{1}{1-\alpha} \log \left(\mathrm{Tr} \left(\rho_A^\alpha \right) \right)$$

for some $\alpha > 0$

Chapter 2

Incomplete tomography

Enormous efforts have thus been made to find procedures for quantum state tomography which require as few as possible copies of the unknown state ρ, are feasible in experimental settings and are applicable as wide as possible. Despite the rich landscape of suggested QST protocols, there are fundamental lower bounds to be respected which limit the performance for any protocol. The lower bound in sample complexity, for example, for any general quantum state within a D-dimensional state-space is $O(D^2)$ [51]. This follows from arguments in quantum information theory which account for the amount of information that needs at least to be extracted for a full state description and how much information is maximally contained in a single measurement. For rank r states this lower bound reduces to only $O(D \cdot r)$. [56] [51]. However, even if it is assumed that these lower-bounds can be practically reached with some yet unknown protocol, it still needs to be considered that D grows exponentially in the number of qubits. Additionally, this exponential scaling not only applies to measurements but to classical computation and storage as well.

Making more physically relevant assumptions, we might consider a certain structure for the state, like an MPS representation. As mentioned in the preceding chapter, the required polynomial number of samples in system size can offer a large boost in efficiency. Although a lot of physical many-body systems do have such a suitable structure, neither is the general quantum state covered nor is the achieved efficiency consistent over different states. [57][69] This makes the performance of protocols generally unreliable. Also, no matter how favourable MPS representations scale. they always scale with system size. This keeps the bottleneck of tomography upright for very large number of qubits or, even worse, qudits.

A less structured but more broadly applicable alternative to MPS tomography is to learn an unknown state by its internal structure using neural networks. However, despite its popular use, the classes of systems which can be efficiently represented is currently not well known. [59] [70] [50]. For both approaches one thus needs to evaluate each application individually to determine whether it is suitable. Also, although potentially cheap in sample complexity, neural networks can be very expensive in computational resources. Overall, the effective use of neural network tomography will depend mainly on a guaranteed consistency in performance and cost for clearly defined classes of states.

Also measuring a full state with low destruction of the measured state using so-called gentle measurements requires in its optimal settings less copies, but only with the cost of an significantly larger error [71] compared to conventional protocols. For larger systems gentle measurements might be far more feasible in practice and therefore a more acceptable alternative [49] to standard measurements, but an approach that can achieve drastically better accuracies is required for most applications.

Thus the amount of ideas to specialize a protocol to gain lower sample complexity has been enormous. However, increasing the assumptions on quantum states, using machine learning tools, employing neural networks, or compressing quantum measurements is not enough anymore for capturing all the information that a general, highly dimensional quantum state contains [14] in targeted NISQ devices. [2] This suggests to rethink the problem and raises the question whether really all the information of a quantum state is required.

The idea of *partial tomography* is a first reaction to this insight. It aims only to estimate certain parts of a density matrix instead of the complete one. The estimated *reduced density matrices* then only represent and yield information on the state of the fixed sets of qubits. Since one generally needs to fix these choices beforehand, this only makes sense if one already has a strong expectation of the resulting state and just wants to confirm this expectation. Or, alternatively, if a desired property is local to a certain essential set of qubits of the system.
Partial tomography has been shown to be efficient and feasible for high dimensional states in a whole series of works with improving sample complexity [33] [63] [72] [73] [74], but the usefulness for obtaining *global* information of the state is highly restricted if the state is not representable in a convenient form. For an arbitrary k-qubit reduced density matrix, we

can then only extract information about k-local observables in expectation. On the other hand, if it has a suitable representation, the reduced density matrices can reconstruct the complete density matrix efficiently as well, and global observables become accesible. [75] [33] [59] Overall partial tomography has become in efficiency and performance a close competitor to protocols that directly estimate target observables. The latter will be of extensive discussion in proceeding sections.

Closely related to partial tomography is the *direct* measurement of single density matrix elements. However, beside the very restricted usefulness of single elements in general, the greatest practical problem has been the normalization for such elements. Doing this properly does effectively require estimation of all matrix elements using this method, which is generally on the same level of complexity as complete tomography methods themselves. [76] The problem for normalization can be illustrated most easily using the direct measurement scheme with photon detectors. Measuring the photon counts on each detector for a single matrix element does not give us information about the relative count with respect to all other unmeasured matrix elements. So for a meaningful capture of information, we would have to retrieve all counts over the complete density matrix and normalize all matrix elements with respect to the total of all these measurements.
Nevertheless, some schemes [77] [78] [76] have been proposed that promise a relatively low sample complexity for single density matrix elements by using weak or strong measurements that feature low-destruction of the measured copies of a state. Also direct measurement schemes are fast since they require no additional post-processing.
Although the formalism is still actively being developed, expectation of observables are far more complex to predict and weak measurements hard to implement. This is while in practice one is often interested in a flexible measurement scheme that can predict certain properties of a system directly.

2.1 Shadow tomography: the problem

To encounter the sketched problems from the preceding section, the suggestion [9] was made to give up on the goal of predicting (parts of) the density matrix on itself. After all, recall that the description of a state was causing the exponential increase of measurements by the exponential increase of required parameters. Rather one would often like to predict only

a few specific properties of a system based on global or local observables, which are however unknown during the construction of the protocol itself to make it universally applicable. With previously discussed methods in quantum state tomography, this would involve predicting the density matrix of the state and taking expectation values for relevant observables. In the usual context this approach clearly imposes one redundant step of first extracting all the information of a system, while one subsequently discards most of it. Equally well one could try to extract expectation values of specific observables directly from the system, but that obviously would also consume a lot of copies of a state if the number of observables is large. We would rather like to implement a protocol that has low sample complexity, has low cost in implementation by a low circuit complexity and can be adapted easily for a particular application.
The problem, originally stated for two-outcome measurements [9], has been named *Shadow Tomography*:

> Given an unknown quantum state with density matrix ρ and M known measurement operators O_i, the goal of **shadow tomography** is to estimate expectation values $\mathrm{Tr}(\rho O_i) = c_i$ with accuracy ϵ for all $i \in \{1, \ldots, M\}$ with collective - success probability $1 - \delta$ and N, the number of copies of state ρ, minimal.

Addressing exactly this problem will be at the core of the remainder of this thesis. While partial tomography, as discussed above, certainly showed promising results, the complexity of such methods has been increased heavily for the benefit of reducing measurement resources. Taking therefore a step back to reconsider more basic methods is of considerable interest for actual applications and convenient for our understanding of the problem.

The name of shadow tomography (further interchangeably used with the term *incomplete tomography*) originates from visualizing the problem as identifying the expectation values, of selected observables that we want to estimate, as the shadow that the quantum state casts on the observables. In particular, we would like to create from measurements a classical description —*the shadow* —of the state which reproduces the behavior of the state at least in expectation of the observables in consideration. Note that such a classical description is more generally optional in yielding expectation values, but it can help to store the observed state for later evaluation.

Notice also that there are some trivial solutions if the sample complexity or the number of targeted expectation values are arbitrarily large. If $M \geq D^2$ and $N \geq D^2/\epsilon^2$ (recall the Hilbert space dimension $D = d^n$) then the set of M observables is a complete, linearly independent measurement set on the Hermitian subspace of $\mathcal{L}(\mathcal{H})$ and there are enough copies to measure them directly. If only $N \geq D^2/\epsilon^2$ holds, then we are still able to conduct full and general quantum state tomography with protocols mentioned in chapter 1, but we have to do it with (another) complete set of measurements to estimate the full density matrix and extract the required expectation values. Another case is when $N \geq M/\epsilon^2$ since then the number of copies are sufficient to measure each of the M desired expectation values directly. Each on a separate copy. This latter solution might seem optimal in general, but when measuring a lot of observables one should mind that some information of different observables might be overlapping. Reuse of information for a lot of overlapping observables is therefore the key advantage of measurement procedures which are applied to the problem of shadow tomography.

The minimal bound on the number of required state copies for estimating the expectation of any observable up to an additive error ϵ, which follows from information theoretic arguments (theorem 16 in [9], theorem 2 in [59]), is $N = \Omega\left(\log\left(M\right)/\epsilon^2\right)$.
In other words, an exponential number of expectations of 2-outcome measurements can be predicted to accuracy ϵ with only a polynomial number of state copies if the bound can be taken tight.

Like for full state tomography, this specific lower bound has therefore nothing to do with how exactly the problem is approached. It rather is an artefact of how much information can fundamentally be extracted from a certain number of measurements. Note that the actual lower bound for any practical situation can be much worse and depends heavily upon the targeted observables (see theorem 2 in [59]) and how measurements are taken. For the references considered here, this mainly refers to the locality and spectrum of observables, as well as the locality of performed measurements. Here the locality is interpreted more broadly than usual to refer to how many qubits, or even how many copies, of the state measurements and observables are taken. For the target observables that apply to multiple copies of a system, we will more frequently refer to them as *non-linear target functions*.
However, the most notable observation on this lower bound is clearly its independence of the system size n i.e. the number of qubits. It demon-

strates that one could benefit enormously from ignoring the steps involving the density matrix ρ in standard state tomography methods. Needless to say, this assumes that one could find any such concrete protocol that scales (nearly) optimal by this lower bound while it is still able to predict all desired observables well enough.

2.2 Shadow tomography protocols

The natural framework of a protocol in incomplete tomography —that is measuring and subsequently estimating the desired expectations via post-processing —stays similar to protocols in complete quantum state tomography. The essential step of first recovering the whole density matrix has, tough, vanished. This is not only beneficial in the amount of information that needs to be extracted, but also allows to optimize the protocol to the amount and type of required information (in form of observables) directly. For example by using pre-processing, appropriate measurements can be constructed for each specific application. The gain is a consistent performance over different situations.

The problem of shadow tomography was also connected to one possible solution [9], in form of a concrete procedure, for the case that expectations of two-outcome measurements are desired. This involves updating an hypothesis state iteratively in a well-known process called *post-selection*. Each iteration only measures which expectation value of the hypothesis state deviates too much from the corresponding expectation of the true state and corrects the hypothesis state accordingly. To reduce the number of required state copies, the measurements are chosen as gentle (weak) measurements to damage the state copies only slightly and extract just enough information in order to conclude whether the deviation falls outside of a certain confidence interval. This confidence interval can be chosen freely with the failure probability δ as defined in the problem above. Only if the hypothesis state deviates more then the confidence interval, the hypothesis state is adjusted with respect to the observable. This subroutine of finding deviating expectation values has been named *gentle search procedure*. While the number of required state copies per iteration scales independently of state-space dimension: $O(\log^4 M/\epsilon^2)$, the number of iterations does not $O(\log D/\epsilon^3)$. [9]

So the total sample complexity does still scale with number of qubits but only polynomial by $O(\log^4 M \log D/\epsilon^5)$. Although this comes not near the

fundamental lower bound, it is still an exceptional advantage over known full state tomography methods.

Later, different larger improvements on this procedure were proposed. Aaronson et al [79] did for example use the concept of differential privacy. When applied to shadow tomography, it allows gentle measurements to be performed differently by applying a well-known concept from computer science: Differential Privacy. It hinges on adding exponential noise to a state copy in order to make it deviate slightly from the original state. Subsequently it is measured together with an untouched copy of the unknown state. This makes it possible to classify the state based on a two-sided threshold. The result is a search procedure for strongly deviating expectations that can replace the previous gentle search procedure. The main advantage of this is that the algorithm can be *online*, which means that the set of targeted measurements in the involved iterative process does not need to be fixed. Measurements can thus in each iteration still be chosen and specified during the run of the procedure. This makes each measurement independent and invariant of the underlying probability distribution. Depending on the number of target expectations M, the system dimension D and the required accuracy ϵ, this algorithm requires at most $O(\log D^2 \log M^2/\epsilon^8)$ state copies. So in terms of sample complexity it is only better in special cases, for example if the number of measurements M is large relative to the system size. Others like [80] have improved the sample complexity with very similar methods to $O(\log d \cdot \log M^2/\epsilon^4)$.

Also other extensions [81], [82] and [83] (experimental demonstration of [81] in [84]) profit from being an online algorithm but via *Probably Approximately Correct* (PAC) learning of quantum states. That is prediction of expectations via machine learning for any set of measurements as sampled from an arbitrary probability distribution. The goal is therefore only to predict most of these sampled measurements correctly rather then all expectations. The sample complexity is then mainly due to the learning of a reasonable model by acceptance or rejection of predictions on randomly sampled measurement operators. Applying the method to the problem of shadow tomography for the first time gave an improved sample complexity in the error dependence of $O(\log M^4 \log D/\epsilon^4)$ [79]) state copies as well. This was improved quickly to a similar scaling as [80], that we mentioned earlier, of $O(\log M^2 \log D/\epsilon^4)$. [83] Even better, the authors prove that if an approximate set of possible states with high internal structure is known, the $\log D$ dependence can be replaced by a much lower factor.

2.3 Experimentally feasible shadow tomography: Classical Shadow Estimation

While the branch of solutions in the preceeding section addresses the original shadow tomography problem [9] tightly and proves the astonishing high gain in comparison with standard full tomography methods, a few different approaches have been initiated that try to generalize the problem and solutions to a physically relevant context. For example, with observables that have an unit operator norm at the most, the space of predictable physical quantities is somewhat restricted. We would rather like to predict any observable or property of a state that could be useful in the lab.
Another key disadvantage of the original approach by Aaronson is that it requires an exponential number of entangled quantum gates by the use of measurements that apply to multiple state copies at once. This is rather difficult to realize practically [59], especially when considering the remaining polynomial scale in the number of qubits. That scaling is much better than the exponential scaling in complete tomography, but it can still become reasonably large when quantum systems are scaled up continuously.

Even before the problem of shadow tomography was stated in its generality, a few protocols were proposed based on the same philosophy but specialized on predicting a single, fixed and physically relevant quantity directly, e.g. fidelity [85] [86] or entanglement (entropy) [87] [88] [89]. Often though, to generalize the application range, we rather want to know a large set of arbitrary quantities or properties. A few different, but relatively similar, approaches have already been presented to extend those methods which are otherwise only applicable to a single, fixed observable.

One of the most extensively studied approaches is *classical shadow estimation* as originally proposed by Huang et al [59]. Although it is a substantially different approach from those presented above, using only single-copy measurements, classical shadow estimation yields an improved, dimensionally independent scaling sample complexity with

$$N = O(log(M) \cdot \max_i \frac{||O_i||^2_{shadow}}{\epsilon^2}) \tag{2.1}$$

state copies (and measurements) required.

Note the dependence on the observables in this expression. This dependence arises from possible variations in the amount of information that is

required for each observable. The norm is called the *shadow norm* and is a valid operator norm defined as:

$$||O||_{shadow} = \max_{\sigma \in \mathcal{D}(\mathcal{H})} [\mathbb{E}_{U \sim \mathcal{U}} \sum_{b \in \{0,1\}^n} \langle b|U\sigma U^\dagger|b\rangle \langle b|U\mathcal{M}^{-1}(O)U^\dagger|b\rangle^2]^{\frac{1}{2}} \tag{2.2}$$

Here the set of unitaries $\mathcal{U} = \{U_i\}_i$ must be topographically complete and have some discrete probability distribution α_{ii} over its elements, such that they can be drawn at random. We denote this also as the *unitary ensemble* $\mathcal{E} := (\alpha, U)$ The bitstrings b (i.e. vectors $\{|b\rangle\}_{b \in \{0,1\}^n}$) just form the computational basis of the n-qubit Hilbert space $\mathcal{H}$.
The shadow norm is a valid Operator norm. That is, it is completely positive, homogeneous and satisfies the triangle inequality. While with linearity of the *quantum channel* $\mathcal{M}$ the first two claims follow trivially, the triangle inequality requires a more direct workout and follows finally from the triangle inequality of the standard 2-norm.

The fundamental idea of classical shadow estimation is captured by the quantum channel $\mathcal{M}$: $\mathcal{M}(\rho) \mapsto \mathbb{E}_{U \in \mathcal{U}, b \in \{0,1\}^n}[U^\dagger|b\rangle\langle b|U]$. It therefore maps the quantum state ρ of an arbitrary system to a classical object since both U and $|b\rangle$ are classical (storable). The expectation and the required tomographical completeness of the unitary ensemble ensure that this linear map is bijective: it is guaranteed that for any two states $\rho \neq \sigma$ there is a unitary U such that $\langle b|U\sigma U^\dagger|b\rangle \neq \langle b|U\rho U^\dagger|b\rangle$. Surjectivity follows trivially from the form of the classical codomain i.e. $U^\dagger|b\rangle$ is always a pure state vector by definition and therefore corresponds to some pure (or even mixed) state with density matrix ρ.

The physical reason for introducing this quantum channel is the representation of the conventional measurement process in an arbitrary basis with a simple notation: First the (yet unknown) quantum state ρ is rotated with a unitary U to the computational basis $\{|b\rangle\}_{b \in \{0,1\}^n}$: $U\rho U^\dagger$. Next the measurement in the computational basis is taken as projective measurements on the vectors of the standard basis. This results in outcome $b \in \{0,1\}^n$ with probability $\langle b|U\rho U^\dagger|b\rangle$, following Born's rule. Finally, as we would like to have the outcome in the original basis of measurement, we have to rotate the outcome vector (given by $|b\rangle$) back using the same unitary to obtain a *classical description* $U^\dagger|b\rangle\langle b|U$.
The defined quantum channel $\mathcal{M}(\rho)$ therefore describes the mapping of the state ρ to its measurement result in expectation over the unitaries U within the chosen ensemble $\mathcal{U}$ and all possible measurement results $b \in$

$\{0,1\}^n$. We observe that the inverse of $\mathcal{M}$ exactly reproduces the state as $\rho = \mathbb{E}_{U\in\mathcal{U},b\in\{0,1\}^n}\mathcal{M}^{-1}(U^\dagger|b\rangle\langle b|U)$. Here, the inverse channel can be applied computationally on a classical device by bijectivity of $\mathcal{M}$. However, physically implementing the inverted channel is not possible as it is not completely positive in its generality. As a quantum channel should always be a CPTP map, we consider it non-physical.- This is reasonable since it involved measurements which are known to be connected to loss of information i.e. irreversibility.

By recalling our goal of predicting a number of expectation values $\langle O_i\rangle$ corresponding to quantum state ρ as given by Born's rule $\langle O_i\rangle = \mathrm{Tr}\,(\rho O_i)$, we see that $\mathrm{Tr}\,(O_i\rho) = \mathbb{E}[\mathrm{Tr}\,(O_i\hat{\rho})]$ with $\hat{\rho} = \mathcal{M}^{-1}(U^\dagger|b\rangle\langle b|U)$ a *classical snapshot*. This classical snapshot could thus be interpreted as a single measurement in the basis represented by a randomly chosen unitary U and measurement outcome b. The array collecting a series of measurements with different U and different outcomes b forms $S(\rho) = \{\hat{\rho}_1, \hat{\rho}_2, ..., \hat{\rho}_N\}$ and is called a *classical shadow*. That is, it captures a certain amount of direct information and structure of the underlying state ρ, depending mainly on the size N of the shadow, that can be used to predict the expectations $\mathrm{Tr}\,(\rho O_i)$. Shadow size has a one-to-one correspondence to the number of measurements and, therefore, measured copies of the state in equation 2.2

To predict single-valued expectations we need to combine our array of snapshots to a single, ideal expectation-valued snapshot $\hat{\rho}$ that we use to calculate any $\mathrm{Tr}\,O_i\hat{\rho}$. The arithmetic mean could be taken. However, the authors of [59] suggest to apply an alternative that is robust from measurement outliers that could have affected the accuracy of the individual snapshots in the shadow. Thereby directly referring to applications in experimental setups. The simple algorithm is called *median-of-means*: First the shadow of size N is partitioned into k sets of N/k snapshots and for each partition the arithmetic mean $\hat{\rho}_{(1)}, ..., \hat{\rho}_{(k)}$ is taken element-wise. For each of these k mean snapshots, the expectation for the desired observables O_i are taken: $\mathrm{Tr}\,O_i\hat{\rho}_l$. Finally the predicted expectation $\hat{o_i}$ is taken as the median of all these k expectation values. As such the influence of outliers on this final prediction becomes exponentially small ([90], [91], [92]) since a median picks the expectation value in such a way that about half of the k expectations would have to deviate towards the outliers to have meaningful effect on the final result. Starting from the k mean values $\{\hat{o_i}^j\}_{j=1,...,k}$, this can be seen more directly by evaluating the probability that more than $k/2$ averages exceed the accuracy ϵ from the desired ex-

pectation: $Pr[|\{|\hat{O}_i - \mathrm{Tr}\, O_i\rho| \geq \epsilon\}| \geq k/2] \leq e^{-k/2}e^{M\epsilon}$. The inequality results from taking a Chernoff bound and shows indeed that this probability becomes exponentially small in k. [93]
Others (e.g. [94],[45]) have remarked that the benefit of median of means for current experimental platforms over a simple arithmetic mean is insignificant. Still in more extended studies of the algorithm, like in [95] under specifically well-behaved noise, this kind of estimator is considered as more important in future practical situations.

It is worth to point out that although this snapshot looks like a density matrix and satisfies the trace condition $\mathrm{Tr}\,(\cdot) = 1$ by construction, it generally is not positive semi-definite. In fact, it turns out that, by comparison with conventional state tomography methods, this is the great advantage of classical shadows [45] since their predictions of expectation values can converge towards the true value under the only constraint of a unit trace. Even if the true state is a state on the boundary of the convex space of density matrices $\mathcal{D}(\mathcal{H})$ e.g. the Bloch sphere. Such a state is often hard to access with traditional state tomography methods since their search space is restricted to the Bloch sphere. In the end, however, the big advantage of having snapshots at all, as opposed to direct post-processing towards certain observable expectations, is that we can take the measurements and store the snapshots completely classically without requiring the associated quantum device at a later time. Also the choice of observables can be taken later, although the procedure can be made more efficient if the observables are known prior to measurement.
Even stronger, the structure of the shadows, involving $U^\dagger|b\rangle$ and $U\rho U^\dagger$, provides us with the ability to use state representations on our classical post-processing device that are more efficient in computational resources. This is true for using MPS representations for states in system that inherit the structure of tensor networks, but even more for so-called *stabilizer states* [59], [26] [20] (or more recently, graph states [96]).

The simplest examples of useful unitary sets —that are in context also demonstrated in [59] —are the n qubit Clifford group $\mathcal{C}^n$ or the n qubit Pauli group $\mathcal{P}^n$ as introduced in section 1.1.4. For both $\mathcal{C}^n$ and $\mathcal{P}^n$ the shadow norm in the lower bound of the sample complexity from equation 2.2 can be replaced by the Hilbert-Schmidt norm $\mathrm{Tr}\,(O^2)$ or a factor $4^k \cdot ||O||_\infty^2$ respectively. [59] Here k is the *locality* (or *weight*) of the observable and $||\cdot||_\infty$ its spectral norm.
The maximum in equation 2.2 then ensures that we extract enough information to even cover the most demanding observable. While for observ-

ables that have small eigenvalues and apply to very few qubits the norm is small, it can increase rapidly for (nearly) global observables or observables with considerably large operator norm. In such cases the system size n can thus still enter indirectly via this shadow norm. Consequently this approach then does not offer any advantage over state tomography.

Furthermore the authors of [59] prove that, for any protocol that uses only single-copy measurements and no additional assumptions on the unknown state, the given bound is information theoretically *optimal*. (see for an off-topic proof supplementary section 7 in [59]) This is true for the n-qubit Clifford unitaries $\mathcal{C}^n$. For other sets of unitaries, like the one represented by the n-qubit Pauli group $\mathcal{P}^n$, the scale in locality can be slightly better. e.g. 3^k instead of 4^k. [59] [35]
Note that this lower bound agrees also with the aforementioned lower optimal bound for the general problem of shadow tomography when multi-copy measurements are used and only 2-outcome POVMs with spectrum in $[0, 1]$ are targeted. Weakening these constraints for classical shadow estimation restricts our abilities to use symmetries and multi-copy structures in the problem, but makes the practical implementation much easier and more broadly applicable to physical systems.

Beside the advantageous lower bound on sample complexity, classical shadow estimation profits from its simplicity. Note that shadow estimation to some extend is just an equivalent to the naive linear inversion method that we introduced with complete quantum state tomography in section 1.2. In fact, choosing an exponential amount of measurements in system size allows one to re-establish the complete density matrix to finite accuracy by reproducing the expectation

$$\mathbb{E}_{U\in\mathcal{U},b\in\{0,1\}^n}[\hat{\rho}] = \mathbb{E}_{U\in\mathcal{U}} \sum_{b\in\{0,1\}^n} [\mathcal{M}^{-1}(U^\dagger|b\rangle\langle b|U)] = \rho$$

The intuition of why the method is now well-suited for the current problem is simply that the amount of required information of the predicted classical shadow is reduced heavily and recycling of information between different pieces of extractable information —i.e. system properties —is possible. This yields the bound discussed above.

Still it is quite remarkable how random sampling actually can reproduce exactly the information that we want without extracting all information. We will give a quick insight on why this is for the most elementary situation of Pauli measurements and observables. For simplicity we also use

just a simple arithmetic mean to find the expectation:
Consider the Bloch sphere representation of the unknown state

$$\rho = \bigotimes_{j=1}^{n} \frac{1}{2}(\mathbb{I} + \alpha_j \sigma_x + \beta_j \sigma_y + \gamma_j \sigma_z) = \bigotimes_{j=1}^{n} \frac{1}{2}(\mathbb{I} + \overrightarrow{n}_j \cdot \overrightarrow{\sigma})$$

with $\overrightarrow{n}_j$ the Bloch vector of the state for each qubit j. We assume a product state but from linearity the following will generalize also to non-product-states. Further take the observable $O = \bigotimes_{k=1}^{n} \sigma_{j_k}^{(k)}$ for $j \in \mathbb{I}, x, y, z$ and a single Pauli basis measurement P with $P^{(k)} \in \{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}$ a single-qubit measurement operator. Importantly, choosing P is conditioned on that it covers information about the observable O, for Pauli measurements that is simply $P^{(k)} = \sigma_{j_k}$ for any k and $j_k \neq \mathbb{I}$. If the observable is locally the identity, the choice of measurement at that position does not matter as we can always locally marginalize the probability distribution that we obtain from the measurement. However, if measurement and observable disagree both with non-trivial Pauli operators, no information can be obtained about the observable at all. We say that a measurement needs to "*hit*" $(O \triangleright P)$ the observable [36].
From the perspective of random measurements in classical shadow estimation that means collecting information for an expectation value by sampling enough measurements that hit the corresponding observable. Finding the total number of required measurements therefore boils down to the statistics of sampling such a P.

The projectors for a measurement P onto the eigenspaces of m different measurement outcomes $b \in \{-1, 1\}^n$ are given by $\Pi_{i \in \{1,m\}} = \bigotimes_{k=1}^{n} \frac{1}{2}(I + b_i^{(k)} P^{(k)})$. Therefore the probability for yielding outcome string b on state ρ on measuring P is $Pr(b|\rho, P) = \mathrm{Tr}\left[\bigotimes_{k=1}^{n} \frac{1}{2}(I + b^{(k)} P^{(k)})\rho\right]$.
Our goal is to find the expectation

$$\mathrm{Tr}\, O\rho = \mathrm{Tr} \bigotimes_{k=1}^{n} \sigma_{j_k}^{(k)} \rho = \mathrm{Tr}\left[\bigotimes_{k=1, \sigma_{j_k}^{(k)} \neq \mathbb{I}}^{n} \sigma_{j_k}^{(k)} \bigotimes_{k=1, \sigma_{j_k}^{(k)} = \mathbb{I}}^{n} \sigma_{\mathbb{I}}^{(k)} \right]$$

Performing then our measurement P on the state allows us to introduce the corresponding outcome string b into this expectation via the above outcome probability and the use of Born's rule: Remark that summing over all possible outcome strings $b \in \{-1, +1\}^n$ will eventually cancel the

term $b^{(k)}P^{(k)}$ in the measurement operator $\frac{1}{2}(I + b^{(k)}P^{(k)})$ for any $P^{(k)}$ and recall $\operatorname{Tr} P^{(k)}\rho = \operatorname{Tr} \frac{1}{2}(b^{(k)} + P^{(k)})\rho = b$.
Using this to rewrite the second and first factor in $\operatorname{Tr} O\rho$ respectively, we get:

$$\operatorname{Tr} O\rho = \sum_{b\in\{-1,1\}^n} \Pi^n_{k=1,\sigma^{(k)}_{j_k}\neq\mathbb{I}} \operatorname{Tr} \frac{1}{2}(P^{(k)} + b^{(k)}) \cdot \Pi^n_{k=1,\sigma^{(k)}_{j_k}=\mathbb{I}} \operatorname{Tr} \frac{1}{2}(\sigma^{(k)}_{\mathbb{I}} + b^{(k)}P^{(k)})$$

Here we used in the first factor the assumption that P hits O to replace the local observable operators by measurement operators since $\sigma^{(k)}_{j_k} \neq \mathbb{I}$. Also note that the second product only picks out those subsystems k with an observable equal to identity, i.e. where the measurement operator does not matter. We therefore recognize the *marginal* probability $\bar{Pr}(b|\rho, P)$ for which we take into account that locally for $\sigma^{(k)}_{j_k} \neq \mathbb{I}$ also the outcome of the measurement is irrelevant. Finally then:

$$\operatorname{Tr} O\rho = \sum_{b\in\{-1,1\}^n} \bar{Pr}(b|\rho, P) \cdot \Pi^n_{k=1,\sigma^{(k)}_{j_k}\neq\mathbb{I}} b^{(k)} = \mathop{\mathbb{E}}_{b\in\{0,1\}} \Pi^n_{k=1,\sigma^{(k)}_{j_k}\neq\mathbb{I}} b^{(k)}$$

with the expectation over the measurement outcome strings b. The most practical way to approximate this expectation is by taking a number of M independent measurements P for which we obtain different outcomes b. An arithmetic mean over only those outcomes that correspond to measurements which hit the desired observable O then yields [36]:

$$Tr(\hat{\rho}O) = \frac{1}{\mu} \sum_{m=1}^{\mu} \Pi^n_{k=1,\sigma^{(k)}_{j_k}\neq\mathbb{I}} b^{(k)}_m$$

with μ the number of hitting measurements out of all measurements M, b_m the outcome string of measurement m and where the sum iterates all hitting measurements m.

Obviously there is no well-defined expectation if no measurement hits the observable. Sampling measurements at random, however, facilitates a reasonable probability of hitting an observable at least once and therefore allows a finite confidence of predicting expectations to a certain accuracy for a finite amount of measurements. Notice that the above is just an abstraction of classical shadow estimation compared to how we introduced it earlier. Since the intermediate step of producing the classical shadow itself is optional, the same justification for the use of a linear inversion scheme applies.

Although we are not directly concerned with predicting non-linear functions in the density matrix ρ in this thesis, it is in the interest of completeness to also mention the capabilities of classical shadow estimation with respect to polynomial features. Common non-linear quantities of interest are for example entropy measures of multipartite entangled systems. e.g. Rényi-entropies as introduced in section 1.3.2 For Rényi-2 entropy there is a quadratic dependence through the purity $\mathrm{Tr}(\rho^2)$.

Analogous to linear properties, the expectation $\mathrm{Tr}(O\rho^2) := \mathbb{E}\ \mathrm{Tr}(O\hat{\rho}^2)$ can be obtained from a classical description for $\hat{\rho}^2$. More accurately, we use the simplification that there exists an $\tilde{O}$ for every observable O such that $\mathrm{Tr}(O\rho^2) = \mathrm{Tr}(\tilde{O}\rho^{\otimes 2})$. [59]
By adopting a statistical tool called *U-statistics*, a snapshot $\hat{\rho}^{\otimes 2}$ can be obtained from the original linear snapshots $\hat{\rho}$ in the protocol.
Essentially U-statistics states that for a set of N observations of a random variable X, $\{X_1, \ldots, X_N\}$, and a *kernel* function $h(X_1, \ldots, X_r)$ of order $r < N$, there exists an unbiased and minimum variance estimator [97]

$$\mathbb{E}[h(X_1, \ldots, X_N)|X_1, \ldots, X_N] = \binom{N}{r}^{-1} \sum_{1 \le i_1 < \ldots < i_N \le N} h(X_{i_1}, \ldots, X_{i_N})$$

This is also called *symmetrization of statistics*. [59]

For classical shadow estimation with quadratic features this means: Given N snapshots $\hat{\rho} = \mathcal{M}^{-1}(U_i^\dagger|b_i\rangle\langle b_i|U_i)$, we obtain the quadratic snapshot by summing over all pairs of linear snapshots:

$$\hat{\rho}^{\otimes 2} = \frac{1}{N(N-1)} \sum_{i \neq j} \mathcal{M}^{-1}(U_i^\dagger|b_i\rangle\langle b_i|U_i) \bigotimes \mathcal{M}^{-1}(U_j^\dagger|b_j\rangle\langle b_j|U_j)$$

for $N(N-1)$ pairs of snapshots. Note that we solely adjust the post-processing of the measurement outcomes, we do not need to change anything on the measurement process itself.
Similar expressions can be found for any other polynomial function of the density matrix and therefore also for any function with a well-defined power expansion.

Obviously there is a cost to this procedure in the variance of the estimator (and thus the required number of linear snapshots N). Compared to linear function prediction, the sample complexity can be lower bounded, like for linear features, by $N \geq \frac{3\,\mathrm{Tr}(O^2)}{\epsilon^2}$ and $N \geq \frac{4^k}{\epsilon^2}$ for global and local Clifford

measurements respectively. The difference to the bounds following from eq. 2.1 is that the norm of an observable on multiple state copies ρ can increase more rapidly and the weight w of observable O is the maximum over the effective weights of the observable O on each of the copies of ρ. From a practical point of view, this still scales much more advantageous than usual state tomography techniques in the lab.

Classical shadow estimation was recently also demonstrated experimentally by Struchalin et al [94] using spatial photon modes. They showed that with classical shadow estimation a clearly more accurate and unbiased prediction of fidelity and other linear observables were obtained where complete state tomography would have failed. However, direct comparison was only shown with Maximum-Likelihood-Estimation which is used practically a lot but which already was known to be biased by its asymptotic optimality, as elaborated in the previous chapter. More advanced protocols and target system properties (non-linear, non-rank-1), that were included in the numerical analysis of [59], were not considered. Also the authors used, in contrast to experiments for other incomplete tomography protocols (e.g. [84]), single photons instead of multipartite systems. This takes away any unique quantum mechanical nature and restricts the assumption of the general benefit for classical shadows. Still it demonstrates the main predictions on accuracy and efficiency for this way of implementation. The consistency, or independence of state-space dimension, was not achieved but this was dominantly claimed to be due to issues in practical implementation of higher dimensional states which resulted in decreasing estimation accuracy with larger system sizes.

Derandomization

The aforementioned information theoretic optimality restricts our ability of improvement in sample complexity strongly. As we target near term feasible solutions to the shadow tomography problem, tightening the assumptions on measurements, observables and possibly occurring quantum states seems therefore the only way to overcome this limit.
A more recent work [36] by the same authors as [59] took the protocol a step further by a well-known process in computer science [98] [99][100] called *derandomization*. Essentially it takes the involved random elements of the protocol away, e.g. random single-copy measurements for classical shadow estimation, by fixing them such that the performance of the protocol is at least as high as for the randomized procedure. [36]

The performance is measured for each possible choice of measurement by a cost function that needs to be minimized. Here the cost function orients itself on our goal to find predictions of expectations for any set of M observables $\{O_m\}_{1\leq l\leq L}$ to accuracy of at least ϵ with high confidence δ. That is, the probability to fail on the desired accuracy limit for any observable O_i under measurement set $P = \{P_i\}$ should satisfy [36]

$$Pr[(\max_{1\leq l\leq L} |\hat{o}_i - \mathrm{Tr}\,(O_i\rho)| \geq \epsilon)|P] \leq \exp\,(-\frac{\epsilon^2}{\mu(O_i,P)}) := Conf[\epsilon, O_i, P] \leq \delta/2 \tag{2.3}$$

Where the first inequality uses a Chernoff bound and $\mu(O_i, P)$ counts the (average) amount of "hitting" measurements in the set P for observable O_i, like defined and used in the preceding subsection. The factor 2 in the last equality originates from the fact that although a normal probability distribution is two-sided around the expectation of 0, we only consider absolute values of deviations. We thus only consider one half of the total confidence interval described by δ.
The actual probability function itself is far more complicated but it suffices to define the confidence $Conf[\epsilon, O, P]$ as explicit upper bounding cost function of measurements and observables. This is called an *pessimistic* upper bound on the probability. [101]

Derandomization then starts by treating each local Pauli operator in the measurement scheme P, $P_i^{(j)}$ for measurement round $1 \leq i \leq N$ and qubit number $1 \leq j \leq n$, as random variable, which can take any operator value from the Pauli group except the identity $\{X, Y, Z\}$. This is the completely randomized situation and corresponds essentially to the original protocol of sampling measurements uniformly. The corresponding confidence bound is easily found since we can consider the only unknown $\mu(O_i, P)$ measurement-wise and qubit-wise: Per local measurement $P_i^{(j)}$ the random draw is between three operators. The *hitting probability* is therefore $1/3$ if observable $O_i^{(j)} \neq \mathbb{I}$. On the other hand, if $O_i^{(j)} = \mathbb{I}$ the exact choice for $P_i^{(j)}$ is irrelevant and the probability of hitting —i.e. the probability of gaining information from the measurement —is one. The measurement operator is then chosen uniformly at random to not waste the opportunity of measurement. Extra gained information can then be used for prediction of other observables, while the information for the local identity observable can be re-obtained from marginalization.
The total hitting probability, and average hitting count $\mu(O_i, P)$, of measurement set P on observable O_i is then $N \cdot 3^{-w_i}$, with w_i the weight of O_i.

So for the randomized case the confidence bound becomes

$$Conf[\epsilon, O, P] = exp(-\frac{\epsilon^2}{\mu(O_i, P)})$$

The idea of derandomization, in particular the method of *conditional expectations* [101], is now to replace random variables $P_i^{(j)}$ by a fixed value $p_i^{(j)}$ such that

$$Conf[\epsilon, O_i, [P_1^1, \ldots, p_i^j, \ldots, P_N^n]] \leq Conf[\epsilon, O_{,i} [P_1^1, \ldots, P_i^j, \ldots, P_N^n]]$$

This can be done be iterating through all possible values for $p_i^{(j)}$, determining the confidence bound $Conf[\epsilon, O_i, P]$ and choosing the $p_i^{(j)}$ with the tightest confidence. The existence of a $p_i^{(j)}$ satisfying the condition is guaranteed since the confidence $Conf[\epsilon, O_i, P]$ for random variable P_i^j is an average over all possible choices. Hence some operator $p_i^{(j)}$ must exist which has a confidence value below or equal to this average. [101] By fixing this value and considering $Conf[\epsilon, O_{,i} [P_1^1, \ldots, P_i^j, \ldots, P_N^n]]$ as the new upper bound on the probability in eq. 2.3, we can repeat the process for any other random measurement operator P_i^j until all random variables are replaced. [36] [101]

The above can also be generalized to more observables by taking a union bound over the probability in eq. 2.3. Combining it with the Chernoff bound for each observable O_i, the same derandomization procedure can be applied by considering the total cost over all observables , i.e. [36]

$$Conf[\epsilon, O, P] := \sum_{i=1}^{N} e^{-\frac{\epsilon^2}{\mu(O_i, P)}}$$

Note that the total number of measurement rounds N has to be manually adjusted to the performance gain of derandomization to make an actual practical difference. For this, [36] considered an adaptive derandomization procedure in which the number of hitting measurements per observable is continuously traced in order to terminate the process as soon as a measurement threshold for all observables is passed. This ensures that the measurement scheme is minimal and complete without tight assumptions on the number of measurements beforehand. Our numerical framework implements this approach as well.

Because the confidence bound is highly dependent on the structure and combinations of observables, no guaranteed performance bound was directly derived by [36] while one might be able derive one for specific sets of local observables. More directly, the guaranteed performance bound must be equal to the randomized procedure from the original protocol since there is theoretically no guarantee that derandomization actually performs better. We only require at least equal performance on the cost function in each step. If, for example, very few (nearly) global observables have an uncommon structure in the full set of observables, derandomization with the confidence cost function is highly unlikely to observe and choose them as measurement operators. Randomized measurements might still have randomly drawn measurements compatible with such observables, obviously at the cost of accuracy on other expectations. In the end this thus boils down to the global-local minima issue in many optimization problems.
In most generic situations, however, derandomization has been shown to yield a much lower bound than the random protocol. [36]

Before derandomization, so-called [35] were introduced by [35] with a very equivalent formalism but with a skewed probability over the random choice between $\{X, Y, Z\}$ on each qubit position and within each measurement round. This can be seen as partial derandomization. Numerical experiments have, however, already demonstrated locally biased shadows to perform worse than derandomization for specific cases. [36]. What might make locally biased shadows nevertheless a better alternative is its still existing partial probabilistic nature. This enables the chance to sample measurements for few observables with rare structures to be non-zero.
So although derandomization has shown that sample complexity of the original protocol can be improved by taking the desired observables into account, we did not gain any specific guarantees or made any changes to the actual prediction process. In fact, the possibly better results with derandomization are not very surprising on themselves. The question thus remains how to change the process for better sample complexities by challenging more fundamental restrictions such as those given by the information-theoretic optimality in [59]. One of the possibilities could be the customization of shadow tomography to specific applications. In the next section we introduce one such application candidate which could be used to benchmark and study the actual experimental use and customizability of incomplete tomography protocols, like classical shadow estimation, in systems of highly quantum mechanical nature..

2.4 Application: Multipartite spatial entanglement in photonic systems

One of the most interesting but directly accessible areas of benchmarking and applying both quantum state tomography and shadow tomography in the near term is quantum optics. The reason is that much of the required transformations, state preparations and measurements are easier to implement in photonics than in, for example, solid state devices. The low influence of decoherence on a photonic system is an important aspect of this. Currently coherence times of tens of microseconds are easily achievable, while extension to multiple hours is expected for this platform. All experimental demonstrations mentioned earlier —e.g. [94] [84] [52] —did use photonic platforms as well for this reason.

Here we concentrate on one particularly interesting experimental setup arising in measurements on systems of 4-OAM photons, that are systems of 4 entangled photons for which we take not the polarization but the spatial orbital angular momentum (OAM) modes in consideration as the primary degrees of freedom for a subsystem. Orbital angular momentum is a spatial property of photons that, in contrast to polarization modes, allows us to have not only two-level (qubit) systems of photons but arbitrarily large and discrete subsystems (qudits) as well.
Well-known natural bases for discrete OAM modes are the Hermite-Gauss (HG_n^m) and Laguerre-Gauss (LG_p^l) modes with indices $m, n, l, p \in \mathbb{N}$. Details on both of these have been discussed extensively in other places [102] [103] [104] and will not be relevant for the current discussion. As of the instructive illustration of orbital angular momentum modes as the gyrating Poynting vector of a beam around its axis of propagation and the consequently emerging picture of helix-shaped wavefronts, we will assume a Laguerre-Gauss (LG_p^l) basis since it supports this visualization with its natural description in cylindrical coordinates and an angular momentum dependent phase factor $\exp(i\phi l)$ [103] [105]. Also we will ignore the index p and focus fully on l which corresponds to the actual azimuthal orbital angular momentum number l of the mode. But one should be aware that this makes LG_p^l modes to an incomplete basis of $\mathcal{H}$.

To reach an effective higher dimensional $D = d^n$ state-space $\mathcal{H}$ we thus increase the dimension d of a single subsystem, rather than increasing the number of subsystems n. By example through increasing the range of possible values of l for the LG basis. Practically such states are often

harder to prepare and to control, the resulting compactness and reduced circuit complexity for such systems, however, might be still be advantageous. Multipartite entanglement of single photon subsystems that have many degrees of freedom allows specifically to encode more information [106] and promises therefore advanced applications in quantum technology like quantum cryptography and information transfer in quantum networks [107]
One application that is particularly of interest in this context is quantum secret sharing between multiple parties. Not only because of the increased information density compared to qubit systems, but also in terms of security. [108]
In the context of this thesis, however, the main interest towards this specific photonic system concerns the study of incomplete tomography protocols in a eventually useful, but practically feasible environment.

2.4.1 Theory

One way to create systems of OAM entangled photons is pair production by parametric down conversion (PDC). It is a non-linear scattering process of an incident pump beam from a laser in an optical, often birefrigent, crystal, which induces occasionally the creation of a photon pair. Photons in the two resulting beams, conventionally called the *signal* and *idler* beams, have then anti-correlated angular momentum modes. The latter following from conservation of angular momentum under the assumption of an gaussian incident beam with no angular momentum. [103] [105] [109]. Also energy conservation holds naturally when an incident photon eventually is converted into a photon pair such that $\frac{1}{\lambda_{pump}} = \frac{1}{\lambda_{signal}} + \frac{1}{\lambda_{idler}}$.
Through the non-linearity of PDC process, however, understanding and creating such systems for more than two photons with discretized degrees of freedom has been a challenge. [110] [105] Still, entanglement of four photons has been experimentally observed [111] and, more recently, entanglement between photon pairs particularly resulting from the process of parametric down conversion (PDC) has been demonstrated and studied. [105] [112].

The production of useful photon pairs with anti-correlated spatial modes hinges on the concept of *phase matching* [103], that is $\overrightarrow{k}_{pump} = \overrightarrow{k}_{signal} + \overrightarrow{k}_{idler}$ for $\overrightarrow{k}$ the wavevectors of each of the beams. This gives rise to different types of PDC, depending on the orientation of the wavevectors ,from

which only *type-I* PDC will be of interest here. Thus we consider the case of all photons from pair production having the same polarization which is orthogonal to the polarization of the incident pump beam. This makes the process polarization independent and enables us to omit it from the dynamical description of the process [103] and, finally, from the description of the resulting entangled four photon state as well.

A recent thesis [103] discusses also the relevance of different kind of pair productions in PDC processes, i.e. the ratio between photon pairs produced by spontaneous parametric down conversion (SPDC) and those resulting from stimulated emission of photon pairs with PDC. Here stimulated emission means, as usual, that a previously created photon pair stimulates the emission of another nearly identical photon pair in the crystal, while spontaneous emission creates pairs at different positions in the crystal independently of the existence of pairs in identical modes. In the case of stimulated emission the two involved pairs are also consequently entangled and it is therefore the most relevant factor in creating entangled four photon states. In practice PDC will be a combination of stimulated and spontaneous pair production.

The difference in production is also represented in the structure of the state for two entangled photon pairs by the angular momentum that each photon in a pair carries. We need to treat photons as indistinguishable bosons, therefore it is not definite which photon occupies which state. To keep notation simple we use the Fock state basis from second quantization i.e. the ket $|n_{k_1} n_{k_2} n_{k_3} ...\rangle$ denotes the number of photons n_{k_i} in the mode with index k_i such that $\sum_i n_{k_i} = n$, the total number of photons. Independent pairs in different OAM modes l_1 and l_2 are then written as $|1_{l_1} 1_{l_1} 1_{\bar{l}_1} 1_{\bar{l}_2}\rangle)$ and pairs of photons in the same mode l give $|2_l 2_{\bar{l}}\rangle$ with $\bar{l} := -l$. Note that when we say a pair is in mode l, we mean that one photon is in the OAM mode with index l while the other is in mode of $-l$.

In consequence we also require a description of the PDC process that is adjusted for second quantization and the angular momentum basis. We introduce therefore the *annihilation operator* $\hat{a}_k$: $\hat{a}_k|n_1...n_k...\rangle = |n_1...(n_k - 1)...\rangle$ and *creation operator* $\hat{a}_k^\dagger$: $ata_k^\dagger|n_1...n_k...\rangle = |n_1...(n_k + 1)...\rangle$ up to normalization.
The relevant non-linear part H_{nl} of the Hamiltonian $H = H_l + H_{nl}$ which describes the interactions within the PDC was found to be: [for an exten-

sive discussion and derivation see chapter 3 & 4 of [103]]:

$$H = \frac{i}{2}\kappa\hbar \sum_{l=-\infty}^{\infty} (\hat{a}_l^\dagger \hat{a}_{\bar{l}}^\dagger - \hat{a}_l \hat{a}_{\bar{l}}) \tag{2.4}$$

States of photon pairs resulting from PDC can then be yielded from expansion of the usual unitary evolution operator $\exp iHt/\hbar$ which is applied to the *vacuum state* $|vac\rangle$. That is the state (usually also denoted $|0\rangle$) in which the system is by definition in its ground state and, in the Fock basis, no mode is occupied by any photon i.e. in total no photons exist.
The expansion is valid because the non-linear contribution, represented by κ above, in the total Hamiltonian is very small. This also leads to a very low pair production rate, in optimal cases on the order of 10^{-6} per incoming photon. [113]. This gives:

$$\begin{aligned}
|\psi\rangle &= \exp(iHt/\hbar)|vac\rangle \approx (1 + \frac{i}{\hbar}Ht - \frac{1}{2\hbar^2}H^2t^2 + ...)|vac\rangle \qquad (2.5)\\
&\propto |vac\rangle + \frac{it}{\hbar}\sum_{l=-\infty}^{\infty}(\hat{a}_l^\dagger \hat{a}_{\bar{l}}^\dagger - \hat{a}_l\hat{a}_{\bar{l}})|vac\rangle \\
&\quad - \frac{t^2}{\hbar^2}\sum_{l_1=-\infty}^{\infty}\sum_{l_2=-\infty}^{\infty}(\hat{a}_{l_1}^\dagger \hat{a}_{\bar{l}_1}^\dagger \hat{a}_{l_2}^\dagger \hat{a}_{\bar{l}_2}^\dagger - \hat{a}_{l_1}^\dagger \hat{a}_{\bar{l}_1}^\dagger \hat{a}_{l_2}\hat{a}_{\bar{l}_2} - \hat{a}_{l_1}\hat{a}_{\bar{l}_1}\hat{a}_{l_2}^\dagger \hat{a}_{\bar{l}_2}^\dagger + \hat{a}_{l_1}\hat{a}_{\bar{l}_1}\hat{a}_{l_2}\hat{a}_{\bar{l}_2})|vac\rangle + ... \\
&= |vac\rangle + \sum_{l=-\infty}^{\infty}\frac{it}{\hbar}|1_l 1_{\bar{l}}\rangle - \frac{t^2}{\hbar^2}\sum_{l_1=-\infty}^{\infty}\sum_{l_2=-\infty}^{\infty}(|1_{l_1}1_{\bar{l}_1}1_{l_2}1_{\bar{l}_2}\rangle - |vac\rangle) + ...
\end{aligned}$$

where we used that $\hat{a}_l|vac\rangle = 0$ for any $l \in \mathbb{N}$. While the zeroth order term, naturally, stays in the vacuum state and creates no photon pair, the first order term is the creation of a single photon pair. The second order term then finally corresponds to the states of interest that correspond to four photons or, equivalently, two photon pairs.

This expression is clearly not normalizable because of the infinite sums that come without proper weights for each of the l modes. Therefore one has to include spatial correlations separately. This has been done in an extensive calculation in [103] but is not further relevant here. Most important to our application of shadow tomography is rather the rough form of the state which can still be distilled from equation 2.5 by selecting now only states where 4 (entangled) photons, i.e. two photon pairs, are involved, this reduces the state (when properly normalized) to:

$$|\psi_4\rangle = \sum_{l_1=-\infty}^{\infty}\sum_{l_2=-\infty, l_2\neq l_1}^{\infty}\alpha_{(l_1,l_2)}|1_{l_1}1_{\bar{l}_1}1_{l_2}1_{\bar{l}_2}\rangle + \sum_{l=-\infty}^{\infty}\beta_l|2_l 2_{\bar{l}}\rangle \tag{2.6}$$

Now the two different terms also unveil the origin of two pairs within the state structure. Photon pairs originating from the same position in the PDC crystal by, primarily, stimulated pair production result in an equal state i.e. the second term. Photon pairs created at different points by spontaneous PDC yield generally different states i.e. the first term.

It is important to remark that in consequence of the no-cloning-theorem [114] the state can not be split up into the two processes as clearly as it may seem. Creation of a pair by stimulated pair production will always also mixture in a term with $|1_{l_1}1_{\bar{l}_1}1_{l_2}1_{\bar{l}_2}\rangle$ since the photon pair from stimulated PDC is forbidden to be copied exactly.

Another property of the state in equation 2.6 becomes apparent when one considers the explicit ket-representation: Since $|2_l2_{\bar{l}}\rangle := \frac{1}{\sqrt{(6)}}(|ll\bar{l}\bar{l}\rangle + |l\bar{l}\bar{l}l\rangle + |\bar{l}\bar{l}ll\rangle + |\bar{l}l\bar{l}l\rangle + |l\bar{l}l\bar{l}\rangle + |\bar{l}ll\bar{l}\rangle)$, we see that in each term there is a binary choice between l and $\bar{l}$. This allows us to identify l as *excitation* in a *Dicke state*, a special type of entangled quantum state, with a total of two excitations. We define a typical Dicke state (which includes the well-known W-state) as

$$|D_m^n\rangle = \frac{1}{N}\sum_{\{\alpha\}}|d_{\{\alpha\}}\rangle \tag{2.7}$$

Here the sum is taken over the sets of indices $\{\alpha\}$ which indicate the position of m excitations in a system of size n, $N = \sqrt{\binom{n}{m}}$ and $|d_{\{\alpha\}}\rangle = \bigotimes_{i\notin\{\alpha\}}|0\rangle_i \bigotimes_{i\in\{\alpha\}}|1\rangle_i$ [115]. Dicke states have many advantages in practice, they can not only be easily distilled to different types of entangled states such as the well-known and maximally entangled GHZ states, but are also less susceptible to photon loss. [115] [105]

2.4.2 Generalizing classical shadows to qudits

On the way of applying classical shadow estimation to our high dimensional photonic systems is the need for its generalization to qudits. Fortunately a lot of the concepts from the protocol apply one-to-one, but there are a few points that do not generalize easily.

Measurement Channel

A central point in finding a relevant mapping between measurement outcomes and classical shadow, as well as in quantifying the performance of

classical shadow estimation, was the quantum channel $\mathcal{M}$. Finding an expression for this channel is dependent upon the precise unitary ensemble that is used for sampling unitaries but will here be restricted to the most relevant cases of the generalised Clifford group $\mathcal{C}_d^n$ ($/\mathscr{C}_d^n$) and the generalised Pauli group P_d^n, also to allow comparison with the qubit case in [59]. Given its similarity we will largely follow the lines of derivation for the qubit case given in supplementary section 5 of [59].

As the quantum channel $\mathcal{M}$ represents a measurement, it maps by definition $\rho \mapsto U^\dagger|\tilde{b}\rangle\langle\tilde{b}|U$ with probability $\langle\tilde{b}|U\rho U^\dagger|\tilde{b}\rangle$. So in expectation

$$\mathcal{M}(\rho) = \mathop{\mathbb{E}}_{\tilde{b}\in\{0,\dots,d-1\}^n, U\in\mathcal{E}} U^\dagger|\tilde{b}\rangle\langle\tilde{b}|U = \mathop{\mathbb{E}}_{U\in\mathcal{E}} \sum_{\tilde{b}\in\{0,\dots,d-1\}^n} \langle\tilde{b}|U\rho U^\dagger|\tilde{b}\rangle \cdot U^\dagger|\tilde{b}\rangle\langle\tilde{b}|U \tag{2.8}$$

Evaluating this expression is usually possible for each of the unitary ensembles in consideration, but often the calculation becomes less cumbersome when considering the well-studied complete group of unitaries $\mathcal{U}(d^n)$. For the Clifford group and its local counterpart, the Pauli group, this is simple because they are both so-called *unitary t-designs*.
A unitary t-design is any such ensemble of unitaries $\mathcal{E} = \{\alpha_U, U_i\}$ with probabilities α_U for which [22]:

$$\mathop{\mathbb{E}}_{U\sim\mathcal{E}} (U^{\dagger\otimes t}) X (U^{\otimes t}) = \sum_{U\sim\mathcal{E}} \alpha_U \cdot (U^{\dagger\otimes t}) X (U^{\otimes t}) = \int_{\mathcal{U}(d^n)} d\eta_{Haar} (U^{\dagger\otimes t}) X (U^{\otimes t}) \tag{2.9}$$

where $d\eta_{Haar}$, the *Haar measure*, is a unique and unitary invariant measure, which assigns a uniform (probability) density to the unitary group $\mathcal{U}$. In words we say that randomly sampling according to distribution $\overrightarrow{\alpha}$ from a unitary t-design $\mathcal{E}$ is equivalent to sampling from the complete unitary group $\mathcal{U}(d^n)$ when the so-called t-fold twirling operation is applied. That is the operation under which an operator $X \in \mathcal{L}(\mathcal{H}^{\otimes t})$ is conjugated by a unitary $U^{\otimes t}$ —where U is randomly drawn from some (possibly infinite) ensemble —and an average over all unitaries in the ensemble is taken. [22] [107] More intuitively we can think of this statement as capturing the first k moments over linear operator $X \in \mathcal{L}(\mathcal{H})$ of the Haar measure $d\eta_{Haar}$ on the unitary group $\mathcal{U}(d^n)$. That is, the higher the value of t, the more detailed the unitary ensemble $\mathcal{E}$ imitates the full unitary group $\mathcal{U}(d^n)$ over the Haar measure $d\eta_{Haar}$ and the more $\mathcal{E}$ approximates it. [22] [116]. While almost any uniformly distributed, complete and linearly independent set of unitaries, including the generators of the Pauli group, can form an unitary 1-design, only very few ensembles are known to be 2- or 3-designs.

This concept is useful in application to eq. 2.8 since:

$$\begin{aligned}\mathcal{M}(\rho) &= \mathop{\mathbb{E}}_{U\sim\mathcal{E}} \sum_{\tilde{b}\in\{0,\dots,d-1\}^n} \langle\tilde{b}|U\rho U^\dagger|\tilde{b}\rangle \cdot U^\dagger|\tilde{b}\rangle\langle\tilde{b}|U \\ &= \mathop{\mathbb{E}}_{U\sim\mathcal{E}} \sum_{\tilde{b}\in\{0,\dots,d-1\}^n} U^\dagger|\tilde{b}\rangle\langle\tilde{b}|(U\rho U^\dagger)|\tilde{b}\rangle\langle\tilde{b}|U = \mathop{\mathbb{E}}_{U\sim\mathcal{E}} U^\dagger\mathscr{T}(U\rho U^\dagger)U\end{aligned} \tag{2.10}$$

Here $\mathscr{T}$ can be interpreted as another quantum channel $\mathscr{T} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H})$. The last equality in eq. 2.10 then follows by using an equivalent characteristic of CPTP maps: the Kraus representation. Given a set of m operators $T_1, \dots, T_D \in \mathcal{L}(H)$ such that they are complete —i.e. $\sum_i T_i^\dagger T_i = \mathbb{I}$ —and mutually orthogonal with respect to the Hilbert-Schmidt inner-product, we can represent a quantum channel $\mathscr{T}$ by its action on any operator $R \in \mathcal{L}(\mathcal{H})$: $\mathscr{T}(R) = \sum_i T_i R T_i^\dagger$ [12].
Choosing an quantum channel $\mathscr{T}$ with $T_i = |i\rangle\langle i|$ for $i \in \{0, \dots, d^n - 1\}$ is then valid since $\{|i\rangle\}_{i\in\{0,\dots,d^n-1\}}$ forms an orthonormal basis and thus indeed $\sum_i |i\rangle\langle i||i\rangle\langle i| = \sum_i |i\rangle\langle i| = \mathbb{I}$ and $\mathrm{Tr}\,|i\rangle\langle i||j\rangle\langle j| = \delta_{ij}$ This yields indeed eq. 2.10.

Recognizing the twirling operation of quantum channel $\mathscr{T}$ over the unitary ensemble $\mathcal{E}$ and assuming $\mathcal{E}$ is at least a 2-design, we can write with eq. 2.9:

$$\mathcal{M}(\rho) = \mathop{\mathbb{E}}_{U\sim\mathcal{E}} U^\dagger\mathscr{T}(U\rho U^\dagger)U = \int_{\mathcal{U}(n)} d\eta_{Haar} U^\dagger\mathscr{T}(U\rho U^\dagger)U = \Phi_{dep}(\rho) \tag{2.11}$$

Here the *depolarization channel* Φ_{dep} is originating from noise characterization in quantum information as a very simple model of noise. [8] Essentially it takes the density matrix of a state and mixes the completely mixed state in with relative amplitude η. Thus

$$\Phi_{dep}(\rho) = \eta\rho + (1-\eta)\frac{\mathbb{I}}{d^n}$$

Loss parameter η is then found as [24]

$$\eta = \frac{d^n F + 1}{d^n + 1} := \frac{d^n(\int_{\mathcal{H}} d\psi\langle\psi|\mathscr{T}(|\psi\rangle\langle\psi|)|\psi\rangle) + 1}{d^n + 1} \tag{2.12}$$

with F the average channel fidelity of $\mathscr{T}$ over whole pure state-space $\mathcal{H}$. For integration one usually uses the Haar measure as it assigns equal probability weights to any pair of states that can be transformed into each other

by a unitary transformation. Therefore one often refers to a uniformly drawn pure state as a *Haar-random state*.
The emergence of this channel in eq. 2.11 is a result obtainable from the integral by a direct calculation [117] or diagrammatically based on representation theory [118].

To make the expression for the depolarization channel complete, we evaluate the integral in eq. 2.12:

$$\begin{aligned}\eta &= \frac{d^n(\sum_{i\in\{0,\dots,d-1\}}\int_{\mathcal{H}} d\psi\langle\psi|i\rangle\langle i|(|\psi\rangle\langle\psi|)|i\rangle\langle i|\psi\rangle)+1}{d^n+1}\\ &= \frac{d^n(\int_{\mathcal{H}} d\psi\sum_{i\in\{0,\dots,d-1\}}\langle i|\left(|\langle i|\psi\rangle|^2|\psi\rangle\langle\psi|\right)|i\rangle)+1}{d^n+1}\\ &= \frac{d^n(\int_{\mathcal{H}} d\psi\sum_{i\in\{0,\dots,d-1\}}\langle i|\left(|\langle i|\psi\rangle|^2|\psi\rangle\langle\psi|\right)|i\rangle)+1}{d^n+1}\\ &= \frac{d^n(\int_{\mathcal{H}} d\psi\,\mathrm{Tr}\,|\psi\rangle\langle\psi|)+1}{d^n+1} = \frac{d^n(\int_{\mathcal{H}} d\psi)+1}{d^n+1} = \frac{1}{d^n+1}\end{aligned} \tag{2.13}$$

Which in the qubit ($d=2$) case indeed agrees with [59]. Here we used that $\sum_{i\in\{0,\dots,d-1\}}|\langle i|\psi\rangle|^2 = 1$ and the (spherical) symmetry of pure state-space $\mathcal{H}$ to conclude $\int_{\mathcal{H}} d\psi = 0$.

Visually this channel corresponds to a contraction of the Bloch sphere of single qubits or, equivalently, shrinking the Bloch vector of a state in its norm while keeping its orientation the same. [8]

Finally we obtain:

$$\mathcal{M}(\rho) = \Phi_{dep} = \frac{1}{d^n+1}\rho + (1-\frac{1}{d^n+1})\frac{\mathbb{I}}{d^n} = \frac{1}{(d^n+1)}(\rho+\mathbb{I}) \tag{2.14}$$

Inverting this linear channel is straightforward:

$$\rho = \mathop{\mathbb{E}}_{\tilde{b}\in\{0,\dots,d-1\}^n, U\in\mathcal{E}} \mathcal{M}^{-1}(U^\dagger|\tilde{b}\rangle\langle\tilde{b}|U) = \mathop{\mathbb{E}}_{\tilde{b}\in\{0,\dots,d-1\}^n, U\in\mathcal{E}} (d^n+1)U^\dagger|\tilde{b}\rangle\langle\tilde{b}|U - \mathbb{I}$$

So, as seen before, sampled measurement arrays $U^\dagger|\tilde{b}\rangle\langle\tilde{b}|U$ can be used to construct classical snapshots $\hat{\rho}$ for the shadow $S(\hat{\rho})$, and to reconstruct expectations $\hat{o} = \mathrm{Tr}\,\rho O = \mathop{\mathbb{E}}_{\tilde{b}\in\{0,\dots,d-1\}^n, U\in\mathcal{E}} \mathrm{Tr}\,\hat{\rho}O$, also for qudits.

Restricted by the 2-design property we return to two unitary ensembles

under consideration: the n-qudit Pauli group $\mathcal{P}_n^d$ and Clifford group C_n^d. By using the separability of tensor products in linear channel $\mathcal{M}^{-1}$, we can then write:

$$\hat{\rho} = \mathcal{M}^{-1} = (d^n + 1)U^\dagger|\tilde{b}\rangle\langle\tilde{b}|U - \mathbb{I} \tag{2.15}$$

$$\hat{\rho} = \bigotimes_{i=1}^{n} \mathcal{M}_i^{-1} = \bigotimes_{i=1}^{n} \Big((d+1)U_i^\dagger|\tilde{b}_i\rangle\langle\tilde{b}_i|U_i - \mathbb{I}\Big) \tag{2.16}$$

for local Pauli measurements $\{U_i\}_{i\in\{1,\dots,n\}}$ with outcomes $\{|\tilde{b}_i\rangle\}_{i\in\{1,\dots,n\}}$ and global Clifford measurements (represented by U) with outcomes $|\tilde{b}\rangle$ respectively.

Variance bounds

For evaluating the performance of classical shadows based on sample complexity in qudit systems, we would like to adapt the qubit performance bound from [59] as well.
We recall that the variance for expectation values is given by

$$\begin{aligned} Var(O_i) &= \langle (O_i - \operatorname{Tr}(O_i\rho))^2\rangle = \langle O_i^\dagger O_i\rangle - \langle O_i\rangle(\operatorname{Tr}(O_i\rho))^* - \langle O_i^\dagger\rangle \operatorname{Tr}(O_i\rho) + |\operatorname{Tr}(O_i\rho)|^2 \\ &= \langle O_i^\dagger O_i\rangle - |\operatorname{Tr}(O_i\rho)|^2 - \operatorname{Tr}(O_i^\dagger\rho)\operatorname{Tr}(O_i\rho) + |\operatorname{Tr}(O_i\rho)|^2 \\ &= \langle O_i^\dagger O_i\rangle - \operatorname{Tr}(O_i^\dagger\rho)\operatorname{Tr}(O_i\rho) \end{aligned} \tag{2.17}$$

In contrast to qubits, we need to respect that generalized Pauli operators are not Hermitian. If O_i is a Pauli string, we need to account for this.

Consider the situation of reconstructing from a sampled set of measurement outcomes $U^\dagger|b\rangle\langle b|U$, we can write the first term as

$$\langle O_i^\dagger O_i\rangle = \langle (\operatorname{Tr} O_i\hat{\rho})^*(\operatorname{Tr} O_i\hat{\rho})\rangle_{\hat{\rho}} = \mathop{\mathbb{E}}_{\tilde{b}\in\{0,\dots,d-1\}^n, U\in\mathcal{E}} \left|\operatorname{Tr} O_i\mathcal{M}^{-1}(U^\dagger|b\rangle\langle b|U)\right|^2$$

Because the measurement channel itself is obviously Hermitian, we can use the defining property of Hermitian linear maps with respect to the Hilbert-Schmidt inner-product to write [59]:

$$\operatorname{Tr}\left(O_i\mathcal{M}^{-1}(U^\dagger|b\rangle\langle b|U)\right) = \operatorname{Tr}\left(\mathcal{M}^{-1}(O_i)U^\dagger|b\rangle\langle b|U\right)$$

So then: [59]

$$
\begin{aligned}
Var(O_i) &= \mathop{\mathbb{E}}_{\tilde{b}\in\{0,\dots,d-1\}^n, U\in\mathcal{E}} \left|\operatorname{Tr}\mathcal{M}^{-1}(O_i)U^\dagger|b\rangle\langle b|U\right|^2 - \operatorname{Tr}(O_i^\dagger\rho)\operatorname{Tr}(O_i\rho) \\
&\leq \mathop{\mathbb{E}}_{\tilde{b}\in\{0,\dots,d-1\}^n, U\in\mathcal{E}} \left|\operatorname{Tr}\mathcal{M}^{-1}(O_i)U^\dagger|b\rangle\langle b|U\right|^2 - \operatorname{Tr}(O_i^\dagger\rho)\operatorname{Tr}(O_i\rho) \\
&= \mathop{\mathbb{E}}_{U\in\mathcal{E}} \sum_{\tilde{b}\in\{0,\dots,d-1\}^n} \langle\tilde{b}|U\rho U^\dagger|\tilde{b}\rangle \left|\operatorname{Tr}\mathcal{M}^{-1}(O_i)U^\dagger|b\rangle\langle b|U\right|^2 - \operatorname{Tr}(O_i^\dagger\rho)\operatorname{Tr}(O_i\rho) \\
&= \mathop{\mathbb{E}}_{U\in\mathcal{E}} \sum_{\tilde{b}\in\{0,\dots,d-1\}^n} \langle\tilde{b}|U\rho U^\dagger|\tilde{b}\rangle \left|\langle b|U(\mathcal{M}^{-1}(O_i)U^\dagger|b\rangle\right|^2 - \operatorname{Tr}(O_i^\dagger\rho)\operatorname{Tr}(O_i\rho) \\
\leq \max_{\sigma\in\mathcal{D}(\mathcal{H})} \mathop{\mathbb{E}}_{U\in\mathcal{E}} &\sum_{ildeb\in\{0,\dots,d-1\}^n} \langle\tilde{b}|U\sigma U^\dagger|\tilde{b}\rangle \left|\langle b|U(\mathcal{M}^{-1}(O_i)U^\dagger|b\rangle\right|^2 - \operatorname{Tr}(O_i^\dagger\rho)\operatorname{Tr}(O_i\rho)
\end{aligned}
\tag{2.18}
$$

Considering first the case of Hermitian observables (e.g. qubit Pauli operators), we can safely ignore the constant term $\operatorname{Tr}(O_i\rho)^2$ and make the first term explicit instead. This is because we only want to find an upper bound on the variance. Constant here means invariant with respect to which measurements (represented by U) are chosen and which outcomes $|b\rangle$ those measurements yield respectively. Under these assumptions, eq. 2.18 is exactly the expression from eq. 2.2, the shadow norm $||O_i||_{shadow}$. So writing $Var(O_i) \leq ||O_i||^2_{shadow}$ implies that estimating the variance boils down to finding an explicit expression for the shadow norm. Dependence on the measurement channel $\mathcal{M}$ requires us to do so separately for any unitary ensemble that we choose for the measurements. Bounding the shadow norm is also where qudits start to deviate from qubits.

For global Clifford measurements we refer back to eq. 2.16 and note the equivalence of the shadow norm when replacing O_i by its traceless counterpart $\bar{O}_i = O_i - \operatorname{Tr}(O_i)\dfrac{\mathbb{I}}{d^n}$, as remarked by [59].
This simplifies the expression for the inverse channel on the observable to $\mathcal{M}^{-1}(\bar{O}_i) = (d^n+1)\bar{O}_i$. So we write [59]:

$$
\begin{aligned}
||O_i||^2_{shadow} \equiv ||\bar{O}_i||^2_{shadow} &= \max_{\sigma\in\mathcal{D}(\mathcal{H})} [\mathbb{E}_{U\sim\mathcal{E}} \sum_{b\in\{0,1\}^n} \langle b|U\sigma U^\dagger|b\rangle \left(\langle b|U(d^n+1)\bar{O}_i U^\dagger|b\rangle\right)^2] \\
&= \max_{\sigma\in\mathcal{D}(\mathcal{H})} \operatorname{Tr}\left[\sigma\, \mathbb{E}_{U\sim\mathcal{E}} U^\dagger|b\rangle\langle b|U \left(\langle b|U(d^n+1)\bar{O}_i U^\dagger|b\rangle\right)^2\right]
\end{aligned}
\tag{2.19}
$$

The form of the expression suggests —like in the preceding section —to use t-design properties to simplify the expression further. The problem now arises that the dependence on unitary U is three-fold, i.e. we must consider the three-fold twirl operation. This implies the use of 3-design characteristics. Unfortunately, while the qubit Clifford group is indeed a 3-design, the generalized Clifford group $\mathscr{C}_n^d$ is generally only a 2-design. [22]. In fact, only Clifford groups for subsystem dimension $d = 2^q; q \in \mathbb{Z}^+$ have been identified as 3-designs. [22].

For those values of d, we can conclude a bound easily while we leave the variance for general d as open question. Here we are mainly interested in gaining insight into the changes in known bounds for qubits. It is expected that a thorough examination of expectation over the generalized Clifford group will yield very similar bounds. This is motivated by the fact that classical shadow estimation was not the first protocol of its kind.

Independently of [59], a yet disregarded work by Morris et al [14] introduced the Selective Quantum State Tomography (SQST) protocol. With an similar approach to partial tomography, they have shown that equivalent bounds can be derived when considering measurement and reconstruction in arbitrary mutually unbiased bases instead of the MUBs induced by the Pauli matrices.
On predicting a single observable, also [119] achieved an equal performance guarantee to global Clifford measurements for classical shadows. Although their method is comparable in structure, they used random single qubit rotations instead and a deviate more clearly on post-processing. Both works suggest an straightforward extension of the bounds obtained through the use unitary t-designs.

Completing the examination of the shadow norm for appropriate d, which allow the use of 3-design properties, we get: [59]:

$$\begin{aligned} ||O_i||^2_{shadow} &= \max_{\sigma\in\mathcal{D}(\mathcal{H})} \operatorname{Tr}\left(\sigma \frac{(d^n+1)^2(\operatorname{Tr}(\bar{O}_i{}^2)\mathbb{I} + 2\bar{O}_i{}^2)}{(d^n+2)(d^n+1)}\right) \\ &= \frac{d^n+1}{d^n+2} \max_{\sigma\in\mathcal{D}(\mathcal{H})} \operatorname{Tr}\bar{O}_i{}^2 + 2\operatorname{Tr}\sigma\bar{O}_i{}^2 \leq 3\operatorname{Tr}(\bar{O}_i{}^2) \leq 3\operatorname{Tr}(O_i^2) \end{aligned} \tag{2.20}$$

where the system size dependent prefactor $0 < \dfrac{d^n+1}{d^n+2} \leq 1$ was ignored since it pushes the estimated upper bound only further down. We refer to

[59] and [118] for details on evaluating the integral over $\mathcal{U}(d^n)$ for $t = 3$ instead of $t = 2$, which leads to the above equality.
To conclude on the inequalities, the relation $||\cdot||_\infty \leq \mathrm{Tr}\,(\cdot)^2$ between the standard operator norm $\mathrm{Tr}\,(\cdot)^2$ and the spectral norm $||\cdot||_\infty$ was used. [12] [59] To unify the two terms in the second to last equality, note that $\max_{\sigma\in\mathcal{D}(\mathcal{H})} \mathrm{Tr}\,(\sigma \bar{O}_i{}^2) = ||\bar{O}_i{}^2||_\infty$.
Although this expression matches exactly with [59] for qubits, and has therefore no direct dependence on subsystem dimension d, the d-dependence reappears indirectly in the Hilbert-Schmidt norm of the observable. Often as worse as d^n.

For non-Hermitian observables the calculation becomes not much more complicated. We note that by maximizing over states σ, the second term in eq. 2.18 $\mathrm{Tr}(O_i^\dagger\sigma)\,\mathrm{Tr}(O_i\sigma)$ can be upper-bounded by $|\lambda_{max}(O_i)|^2$ and can therefore be ignored as a constant as well. For the shadow norm we thus, again, use only the first term. Equally we observe that because classical shadows have unit trace by construction, i.e.

$$\mathrm{Tr}(O_i^\dagger\hat{\rho}) - \mathrm{Tr}(O_i\rho)^* = \mathrm{Tr}(\tilde{O}_i{}^\dagger\hat{\rho}) - \mathrm{Tr}(\tilde{O}_i\rho)^*$$

for $\tilde{O}_i^\dagger = O_i^\dagger - \dfrac{\mathrm{Tr}(O_i^\dagger)}{d^n}\mathbb{I}$, the variance of O_i is independent of the trace of $O_i^\dagger$ as well [59]. That is, $||O_i||^2_{shadow} = ||\tilde{O}_i||^2_{shadow}$, also in the non-Hermitian case.

We again proceed by assuming 3-design compatible values of d and by applying eq. 12 from [118] on eq. 2.19 now explicitly:

$$\begin{aligned}
||O_i||^2_{shadow} &= \max_{\sigma\in\mathcal{D}(\mathcal{H})} \mathrm{Tr}\,(\sigma\frac{(d^n+1)^2(\mathrm{Tr}\,(\bar{O}_i\bar{O}_i{}^\dagger)\mathbb{I} + \bar{O}_i\bar{O}_i{}^\dagger + \bar{O}_i{}^\dagger\bar{O}_i)}{(d^n+2)(d^n+1)}) \\
&= \frac{d^n+1}{d^n+2}\max_{\sigma\in\mathcal{D}(\mathcal{H})} \mathrm{Tr}\,(\bar{O}_i\bar{O}_i{}^\dagger) + \mathrm{Tr}\,(\sigma\bar{O}_i\bar{O}_i{}^\dagger) + \mathrm{Tr}\,(\sigma\bar{O}_i{}^\dagger\bar{O}_i) \qquad (2.21) \\
&\leq 3\,\mathrm{Tr}\,(\bar{O}_i\bar{O}_i{}^\dagger) \leq 3\,\mathrm{Tr}\,(\bar{O}_i\bar{O}_i{}^\dagger) \qquad (2.22)
\end{aligned}$$

since $O_i^\dagger O_i = O_i O_i^\dagger$ is always Hermitian. Further the same arguments as for eq. 2.19 apply.

For local Pauli measurements, we reconsider the shadow norm eq. 2.2 with the corresponding local inverse channel from eq. 2.16. For Hermitian

observables O_i this means:

$$||O_i||^2_{shadow} = \max_{\sigma\in\mathcal{D}(\mathcal{H})} \mathbb{E}_{U_i\sim\mathcal{E}} [\left(\bigotimes_{j=1}^{n} \sum_{b\in\{0,\dots,d-1\}} \langle b_j|U_j \right) \sigma \left(\bigotimes_{j=1}^{n} \sum_{b\in\{0,\dots,d-1\}} U_j^\dagger|b_j\rangle \right) \cdot \left(\left(\bigotimes_{j=1}^{n} \sum_{b\in\{0,\dots,d-1\}} \langle b_j|U_j \right) \Phi_{dep}^{\otimes n}(O_i) \left(\bigotimes_{j=1}^{n} \sum_{b\in\{0,\dots,d-1\}} \sum_{b\in\{0,\dots,d-1\}} U_j^\dagger|b_j\rangle \right) \right)^2] \tag{2.23}$$

The problem changes only slightly compared to global Clifford measurements since the form of the shadow norm is similar. Through the locality of measurement unitaries $\{U_j\}_{j\in\{1,\dots,n\}}$ and outcomes $\{|\tilde{b}_j\rangle\}_{j\in\{1,\dots,n\}}$ we, however, need to distinguish between the most general case of global observables O_i that operate on at most w qudits and w-qudit tensor products of local observables $O_i = \bigotimes_{j=1}^{n} O_i^{(j)}$. [59]
The latter case means $O_i = (\bigotimes_{j=1}^{w} O_i^{(j)}) \otimes \mathbb{I}^{\otimes(n-w)}$ and makes the shadow norm separable. Moreover, from eq. 2.23 and $\Phi_{dep}(\mathbb{I}) = \mathbb{I}$, the equivalence of the shadow norm for the n-qubit observable O_i and the w-qubit non-trivial part follows: $||O_i||^2_{shadow} = ||(\bigotimes_{j=1}^{w} O_i^{(j)})||^2_{shadow}$
Together this yields:

$$\begin{aligned} ||O_i||^2_{shadow} = \max_{\sigma\in\mathcal{D}(\mathcal{H})} \mathrm{Tr}[\sigma \bigotimes_{j=1}^{w} \mathbb{E}_{U_j\sim\mathcal{E}} \sum_{b_j\in\{0,\dots,d-1\}} U_j^\dagger|b_j\rangle\langle b_j|U_j \\ \cdot \left(\langle b_j|U_j \left((d+1)O_i^{(j)} - \mathbb{I} \right) U_j^\dagger|b_j\rangle \right)^2] \end{aligned} \tag{2.24}$$

Also in this case we encounter the problem of a three-fold dependence on the local unitaries U_j and therefore a requirement for the unitary ensemble $\mathcal{E}$ to be a 3-design. Neither for the generalized single-qudit Clifford group nor for the generalized Pauli group this is the case unless the dimension d is a power of two. [22] For these subsystem dimensions we proceed as before by using the explicit form for the expectation over the single-qudit

Clifford group $\mathscr{C}_1^d$ from [59] and [118]. We find:

$$
\begin{aligned}
||O||^2_{shadow} &= \max_{\sigma\in\mathcal{D}(\mathcal{H})} \mathrm{Tr}\left[\sigma \bigotimes_{j=1}^{w} (d+1)^2 \frac{\mathrm{Tr}\,(O^{(j)})^2\mathbb{I} + 2(O^{(j)})^2}{(d+2)(d+1)}\right] \\
&= \left(\frac{(d+1)}{(d+2)}\right)^w \max_{\sigma\in\mathcal{D}(\mathcal{H})} \mathrm{Tr}\left[\sigma \bigotimes_{j=1}^{w} \mathrm{Tr}\,((O^{(j)})^2)\mathbb{I} + 2(O^{(j)})^2\right] \\
&= \left(\frac{(d+1)}{(d+2)}\right)^w \max_{\sigma\in\mathcal{D}(\mathcal{H})} \left[\Pi_{j=1}^{w} \mathrm{Tr}\,((O^{(j)})^2)) + 2\,\mathrm{Tr}\,(\sigma \bigotimes_{j=1}^{w} (O^{(j)})^2)\right] \\
&= \left(\frac{(d+1)}{(d+2)}\right)^w \max_{\sigma\in\mathcal{D}(\mathcal{H})} \left[\Pi_{j=1}^{w} \mathrm{Tr}\,((O^{(i)})^2)) + 2\,\mathrm{Tr}\,(\sigma \bigotimes_{j=1}^{w} (O^{(j)})^2)\right] \\
&\leq \left(\frac{(d+1)}{(d+2)}\right)^w \Pi_{j=1}^{w} (\mathrm{Tr}\,((O^{(j)})^2)) + 2||(O^{(j)})^2||_\infty) \leq (d+1)^w \cdot ||\tilde{O}||^2_\infty
\end{aligned}
\tag{2.25}
$$

Additionally to the factorization of the spectral norm under tensor products $||P \bigotimes Q||_\infty \leq ||P||_\infty \cdot ||Q||_\infty$ [12], we used the same identities as for eq. 2.20. This is indeed consistent with the qubit bound $(d+1)^w \cdot ||\tilde{O}_i||^2_\infty$ for $\{O^{(j)}\}_{j=1,\ldots,w}$ Pauli operators. Note that the actual position of the w non-identity Pauli operators is irrelevant to the shadow norm.

If the observable O_i is global, it can nevertheless always be expanded into a sum of tensor products of observables on individual qubits $\tilde{O}_k$. So $O_i = \sum_k c_k \tilde{O}_k$. The expression in eq. 2.23 is then not separable and the proof of a bound on the shadow norm becomes somewhat more involved. Since the derivation deviates only on few points for qudits (particularly in eq. S59 in [59]), we leave the proof to section S5C of [59]. Note that because the Pauli group is isomorphic to tensor products of the single-qudit Clifford group, the arguments are solely based on the assumption that the local observables are represented by Pauli operators.
Qubits yield a scaling of the shadow norm by $4^w \cdot ||\tilde{O}_i||^2_\infty$ [59], where the factor 4^w originates from counting the number of non-trivial Pauli strings of length w. In the proof this follows from a binomial series $\sum_{q \triangleright s} 3^{|q|} = 3^w \sum_{i=0}^{k} \binom{k}{i} \left(\frac{1}{3}\right)^i = 4^w$ for fixed measurement Pauli string $s \in \{X, Y, Z\}^w$, $|q|$ the weight of observable Pauli string q and $\triangleright$ the condition of s to hit q. (see section 2.3). So in general, for a cardinality $d^2 - 1$ of $\mathcal{W}_d^n \backslash \mathbb{I}$, we can

replace $(3+1)^w = 4^w$ by $(d^2-1+1)^w = (d^2)^w$. It is important to remark the reappearance of dependence on the 3-design property in the proof. A derivation independent of this property has been given by [35] for qubits, but the same generalization to qudits holds there as well. It is assumed, however, that local measurements are solely taken in the Pauli basis. This is also the assumption we took in the numerical experiments due to the same current restriction of derandomization to Pauli measurements.[36]

For non-Hermitian local and global observables $O_i = \tilde{O}_i \otimes \mathbb{I}^(n-w)$, all steps follow equivalently and yield a bound $||O||^2_{shadow} \leq (d+1)^w \cdot ||\tilde{O}_i||^2_\infty$ and $||O_i||^2_{shadow} \leq (d)^w \cdot \mathrm{Tr}\,(\tilde{O}_i \tilde{O}_i{}^\dagger)$ respectively instead.

Finally to find the variance and sample complexity over M observables, instead of a single observable, we can apply a union bound. That is [8]

$$Pr[\bigcup_{m=1}^{M} (\hat{o} - \mathrm{Tr}\,(O_i\rho)| \geq \epsilon)] \leq \sum_{m=1}^{M} Pr[\hat{o} - \mathrm{Tr}\,(O_i\rho)| \geq \epsilon] \leq M\frac{Var(O_i)}{\epsilon^2\delta} \propto N \tag{2.26}$$

So the probability to fail in collective prediction of M observables to accuracy ϵ is upper bounded by the sum of individual probabilities.
Here the probability over M observables can be interpreted as the probability that the maximum deviation over all observables is exceeding the accuracy ϵ. The second inequality is a simple Chebyshev bound with δ a confidence parameter. [8]. By using medians of means prediction, rather than an arithmetic average, the linear scaling in M can eventually be reduced to a $log(M)$ dependence and the δ-dependence be removed. [59]

Putting everything together we find that classical shadow estimation with global Clifford measurements enables us to predict M observables $O_i = \left(\otimes_{j=1}^{w} O_i^{(j)}\right) \otimes \mathbb{I}_d^{\otimes(n-w)}$ with an accuracy of at least ϵ using

$$N \geq log(M) \cdot 3 \cdot \max_i \frac{\mathrm{Tr}\,(O_i^\dagger O_i)}{\epsilon^2} = log(M) \cdot 3 \cdot \max_i \frac{\mathrm{Tr}\,(O_i^2)}{\epsilon^2}$$

copies of quantum state ρ for qudits, like for qubits. The latter equality obviously only holds for Hermitian observables. Equally for local Pauli measurements we obtain a sample complexity of

$$N \geq log(M)\frac{(d)^w}{\epsilon^2} \cdot \prod_{j=1}^{w} \mathrm{Tr}\left((O_i^{(j)})^\dagger O_i^{(j)}\right)$$

for observables that operate on at most w qudits. Hermitian observables $O_i = \tilde{O}_i \otimes \mathbb{I}^{(}n - w)$ can be further upper-bounded by and

$$N \geq log(M)\frac{(d^2)^w}{\epsilon^2} \cdot ||\tilde{O}_i||_\infty^2$$

respectively. If the observable can be decomposed into tensor products of single-qudit observables, the scaling in subsystem dimension d is reduced significantly by a sole indirect dependence in the maximal Hilbert-Schmidt norm of the local observables. Also note that an upper-bound implies only worst-case scaling.

Chapter 3

Numerical experiments

Based on the collected insights from shadow tomography, numerical experiments were performed to illustrate the practical advantage of classical shadow estimation over quantum state tomography. Subsequently for experimental relevant situations, such as spatially entangled 4-photon systems described in the last chapter, we assess performance under experimentally feasible conditions. First reproduction of some basic but key results from earlier studies are demonstrated.

3.1 Framework

A central objective in designing the code structure were flexibility and re usability. Hardcoding the targeted systems was prevented as much as possible while classical computational resources were equally addressed. As such, a modular program was constructed using the Julia language [120] to yield highest possible performance in compromise with readability and compactness of code. Performance is important because of the exponential amount of information about the true quantum state that needs to be processed and stored when simulating a measurement protocol. After measurements are performed, the computational complexity reduces to actual specifications of the post-processing step in the specific protocol. Note that under experimental conditions only the latter performance specification is relevant since only the design of a measurement scheme and the post-processing of measurement outcomes are required. For numerical comparison with an experiment, one can adjust the code to work with efficient representations that fit to possibly known structures of the states in consideration. We refer to [121] for the open-source code, basic usage

instructions and example experiments for our implementation.
The framework was built from, and designed around, the minimal open source Python implementation of Huang et al [122]. This includes the randomized and derandomized sampling procedures for local Pauli measurements as proposed in [36], as well as the shadow prediction method for local observables from [36].

We remark that during finalization of this thesis, a new complete implementation for the Qiskit-based Python framework Pennylane [123] was published. It implements the general and original protocol in [59] such as it was presented in this thesis. However, it does not include a generalization to qudit systems.

Before we describe general methods in the different steps of the numerical framework, we start by introducing a useful formalism which was mainly used for the benchmarking simulations.

3.1.1 Stabilizer formalism

One tool that we have been focussing on to speed up performance of the numerical framework is the *stabilizer formalism* as first introduced by Gottesman [20] in 1997 in the context of quantum error correcting codes [124] [125] [8]. The formalism allows to simulate a particular subset of quantum states, known as *stabilizer states*, and their evolution efficiently on a classical computer. That is instead of the usual exponential computational cost of simulating quantum circuits, the cost is only polynomial in system size. This result is known as the *Gottesman-Knill theorem* [124] [26]. Importantly it restricts evolution to unitary evolution by elements in the Clifford group from section 1.1.4, measurements in the computational basis and Clifford operations conditioned on measurement outcomes. Thus any circuit which consists only of Hadamard, Phase and CNOT gates, as well as measurements, can be simulated efficiently on classical devices.

Although this set of operations, generating the Clifford group, is incomplete in perspective of universal quantum computation, some fundamental quantum algorithms in quantum information processing rely only on Clifford operations. For example quantum teleportation [126] and superdense coding [8]. Also interesting states like the GHZ-state and cluster states can be prepared and studied. Other well-known quantum algorithms like Shor's algorithm, Grover's algorithm or the Deutsch-Jozsa al-

gorithm require additionally universal quantum gates like the Toffoli gate, also known as the 3-qubit CCNOT gate. Adding only a single such universal gate to the generators of the Clifford group allows us to approximate any operation on universal quantum devices. Only circuits with few non-Clifford gates can be approximated in classical simulations. [127] Nevertheless, our choice to restrict some numerical demonstrations to the stabilizer formalism is mainly motivated by the capabilities to simulate data acquisition protocols on classical devices for high dimensional state-spaces.

A *stabilizer* of a state $|\psi\rangle$ is any such unitary operator $X \in \mathcal{L}(\mathcal{H})$ for which $X|\psi\rangle = |\psi\rangle$. It is easy to see that those operators form a group $Stab(|\psi\rangle)$ since the identity operator I, the product XX': $(XX')|\psi\rangle = X(X'|\psi\rangle) = X|\psi\rangle = |\psi\rangle$ and the inverse $X^{-1} = X^\dagger$: $X^{-1}|\psi\rangle = X^\dagger|\psi\rangle = X^\dagger X|\psi\rangle = |\psi\rangle$ stabilize any state $|\psi\rangle$ for any $X, X' \in Stab(|\psi\rangle)$. Representing the state $|\psi\rangle$ uniquely with a generating set of such a finite group instead of a collection of complex amplitudes would benefit from the rich field of group theory and forms the basis of the stabilizer formalism. Unfortunately, the generating set of a group $Stab(|\psi\rangle)$ for an arbitrary quantum state $|\psi\rangle$ still turns out to scale exponentially in system size and would thus change notation but not processing cost on a classical device. When restricting ourselves to highly structured set of states, we can however consider a considerably smaller subgroup $S(|\psi\rangle) = Stab(|\psi\rangle) \cap \mathcal{P}^n$ [26] i.e. the intersection with the n-qubit Pauli group $\mathcal{P}^n$ as introduced in section 1.1.4.
The state $|\psi\rangle$ can then be represented by only 2^n stabilizers that form an abelian group generated by $O(n)$ stabilizers. [26] The latter follows since any group with cardinality $|G|$ is guaranteed to be representable uniquely by $log_2(|G|)$ generating elements [8], any other stabilizer arises from the product between a pair of generators. This stabilizing group must be abelian since only commuting operators have common eigenvectors and therefore have a non-empty intersection in Hilbert space $\mathcal{H}$ that corresponds to the stabilized state $|\psi\rangle$. Henceforth we will refer with *stabilizer* to elements in the generating set.

This then indeed reduces the cost of representing a state with an exponential number of density matrix elements to a linear scaling number of generators for the stabilizer group. Also simulating the evolution of a state on a classical machine becomes as easy as keeping track of the corresponding stabilizers of the state. Unitary evolution by U, for example, simply involves updating $X \mapsto UXU^\dagger$ since $U|\psi\rangle = UX|\psi\rangle = (UXU^\dagger)U|\psi\rangle$ so $UXU^\dagger$ stabilizes $U|\psi\rangle$.

The formalism was improved by Aaronson and Gottesman [26] in 2004 by also keeping a record of the *de*stabilizers of a state. Destabilizers are mutually commuting Pauli strings that complete the set of stabilizers to form together a generating set of the $\mathcal{P}^n$. Every destabilizer has a corresponding anticommuting stabilizer, while it commutes with all other stabilizers. Adding information about destabilizers improves computation efficiency of important subroutines by simplifying them. This is relevant since although unitary evolution is straightforward to implement, other operations like measurements are not. Adding destabilizers improves the runtime of measurements from $O(n^3)$ to $O(n^2)$.[26] The required storage is obviously increased but still only grows linearly in system size.

Although the information one can yield from the stabilizer formalism is restricted in more general cases, the ability of capturing even highly entangled states ,like the maximally entangled GHZ state (see figure 1.2), is remarkable. In consequence of the ability to classically simulate such systems, however, somewhat weakens the common conception of entanglement as main ingredient of quantum advantage. One namely identifies the additional assumption for correct exploitation of entanglement in order to actually outperform classical devices. [126] This aspect is also reflected in the kind of states that fall outside the range of stabilizer states, these are generally known as *magic states* but are actually far more diverse in the way they differ from stabilizer states. See [128] for an extensive discussion.

To translate the concept of stabilizers to an implementable algorithm, we exploit the defining normalizing property of the Clifford group $\mathcal{C}^n$ with respect to the Pauli group $\mathcal{P}^n$ as we introduced in equation 1.3 of section 1.1.4. Applying any Clifford unitary to a tensor product of n Paulis P with a certain phase $e^{i\pi k}$ yields another Pauli string P' with some other phase $e^{i\pi k'}$, so $e^{i\pi k} \cdot CPC^\dagger = e^{i\pi k'} P'$. With the way how stabilizers transform we can thus see that we consistently stay within the chosen restricted set of stabilizers from the Pauli group if we apply only Clifford operations. Here we follow the definition of [26] of a stabilizer state as any such state that can be obtained from applying Cliffords to the vacuum state $|0\rangle^{\otimes n}$. The stabilizer group of the vacuum state is trivially generated by $\{Z_1, .., Z_n\}$ with Z_i the identity on all qubits except a Pauli $Z = \sigma_z$ on qubit i. Applying Hadamard, phase and CNOT gates with U locally then transforms the generators to the Pauli operators which stabilize the new state $U|\psi\rangle$. Bookkeeping of the exact Clifford operations allows for yet another representation of the state without keeping track of the complex amplitudes in

the computational basis.

The key idea is now that unitaries can be uniquely defined by their action on Pauli strings since for n qubits [23]:

$$\bar{X}_j := UX_jU^\dagger = (e^{i\pi})^{r_j} \bigotimes_{i=1}^{n} X_i^{\alpha_{ji}} Z_i^{\beta_{ji}} \tag{3.1}$$
$$\bar{Z}_j := UZ_jU^\dagger = (e^{i\pi})^{s_j} \bigotimes_{i=1}^{n} X_i^{\gamma_{ji}} Z_i^{\delta_{ji}}$$

for $1 \leq j \leq -n$, bit-strings $r, s \in \mathbb{Z}_2^n$ and matrices $\alpha, \beta, \gamma, \delta \in \mathbb{Z}_2^{n\times n}$. Fixing these parameters defines then the Clifford U, including the corresponding stabilizer state, uniquely: Let $U, U' \in \mathcal{C}^n$ such that for fixed Pauli operator $P = X^x Z^z \in \mathcal{P}^n$ we have $UPU^\dagger = U'PU'^\dagger$. Then $U = U'$ since $U'^\dagger UP = PU'^\dagger U$ forces $U'^\dagger U = \mathbb{I}$ up to a global phase. The latter follows because only the identity commutes with all Pauli operators P. [28]
This description includes also the action on the Pauli-Y operator since $\sigma_x\sigma_z = -i\sigma_y = -iY$ which gives with the above relations:

$$\begin{aligned} UY^{(j)}U^\dagger &= -i(UX^{(j)}Z^{(j)}U^\dagger) = -i(UX^{(j)}U^\dagger UZ^{(j)}U^\dagger) \\ &= -i((e^{i\pi})^{r_j} \bigotimes_{i=1}^{n} X_i^{\alpha_{ji}} Z_i^{\beta_{ji}})(e^{i\pi})^{s_j} \bigotimes_{k=1}^{n} X_k^{\gamma_{jk}} Z_k^{\delta_{jk}}) \end{aligned}$$

Equally the action of U on any other Pauli string can be found this way from the logical operators $\bar{X}_j$ and $\bar{Z}_j$. These logical operators can both encode the action of a unitary, and therefore the unitary itself, but also the state that is stabilized by $\bar{Z}_j$ for $1 \leq j \leq n$ since those Pauli strings corresponded to the vacuum state before the unitary was applied.

The choice of parameters $(r, s, \alpha, \beta, \gamma)$ can, however, not be arbitrary since mapping from a set of stabilizers to another set of stabilizers needs to preserve commutation relations. Otherwise the resulting stabilizers would not form a stabilizer group anymore. Preserving commutation relations is equivalent to preserving a particular bilinear form called the symplectic inner product. The tuples of vectors $(\overrightarrow{\alpha_i}, \overrightarrow{\beta_i}$ and $(\overrightarrow{\gamma_i}, \overrightarrow{\delta_i})$, from the right hand side of equation 3.1, each just define a complete n-qubit Pauli string. The symplectic inner product $\bigodot$ of these two Pauli strings then just tells us whether they commute or anticommute:

$$(\overrightarrow{\alpha_i}, \overrightarrow{\beta_i}) \bigodot (\overrightarrow{\gamma_j}, \overrightarrow{\delta_j}) = \langle \overrightarrow{\alpha_i}, \overrightarrow{\delta_j} \rangle \bigoplus \langle \overrightarrow{\beta_i}, \overrightarrow{\gamma_j} \rangle = \begin{cases} 1, & \text{iff } (\alpha_i, \beta_i) \text{ commute} \\ 0, & \text{iff } (\overrightarrow{\alpha_i}, \overrightarrow{\beta_i}) \text{ anticommute} \end{cases}$$

with $\bigoplus$ modulo 2 addition.

Fortunately compact representations for the parameter set $\{\alpha, \beta, \gamma, \delta\}$ exist which preserve symplectic inner products between stabilizers. It is known as the Symplectic group $Sp(2n, \mathbb{Z}_2^{2n})$ of size $2n$ over the finite field $\mathbb{F} = \mathbb{Z}_2^{2n}$, and contains block matrices of the form

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1n} & z_{11} & \dots & z_{1n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{2n1} & \dots & x_{2nn} & z_{2n1} & \dots & z_{2nn} \end{pmatrix} \in \mathbb{Z}_2^{2n \times 2n} \tag{3.2}$$

with $x_{ij}, z_{ij} \in \{0, 1\}$ such that for Pauli string in row i:$P_i = \bigotimes_{k=1}^{n} X_k^{x_{ik}} Z_i^{z_{ik}}$. [26] [23] Elements, like eq. 3.2, of the symplectic group must be a *symplectic matrix S* satisfying:

$$S\Omega S^T = \Omega; \text{for } \Omega = \bigoplus_{i=1}^{n} \begin{pmatrix} 0 & \mathbb{I} \\ \mathbb{I} & 0 \end{pmatrix} \tag{3.3}$$

where Ω is the matrix representation of the symplectic inner-product. [23] The phase parameters r and s are not constrained and therefore stored separately in a single phase vector $\begin{pmatrix} \overrightarrow{r} \\ \overrightarrow{s} \end{pmatrix} \in \mathbb{Z}_2^{2n}$.
When we use symplectic matrices to encode unitaries by their action or stabilizer states through their generating unitaries, we call them *tableau* (or check matrix). It is easy to see that the initial state, the vacuum state, would correspond to the $(2n \times 2n)$-identity matrix. Further we see from the definition of $(\alpha, \beta, \gamma, \delta)$ and eq. 3.2 that the first n rows in a tableau correspond to logical operators $\bar{X}_j$ and therefore the destabilizers, while the last n rows define the logical operators $\bar{Z}_j$ corresponding here directly to the stabilizers.
Beware that in our implementation we define a tableau as the transpose of eq. 3.2 because of the unusual order of dimensions for arrays in the Julia language.

Note that also mixed states can be represented by stabilizer states through purification: Any mixed state can be represented by a non-unique pure state on a larger system. Obtaining such a purification in form of a tableau for a given mixed state, simulating the desired stabilizer circuit on the larger system and tracing out the previously added ancillas, results in a state that is completely equivalent to the state one would have obtained by direct operation of the circuit on the mixed state. [26]

The quadratic, instead of exponential, growth of the tableau size with number of qubits n appears also here beneficial to the density matrix. In fact, all operations can be implemented via some operation on this $(2n \times 2n)$-bit array and the phase vector. For the tableau, for example, Clifford gates can be easily applied by elementary column operations. While the Hadamard gate interchanges the X and Z tableau values for a single qubit a, the phase gate adds the column of values for X to the column vector of Z of the same qubit. [129] As the exact routines of the individual operations are quite detailed and neither very insightful, nor of further interest here, we refer for instructions and derivations to [26] Fortunately the framework as proposed by Aaronson and Gottesman has been extensively studied in the literature and implementations for various languages readily existed. For qubits we combined two incomplete Julia frameworks, Juqst.jl [130] and QuantumCliffords.jl [131] to implement tableaus and basic operations on them.

Measurements

Measurements, however, are a special case since they generalize non-trivially to higher subsystem dimensions d. Therefore we will give a brief overview of their implementation as is described in [124] [26]
We consider within this framework solely projective measurements on pure states. Also we have performed measurements qubit- and qudit-wise. By simulating global or local Clifford unitaries on the tableau that corresponds to the original state $|\phi\rangle$, we obtain the state tableau in the computational basis $|\psi\rangle = U|\phi\rangle$. The following procedure is repeated over each qubit a of the system by focussing on the Pauli operators given by columns a and $a + n$ (eq. 3.2), but beware that any operation is applied to the complete tableau:

For qubits, computational basis measurements are based on whether the stabilizers $S(|\psi\rangle)$ of state $|\psi\rangle$ commute or anticommute with the measurement operator. If we have a set of measurement operators $\{M_i\}_i$, with outcome i labelling the corresponding eigenvalue λ_i, the measurement coincides with the projection $P_\pm = (\mathbb{I} + \lambda_i M_i)$. In the computational, i.e. Pauli-Z, basis, we thus apply the projection $P = (\mathbb{I} \pm Z)$ depending on the measurement outcome ± 1.

If there exists a stabilizer V that anticommutes with the measurement operator $M_i = \pm Z$, $\{M_i, V\} = 0$, the measurement is incompatible with the current state. When there are multiple options for choosing V, we define

V to be the first incompatible stabilizer in the tableau. In this case the measurement outcome is random. By the fact that Y, Y, Z represent three different MUBs, we know that all stabilizers V —which correspond to a different eigenbasis than the measurement operator —must have eigenvectors that are all equidistant to all eigenspaces of M_i. The probability to project with P into the eigenspace of outcome i is thus the same for all i. Therefore we sample the outcome uniformly at random.

Next the post-measurement state must be obtained in preparation for the measurement on the next qubit by simulating the collapse of the state according to measurement M_i with outcome i. This is called *post-selection*. The essential task is to find (de)stabilizers that are compatible with measurement operator M_i such that when repeating the measurement on the same qubit, no stabilizer V can be found. This is done by overwriting the destabilizer corresponding to V by V and replacing the stabilizer V itself by the measurement operator M_i. Additionally we post-select on the specific outcome (phase) by storing the outcome as the phase of the new stabilizer M_i in the phase vector of the stabilizer state $|\psi\rangle$. Finally, by considering all other stabilizers and destabilizers W_j that anticommute with M_i, we can make them compatible as well by replacing them one-by-one with the product of Pauli operators VW_i. Since $[M_i, (VW_i)] = (-VM_i)W_i - (V(-M_iW_i) = \mathbf{0}$. Returned are the outcome and post-selected state.

If no incompatible stabilizer V can be found in the first step, then the state is already its own projection on the measurement basis and the outcome is deterministic. We thus only need to read out the outcome i. Here it will be useful to consider the full tableau, rather than only the columns of qubit a. A superscript (i) will denote the operator which acts trivially on all qubits but i.
Because the stabilizers form a generating set for the stabilizer group $S(|\psi\rangle)$ and all of them commute with the measurement operator $Z^{(a)}$, we can always write $Z^{(a)}$ as a product of all stabilizers. Translated to the symplectic tableau, this corresponds to the sum

$$\sum_{q=1}^{n} C_q \overrightarrow{S_{(n+q)}} = (e^{i\pi})^h \overrightarrow{Z}^{(a)}$$

for $C_q \in [0, 1]$, a phase $h \in \{0, 1\}$, $\overrightarrow{Z}^{(a)}$ the symplectic row vector of $Z^{(a)}$ and $\overrightarrow{S_{(n+q)}}$ the $(n+q)$th row vector in the tableau.

Starting from the measurement operator with phase $h = 0$, corresponding to the outcome $+1$, $M = Z^{(a)}$, we want to know which phase h the stabilizers force on this Pauli operator. Thus which $M_i = \pm Z$ follows. The coefficients C_q can be obtained from recalling one of the defining properties of destabilizers, namely that each destabilizer anticommutes with its stabilizer but commutes with all other generating stabilizers in $S(|\psi\rangle)$. As indicated in the preliminaries, we consider a symplectic inner-product of 0 to represent commutation between Pauli operators.
So then $\overrightarrow{S_{(q')}} \bigodot \overrightarrow{S_{(n+q)}} = 1 - \delta_{q'q}$ and therefore

$$\overrightarrow{S_r} \bigodot \left(\sum_{q=1}^{n} C_q \overrightarrow{S_{(n+q)}} \right) = \sum_{q=1}^{n} C_q \left(\overrightarrow{S_r} \bigodot \overrightarrow{S_{(n+q)}} \right) = C_q = \left(\overrightarrow{S_r} \bigodot \overrightarrow{Z}^{(a)} \right) \in \{0, 1\}$$

This reduces the task to multiplying all destabilizers $P(\overrightarrow{S_r})$ in the tableau for which $C_q \neq 0$, including their phases from the separate phase vector. The phase corresponding to the product then agrees with the phase h of $Z^{(a)}$ and therefore defines the measurement outcome.
As the measurement was deterministic, the state stays unaffected and no post-selection is required. We finally return the measurement outcome and state for the next measurement.

Generalizing to qudits

For qudits the reasoning from the qubit case can be applied completely in analogue but using the generalized Pauli group $\mathcal{P}_d^n$ and generalized Clifford group $\mathcal{P}_d^n$ from section 1.1.4. However, not only the formalism itself generalizes easily, also the efficient simulation of qudit stabilizer systems (i.e. the Gottesman-Knill theorem) was shown to be generalizable. [29]

The starting point is again to identify stabilizers from $Stab(|\psi\rangle)$, for a state $|\psi\rangle$ of any system with n d-dimensional subsystems, which are contained in the intersection with the Weyl-Heiserberg group $\mathcal{W}_d^n$ and form a maximal abelian subgroup of the latter. We defined a Pauli operator from the Weyl-Heisenberg group by a vector $\overrightarrow{a} \in \mathbb{Z}_d^{2n}$ and a phase $\tilde{\omega}^{\alpha}$ with $\alpha \in \mathbb{Z}_{2d}$. We exploit the normalization property to write [23] [29]:

$$\begin{aligned} \bar{X}_j &:= U X_j U^\dagger = (\tilde{\omega})^{r_j} \bigotimes_{i=1}^{n} X_i^{\alpha_{ji}} Z_i^{\beta_{ji}} \\ \bar{Z}_j &:= U Z_j U^\dagger = (\tilde{\omega})^{s_j} \bigotimes_{i=1}^{n} X_i^{\gamma_{ji}} Z_i^{\delta_{ji}} \end{aligned} \tag{3.4}$$

As we did for eq. 3.1 as well.

Also here the parameters completely characterize the unitary U with the only difference that now $\overrightarrow{r}$, $\overrightarrow{s} \in \mathbb{Z}_d^{2n}$ and $\alpha, \beta, \gamma, \delta \in \mathbb{Z}_{2d}^{2n \times 2n}$.
Finally we obtain again, by enforcing invariance of the symplectic inner product between Pauli strings, the symplectic matrix ("Tableau") $T = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ and phase vector $\overrightarrow{h} = \left| \begin{matrix} \overrightarrow{r} \\ \overrightarrow{s} \end{matrix} \right|$.

In this general context, it follows also immediately from the fact that the symplectic group is isomorphic with $Sym(2n, \mathbb{Z}_d^n) \cong \mathcal{C}_d^n / \mathcal{P}_d^n$. [30] [23]
This means we get a representation of d-independent size, but with dit- instead of bit-valued entries. Also we need to keep track of the phase base $\tilde{\omega}$ for our specific value of d. Useful operations between Cliffords $U(T, h)$ and Pauli operators $P(\overrightarrow{a}, \alpha)$ are easily generalized from combining the commutation relations for Pauli operators with the symplectic description of actions for Cliffords, as done in [28]. As yet no framework for simulation of generalized Clifford circuits existed for Julia, the qudit implementation of the stabilizer formalsim had to be written from ground up. This involved for a standardized storage and syntax also the introduction of two structures for Clifford/State tableaus and Pauli operators respectively. To ensure consistency and compatibility with the qubit case, we followed [28] for basic Tableau-based operations and generalized the measurement procedure from [26], as we now describe briefly.

For measurements on qudit systems, we have to respect more complex commutation relations, different projectors on the computational basis and tableaus (phase vectors) with values modulo d ($2d$). Like before, we can use the toolset from the stabilizer formalism to rotate a state $|\phi\rangle$ to the computation bases $|\psi\rangle = U|\phi\rangle$. From the d eigenvalues, we have d possible measurement outcomes with measurement operators M_x and corresponding single-qudit projectors $P = |x\rangle\langle x| = \mathbb{I} + M_x$. However, by the order of generalized Pauli operator Z, $M_x \neq Z$. Rather we have to obtain M_x separately. We recall that $Z_d|j\rangle = \omega^j|j\rangle$. So $|x\rangle\langle x| = \sum_{k=0}^{d-1} \omega^{-j \cdot k} Z^k$ by a discrete inverse Fourier transform. Here the exponent is taken modulo $2d$. We thus have $M_x = \sum_{k=1}^{d-1} \omega^{-j \cdot k} Z^k$.

The idea of compatibility with measurement operators, however, stays the same. Instead of checking for an anticommuting stabilizer V of $|\psi\rangle$ with M_x, we now look for stabilizers V which do not commute with one

of the terms in M_x. That is, additionally to iterating over each individual qudit in the system, we also perform the measurement procedure per term of M_x. Beside the convenient equivalence with the qubit case in the subsequent measurement steps, this is also necessary to read out the measurement outcomes on a unique basis. In practice, we calculate the unique set of phases from the Fourier transform corresponding to each basis state $|x\rangle$, we measure the phases corresponding to each term of M_x and finally match the string of phases to the dictionary values to identify the final outcome $|x\rangle$ on each qubit. We proceed for qubit a and M_x term $\left(Z^{(a)}\right)^p$.

The difference between deterministic and random measurements slightly different from the commutation relation between generalized Pauli Operators. From checking for non-commuting stabilizers V with $\left(Z^{(a)}\right)^p$, we conclude again a deterministic outcome if such a V does not exist. Postselection is not required and reading out the measurement outcome goes analogous to the qubit case:

If $\left(Z^{(a)}\right)^p$ commutes with all stabilizers, we can construct it from a product of the stabilizer generators in the tableau. This equivalently corresponds to a sum $\sum_{q=1}^{n} C_q \overrightarrow{S_{(n+q)}} = \tilde{\omega}^h \overrightarrow{\left(Z^{(a)}\right)^p}$. Notice the slight difference to the qubit case, where ω instead of $\tilde{\omega}$ was used. We can use the same argument on the symplectic inner-product as before to notice that $C_q = 0$ if and only if the qth destabilizer commutes with $\left(Z^{(a)}\right)^p$. Therefore, when calculating the phase h from multiplying stabilizers, we do not include stabilizers with coefficient $C_q = 0$. However since C_q is no longer just the set $\{0, 1\}$ but rather any phase modulo $2d$, we have to weight the sum and the product of Pauli operators. This can be done by adding the phase C_q to the qth element in the phase vector of the state. Finally we can read out the phase of the product to obtain the measurement outcome for $\overrightarrow{\left(Z^{(a)}\right)^p}$.

On the other hand if a non-commuting stabilizer V exists, the outcome is random. For arbitrary d, the reasoning for qubits is not directly applicable anymore since the operators from the Weyl-Heisenberg group, beyond X and Z, no longer necessarily give rise to mutually unbiased bases. For $d = p^\alpha$ with p a prime, it has been proven that at least $X, Z, XZ, XZ^2, \ldots, XZ^{d-1}$ have mutually unbiased eigenbases. [132]
Whether this is true for arbitrary d is an open problem. Conceptually the

property of MUBs can be bypassed easily by keeping track of all possible outcomes and their probabilities for each possible operator. If multiple non-commuting stabilizers V existed, possibly with different probability distributions, we can just apply post-selection like we did in the qubit case. For larger dimensions d or number of qudits n this can become computationally intense, even if probability distributions are in advance. Even more this was the case on the usual computational resources available during the preparation of this thesis.
An additional issue is that definitions of meaningful probability distributions become ambiguous for non-degenerate eigenvalues, which appear regularly.

Therefore we considered alternative approaches. During examination of distributions up to $d = 10$, we noticed that choosing an uniform distribution approximates the outcome distributions generally very well. Only few distributions excluded some measurement outcomes but were even then uniformly distributed for all non-zero probabilities. Since we repeat the measurement procedure over all terms of M_x and try to match the resulting phase string to an unique outcome for qubit a, this approach comes not without problems. We handled this by always choosing the closest unique outcome for qubit a. Since we can generally take (in simulations) large number of measurements into account, such errors average out easily. Especially with median-of-means estimation or other calibration techniques like from [95]. Remark also that this only concerns the numeric simulation of the protocol and not its actual application in the lab.
After post-selection, we return the random outcome and the post-selected state.

Exactly the same measurement procedure can also be used to calculate a deterministic quantity like the inner products between states which are represented by stabilizer tableaus. We use that for the inner-product $\langle \phi, \psi \rangle$, we can write $|\phi\rangle = U|0\rangle$ for some unitary U and vacuum state $|0\rangle$. We then write $\langle \phi, \psi \rangle = (U|0\rangle)^\dagger |\psi\rangle = \langle 0|U^\dagger|\psi\rangle$. So applying $U^\dagger$ to $|\psi\rangle$ gives a state $U^\dagger|\psi\rangle$ for which we want to know the overlap with the vacuum state i.e. the state stabilized by the set generated from all local Pauli-Z operators. This is roughly equivalent to a deterministic version of a computational basis measurement. [133]
We initially assume a full overlap/inner-product between the states and include extra factors based on the measurement routine. Instead of sampling a random outcome in the random measurement case, we observe that a non-commuting stabilizer of the state $U^\dagger|\psi\rangle$ with a local Pauli-Z

operator is equivalent to a partial overlap between the states. While for qubits this is simply a factor $1/\sqrt{2}$, it is an overlap of $\cos\frac{\pi k}{2d}$ for qudits with k the commutation phase (i.e. symplectic inner product) between the Pauli-Z and non-commuting stabilizer. If the deterministic measurement case is applicable, the overlap is either complete or non-existent. the choice is made based on the outcome that a measurement would have produced. If the measurement outcome is different from 0 (modulo $2d$), the inner-product between the states is zero, otherwise it is left unchanged. This is a relatively lightweight method to find the inner-product, although time complexity scales, like for measurements in general, as $O(n^3)$ for qubits. [133] [26]

Ultimately we remark that even more compact representations than stabilizers do exist for similar quantum systems, the graph state formalism. Graph states are almost identical to stabilizer states but allow an even better compression of information on the state. [96]

3.1.2 Implementing simulations on qudits

For simulating and benchmarking the performance for photonic systems with inclusion of OAM spatial modes, we also needed to adjust the simulation framework as a whole to allow for qudit type systems. The main adjustments consisted of implementing generalized Pauli and Clifford operations into the procedures of state preparation, sampling of measurement schemes and feature prediction. Also, while handling states like those from the photonic experiment, the use of the stabilizer formalism becomes impossible because the efficiency of the representation is no longer given. This forced us to use exponentially scaling matrices and arrays in those cases While one might be able to find other efficient representations, like an MPS representation, for certain kinds of systems, our goal was to keep the possible inputs to the framework as general as possible. Through the modularity, however, it should be possible to implement other suitable representations as well. The stabilizer formalism mainly serves the possibility to explore the capabilities of incomplete tomography protocols.

State preparation

Preparing the quantum state in a suitable format for the measurement and feature prediction routines was dedicated to a separate module to allow flexibility in representations.

While generating density matrices as complex-valued 2D-arrays is rather expensive for large system sizes, and even more for large subsystem dimensions d, it is obviously the most straightforward way to do so. Dicke states were prepared according to eq. 2.7 in the computational basis and used to obtain the total 4-photon OAM state in eq. 2.6 from the PDC process. Although this is a pure state, we converted it to a density matrix to exploit full capabilities in the subroutines of the simulation.

Also implementing random stabilizer states is rather simple since we can sample them once by taking an empty tableau and phase vector (the vacuum state) on which we apply a random Clifford circuit. The number of applied gates was randomly drawn as well with lower bound of $d \cdot n$ and an fixed upper bound of 400 gates. (suitable for $d \leq 10$) The upper bound is important because sampling truly uniformly from the space of stabilizer states is rather difficult with this naive approach. With the low number of generating Clifford gates, we often sample equal or equivalent states and are unlikely to sample states that require very specific preparation circuits. This problem is due to the non-unique mapping of circuits to quantum states. By sampling a large number of gates we can compensate this because the probability to sample unique and inequivalent circuit becomes smaller. Through the inclusion of CNOT gates, we then also gain the full entanglement capabilities reachable with the stabilizer formalism. By applying specific gates, like those needed for the GHZ state (see figure 1.2), we can of course also obtain particular fixed states from the vacuum tableau.

Sampling pure states as numeric array uniformly at random is, in contrast, an own field of study. [134] To obtain random normalized state vectors, we took the route of picking random columns (or rows) of uniformly sampled unitaries, albeit we realized later that this is actually not advantageous to normalizing a random vector of normal distributed complex entries. [134] While a unitary matrix parametrization is quickly derived, sampling just the parameters uniformly at random does not respect the non-uniform distribution density of unitaries in $\mathcal{U}(d^n)$. In fact, the dependence of elements in a unitary matrix on each other is creating complexities. [135] The unitary invariant measure that we use to define a (probability) density on $\mathcal{U}(d^n)$ is the earlier encountered Haar-measure. The key result from random matrix theory is that by the uniqueness and invariance under unitary operations of the Haar measure, we can generate a Haar-random unitary by sampling a random complex matrix Z and orthogonalizing its column-space. The entries of Z are normally distributed and make Z a ran-

dom sample from the so-called Gimbre-ensemble. The well-known Gram-Schmidt procedure enables the orthogonalization of Z to a unitary matrix Q. We summarize this into a concept called the QR-decomposition since $Z = QR$ with R a certain upper-triangular matrix. [135] [136] This latter can be understood from the typical orthogonalization strategy with an iterative routine in which every subsequent iteration involves an additional, readily orthogonal vector. This implies the triangular form. Fortunately, obtaining a random Z is easy and applying the QR-decomposition is a readily built-in method in Julia. The problem, however, is that R and Q are not unique. Repeating this process therefore with other implementations of the QR-decomposition might result in different decompositions. To demand uniqueness of Q, given Z, we apply a transformation to Q based on R. This is the diagonal matrix of R with normalized entries. For mathematical details we refer to [135] and [136].

Pure states can then be obtained by picking a (random) column of a Haar-random unitary. For enabling sampling of mixed states, a pure state on a larger system is generated and reduced to the desired state-space using the partial trace. Alternatively, a smaller random unitary matrix can directly be used with a method described in [134].

Generating data acquisition schemes

While benchmarking the simulation is not very sensitive to the actual randomness of a single pure trail state, performing genuine random measurements is of high relevance to capture all the details of quantum states equally well. Within the stabilizer formalism, for example, randomly sampling Clifford gates is rather exhaustive when trying to approximate a uniform distribution. This is seen from the cardinality of the quotient Clifford group $|\mathcal{C}_d^n| = d^{n^2+2n}\Pi_{i=1}^n(d^{2i}-1)$ which increases rapidly. [24] By the finite order of each generator of $\mathcal{C}_d^n$ and equivalence of certain combinations of Clifford gates, we often end up with a biased sample. Increasing the sample size exponentially was computationally not achievable with the way our stabilizer tableaus were constructed since we can only find the tableau representation by explicitly applying each gate individually in succession. This is unproblematic for a single measurement, but not for number of measurements on the order of 3 or 4 orders of magnitude. We rather require a random circuit that has immediately minimum depth. Sampling from $\mathcal{C}_d^n$ directly is practically also not feasible since one would need to generate and label each Clifford in advance.

Far more economic solutions have been given by [23] and [129] through exploitation of just exactly the same efficient representation, by a tableau and phase vector, as we were already using within the stabilizer formalism. The explicit symplectic constraint and direct interpretation as action of the Clifford on Pauli operators 3.1) allows to select an element of $\mathcal{C}_d^n$ without additional knowledge about other elements in the group and can be easily translated to a unique direct representation if required.

Generating a symplectic matrix is based on the commutation rules, i.e. the symplectic inner-product identities, of Pauli operators which Aaronson et al (proposition 3, [26]) summarized for (de)stabilizers of any stabilizer state. The key constraint of which is anticommutation of a stabilizer with its corresponding destabilizer. That is, a non-zero inner-product between their corresponding row vectors in the tableau. If we sample valid (i.e. non-identity) Pauli strings with a random phase while obeying this commutation rule, we create a proper formatted tableau. The phase vector stays completely unconstrained. [129]

If a circuit representation of the Clifford is required, we apply a process called "*sweeping*" to the tableau. In essence it is just Gaussian elimination as known from linear algebra. This is because elementary row operations are preserving the symplectic inner-product as well. This has to be true since applying Clifford gates to a tableau is equivalent to applying elementary column operations and should result in a valid, symplectic tableau as well. So by keeping track of what operations we apply to arrive at the vacuum tableau (represented by an identity matrix), we can convert the procedure to a series of local Clifford gates and, therefore, a quantum Circuit.

Our implementation follows the simplified procedure in [129] which does so by initializing an empty tableau, pairwise random generation of a stabilizer and destabilizer pair, application of a fixed sweeping algorithm [129] to solely these two rows in the tableau and repetition of the last few steps for all n stabilizer-destabilizer pairs. We thus fill and sweep the tableau from top to bottom with two rows in each iteration. By equivalence of two tableaus on interchanging two rows, this indeed is equivalent to creating and sweeping a tableau with destabilizers in the upper half and stabilizers in the lower half. The independence between pairs of rows is here, although surprising, unproblematic. But, it comes with the cost of creating a circuit on the full n-qubit register that might still have suboptimal depth compared to more advanced algorithms. In contrast to our earlier naive sampling of Clifford gates, however, it scales beneficial by $O(n \log n)$

in circuit depth [129] and $O(n^2)$ in time complexity (for qubits).
Note also that sweeping a tableau actually corresponds to the inverse (or adjoint) Clifford operation. To avoid an inversion step, we consider this adjoint operation as the final, randomly sampled Clifford. Repeating the complete routine for each required measurement is then efficient and uniform in distribution.

Remark also that the procedure was only described for qubits in [129] and had be generalized by mainly changing the conditions for non-binary tableau entries. This was rather straightforward since the qubit procedure was detailed enough in the intended tableau format. We refer to the comments in our implementation [121] for the detailed adjustments.

In the non-stabilizer mode of our framework, we can use the circuit representation from the sweeping process to convert the Clifford tableau to its matrix-representation, i.e. by applying the Clifford circuit to an identity matrix. This is very computationally very costly for high dimensions d^n and therefore one of the large bottlenecks in the framework.

For local measurements the above procedure can not be applied since the two-qudit CNOT gate must be applicable. Huang et al [122], however, provided a simple implementation for local Pauli basis measurements for randomized and derandomized approaches from the preceding chapter. Qudit generalization was trivial for those methods as well.
Finally we also mention our implementation of the partial derandomized sampling method of locally-biased shadows from [35]. However, during finalization of this thesis, it is only available for qubits in non-stabilizer settings.

Measurement and feature prediction

While measurements in the stabilizer formalism were discussed in section 3.1.1 and 3.1.1, we wrote simple alternatives for the raw array representation of density matrices of non-stabilizer states and Cliffords. The main issue in this case has been the reconstruction of the full unitary matrix associated to the specific measurement and rotating the state into the computational basis with this matrix. In contrast to feature prediction, the measurement can not be separated into single-qudit measurements unless the state is a product-state and the measurements are local single-qudit operations. Measurements are therefore one of the few points in the framework where the matrix representation is a bottleneck relatively quickly when

increasing the number of qudits n and subsystem dimensions d. Because most of the occurring unitary operations have matrix representation with a lot zeros by their local action, we used the built-in type for sparse arrays in Julia to save memory. However, by the highly restricted set of operations that can be applied to sparse arrays, the full potential could not be used throughout the complete framework.

After rotation of the state to the computational basis, the diagonal of the resulting density matrix is used as probability vector for a categorical distribution to sample a random measurement outcome. This corresponds to a simulated computational basis measurement. [29] Note that from this moment, the state has become a product state as a computational basis vector. If local Clifford or Pauli measurements were used, this means that feature prediction can be executed qudit-wise instead. For global measurements the rotation from the computational basis back to the actual desired measurement basis has to be performed with the full adjoint of the corresponding unitary matrix.

To unify the procedure of feature prediction for the stabilizer and matrix representations of quantum states into a single routine, it was ensured that the different measurement simulations result in a common output format for measurement outcomes. Although less efficient in storage compared to single integers, outcomes were stored as computational basis vectors to use them directly for prediction of local observables. Beside reducing the complexity of the numerical framework as a whole, this allowed us to ensure the same efficiency and performance for both settings.

Following [59] we would usually proceed from obtaining outcomes $|b\rangle$ by partitioning the snapshots $\mathcal{M}^{-1}(U^\dagger|b\rangle\langle b|U)$ into k sets, over each of which we take the mean element-wise, calculate the expectation of that mean shadow and take the median over all mean expectations. However, this would force us to use full-scale complex arrays with non-seperable global Cliffords U in the most general case.
Therefore we rather shortcuted the procedure by rewriting

$$\langle \operatorname{Tr} \hat{\rho} O \rangle_{\hat{\rho}} = \langle \operatorname{Tr} \left(O \mathcal{M}^{-1}(U^\dagger|b\rangle\langle b|U) \right) \rangle_{\hat{\rho}}$$

$$= (d^n+1) \cdot \langle \operatorname{Tr} \left(O \left(U^\dagger|b\rangle\langle b|U \right) \right) \rangle_{\hat{\rho}} - \operatorname{Tr} O = (d^n+1) \cdot \langle \langle b|UOU^\dagger|b\rangle \rangle_{\hat{\rho}} - \operatorname{Tr} O$$

From separability of the quantum channel for local measurements (eq. 2.16) the same follows, but written as a computationally beneficial product

over all individual qudits instead:

$$\Pi_{i=1}^{n}\left[(d+1)\cdot\langle\langle b_i|U_iO^{(i)}U_i^\dagger|b_i\rangle\rangle_{\hat{\rho}}\right]-\Pi_{i=1}^{n}\operatorname{Tr}O^{(i)}$$

Since we used only (local) Pauli strings as observables for our implementation and applications thereof, we pre-calculated the trace of every Pauli operator during framework initialization to obtain $\operatorname{Tr}O=\Pi_{i=1}^{n}\operatorname{Tr}O^{(i)}$ at this point conveniently.

To calculate the expectations from the measurement outcomes $|b\rangle$ now efficiently, we exploit the normalization property of Clifford unitaries once again to convert $UOU^\dagger$ and $U_iO^{(i)}U_i^\dagger$ (with O,$O^{(i)}$ Pauli operators) to single Pauli operators with a certain phase. This allows us to completely circumvent the global structure of U that would force us usually to construct the full unitary matrix for non-Clifford circuits. The resulting Pauli string $UOU^\dagger=\tilde{\omega}^pP$ is finally separable and leads to $\langle b|UOU^\dagger|b\rangle=\langle b|\left(\Pi_{i=1}\tilde{\omega}^{p_i}P^{(i)}\right)|b\rangle=\Pi_{i=1}\left[\tilde{\omega}^{p_i}\langle b_i|P^{(i)}|b_i\rangle\right]$ for $|b\rangle=\bigotimes_{i=1}^{n}|b_i\rangle$ in the global measurement case. The local case is completely analogous, although the expectation was readily locally computable without the assumption of (local) Cliffords. Since the conjugation of $O^{(i)}$ by U_i involves matrix multiplication as well, the advantage of this method compared to direct computation of $U_iO^{(i)}U_i^\dagger$ depends on how the values of n and d compare.

Finally, for the readout of the diagonal component corresponding to local measurement outcome $|b_i\rangle$, we reconstructed the matrix representation of single-qudit Pauli operator $P^{(i)}$ and slices out entry b_i on the diagonal. This is reasonably inefficient, especially if d increases. The economical way would be to query the stabilizer tableau for $P^{(i)}|b_i\rangle$ and to calculate its inner-product with the tableau for $|b_i\rangle$. This is possible independent of whether the actual probed system state is a stabilizer state. Unfortunately did we not finish a reliable implementation of the inner-product procedure for $d\geq 3$, which was, as described earlier, based on the stabilizer measurement procedure. Due to an minor, yet unexplained, bug in the phase calculation of the measurement outcome, we could not use it for estimation of the tableau inner-products. As to obtaining the actual measurement outcomes $|b\rangle$, this effect was largely insignificant for feature prediction with a large number of measurements.

Median of means can be applied at the very end by interpreting the expectation $\langle\cdot\rangle_{\hat{\rho}}$ as the mean on one of k partitions of the measurement scheme.

After applying the channel to the expectation as stated above, the median is taken of the resulting k final expectations to find the output estimation $\hat{o}$ for the expectation of O. For our numerical experiments, we used, however, always the arithmetic means ($k = 1$) instead of median of means. Both because of simplicity and general insignificance on the final result, especially in noise-less simulations. [94]

An even more efficient and direct way of predicting expectations for local observables under the use of local Pauli measurements was proposed and implemented by Huang et al [36] [122] as part of the derandomized protocol of classical shadows. The routine works nevertheless completely agnostic to the way measurements are sampled as long as only Pauli strings are used to represent observables and measurements.

Based on our discussion in section 2.3, for example, the concept of compatible measurements and observables is used. While iterating over the desired set of observables, compatibility with each of the measurements in the sampled measurement scheme is checked. That is, whether a measurement "hits" the observable. If for a fixed observable O a measurement P is accepted as compatible, we add the eigenvalue corresponding to the outcome of P to a cumulative sum. When the final element in the measurement scheme is passed, an arithmetic mean is taken over the sum and returned as estimation of the observable expectation. Median of means estimation is not required since the contribution from each measurement outcome to the expectation is now immediately scalar-valued and discrete. If no measurement contained information about an observable, i.e. no measurement P hits observable O, the expectation is assumed to be zero.

For qudits this procedure works equally well as for the original proposition for qubits, but because the eigenvalues are no longer ± 1 and not equal for all generalized Pauli operators, we need to compute the eigenvalues of each local measurement operator in advance. Although this involves only the computation of eigenvalues for d^2 $d \times d$-matrices once, it becomes slightly more demanding compared to qubits in the initialization phase of the framework.
Note that although we discuss classical shadow estimation, the advantageous aspect of this direct prediction method is the avoidance of creating the classical shadow explicitly. The benefits, however, still persist. If archiving of measurement data for future feature prediction is required, one can still do so along the way, albeit the classical storage without stabilizer tableaus might not be economical anymore.

Beside prediction of local observables, we also provided the option for fidelity prediction between the state ρ of the system in consideration and an arbitrary comparison state σ. Fidelity as observable differs from most other relevant features of quantum systems in that it can not be decomposed or approximated by a set of local observables, like Pauli strings. [59] Therefore we had to include a separate prediction method for fidelity prediction. Note that by construction, classical shadows require only adjustments in the post-processing step. Measurement results from local observable prediction can completely be reused.

Based on our definition for fidelity in definition 1.3.1, we created a function to evaluate fidelity for both pure and mixed state separately. Also here the choice between a stabilizer and matrix representation can lead to an efficient and less efficient implementation. However, since the definition of $|\langle\psi|\phi\rangle|^2$ for pure states requires us to evaluate an inner-product between the state vectors, we could not implement the economic stabilizer method for non-qubit systems in a stable manner.
This forced us to use the matrix representation in general. To enable prediction of mixed state fidelities, we used the expression for mixed states $(\mathrm{Tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}})^2$ from definition 1.3.1. The desired fidelity between the pseudo density matrices of classical snapshots $\hat{\rho}$ and an arbitrary state σ can then be evaluated by preparing classical snapshots $\hat{\rho}$ as matrices explicitly for every performed measurement. We subsequently extract k mean snapshots by partitioning the shadow into k parts and compute the fidelity of each of those mean snapshots with σ. The median over the k mean fidelities becomes then our final prediction again.

3.2 Benchmarking Classical Shadow estimation

Following up the description of our implementation, in this section we finally present our simulation results from reproducing simulations from [59], benchmarking our generalization to qudits and applying it to multipartite entangled systems of photons.

3.2.1 Basic feature prediction

As our implementation of classical shadows was not written within an existing and stable simulation framework for quantum many-body systems or quantum circuits, we needed to verify the framework thoroughly.

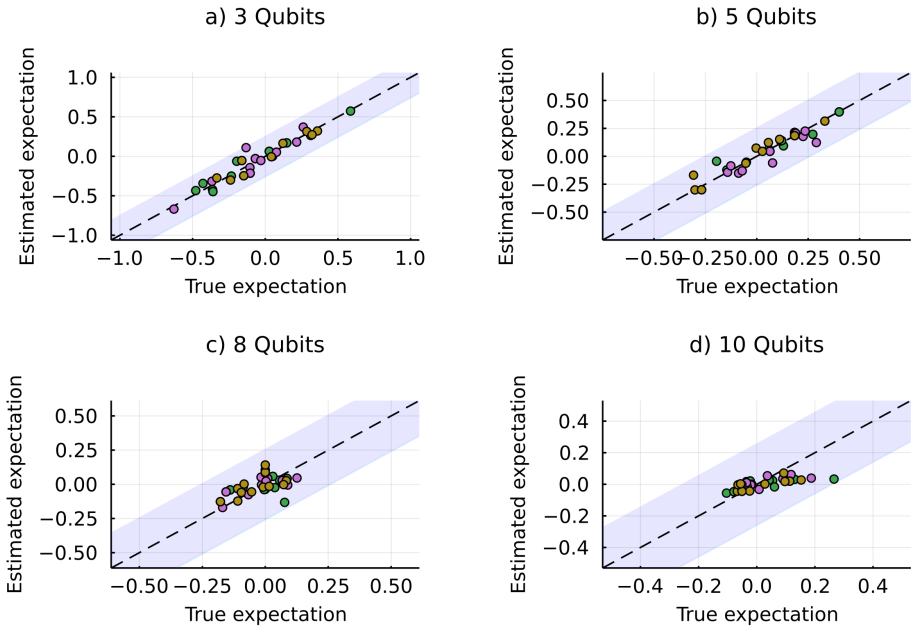

Figure 3.1: *True vs. predicted expectation of random 3-local observables for different number of qubits ($d = 2$). Marker colors indicate different simulation repetitions of predicting 12 observables with $2.8 \cdot 10^3$ random local Pauli measurements. Dashed diagonals show perfect match while blue shaded areas mark the performance bound for classical shadows from section 2.4.2, i.e. the maximal expected deviation.*

To do this, we wrote a few example experiments, most of which can be found at [121] as well.
Most relevant to these experiments was randomization of probed states in order to make the verification agnostic to exact properties of predictable quantum systems and to cover the average case. So the provided examples neither represent the worst case nor the most beneficial situation.

To obtain a first qualitative impression of the prediction capabilities of classical shadows and to judge estimated values based on the performance bound from section 2.4.2, we plotted a few predictions of random observables against their true value. For 3, 5, 8 and 10 qubits we see in figure 3.1 indeed that predictions follow the optimal diagonal line closely. All points, including measurements from additional simulations up to 12 qubits, fall within the maximum error given by the performance bound.

The reason for the close contraction of points around the expectation of

zero for increasing number of qubits, is the uniform statistical distribution for eigenvalues in sampling the observables. Their mean is zero and the more qubits we add, the larger is the set of possible observables and eigenvalues which distribute themselves around this mean. This also means that the dispersion of points becomes more fine-grained and compact because the associated normal distribution becomes narrower.

Another observation is the tilt of the rough gradient line through all the points towards the horizontal. That is, the estimated value for truly non-zero expectations progressively vanishes. This is due to the scaling of the qubit register with a constant number of predicted observables. Because the number of n-qubit Pauli strings is increasing, a random measurement P becomes less likely to hit a single random observable O. Incompatible measurements contribute no information for the prediction of the expectation of O and therefore less measurement information per observable is available. Beside the zero mean of observables in this experiment, we earlier also set the rule of predicting an expectation of zero whenever there are no hitting measurements for O. This means that, on average, the prediction of a non-zero expectation approaches zero. which clearly occurs for 10 qubits in figure 3.1. To prevent such behaviour in actual application of the measurement protocol, we can simply increase the number of randomly sampled measurements. This is no disadvantage per se since overcomplete sets of measurements would increase exponentially in n, while this adjustment scales linearly in the worst case.

Moreover, by the information theoretic optimal bound presented with classical shadow estimation in [59], we know that there must exist a variant of the protocol in which modifications to the sample size are not required for increasing n. One existing example is derandomization. Basically it makes the protocol independent of how much different choices can be made for single measurements through applying a bias towards informationally relevant measurements

In figure 3.2 we compare the randomized and derandomized approaches by plotting an overview of the absolute prediction error with our framework over different values for the number of qubits n and subsystem dimension d. For this we repeated the prediction of 80 random observables on random pure quantum states with $4 \cdot 10^4$ local Pauli measurements per run, and averaged subsequently over 16 and 6 prediction rounds for the randomized and derandomized protocol respectively. Here the number of samples $4 \cdot 10^4$ for the derandomized procedure was technically not fixed

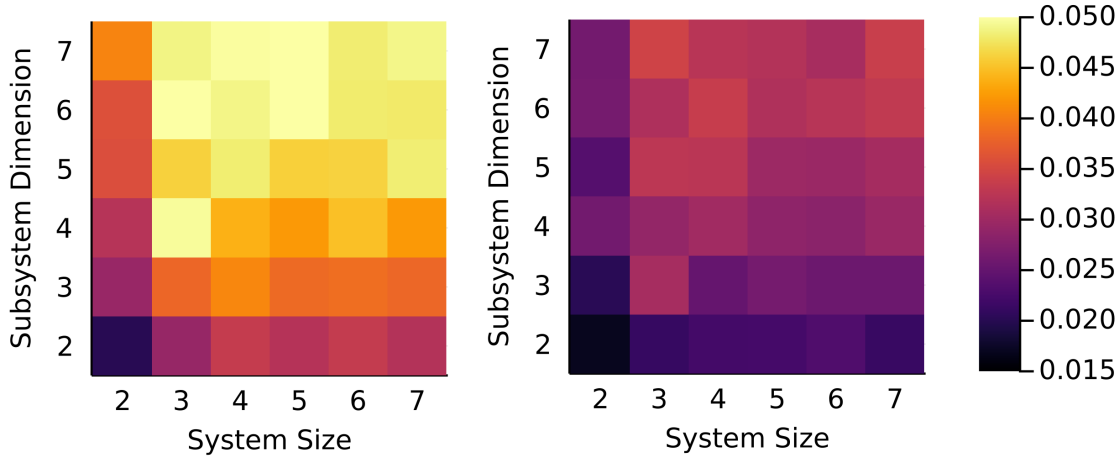

Figure 3.2: *Absolute deviation of prediction to true expectation for various system sizes n and subsystem dimensions d. A total of 80 random (max. 3-local) observables are predicted with the (original) randomized protocol (left) and its derandomized counterpart (right). Left, each square is the average of 16 repetitions with $4 \cdot 10^4$ local projective measurements. On the right, each square is the average of 6 repetitions with at least 80 local projective measurements per observable.*

since —as mentioned in section 2.3 —we rather fix the minimum number of measurements per observable that we want to achieve. To allow some comparison to the randomized measurements, we fixed it to $\frac{4 \cdot 10^4}{80} = 50$ compatible measurements per observable. This is not completely equivalent because the 50 samples per observable is a minimum value. In general more measurements than $4 \cdot 10^4$ are sampled. We, however, consider this as insignificant. In particular for $4 \cdot 10^4$ randomized measurements, we got additional derandomized measurements on the order of at most $9 \cdot 10^2$.

The number of repetitions per cell was chosen to the maximum of available computational resources so that the fluctuations in the grid are reduced and, therefore, a sharper view on the trend between different parts of the grid can be obtained. The difference in the number of runs between the two simulation variants reflects the additional computational complication of derandomizing a random set of measurements. To ensure a reasonable runtime, we therefore cut the number of repetitions for the derandomized variant with respect to the randomized protocol.

Most apparently figure 3.2 shows in direct comparison that indeed derandomization can have a significant advantage with respect to accuracy or,

equivalently, sample complexity. Note that the uniform distribution of observables, however, does also imply that this is a comparably undemanding case for derandomization. In actual applications, derandomization might often be even less significant in improving on completely random measurements.

Furthermore we observe that both grids show an inconsistent drop in error for the complete column of system size 2. This is due to an extension of the data from simulations concerning only values of $n > 2$. Because higher values of n would have enabled random observables of increasing locality with negative effect on the prediction accuracy, we restricted the simulations to sample observables of maximum weight 3 throughout. By the incompatibility of this choice for $n = 2$, we reduced this maximum weight to 2 for solely $n = 2$. In the variance bound of section 2.4.2, this reflects a drop from a factor 4^3 to 4^2 in the variance. Although this scaling only refers to an upper bound, it is clearly noticeable in figure 3.2.

Also other dependencies on n and d can be recognized in both plots. Most apparent is the strong dependence on the subsystem dimension d for constant n. In the case of local Pauli measurements, we found in section 2.4.2 a factor $(d^2)^w$ in the variance upper bound for observables with weight w. Despite the difficulty to fit the trend in the simulation results to any definite basic function, we can perceive an nearly linear relation qualitatively in the case of randomized measurements. The plot for the derandomized protocol shows a more faint relation between cells because fluctuations are more influential. Still, the contrast between the very top and bottom of the grid indicates some very similar dependence.

From the variance we also concluded that the system size n should not influence the performance of classical shadows. As we have seen for figure 3.1, this is to be taken with care and assumes saturation of the lower bound of sample complexity. Also it still underlies numerical and statistical fluctuations. For both plots we can recognize this system size independence reasonably well, but a higher number of experiment repetition and a more thorough evaluation will be needed to confirm this.

One way is to split the evaluation of the n- and d-dependence into separate experiments to use more computational power for each of them. In figure 3.3 we did consider the root mean square error (RMSE) of predicting 80 random separable observables as function of n and d. The choice for the RMSE was made to have a measure with a little more robustness than

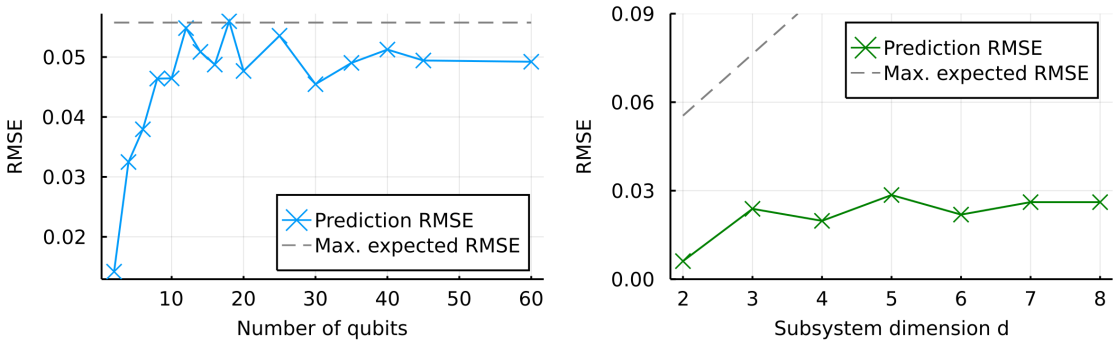

Figure 3.3: *Root-Mean-Square-Error (RMSE) for various system sizes n (left) and subsystem dimensions d (right). Each point is a single run with 10^3 random projective measurements. Dashed lines indicate the maximum expected RMSE*

the absolute error used earlier, but serves no other particular benefit. For each prediction round, 10^4 measurements were used. Using the gained computational resources for the extension of the considered range of subsystem dimensions was not possible as the routines in our framework, also in the case of stabilizers, require the construction of a polynomial number of polynomial sized arrays of complex numbers. This made the initialization in its current format infeasible for $d > 8$ on a usual computing device. We expect, however, that more economic solutions for problematic scaling procedures can be designed for the stabilizer formalism.

This is in contrast to the system size n which we found to scale reasonably well for qubits such that we found that system sizes on the order of 150 can be reached within a reasonable timeframe. This comes nowhere near the original demonstrated system sizes for the stabilizer formalism [26] or even the closed-source implementation in [59], but gives for the experiment in consideration already a sufficient insight into the development of accuracy with increasing system sizes. In light of the maximum Hilbert space dimension we reached in figure 3.2, taking for example $n = 7$ and $d = 7$ would require about 11 TB of RAM on a classical computer in direct array-based computations.

On the left figure 3.3 we took a single sweep over various qubit system sizes ($d = 2$) up to 60 qubits and plotted the RMSE for predicting the 80 random Pauli strings. The simulations for each point were performed only once because of the wide range in the number of qubits.
We obtained now indeed a more conclusive plot of the relation between

accuracy and system size n. While for $n < 10$ the error starts very low and increases rapidly, we obtain the expected behaviour of the accuracy which fluctuates around a constant value and seems to converge for large n. Notice also that the upper bound on the RMSE is strictly respected by the numerics. At a few points the values get dangerously close, while the actual line of convergence is located somewhat below the upper bound.

As mentioned before, the behaviour for low number of qubits is largely dominated by the restricted size of the set of Pauli strings. Adding a few qubits, increases this size significantly, which on its own has a large effect on the probability that measurements hit an observable.
For larger number of qubits ($n > 10$), this changes insignificantly. Especially if the locality of the observables (here 3-local) is very low compared to n (and constant). Then the prediction variance indeed becomes independent of system size as expected. This is exactly what we see in the figure.

On the right hand side of figure 3.3 we consider the same experiment but with constant $n = 3$ instead and varying subsystem dimension d. Also here the dashed line denotes the performance bound from section 2.4.2 and is indeed not constant anymore. We recall our conjecture that while the d^2 dependent expression (i.e. linear in the plot) for the variance does give an theoretical upper bound on the variance, we expect the actual scaling to be much more advantageous towards prediction accuracy. A tighter upper is, however, still to be found.
The plot of simulation results seems to support this, at least for low values of d. The first point for qubits coincides with the high accuracy from the left plot, but increases slightly to a series of almost equally levelled points. The trendline has a slight positive gradient but suffers from too much fluctuations to be conclusive. The prediction error for higher values of d needs to explored further since a deviating behaviour for low d, like for the n-dependence, can not be ruled out based on solely these results. For this,however, our framework and computational resources were not optimized and would require a further improvement in performance restricting aspects like memory allocations.

Thus far we only considered the prediction of local Pauli measurements on random quantum states. However, our framework also supports global Clifford measurements which enables us to predict also non-local properties of quantum systems. Sometimes more efficiently than with local measurements. Note that this, in practice, is more demanding on quan-

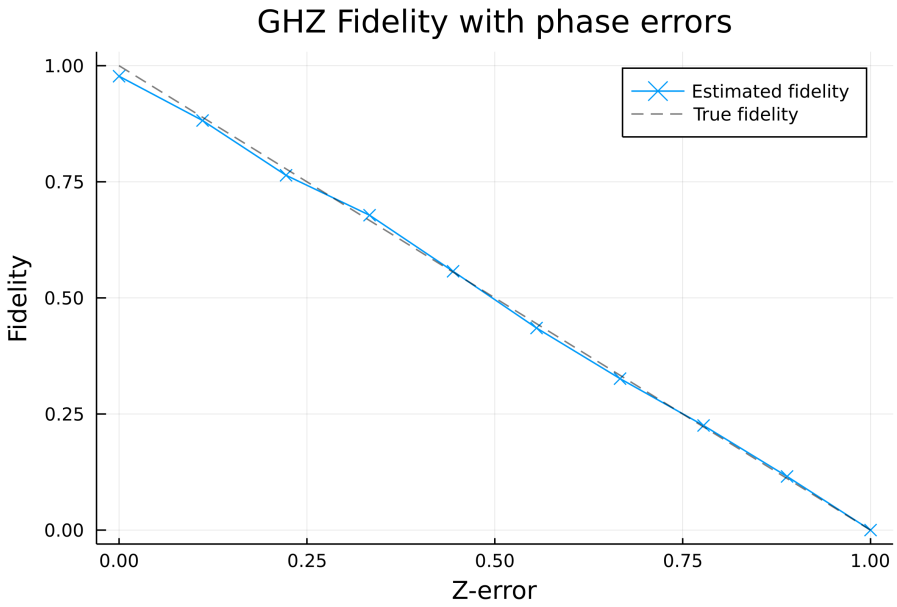

Figure 3.4: *Fidelity of GHZ state prepared with probabilistic phase flip (Z-error) given by Bernoulli parameter p (eq. 3.5). In total $6 \cdot 10^3$ Global random Clifford measurements are taken in a single run for each of the ten datapoints. The optimal and expected relation is indicated by the dashed line.*

tum hardware since physical realizations of global Cliffords require many entangled quantum gates. [59] [2]

One way to probe the reliability in predicting features with global observables, is to test the protocol on fidelity estimation. Since fidelity is the measure of overlap between two states, it is often used to verify quantum hardware components, like physical quantum gates and channels. Possible errors, through e.g. decoherence or thermalization, can then be detected, characterized and corrected using fidelity. [85]

Similar to [59] we have, therefore, examined the use of the classical shadow estimation protocol for detecting errors in the preparation of a quantum state. If we, for example, take the highly entangled GHZ state and mix in an impurity with a certain probability, we can simulate occasional errors from a noisy environment. In particular, we consider the case of random phase flips. This corresponds to a classical statistical mixture and yields the mixed state: [59]

$$p \cdot |GHZ_+\rangle\langle GHZ_+| + (1-p) \cdot |GHZ_-\rangle\langle GHZ_-| \tag{3.5}$$

with p the probability of preparing the state with a flipped phase and $|GHZ_{\pm}\rangle = \frac{1}{\sqrt{2}}(|000\rangle \pm |111\rangle)$ as usual. Because the Pauli-Z operation corresponds to a phase flip for qubits, this is also called an *Z-error*.

In the numerical implementation this means that for every single measurement a Bernoulli trail with parameter p is performed to obtain the measurement state, i.e $|GHZ_{+}\rangle$ or $|GHZ_{+}\rangle$. For the direct calculation of the true fidelity value, the complete mixed density matrix is constructed and measured.

In figure 3.4 we plotted the estimated fidelity for a range of probabilities p. The dashed anti-diagonal represents the expected true fidelities. Per datapoint a single round of $6 \cdot 10^3$ global Clifford measurements is performed. We see that indeed the measurements follow the measured and theoretical truth with few deviations, especially for low Z-error probabilities. Those deviations surprisingly, however, turned out to be not due to typical statistical fluctuations through the use of finite measurement statistics in sampling the classical snapshots. Rather there seem to occur static systematic errors from the measurement simulation as well. Although we readily reported such deviations for the simulation of stabilizer measurements, they seem to appear equivalently but with less impact in the array-based part of the framework. Also here we suspect a phase mismatch or numeric accuracy problem during measurements.
Overall the predictions coincide with the expected linear decrease in fidelity very well and reproduce therefore also figure 2 in [59] in the $n = 3$ case.

Finally, since we are mainly concerned with reducing sample complexity, we looked at the regular fidelity of a qubit GHZ ($|GHZ_{+}\rangle$) state when predicted with classical shadow estimation for different number of measurements. Here each point in figure 3.5 is the average over 4 runs. Maximal expected errors are indicated by shaded areas around the predictions for probed system sizes of 2 to 4 qubits.

While we see huge fluctuations for low numbers of measurements (< 400), the predicted fidelity converges towards a horizontal line at the level of the true fidelity $F = 1.0$ for more than $4 \cdot 10^3$ samples. While for 2 and 3 qubits the predictions seem to match up well, the case of $n = 4$ seems to overestimate the fidelity. From the data it does not become clear whether this

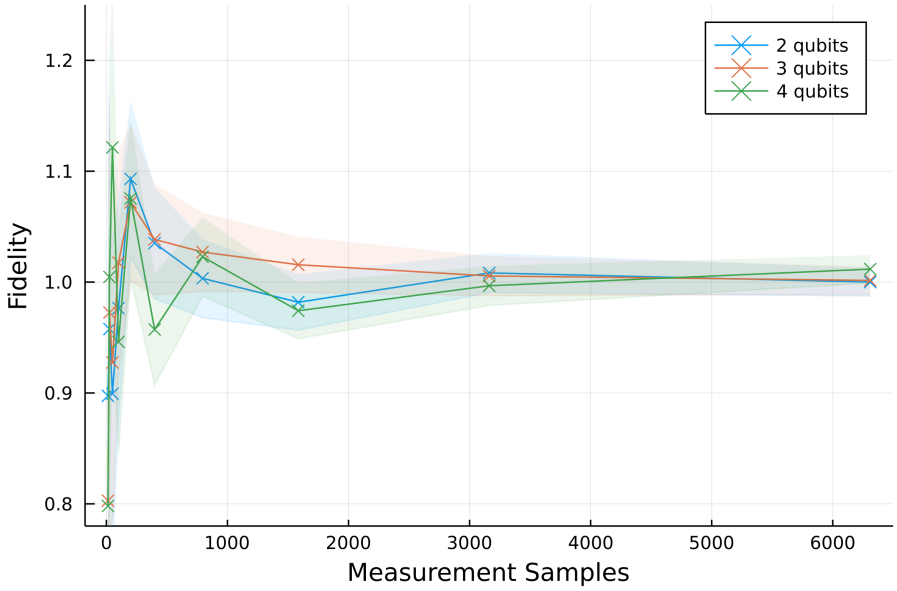

Figure 3.5: *Fidelity of a pure GHZ state with its own prediction for increasing number of random Clifford measurements. Each prediction was averaged over 4 measurement rounds. The equal coloured shaded area around each line indicates the maximal absolute deviation within which the expected fidelity ($F = 1.0$) should fall.*

is due to a statistical deviation or whether convergence occurs when more samples are included. Within the current plot we see that, starting with a saturation value of 500 measurements, the true fidelity value stays consistently in the maximal deviation bounds for all three system sizes. Even if we choose $n = 4$ and take more than 6000 measurements into account, the truth of $F = 1.0$ is still located on the boundary of the maximal error bound for the prediction. Therefore we still expect this to be a statistical outlier.

A final judgement depends on how the prediction for $n = 4$ further develops, we might be in need to recalibrate our prediction routine. This should generally not be required, but could be a reasonable method to force correct offsets or physical expectation values. For this we only need a few more measurements than for classical shadow estimation itself to find optimal parameters and do the final calibration. In [94], classical shadows are also shown to converge slower, but with physically meaningful values, by projecting each measurement snapshot $\hat{\rho}$ to a positive-semi-definite version of itself.
On the other hand figure 3.5 shows also the key benefit of classical shadows to allow estimates which are non-physical expectation values. Especially for eigenvalues that are located near physical boundaries (like $F = 1.0$ for fidelities), predictions can be made very rough for only few measurements. By increasing the number of measurements, the expectation value gets tighter towards the truth. This is exactly what we observe for the fluctuations around $F = 1.0$ as well. This is where a lot of current estimation methods fail when trying to converge to such expectations. [45].

In the end, figure 3.5 demonstrates the prediction capabilities of classical shadows for fidelities sufficiently. It also agrees with the simulation and experimental results from [94], which includes the above fidelity prediction with classical shadows as well.
Many of the experiments on predicting local observables with local measurements required less measurements to reach the same accuracy, but the example of fidelity in our two simulations shows that classical shadows can be feasible and efficient for global, but non-separable, observables as well.

3.2.2 Predicting observables in entangled photonic systems

Demonstrating effectiveness and accuracy of a measurement protocol, like classical shadow estimation, is of interest to compare it with other methods. But even more important is to demonstrate its performance for particular useful tasks, such as characterization of the four-photon entangled system that we suggested as an application in section 2.4.1. For states that are no stabilizer states, the demonstrations of classical shadow estimation have yet been scarce.

Many interesting features of quantum many-body systems can be decomposed into local single-qudit observables which we have shown in the preceding section, in addition to [59] and [94], to be efficiently predictable with classical shadows. In near term applications, most potential can be seen in improving protocols based on random measurements and in realizing required quantum hardware. Currently protocols are mostly only feasible when local measurements are used.
By employing local Pauli measurements, we specifically can use our framework to simulate predictions on such systems by just appending customized options in the state generation and observable prediction controller for specific states and observables in the experiment.

Like for the theory part, we restrict ourselves to the discussion in [105]. One basic feature that was used to verify entanglement in the experimental setting of [105] is the simple entanglement witness $W_4^{(2)}$ from [137]. Based on local (qubit) spin operators, it can verify entanglement in multipartite systems.

In the first place, an *entanglement witness* is a tool, mostly stated in the form of an inequality involving a measurable observable, which is able to detect (non-)separability of a quantum state directly by various properties. This can be imagined as a classification model in state-space.
Here $W_4^{(2)}$ does this specifically for n-qubit Dicke states with $\frac{n}{2}$ excitations. This is of interest since the the state created from PDC is given by eq. 2.6. The second term is of most relevant in the consideration of the system as it constitutes the genuine entangled four photon state. As pointed out, this term can be interpreted as a Dicke state (see eq. 2.7) where we label $|\bar{l}\rangle$, $|l\rangle$ as the ground state and excited state respectively, for example. Note that the exact value of l on this specific term of the total state does not matter for this description.

Concretely, the entanglement witness is given by

$$\mathcal{W}_4^{(2)} = \sum_{i \in \{X,Y\}} \langle J_i^2 \rangle$$

with J_i the global spin operators i.e. $J_i = \frac{1}{2} \sum_{k=1}^{n} J_i^{(k)}$ in the qubit case. We can use this notation of total angular momentum since the total orbital angular momentum must be zero anyway. [137][105]
The witness can first of all detect entanglement in general by violation of the system-specific inequality $\langle \mathcal{W}_4^{(2)} \rangle \leq \frac{4}{2} \left(\frac{4}{2} + \frac{1}{2} \right) = 5$ and genuine entanglement between the 4 photons if $\langle \mathcal{W}_4^{(2)} \rangle \leq \frac{7}{2} + \sqrt{3} \approx 5.23$ is not satisfied. [137][105] Both follow from arguments about the eigenvalues of spin operators of symmetric states, like the Dicke state.

In preparation to the numeric experiments, we implemented the general Dicke state $|D_m^n\rangle$ from eq. 2.7 in array-mode —since Dicke states are no stabilizer states —and created a custom option to extract all the local observables needed to find the expectation value of $\mathcal{W}_4^{(2)}$. The latter had to be done within qubit mode ($d = 2$) since finding a generalizing representation of this entanglement witness for qudits turned out to be not straightforward within the structure of our framework. This also forced us to only run predictions on the raw Dicke state. Nevertheless we included the PDC state as option in the framework. With a more thorough examination of this and other entanglement witnesses, the appropriate modifications to the framework could enable the simulation of the original state as well.

Resulting from our simulations, figure 3.6 demonstrates the prediction of this basic entanglement witness $\mathcal{W}_4^{(2)}$ for the Dicke state $|D_2^4\rangle$ by varying the size of the classical shadow, i.e. number of measurements, for the prediction in order to see the performance of the protocol with respect to sample complexity. Each datapoint is the average value over 5 independent runs. The maximum error bound of predicting 32 single qubit observables, from section 2.4.2, has been indicated by the blue region. Notice that also here the measured true value of the witness always falls within this maximal expected error when the number of measurements passed a certain saturation value of about 500 samples.

For shadows of size < 2000 the witness value still fluctuates heavily, especially before reaching the above mentioned saturation value. In contrast when increasing the sample size further, we see that indeed the predicted

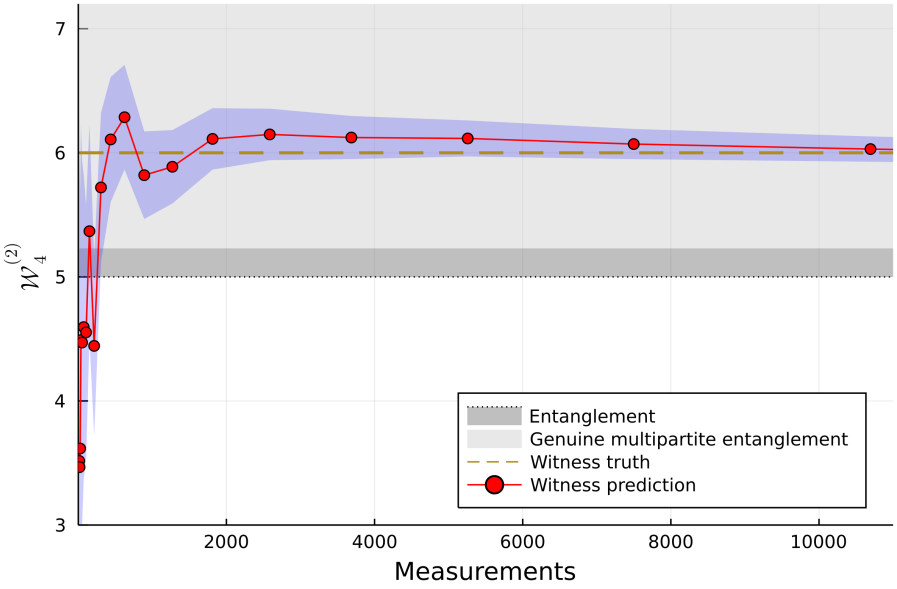

Figure 3.6: *Prediction of the $\mathcal{W}_4^{(2)}$ entanglement witness of a 4 qubit $|D_2^4\rangle$ Dicke state with 2 excitations using classical shadow estimation with varying numbers of local Pauli measurements. Each point was averaged over 5 consecutive rounds. Gray regions indicate either entanglement or genuine multipartite entanglement based on the criteria from [137]. Other regions correspond to fully separable states. Satisfaction of the propagated maximum expected prediction error specified by the blue shaded area.*

value converges to the directly measured value smoothly from above. We remark that we indeed expected a expectation value of 6 for the $|D_2^4\rangle$ Dicke state [105], which is predicted by the protocol sufficiently. The plot was cut off early for showing details, but up to the investigated value of $6 \cdot 10^4$ measurements, no quantitative deviation larger than the last appearing point in the figure were estimated.

We can interpret the predicted expectation values by the given inequalities of the witness. These have also been marked in the figure 3.6 by the gray shaded regions. Since the value of 6 lies significantly above both the entanglement threshold and the threshold for genuine multiparitte entanglement, we have indeed detected genuine entanglement for this Dicke state. The prediction reached the corresponding region quite quickly for more than 500 measurements and remained far above the threshold for all sample sizes exceeding the saturation. If the truth of the entanglement witness would have been closer to the threshold values, a conclusion would be hard to make like for direct measurements in practice as well. However, through the smooth convergence, a tight performance bound and a reasonable method for mitigation of measurement noise, one might be more confident to judge the (non-)existence of (genuine multipartite) entanglement.

Many other robust and experimentally feasible entanglement witnesses exist, albeit they are often more demanding in practice. We demonstrated that the witness $\mathcal{W}_4^{(2)}$ can be estimated sufficiently well by classical shadows. However, comparing to direct measurement of 32 observables to a certain accuracy, the advantage is rather small. Huang et al [59] came to the same conclusion in considering the application to variational quantum eigensolvers. The actual advantage of classical shadows becomes practically only apparent when considering either a large system size ($> 10^2$) or a large number of observables. For the subsystem dimension we recall that the benefit of classical shadows depends heavily on how tight the variance bound with the d^2 scaling really is. If the actual dependence of d in general is linear —or (almost) non-existent as in figure 3.3 —classical shadows might be favourable for low system sizes as well.

States resulting from pair production in PDC, however, have a reasonably small system size since the production rate diminishes quickly for higher number of photon pairs. Also when characterizing the whole system, we might have some higher order terms with higher subsystem dimension by

the value of l. But also then the coefficients in the superposition drop very fast to zero for increasing l. [103].
When considering such a system, we thus get mainly a benefit if we want to predict many local observables at once. This is more often the case than the above example might suggest. For instance, when obtaining expectations of (the variance of) Hamiltonians [59] and more complex sets of entanglement witnesses, or quantifying subsystem entanglement using Renyi-α-entropies. In terms of entanglement witnesses, we could also pinpoint the exact type of entanglement even further down by prediction a whole set of entanglement witnesses at once. The same is true for even more fundamental classifying observables that probe the state-space dimension of entangled multipartite systems, so-called dimensionality witnesses. [138] [139]

But even if the set of desired observables turns out to be small, we should keep in mind that there are only very rare cases where choosing classical shadows over direct measurements is actually a significant disadvantage, especially with derandomization. Moreover, other aspects of classical shadows can also be beneficial, like the ability to store a classical snapshot for time-delayed estimation for observables that can even be chosen after the measurements.

A prominent example [115] for a more robust entanglement witness on the considered type of systems has been mentioned in [105] as well. The witness I_m^n from [115] is more robust to noise, albeit requiring more different expectation values. It is based on permutation invariance of the state with respect to bipartitions of the system. That is, considering a specific element of the density matrix of our state, $|\langle\alpha|\rho|\beta\rangle|$, we can rewrite it under the assumption of different bipartitions using the Cauchy-Schawrz inequality. If we then define I_m^n as the subtraction of the expressions of all possible bipartitions from the original matrix element, any valid bipartition in the state ρ will pull the I_m^n below zero. Therefore the inequality $I_m^n \leq 0$ holds for any biseparable state ρ, but is violated for genuine multipartite entangled states.[115]

A basic template for direct calculation of this witness has been built into the feature prediction controller of our framework. However, yet with mixed benchmarks results. Both in post-processing the expectation values to an unbiased estimation of the witness value, as well as in handling the large number of observables by scheduling compatible measurements for simultaneous measurement.

For this specific system —also in regard to the complete mixed PDC state instead of the pure Dicke state —further numerical investigation is required to obtain a better impression and overview on the effectiveness of classical shadow estimation. On the one hand in context of this photonic experiment, but even more for general systems that are restricted to small system sizes and subsystem dimensions. This would be for assessing more robust and practical entanglement witness, but also for other interesting properties that we mentioned above. Recall that classical shadow estimation can be quite useful by its sensitivity to the structure of observables.

Another next step would be the actual incorporation of experimental measurement data as a more practical relevant survey of classical shadows —supplementary to the basic experimental study by [94] —in promising near term applications and environments. The four photon OAM entangled system could certainly be one of such interesting and practically feasible systems.

Chapter 4

Conclusion & Outlook

Quantum state estimation is a powerful tool to fully reconstruct quantum states and consequently gain information about their properties. This approach, however, has faced ultimate limits by the emerging ability to prepare and control systems of vastly increasing state-spaces. A lot of measurement repetitions are needed, which restricts the realization of scalable but noise-resilient quantum devices.

In this thesis, we gave an overview of the basics and current state-of-art in quantum state estimation. We investigated key ideas in bypassing information theoretical lower bounds for QST using incomplete tomography and settled our focus onto the specific problem of shadow tomography. By introducing the concept of classical shadows from recent efforts in designing near term and experimentally feasible data-acquisition protocols, we emphasized the relevance of universal but resource efficient solutions. Both with respect to classical computational resources, i.e. in preparing and post-processing measurements ,as well as in terms of quantum hardware requirements in the NISQ era. Here the latter involves mainly the design of practically achievable quantum gates and circuits.
We gave a brief overview of emerging variants of classical shadows in the literature and presented a generalization of the original protocol to qudit many-body systems. For the purpose of benchmarking classical shadow estimation outside the qubit regime for various systems and applications, we developed a modular numerical framework that implements the generalized protocol from state preparation to feature prediction. An existing generalization of the stabilizer formalism ensured high flexibility in state-space dimension during benchmarking and enabled an efficient use of computational resources and storage on the simulating classical device.

In numerical experiments, our framework showed consistency with basic benchmarks by [59] [94] [35] on global measurements (i.e. for GHZ fidelity estimation), as well as on local Pauli measurements. We evaluated both cases on the scaling of prediction accuracy with respect to the system size of qubit systems.
In terms of derandomization, however, our randomized benchmark shows a smaller benefit compared with the specialized example in [36]. We suspected either the focus on a single problem in [36] or errors from a phase mismatch in the measurement simulation procedure that we reported earlier.
In contrast, our derived upper bounds on the variance and sample complexity on qudit systems were consistently obeyed throughout our numerical experiments. It is an open question whether there is a tighter bound in the case of local Pauli measurements on qudit systems than our derived d^2 dependent expression. Based on the proof of information theoretic optimality for qubits in [59], this can however be expected.

We thus conclude that classical shadows, also in the generalized form presented here, can efficiently predict many properties of quantum systems. Although this assumes an appropriate combination of system and desired observables.
Deciding on shadow tomography becomes less obvious if the universality over many different features or the advantage of classical storable snapshots do not matter to the experimenter. The choice of randomized shadow tomography methods over conventional direct measurement (or complete tomography) approaches is then mainly a matter of practical limitations on the specific experimental platform. For classical shadows, this mainly concerns the unitary rotations and measurements that can be performed.
For the system of four spatially entangled photons, such as considered here, we mainly profit from the photonic platform, which has shown many capabilities in these aspects for the near term. Classical shadows can also be beneficial for such low dimensional systems if one wishes to extract a lot, but not all, information on the state of a system. Additionally, the protocol simplifies the measurement process of directly measuring multiple observables and makes the post-processing computationally less demanding since, in contrast to the commonly used MLE technique, no optimization is required.

The exact benefit for individual applications is still to be investigated with more extensive simulations. But even more importantly, benchmarks based on actual measurement data are required to find the current experimental

benefits of this and other incomplete tomography protocols. Although prediction on measurement data is already possible in our framework, the current setting is expected to be quite vulnerable to noise. Only a few methods have yet been proposed for noise mitigation within the classical shadows framework. Further study of the efficient integration of noise suppression into the measurement procedure would make the protocol more feasible in very noisy environments. Also, performance bounds in those situations have to evaluated in order to compare methods in an equal manner.
When considering other platform candidates of NISQ devices, an extension of the simulation framework to include other state representations, like MPS representations on large scale systems, could be interesting.

Improving the method of classical shadows on itself will also relevant for future directions. While upper bounds for the original protocol were proven information-theoretically optimal, this always represents the worst case. Specific systems and observables can be expected to do much better. Our simulation results of random states demonstrate this very clearly. By, for example, adjusting the type of unitary ensemble, we could take the explicit structure of the state into account. The suggested photonic application of this thesis, for example, might have an inconvenient total state structure. But optimizing the type of measurements to a highly structured Dicke state with $n/2$ excitations could ultimately improve performance, deliver a tighter upper bound on the sample complexity, and might even be of general interest regarding Dicke states for quantum information. Equally we could take the targeted features explicitly into account when scheduling measurements.
Derandomization of classical shadows already took a step in this kind of protocol customization by explicitly taking (sub)structures of desired observables into account. In contrast to other handcrafted methods of feature prediction for specific applications, we expect that customization of yet very general methods, like shadow tomography, could standardize and simplify the design of protocols for specific applications.
One other way of improvement would be to allow requirements for more advanced classical or quantum hardware. Simultaneous measurement of multiple state copies is, for example, demanding in that it requires a quantum memory for caching multiple state copies. However, this is not realistic for experimental realization in the near term.

In the end, the success of incomplete tomography protocols, like classical shadow estimation, could be a vital tool in developing and operating

NISQ devices. This includes verification of quantum hardware and estimation of numerous expectations involved in VQEs. Also beyond the NISQ era, efficient data acquisition stays relevant to all kinds of problems in quantum computing and outside. In more fundamental research, probing of quantum systems is relevant as well. This can be as simple as detecting or quantifying entanglement in a system to distinguishing whole quantum phases of matter. The suggested application in this thesis is an example of the former task.
In the big picture, it is still difficult to say what quantum technology will really bring about. From enhancements of the technology of today, with e.g. quantum metrology, towards complete quantum computers and networks themselves. A lot can be imagined. But for whatever might emerge, novel measurement protocols will for sure play an essential role on the path to it.

Acknowledgements

I thank my supervisor Jordi Tura for his support and advice throughout the project, as well as Wolfgang Löffler as 2nd reader for providing knowledge and details on ongoing experimental work and needs in his group, and for raising my interest into quantum optics.

Last but not least, I would like to thank Paul Erker and Marcus Huber from IQOQI Vienna for an interesting discussion about their paper on experimentally feasible entanglement witnesses. [115]

Bibliography

[1] F. Arute et al., *Quantum supremacy using a programmable superconducting processor*, Nature **574**, 505 (2019).

[2] J. Preskill, *Quantum computing in the NISQ era and beyond*, Quantum **2**, 79 (2018).

[3] J. B. Altepeter, D. F. V. James, and P. G. Kwiat, *4 Qubit Quantum State Tomography*, in *Quantum State Estimation*, chapter 4, pages 113–145, Springer, 2004.

[4] I. L. Chuang, M. A. Nielsen, I. L. Chuangts, and M. A. Nielsentp, *Prescription for experimental determination of the dynamics of a quantum black box*, JOURNAL OF MODERN OPTICS **44**, 2455 (1997).

[5] P. Hassanzadeh, *Towards the quantum-enabled technologies for development of drugs or delivery systems*, Journal of Controlled Release **324**, 260 (2020).

[6] S. Wei, H. Li, and G. Long, *A Full Quantum Eigensolver for Quantum Chemistry Simulations*, Research **2020**, 1 (2020).

[7] D. T. Smithey, M. Beck, J. Cooper, M. G. Raymer, and A. Faridani, *Complete experimental characterization of the quantum state of a light mode via the Wigner function and the density matrix: application to quantum phase distributions of vacuum and squeezed-vacuum states Related content Conceptions of quantum optical phas*, Physica Scripta **T48**, 35 (1993).

[8] M. a. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, Cambridge University Press, 10 edition, 2011.

[9] S. Aaronson, *Shadow tomography of quantum states*, SIAM Journal on Computing **49**, 325 (2020).

[10] W. Kozlowski and S. Wehner, *Towards Large-Scale Quantum Networks*, in *Proceedings of the Sixth Annual ACM International Conference on Nanoscale Computing and Communication*, pages 1–7, New York, NY, USA, 2019, ACM.

[11] G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Theoretical framework for quantum networks*, Physical Review A **80**, 022339 (2009).

[12] S. Fehr, *Lecture notes on Quantum Computation*, `https://homepages.cwi.nl/{~}fehr/QIT2019/QComp.pdf`, 2019, [Online; accessed 16-June-2021].

[13] S. Fehr, *Lecture Notes on Quantum Information Theory - Chapter 6*, `https://homepages.cwi.nl/{~}fehr/QIT2019/Chap6.pdf`, 2019, [Online; accessed 16-June-2021].

[14] J. Morris and B. Dakicdakic, *Selective quantum state tomography*, arXiv (2020).

[15] T. Durt, B.-G. Englert, I. Bengtsson, and K. Zyczkowski, *On mutually unbiased bases*, International Journal of Quantum Information **08**, 535 (2010).

[16] L. Garcia-Alvarez, A. Ferraro, and G. Ferrini, *From the Bloch Sphere to Phase-Space Representations with the Gottesman-Kitaev-Preskill Encoding*, in *From the Bloch Sphere to Phase-Space Representations with the Gottesman-Kitaev-Preskill Encoding*, pages 79–92, 2021.

[17] P. Van Der Vaart, *Statistical Methods for Quantum State Estimation*, Bachelor's thesis, Leiden University, 2019.

[18] R. A. Bertlmann and P. Krammer, *Bloch vectors for qudits*, Journal of Physics A: Mathematical and Theoretical **41**, 235303 (2008).

[19] J. Preskill, *Lecture Notes for Physics 229: Quantum Information and Computation*, `https://www.lorentz.leidenuniv.nl/quantumcomputers/literature/preskill{_}1{_}to{_}6.pdf`, 1998, [Online; accessed 21-June-2021].

[20] D. Gottesman, *Stabilizer Codes and Quantum Error Correction*, Doctoral thesis, California Institute of Technology, 1997.

[21] J. Eisert, *Entangling power and quantum circuit complexity*, (2021).

[22] Z. Webb, *The Clifford group forms a unitary 3-design*, (2015).

[23] R. Koenig and J. A. Smolin, *How to efficiently select an arbitrary Clifford group element*, Journal of Mathematical Physics **55**, 122202 (2014).

[24] M. Jafarzadeh, Y.-D. Wu, Y. R. Sanders, and B. C. Sanders, *Randomized benchmarking for qudit Clifford gates*, New Journal of Physics **22**, 063014 (2020).

[25] J. Tolar, *On Clifford groups in quantum computing*, Journal of Physics: Conference Series **1071**, 012022 (2018).

[26] S. Aaronson and D. Gottesman, *Improved simulation of stabilizer circuits*, Physical Review A - Atomic, Molecular, and Optical Physics **70**, 052328 (2004).

[27] D. M. Appleby, I. Bengtsson, S. Brierley, M. Grassl, D. Gross, and J.-A. Larsson, *The monomial representations of the Clifford group*, (2011).

[28] E. Hostens, J. Dehaene, and B. De Moor, *Stabilizer states and Clifford operations for systems of arbitrary dimensions and modular arithmetic*, Physical Review A **71**, 042315 (2005).

[29] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, *Negative quasi-probability as a resource for quantum computation*, New Journal of Physics **14**, 113011 (2012).

[30] J. M. Farinholt, *An ideal characterization of the Clifford operators*, Journal of Physics A: Mathematical and Theoretical **47**, 305303 (2014).

[31] A. Vourdas, *Quantum systems with finite Hilbert space*, Reports on Progress in Physics **67**, 267 (2004).

[32] E. Toninelli, B. Ndagano, A. Vallés, B. Sephton, I. Nape, A. Ambrosio, F. Capasso, M. J. Padgett, and A. Forbes, *Concepts in quantum state tomography and classical implementation with intense light: a tutorial*, Advances in Optics and Photonics **11**, 67 (2019).

[33] X. Bonet-Monroig, R. Babbush, T. E. O'Brien, and T. E. O'brien, *Nearly Optimal Measurement Scheduling for Partial Tomography of Quantum States*, Physical Review X **10**, 031064 (2020).

[34] O. Crawford, B. van Straaten, D. Wang, T. Parks, E. Campbell, S. Brierley, B. Van Straaten, D. Wang, T. Parks, E. Campbell, and S. Brierley, *Efficient quantum measurement of Pauli operators in the presence of finite sampling error*, Quantum **5**, 385 (2021).

[35] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo, *Measurements of Quantum Hamiltonians with Locally-Biased Classical Shadows*, (2020).

[36] H.-Y. Huang, R. Kueng, and J. Preskill, *Efficient estimation of Pauli observables by derandomization*, arXiv (2021).

[37] A. Royer, *Measurement of Quantum States and the Wigner Function*, Technical Report 1, 1989.

[38] E. C. Kemble, *Fundamental Principles of Quantum Mechanics with Elementary Applications*, in *Fundamental Principles of Quantum Mechanics with Elementary Applications*, page 7, Dover Publications, 1937.

[39] B. D'Espagnat, *Conceptual Foundations of Quantum Mechanics*, CRC Press, 2nd editio edition, 1976.

[40] K. E. Cahill and R. J. Glauber, *Density Operators and Quasiprobability Distributions*, Phys. Rev. **177**, 1882 (1969).

[41] H. Häffner, W. Hänsel, C. F. Roos, J. Benhelm, D. Chek-Al-Kar, M. Chwalla, T. Körber, U. D. Rapol, M. Riebe, P. O. Schmidt, C. Becher, O. Gühne, W. Dür, and &. R. Blatt, *Scalable multiparticle entanglement of trapped ions*, Nature **438**, 643 (2005).

[42] A. Acharya, T. Kypraios, and M. Guta, *A comparative study of estimation methods in quantum tomography*, Journal of Physics A: Mathematical and Theoretical **52**, 234001 (2019).

[43] R. Schmied, *Quantum state tomography of a single qubit: comparison of methods*, Journal of Modern Optics (2016).

[44] R. Blume-Kohout, *Optimal, reliable estimation of quantum states*, New Journal of Physics **12**, 043034 (2010).

[45] J. M. Lukens, K. J. H. Law, and R. S. Bennink, *A Bayesian analysis of classical shadows*, arXiv (2020).

[46] S. T. Flammia, D. Gross, Y. K. Liu, and J. Eisert, *Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators*, New Journal of Physics **14**, 95022 (2012).

[47] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, *Sample-optimal tomography of quantum states*, IEEE Transactions on Information Theory , 1 (2017).

[48] M. Guta, J. Kahn, R. Kueng, and J. A. Tropp, *Fast state tomography with optimal error bounds*, arXiv (2018).

[49] L. Calderaro, G. Foletto, D. Dequal, P. Villoresi, and G. Vallone, *Direct Reconstruction of the Quantum Density Matrix by Strong Measurements*, Phys Rev Lett. **121**, 230501 (2018).

[50] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, *Reconstructing quantum states with generative models*, Nature Machine Intelligence **1**, 155 (2019).

[51] R. O'Donnell and J. Wright, *Efficient quantum tomography*, Proceedings of the Annual ACM Symposium on Theory of Computing **STOC '16**, 899 (2016).

[52] C. Schwemmer, G. Toth, A. Niggebaum, T. Moroder, D. Gross, O. Gühne, and H. Weinfurter, *Experimental Comparison of Efficient Tomography Schemes for a Six-Qubit State*, Phys. Rev. Lett. **113**, 40503 (2014).

[53] M. S. Kaznady and D. F. V. James, *Numerical strategies for quantum tomography: Alternatives to full optimization*, Physical Review A **79**, 22109 (2009).

[54] R. Nehra, M. Eaton, C. González-Arciniegas, M. S. Kim, T. Gerrits, A. Lita, S. W. Nam, and O. Pfister, *Generalized overlap quantum state tomography*, Phys. Rev. Research **2**, 42002 (2020).

[55] H. K. Ng and B. G. Englert, *A simple minimax estimator for quantum states*, International Journal of Quantum Information **10**, 1 (2012).

[56] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, *Quantum State Tomography via Compressed Sensing*, Phys. Rev. Lett. **105**, 150401 (2010).

[57] B. P. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos, *Efficient tomography of a quantum many-body system*, Nature Physics **13**, 1158 (2017).

[58] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, Y.-K. K. Liu, R. Somma, O. Landon-Cardinal, Y.-K. K. Liu, and D. Poulin, *Efficient quantum state tomography*, Nature Communications **1**, 149 (2010).

[59] H. Y. Huang, R. Kueng, and J. Preskill, *Predicting many properties of a quantum system from very few measurements*, Nature Physics **16**, 1050 (2020).

[60] V. N. Ivanova-Rohling, G. Burkard, and N. Rohling, *Quantum state tomography as a numerical optimization problem*, arXiv (2020).

[61] M. Mohammadi, A. M. Branczyk, and D. F. V. James, *Fourier-transform quantum state tomography*, Phys. Rev. A **87**, 12117 (2013).

[62] C. A. Riofrío and C. A. R. Almeida, *Continuous Measurement Quantum State Tomography of Atomic Ensembles*, (2011).

[63] J. Cotler and F. Wilczek, *Quantum Overlapping Tomography*, Phys. Rev. Lett. **124**, 100401 (2019).

[64] B. Qi, Z. Hou, Y. Wang, D. Dong, H.-S. Zhong, L. Li, G.-Y. Xiang, H. M. Wiseman, C.-F. Li, and G.-C. Guo, *Adaptive quantum state tomography via linear regression estimation: Theory and two-qubit experiment*, npj Quantum Information **3** (2017).

[65] J. G. Titchener, M. Gräfe, R. Heilmann, A. S. Solntsev, A. Szameit, and A. A. Sukhorukov, *Scalable on-chip quantum state tomography*, npj Quantum Information **4** (2018).

[66] C. H. Baldwin, I. H. Deutsch, and A. Kalev, *Strictly-complete measurements for bounded-rank quantum-state tomography*, Phys. Rev. A **93**, 052105 (2016).

[67] B. I. Bantysh, A. Y. Chernyavskiy, and Y. I. Bogdanov, *Comparison of Tomography Methods for Pure and Almost Pure Quantum States*, JETP Letters **111**, 512 (2020).

[68] F. G. S. L. Brandão, R. Kueng, D. S. França, D. Stilck França, and D. S. França, *Fast and robust quantum state tomography from few basis measurements*, arXiv (2020).

[69] U. Schollwöck, *The density-matrix renormalization group in the age of matrix product states*, Annals of Physics **326**, 96 (2010).

[70] S. Ghosh, A. Opala, T. Matuszewski Michaland Paterek, T. C. H. Liew, M. Matuszewski, T. Paterek, and T. C. H. Liew, *Reconstructing quantum states with quantum reservoir networks*, arXiv (2020).

[71] L. Maccone and C. C. Rusconi, *State estimation: A comparison between direct state measurement and tomography*, Physical Review A **89**, 022122 (2014).

[72] T. J. Evans, R. Harper, and S. T. Flammia, *Scalable Bayesian Hamiltonian learning*, (2019).

[73] Z. Jiang, A. Kalev, W. Mruczkiewicz, and H. Neven, *Optimal fermion-to-qubit mapping via ternary trees with applications to reduced quantum states learning*, Quantum **4**, 276 (2020).

[74] A. Zhao, N. C. Rubin, and A. Miyake, *Fermionic partial tomography via classical shadows*, (2020).

[75] T. Xin, D. Lu, J. Klassen, N. Yu, Z. Ji, J. Chen, X. Ma, G. Long, B. Zeng, and R. Laflamme, *Quantum State Tomography via Reduced Density Matrices*, Phys Rev Lett. **118**, 020401 (2017).

[76] T. Feng, C. Ren, and X. Zhou, *Direct Measurement of Arbitrary Quantum Density Matrix using Phase-Shifting Technique*, (2021).

[77] J. S. Lundeen, B. Sutherland, A. Patel, C. Stewart, and C. Bamber, *Direct measurement of the quantum wavefunction*, Nature **474**, 188 (2011).

[78] J. S. Lundeen and C. Bamber, *Procedure for Direct Measurement of General Quantum States Using Weak Measurement*, Physical Review Letters **108**, 070402 (2012).

[79] S. Aaronson and G. N. Rothblum, *Gentle measurement of quantum states and differential privacy*, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 322–333, New York, NY, USA, 2019, ACM.

[80] C. Badescu and R. O'Donnell, *Improved quantum data analysis*, (2020).

[81] S. Aaronson, X. Chen, E. Hazan, S. Kale, and A. Nayak, *Online learning of quantum states*, J. Stat. Mech , 124019 (2019).

[82] Y. Chen and X. Wang, *More Practical and Adaptive Algorithms for Online Quantum State Learning*, (2020).

[83] S. Arunachalam, Y. Quek, and J. Smolin, *Private learning implies quantum stability*, (2021).

[84] A. Rocchetto, S. Aaronson, S. Severini, G. Carvacho, D. Poderini, I. Agresti, M. Bentivegna, and F. Sciarrino, *Experimental learning of quantum states*, Science Advances **5** (2019).

[85] S. T. Flammia and Y.-K. Liu, *Direct Fidelity Estimation from Few Pauli Measurements*, Physical Review Letters **106**, 230501 (2011).

[86] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, *Practical Characterization of Quantum Devices without Tomography*, Physical Review Letters **107**, 210404 (2011).

[87] F. G. S. L. Brandão, *Quantifying entanglement with witness operators*, Physical Review A **72**, 022310 (2005).

[88] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, *Probing Rényi entanglement entropy via randomized measurements*, Science **364**, 260 (2019).

[89] D. A. Abanin and E. Demler, *Measuring Entanglement Entropy of a Generic Many-Body System with a Quantum Switch*, Physical Review Letters **109**, 020504 (2012).

[90] N. Alon, Y. Matias, and M. Szegedy, *The Space Complexity of Approximating the Frequency Moments*, Journal of Computer and System Sciences **58**, 137 (1999).

[91] Y. Zhang and P. Liu, *Median-of-means approach for repeated measures data*, Communications in Statistics - Theory and Methods , 1 (2020).

[92] M. R. Jerrum, L. G. Valiant, and V. V. Vazirani, *Random generation of combinatorial structures from a uniform distribution*, Theoretical Computer Science **43**, 169 (1986).

[93] H.-Y. Huang and R. Kueng, *Predicting Features of Quantum Systems from Very Few Measurements*, (2019).

[94] G. I. Struchalin, Y. A. Zagorovskii, E. V. Kovlakov, S. S. Straupe, and S. P. Kulik, *Experimental Estimation of Quantum State Properties from Classical Shadows*, PRX QUANTUM **2**, 10307 (2021).

[95] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, *Robust shadow estimation*, arXiv , 1 (2020).

[96] S. Anders and H. J. Briegel, *Fast simulation of stabilizer circuits using a graph-state representation*, Physical Review A **73**, 022334 (2006).

[97] A. DasGupta, *U-statistics*, in *Asymptot. Theory Stat. Probab.*, chapter 15, pages 225–234, Springer New York, New York, NY, 2008.

[98] J. Matoušek, *Derandomization in Computational Geometry*, in *Handbook of Computational Geometry*, edited by J.-R. Sack and J. Urrutia, pages 559–595, Elsevier, Amsterdam, 2000.

[99] A. E. Andreev, A. E. F. Clementi, and J. D. P. Rolim, *A new general derandomization method*, Journal of the ACM **45**, 179 (1998).

[100] B. Chor and O. Goldreich, *On the power of two-point based sampling*, Journal of Complexity **5**, 96 (1989).

[101] N. Harvey, *Lecture 16: Derandomization: Method of Conditional Expectations, Method of Pessimistic Estimators*, `https://nickhar.wordpress.com/2012/03/05/lecture-16-derandomization-method-of-conditional-expectations-method-of-pessim` 2012, [Online; accessed 16-June-2021].

[102] I. Kimel and L. Elias, *Relations between Hermite and Laguerre Gaussian modes*, IEEE Journal of Quantum Electronics **29**, 2562 (1993).

[103] M. Poortvliet, *High Gain Spatial Photon Correlations*, Bachelor thesis, Leiden University, 2020.

[104] M. Krenn, M. Malik, M. Erhard, and A. Zeilinger, *Orbital angular momentum of photons and the entanglement of Laguerre-Gaussian modes*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **375**, 20150442 (2017).

[105] B. C. Hiesmayr, M. J. A. De Dood, and W. Löffler, *Observation of Four-Photon Orbital Angular Momentum Entanglement*, (2016).

[106] J. P. Torres, A. Alexandrescu, and L. Torner, *Quantum spiral bandwidth of entangled two-photon states*, Physical Review A **68**, 050301 (2003).

[107] O. D. Matteo and C. S. Qic, *A short introduction to unitary 2-designs*, `http://glassnotes.github.io/OliviaDiMatteo_Unitary2Designs.pdf`, 2014, [Online; accessed 16-June-2021].

[108] I.-C. Yu, F.-L. Lin, and C.-Y. Huang, *Quantum secret sharing with multilevel mutually (un)biased bases*, Physical Review A **78**, 012344 (2008).

[109] L. Beltran, G. Frascella, A. M. Perez, R. Fickler, P. R. Sharapova, M. Manceau, O. V. Tikhonova, R. W. Boyd, G. Leuchs, and M. V. Chekhova, *Orbital angular momentum modes of high-gain parametric down-conversion*, Journal of Optics **19**, 044005 (2017).

[110] M. Krenn, M. Huber, R. Fickler, R. Lapkiewicz, S. Ramelow, and A. Zeilinger, *Generation and confirmation of a (100 x 100)-dimensional entangled quantum system*, Proceedings of the National Academy of Sciences **111**, 6243 (2014).

[111] W. Wieczorek, C. Schmid, N. Kiesel, R. Pohlner, O. Gühne, and H. Weinfurter, *Experimental Observation of an Entire Family of Four-Photon Entangled States*, Physical Review Letters **101**, 010503 (2008).

[112] M. Malik, M. Erhard, M. Huber, M. Krenn, R. Fickler, and A. Zeilinger, *Multi-photon entanglement in high dimensions*, Nature Photonics **10**, 248 (2016).

[113] M. Bock, A. Lenhard, C. Chunnilall, and C. Becher, *Highly efficient heralded single-photon source for telecom wavelengths based on a PPLN waveguide*, Optics Express **24**, 23992 (2016).

[114] W. K. Wootters and W. H. Zurek, *A single quantum cannot be cloned*, Nature **299**, 802 (1982).

[115] M. Huber, P. Erker, H. Schimpf, A. Gabriel, and B. Hiesmayr, *Experimentally feasible set of criteria detecting genuine multipartite entanglement in <math display="inline"> <mrow> <mi>n</mi> </mrow> </math> -qubit Dicke states and in higher-dimensional systems*, Physical Review A **83**, 040301 (2011).

[116] H. Zhu, R. Kueng, M. Grassl, and D. Gross, *The Clifford group fails gracefully to be a unitary 4-design*, (2016).

[117] M. A. Nielsen, M. J. Bremner, J. L. Dodd, A. M. Childs, and C. M. Dawson, *Universal simulation of Hamiltonian dynamics for quantum systems with finite-dimensional state spaces*, Physical Review A **66**, 022317 (2002).

[118] D. Gross, F. Krahmer, and R. Kueng, *A Partial Derandomization of PhaseLift Using Spherical Designs*, Journal of Fourier Analysis and Applications **21**, 229 (2015).

[119] M. Paini, A. Kalev, D. Padilha, and B. Ruck, *Estimating expectation values using approximate quantum states*, Quantum **5**, 413 (2021).

[120] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, *Julia: A Fresh Approach to Numerical Computing*, SIAM Review **59**, 65 (2017).

[121] T. Rothe, *QfeatQ: a Quantum feature prediction framework*, `https://github.com/scipham/QfeatQ`, 2021, [Online; accessed 25-June-2021].

[122] H.-Y. R. Huang, *Predicting Properties of Quantum Many-Body Systems*, `https://github.com/momohuang/predicting-quantum-properties`, 2021, [Online; accessed 16-June-2021].

[123] R. Doolittle and W. Brian, *Classical Shadows*, `https://pennylane.ai/qml/demos/tutorial{_}classical{_}shadows.html`, 2021, [Online; accessed 21-June-2021].

[124] D. Gottesman, *The Heisenberg Representation of Quantum Computers*, (1998).

[125] J. Roffe, *Quantum error correction: an introductory guide*, Contemporary Physics **60**, 226 (2019).

[126] M. E. Cuffaro, *On the Significance of the Gottesman-Knill Theorem*, The British Journal for the Philosophy of Science **68**, 91 (2017).

[127] S. Bravyi and D. Gosset, *Improved Classical Simulation of Quantum Circuits Dominated by Clifford Gates*, Physical Review Letters **116**, 250501 (2016).

[128] D. Andersson, I. Bengtsson, K. Blanchfield, and H. B. Dang, *States that are far from being stabilizer states*, Journal of Physics A: Mathematical and Theoretical **48**, 345301 (2015).

[129] E. van den Berg, *A simple method for sampling random Clifford operators*, (2020).

[130] R. Harper, *rharper2/Juqst.jl*, `https://github.com/rharper2/Juqst.jl`, [Online; accessed 16-June-2021].

[131] S. Krastanov, *QuantumClifford.jl*, `https://github.com/Krastanov/QuantumClifford.jl`, [Online; accessed 16-June-2021].

[132] Bandyopadhyay, Boykin, Roychowdhury, and Vatan, *A New Proof for the Existence of Mutually Unbiased Bases* , Algorithmica **34**, 512 (2002).

[133] M. Rijlaarsdam, *Improvements of the classical simulation of quantum circuits*, Master thesis, Delft University of Technology, 2020.

[134] J. Maziero, *Random Sampling of Quantum States: a Survey of Methods*, Brazilian Journal of Physics **45**, 575 (2015).

[135] F. Mezzadri, *How to generate random matrices from the classical compact groups*, (2006).

[136] M. Ozols, *How to generate a random unitary matrix*, (2009), [Online; accessed 16-June-2021].

[137] G. Toth, *Detection of multipartite entanglement in the vicinity of symmetric Dicke states*, Journal of the Optical Society of America B **24**, 275 (2007).

[138] N. Brunner, S. Pironio, A. Acin, N. Gisin, A. A. Méthot, and V. Scarani, *Testing the Dimension of Hilbert Spaces*, Physical Review Letters **100**, 210503 (2008).

[139] M. Hendrych, R. Gallego, M. Micuda, N. Brunner, A. Acin, and J. P. Torres, *Experimental estimation of the dimension of classical and quantum systems*, Nature Physics **8**, 588 (2012).