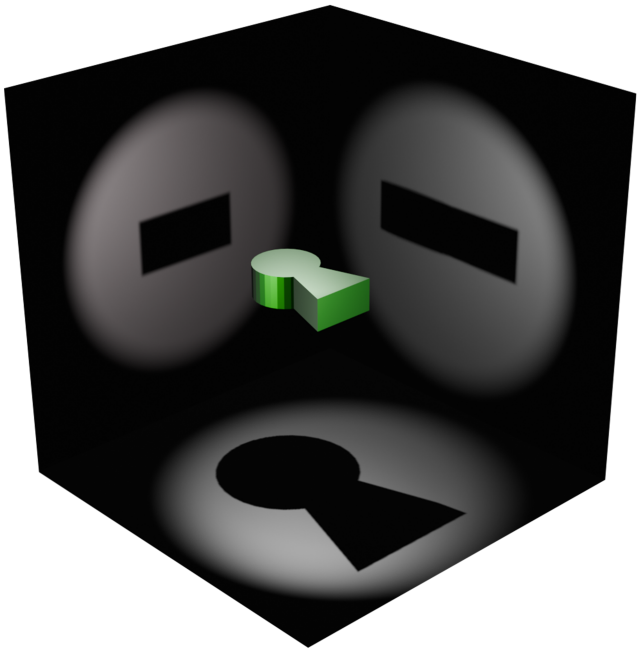
UNIVERSITÀ DEGLI STUDI DI PAVIA
DOTTORATO DI RICERCA IN FISICA - CICLO XXXVIII

Novel Approaches to Quantum State Tomography

Diego Maragnano



Tesi per il conseguimento del titolo

UNIVERSITÀ
DI PAVIA

Università degli Studi di Pavia
Dipartimento di Fisica
Dottorato di Ricerca in Fisica - Ciclo XXXVIII

Novel Approaches to Quantum State Tomography

Diego Maragnano

Submitted to the Graduate School of Physics of the University of Pavia in partial fulfillment of the requirements for the degree of

DOTTORE DI RICERCA IN FISICA
DOCTOR OF PHILOSOPHY IN PHYSICS

Supervisors
Prof. Marco Liscidini
Prof. Fabio Sciarrino

Non essere indegni di ciò che ci accade.
Gilles Deleuze

Diego Maragnano
Novel Approaches to Quantum State Tomography
Pavia, Italy (2025)

Contents

Abstract

Quantum state tomography (QST) aims at reconstructing the representation of a quantum state from the measurement of a sufficiently large number of observables. Traditionally, in QST the number of required observables grows exponentially with the dimension of the Hilbert space associated with the system under consideration, unless a priori knowledge of the state is available. In practice, QST becomes infeasible for systems of limited size. Thus, it is necessary to develop novel methods that are easy to implement and allow one to reconstruct high-dimensional states. The focus of this PhD research is to improve the efficiency of QST protocols. The main goal is to develop QST approaches that do not make any assumptions about the state to be reconstructed, reduce the number of required observables, are computationally efficient, and are experimentally feasible. The core idea is to introduce a threshold parameter that allows an efficient trade-off between the number of observables and the accuracy of the state reconstruction. Threshold quantum state tomography (tQST) is the first protocol based on this concept, tailored to systems composed of qubits that can be measured via projective measurements, and the state is reconstructed via maximum likelihood estimation. In many cases, including those of interest for technological applications, tQST works and accurately reconstructs the density matrix. We leveraged the feasibility of the protocol to numerically validate it on computer-generated data and experimentally test it on two different platforms, a superconducting and a photonic one. We show that tQST can drastically reduce the resources required for state reconstruction, but it can also be used to obtain an approximate density matrix by further reducing the number of measurements and the experimental efforts. State reconstruction can also be performed using tools other than maximum likelihood estimation. Furthermore, the tQST protocol might allow us to estimate quantities that are functions of the density matrix elements, such as purity, without reconstructing the entire state. We investigated this possibility through an original deep learning model that leverages the specific symmetries of the density matrix to enhance its performance. We then develop the enhanced compressive threshold QST (ECT-QST) protocol, specifically designed for systems composed of qudits where one can implement measurement settings. As an extension of tQST, ECT-QST is tailored for multiplexing platforms where one can perform projective measurements on all the basis vectors of a given setting. We numerically validate and experimentally test ECT-QST on the same platforms as tQST. Finally, moving beyond application matters, we turn to deeper questions and examine a problem at the crossroads of information theory and energetics, which are at the core of modern physics. We explore ECT-QST and QST through the lens of the quantum Maxwell demon, directly comparing their roles in the fundamental challenge of quantum work extraction.

List of Publications

Peer-reviewed publications

- Daniele Binosi, Giovanni Garberoglio, **Diego Maragnano**, Maurizio Dapor, and Marco Liscidini, "*A tailor-made quantum state tomography approach.*", **APL Quantum** 1, 036112 (2024) [1].
- Giovanni Garberoglio, Maurizio Dapor, **Diego Maragnano**, Marco Liscidini, and Daniele Binosi, "*Enhanced compressive threshold quantum state tomography for qudit systems.*", **Physical Review A** 111, 032436 (2025) [2].
- Diego Maragnano, Claudio Cusano, and Marco Liscidini, "*A permutation-equivariant deep learning model for quantum state characterization.*", **APL Machine Learning** 3, 026110 (2025) [3].
- Eugenio Caruccio, **Diego Maragnano**, Giovanni Rodari, Davide Picus, Giovanni Garberoglio, Daniele Binosi, Riccardo Albiero, Niki Di Giano, Francesco Ceccarelli, Giacomo Corrielli, Nicolò Spagnolo, Roberto Osellame, Maurizio Dapor, Marco Liscidini, and Fabio Sciarrino, "*Experimental verification of Threshold Quantum State Tomography on a fully-reconfigurable photonic integrated circuit*", **npj Quantum Information** 11, 173 (2025) [4].

Papers in preparation

- Eugenio Caruccio, **Diego Maragnano**, Giovanni Rodari, Davide Picus, Giovanni Garberoglio, Daniele Binosi, Riccardo Albiero, Niki Di Giano, Francesco Ceccarelli, Giacomo Corrielli, Nicolò Spagnolo, Roberto Osellame, Maurizio Dapor, Marco Liscidini, and Fabio Sciarrino, "*Experimental verification of Enhanced Compressive Threshold Quantum State Tomography on a fully-reconfigurable photonic integrated circuit*" [5].
- Giacomo Guarnieri, **Diego Maragnano**, Matteo Piccolini, and Marco Liscidini, "*Quantum work extraction from only partial information and finite resources*" [6].

Conference proceedings

- **Diego Maragnano**, Daniele Binosi, Giovanni Garberoglio, Claudio Cusano, Maurizio Dapor, and Marco Liscidini, "*Threshold quantum state tomography.*" In 2023 23rd International Conference on Transparent Optical Networks (**ICTON**), pp. 1-4. IEEE, 2023 [7] (**oral presentation**).
- Daniele Binosi, **Diego Maragnano**, Giovanni Garberoglio, Maurizio Dapor, and Marco Liscidini, "*Threshold Quantum State Tomography.*" In 2024 Conference on Lasers and Electro-Optics (**CLEO**), pp. 1-2. IEEE, 2024 [8] (**oral presentation**).

List of Figures

List of Tables

Introduction

Quantum technologies have emerged as one of the most promising frontiers in modern science and engineering, offering revolutionary capabilities in information processing, secure communication, and precision measurements. These advances rely on precise manipulation of quantum states. A fundamental requirement for the reliable operation of any quantum device is the accurate characterization of the underlying quantum states. The representation of a quantum system is typically reconstructed from the measurement of a sufficiently large number of observables. This process consumes multiple identical copies of the system. This procedure is known as quantum state tomography (QST) and occupies a central role in the theoretical and experimental development of quantum technologies. In the context of quantum computing, QST enables the verification of quantum states produced by quantum algorithms. In quantum communication, tomographic methods provide means to validate entanglement distribution and ensure the correctness of protocols such as quantum key distribution and quantum repeaters. Beyond these, applications in quantum metrology and quantum simulation also depend on accurate state reconstruction, where QST contributes to the validation of enhanced sensing schemes and the exploration of many-body quantum systems.

The significance of QST is accompanied by substantial challenges. The number of observables required for complete state characterization scales exponentially with the dimension of the Hilbert space, making conventional tomographic approaches infeasible for large systems. This number can be reduced in some cases by introducing some assumptions about the state to be reconstructed, such as being pure or having a matrix product structure. Moreover, the computational burden of processing the measurement outcomes and reconstructing the state can further limit the realization of QST to relatively small-size systems. These limitations motivated the development of alternative strategies, such as compressed sensing techniques, Bayesian estimation frameworks, and, more recently, machine learning–based approaches. Each of these methods aims to balance accuracy, efficiency, and robustness in the face of practical constraints.

To address these limitations, in this PhD research, we investigate innovative and efficient QST protocols. The primary objective is to develop QST methods that achieve some desirable features, that is, do not assume any particular form of the state to be reconstructed, reduce the number of required observables, and ensure computational efficiency along with experimental feasibility. Central to this approach is the introduction of a threshold parameter that provides a balanced compromise between the accuracy of state reconstruction and the number of observables needed. The threshold quantum state tomography (tQST) protocol is the first based on this principle, designed for systems made up of qubits where one can implement single projective measurements, with state reconstruction carried out through maximum likelihood estimation. After the numerical

validation using synthetic data, we experimentally demonstrated the efficacy of tQST on superconducting and photonic platforms. We then broadened the field of applicability of tQST in several directions. First, we embedded a physics-inspired deep learning model into the tQST protocol, and extended it to the more general problem of quantum state characterization. Second, we generalized tQST to multiplexing platforms where measurement settings can be implemented on systems composed of qudits, resulting in the enhanced compressive threshold QST (ECT-QST) protocol. Finally, moving beyond practical applications, we turn to the more profound problem of quantum work extraction, at the intersection of quantum mechanics and thermodynamics. In this context, we examine ECT-QST and QST through the perspective of a realistic quantum Maxwell demon that deals with finite resources, comparing their respective roles in estimating the ergotropy.

In the first chapter of this manuscript, we recall the fundamental elements of quantum mechanics, set the main definitions regarding QST that will be used throughout the manuscript, and review the literature and the state of the art of QST. In the second chapter, after defining the desiderata of any QST protocol, we introduce the workflow of the tQST protocol and illustrate each step in detail. Furthermore, we provide numerical evidence and experimental results on the two platforms mentioned above. In the third chapter, we analyze how a physics-inspired deep learning model can fit into the tQST protocol and how it can be leveraged to characterize quantum states. In the fourth chapter, we conduct a step-by-step investigation of the workflow of ECT-QST, along with numerical validation and experimental results on the same platforms as tQST. Finally, in the fifth chapter we report the comparison of ECT-QST and QST within the framework of quantum thermodynamics.

Chapter 1

Quantum Mechanics and Quantum State Tomography

In this chapter, we review the essentials of quantum mechanics, including postulates, density matrix formalism, and relevant examples of quantum states, measurement, and channels. Finally, we introduce quantum state tomography (QST) and review the literature about it, from pioneering papers to modern and more advanced concepts and tools. References for the first part of the chapter are [9] and [10].

1.1 Postulates of Quantum Mechanics

We start by presenting the von Neumann formulation of quantum mechanics for closed systems. The postulates are usually sorted according to the fundamental steps of any prototypical quantum experiment: state preparation, time evolution, and measurement.

Postulate 1: Associated with any isolated physical system is a complex Hilbert space known as the state space of the system. The system is fully described by a state vector, which is a unit vector in the state space of the system.

Postulate 2: The evolution of a closed quantum system is described by a unitary transformation. The state $|\psi(t_0)\rangle$ of the system at time t_0 is related to the state $|\psi(t)\rangle$ at time t by a unitary operator U that depends only on times t and t_0:

$$|\psi(t)\rangle = U(t, t_0)\,|\psi(t_0)\rangle\,. \tag{1.1}$$

The second postulate has some relevant consequences. First, the time-reverse evolution operator from t to t_0 is given by:

$$U(t, t_0) = U^{-1}(t_0, t) = U^{\dagger}(t_0, t)\,. \tag{1.2}$$

Second, the set of unitary transformations is a group:

$$U(t_2, t_1)\,U(t_!, t_0) = U(t_2, t_0)\,, \quad t_2 > t_1 > t_0. \tag{1.3}$$

Finally, in both the case of undriven and externally driven system, which differentiates between isolated and closed, the dynamics is generated by a Hermitian operator $H = H^{\dagger}$

(which in the driven case is $H(t)$). In the undriven case, the system is invariant under time translations: $U(t,t_0) = U(t-t_0) \equiv U_t$, and the Stone's theorem guarantees the existence and uniqueness of a Hermitian (possibly unbounded) operator $H = H^\dagger$ that is the generator of the evolution:

$$U_t = e^{-\frac{i}{\hbar}Ht}, \tag{1.4}$$

with $\hbar = 6.626 \times 10^{-34}\,\mathrm{J\,s}$ the Planck's constant. From (1.4) we can derive the following differential equation for $|\psi(t)\rangle$:

$$i\hbar\frac{\partial\,|\psi(t)\rangle}{\partial t} = H\,|\psi(t)\rangle\,, \tag{1.5}$$

which is the time-independent Schrödinger equation. In the time-dependent case the dynamics is no more invariant under time translations, Stone's theorem does not hold anymore and more sophisticated tools are required to show that any $U(t,t_0)$ has a unique self-adjoint generator $H(t) = H^\dagger(t)\ \forall t$.

Postulate 3: Any physical observable A is associated with a Hermitian operator $\hat{A}$ acting on the Hilbert space of the system.

We can use the spectral theorem to express any observable in diagonal form. Assuming a discrete spectrum of $\hat{A}$ for simplicity, we have:

$$\hat{A} = \sum_j a_j\,|a_j\rangle\langle a_j| = \sum_j a_j \Pi_{a_j}, \tag{1.6}$$

with $\hat{A}\,|a_j\rangle = a_j\,|a_j\rangle$, $a_j \in \mathbb{R}$, and $\Pi_{a_j} = |a_j\rangle\langle a_j|$. The eigenvectors of a Hermitian operator form an orthonormal basis, $\langle a_i|a_j\rangle = \delta_{ij}$, with δ_{ij} the Kronecker delta, and Π_{a_j} is a projector, $\Pi_{a_j}^2 = \Pi_{a_j}$, and satisfy the completeness relation $\sum_j \Pi_{a_j} = \mathbb{I}_D$. Here, $\mathbb{I}_D$ is the D-dimensional identity operator, with D the dimension of the Hilbert space associated to the system according to Postulate 1. In the followings we may omit the hat symbol to denote operators when it is clear from the context.

Postulate 4: The (only) possible outcomes of a measurement of an observable A are the eigenvalues of the corresponding Hermitian operator $\hat{A}$.

The fourth postulate implies that if the spectrum of an observable A is discrete, then all its measurement outcomes can only take discrete values.

Postulate 5 (Born rule): When a measurement of an observable A is performed on a system with state $|\psi\rangle$, the probability of obtaining the outcome a_i is given by:

$$P(a_i) = |\langle\psi|a_i\rangle|^2 = \langle\psi|a_i\rangle\,\langle a_i|\psi\rangle = \langle\psi|\Pi_{a_i}|\psi\rangle\,. \tag{1.7}$$

Observing that $P(a_i) \geq 0\ \forall i$, and that the completeness relation implies that $\sum_i P(a_i) = 1$, we conclude that $\{P(a_i)\}$ is a proper probability distribution.

The intrinsic probabilistic nature of quantum mechanics associated with measurement outcomes becomes apparent whenever a quantum system is not prepared in an eigenstate

$|a_i\rangle$ of a certain observable A, but rather in a generic state that can be seen as a superposition of eigenstates of A, that is, $|\psi\rangle = \sum_i c_i |a_i\rangle$. The outcome of the measurement of an observable A on a quantum system therefore becomes a random variable distributed according to the Born rule. This means that, upon repeating the same experiment N times, one would obtain a series of outcomes $(x_1, x_2, \ldots, x_N)$ and the probability $P(a_i)$ is equal to the fraction of times $\nu_{a_i} = N_{a_i}/N$ that the outcome $x_i = a_i$ appears. We can then characterize the statistics of A with some metrics such as the mean, also called average or expectation value:

$$\langle A\rangle_\psi = \langle\psi|A|\psi\rangle = \sum_i a_i |\langle\psi|a_i\rangle|^2 = \sum_i a_i\, P(a_i)\,, \tag{1.8}$$

and the variance:

$$\mathrm{Var}(A)_\psi = \langle A^2\rangle_\psi - \langle A\rangle_\psi^2 = \sum_i a_i^2 P(a_i) - \left[\sum_i a_i P(a_i)\right]^2 \tag{1.9}$$

1.1.1 The qubit

We now exemplify the postulates just introduced with the simplest quantum mechanical system we can consider, the **qubit**. A qubit is a 2-level carrier of information associated with a two-dimensional Hilbert space $\mathcal{H} \simeq \mathbb{C}^2$. All linear bounded operators are represented by linear operators on $\mathbb{C}^2$, that is, 2×2 complex matrices, while states are represented as 2-dimensional complex column vectors. Let $\{|0\rangle, |1\rangle\}$ be the standard orthonormal basis of $\mathcal{H}$:

$$|0\rangle = \begin{pmatrix}1\\0\end{pmatrix}, \quad |1\rangle = \begin{pmatrix}0\\1\end{pmatrix}. \tag{1.10}$$

An arbitrary state vector $|\psi\rangle$ of a qubit can be written as:

$$|\psi\rangle = \begin{pmatrix}a\\b\end{pmatrix} = a\,|0\rangle + b\,|1\rangle\,, \tag{1.11}$$

with $a, b \in \mathbb{C}$. The condition that $|\psi\rangle$ is a unit vector, also known as the normalization condition of the state vector, implies that:

$$\langle\psi|\psi\rangle = 1 = |a|^2 + |b|^2. \tag{1.12}$$

The following are paradigmatic examples of qubit states:

$$\begin{aligned}
&|0\rangle \equiv |z_+\rangle \equiv |H\rangle = \begin{pmatrix}1\\0\end{pmatrix} && |1\rangle \equiv |z_-\rangle \equiv |V\rangle = \begin{pmatrix}0\\1\end{pmatrix}\\
&|+\rangle \equiv |x_+\rangle \equiv |D\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} && |-\rangle \equiv |x_-\rangle \equiv |A\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}\\
&|y_+\rangle \equiv |R\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\i\end{pmatrix} && |y_-\rangle \equiv |L\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-i\end{pmatrix}
\end{aligned} \tag{1.13}$$

Qubit observables are represented as 2×2 complex Hermitian matrices. For instance, consider the **Pauli operators**, or **Pauli matrices**, the representation of which in the standard basis is:

$$\sigma_x \equiv X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y \equiv Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z \equiv Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.14}$$

The eigenvalues of each Pauli operator are $\{+1, -1\}$, and the corresponding eigenstates are the states reported in (1.13). In particular, $\{x_+, x_-\}$, $\{y_+, y_-\}$, and $\{z_+, z_-\}$ are the eigenvectors of X, Y and Z with eigenvalues $\{+1, -1\}$, respectively. Equivalently, the Pauli operators satisfy the following eigenvalue equations:

$$\begin{aligned} X\left|x_\pm\right\rangle &= (\pm 1)\left|x_\pm\right\rangle \\ Y\left|y_\pm\right\rangle &= (\pm 1)\left|y_\pm\right\rangle \\ Z\left|z_\pm\right\rangle &= (\pm 1)\left|z_\pm\right\rangle \end{aligned} \tag{1.15}$$

Qubit Hamiltonians are observables too. As an example, consider the following Hamiltonian:

$$H = \hbar\omega\sigma_x = \hbar\omega \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{1.16}$$

with ω a constant frequency. The eigenvectors $\{|k\rangle\}$ of H are the same of σ_x, and the corresponding eigenvalues $\{E_k\}$ are the same of σ_x multiplied by $\hbar\omega$. The time-evolution operator is then given by:

$$\begin{aligned} U &= \sum_k e^{-\frac{i}{\hbar}\omega_k t}\left|k\right\rangle\!\left\langle k\right| = e^{i\omega t}\left|x_-\right\rangle\!\left\langle x_-\right| + e^{-i\omega t}\left|x_+\right\rangle\!\left\langle x_+\right| \\ &= e^{i\omega t}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \end{pmatrix} + e^{-i\omega t}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \end{pmatrix} \\ &= \frac{e^{i\omega t}}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \frac{e^{-i\omega t}}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \cos\omega t & -i\sin\omega t \\ -i\sin\omega t & \cos\omega t \end{pmatrix} \end{aligned} \tag{1.17}$$

having defined $\omega_k = E_k/\hbar$. According to postulate 4, the possible outcomes of a measurement of one of the Pauli operators are +1 and -1, while those of Hamiltonian (1.17) are $+\hbar\omega$ and $-\hbar\omega$. Finally, the probability of obtaining the outcome -1 when measuring σ_y on the state $|x_+\rangle$ is given by the Born rule:

$$\begin{aligned} P\left(-1\right|\left|x_+\right\rangle) &= \left|\langle x_+|y_-\rangle\right|^2 \\ &= \left|\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}\right|^2 \\ &= \frac{1}{4}|1+i|^2 = \frac{1}{2}. \end{aligned} \tag{1.18}$$

1.2 Modern Reformulation of Quantum Mechanics

1.2.1 The statistical operator

Postulate 1, and hence the use of a ket to describe the state of a quantum system, works when we have an ideal source that prepares a quantum state with perfect consistency. However, we can actually prepare a state only within a certain error, which might reflect either some technological limitations or some fundamental lack of detailed knowledge of such a state. We are then led to consider a more general situation in which the source can prepare the system in a state within a given set of states $\{|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_N\rangle\}$, each of which has a corresponding probability (relative frequency) $\{p_1, p_2, \ldots, p_N\}$. If we carry out an experiment where we measure an observable A many times, we will find an average $\langle\psi_k|A|\psi_k\rangle$ conditioned on the state being $|\psi_k\rangle$. The resulting expectation value is then given by a weighted average:

$$\begin{aligned}\langle A\rangle &= \sum_k p_k \langle\psi_k|A|\psi_k\rangle = \sum_k p_k \langle\psi_k| \left(\sum_j |j\rangle\langle j|\right) A|\psi_k\rangle \\ &= \sum_j \langle j|A\left(\sum_k p_k |\psi_k\rangle\langle\psi_k|\right)|j\rangle \\ &= \mathrm{Tr}\left(A\rho\right),\end{aligned} \tag{1.19}$$

where we have introduced the **statistical operator**, or **density operator**:

$$\rho = \sum_k p_k |\psi_k\rangle\langle\psi_k|, \tag{1.20}$$

with $p_k \geq 0 \ \forall k$ and $\sum p_k = 1$. From (1.20) it follows that the density operator is Hermitian, that is, $\rho = \rho^\dagger$. Equation (1.20) is the most general form to represent a quantum system that is compatible with the Born rule, as formally proven in Gleason's theorem.

We now provide a more mathematical definition of a quantum state. Let $\mathcal{B}(\mathcal{H})$ be the set of linear and bounded operators on a Hilbert space $\mathcal{H}$. The set of all operators in $\mathcal{B}(\mathcal{H})$ with finite trace is denoted as $\mathcal{T}(\mathcal{H})$ and is referred to as the set of trace-class operators. Then, a state is represented by a Hermitian, positive semi-definite, and with trace equal to 1, operator called statistical operator, that is, an element of the set:

$$\mathcal{S}(\mathcal{H}) = \left\{\rho \in \mathcal{T}(\mathcal{H}) \mid \rho = \rho^\dagger, \rho \geq 0, \mathrm{Tr}(\rho) = 1\right\}. \tag{1.21}$$

The set $\mathcal{S}(\mathcal{H})$ is a convex subset of $\mathcal{T}(\mathcal{H})$, which implies that a convex combination of states is still a state:

$$\sum_i p_i\rho_i \in \mathcal{S}(\mathcal{H}), \quad \sum_i p_i = 1, p_i \geq 0 \forall i. \tag{1.22}$$

The extremal points of $\mathcal{S}(\mathcal{H})$, that is, one-dimensional projectors of the form $\rho = |\psi\rangle\langle\psi|$, are called **pure states**. Any other state that can be written as a convex combination of pure states is called a **mixed state**.

Let us analyze the density operator in more detail. Given a Hilbert space of (finite) dimension D along with an orthonormal basis $\{|i\rangle\}_{i=1}^D$, we can find a representation of

the density operator in the given basis via $\rho_{ij} = \langle i|\rho|j\rangle$. The representation of the density operator in that basis is thus a Hermitian, trace-one, and positive semi-definite $D \times D$ complex matrix called **density matrix**. A density matrix is described by $D^2 - 1$ real independent parameters: $D - 1$ real numbers to characterize the diagonal elements due to normalization, and two real numbers (one for the real part and one for the imaginary part) for each of the $D(D-1)/2$ off-diagonal elements above or below the diagonal, for the others are determined by $\rho = \rho^\dagger$. This number grows quadratically with D, and is also equal to the number of independent observables required in principle to determine the density matrix that represents the state of a quantum system.

The diagonal elements $\{\rho_{ii}\}_{i=1}^{D}$ of the density matrix are called populations, while the off-diagonal elements are called coherences. On the one hand, we can directly determine the diagonal elements by measuring the expectation value of the projectors onto the elements of the basis chosen to represent the density matrix:

$$\rho_{kk} = \langle k|\rho|k\rangle = \mathrm{Tr}\left(\rho\,|k\rangle\langle k|\right) = \mathrm{Tr}\left(\rho\Pi_k\right). \tag{1.23}$$

On the other hand, we cannot directly retrieve the off-diagonal elements with projective measurements. Indeed, if we consider the following observable with only two non-zero elements at positions (i,j) and (j,i), each equal to $1/2$:

$$O_{mn}^{(\mathrm{Re})} = \begin{cases} \frac{1}{2}, & \text{if } (m,n) = (i,j) \text{ or } (m,n) = (j,i), \\ 0, & \text{otherwise.} \end{cases} \tag{1.24}$$

and compute the following expectation value:

$$\begin{aligned} \mathrm{Tr}(\rho O_{mn}^{(\mathrm{Re})}) &= \sum_{m,n} \rho_{mn} O_{nm} = \rho_{ij} O_{ji} + \rho_{ji} O_{ij} \\ &= \rho_{ij} \cdot \frac{1}{2} + \rho_{ji} \cdot \frac{1}{2} = \rho_{ij} \cdot \frac{1}{2} + \rho_{ij}^* \cdot \frac{1}{2} = \frac{1}{2} \cdot \left(\rho_{ij} + \rho_{ij}^*\right) \\ &= \frac{1}{2} \cdot 2\mathrm{Re}\left(\rho_{ij}\right) = \mathrm{Re}\left(\rho_{ij}\right), \end{aligned} \tag{1.25}$$

it actually returns the real part of the off-diagonal element ρ_{ij}. However, $O_{mn}^{(\mathrm{Re})}$ is not a projective operator, for $\left[O_{mn}^{(\mathrm{Re})}\right]^2 \neq O_{mn}^{(\mathrm{Re})}$. A similar line of reasoning can be followed for the observable:

$$O_{mn}^{(\mathrm{Im})} = \begin{cases} \frac{i}{2}, & \text{if } (m,n) = (i,j) \\ -\frac{i}{2} & \text{if } (m,n) = (j,i), \\ 0, & \text{otherwise.} \end{cases} \tag{1.26}$$

for which the following result holds:

$$\begin{aligned} \mathrm{Tr}(\rho O_{mn}^{(\mathrm{Im})}) &= \sum_{m,n} \rho_{mn} O_{nm} = \rho_{ij} O_{ji} + \rho_{ji} O_{ij} \\ &= \rho_{ij} \cdot \left(-\frac{i}{2}\right) + \rho_{ji} \cdot \left(\frac{i}{2}\right) = -\rho_{ij} \cdot \frac{i}{2} + \rho_{ij}^* \cdot \left(\frac{i}{2}\right) = -\frac{i}{2}\left(\rho_{ij} - \rho_{ij}^*\right) \\ &= -\frac{i}{2} \cdot \left[2i\ \mathrm{Im}\left(\rho_{ij}\right)\right] = \mathrm{Im}\left(\rho_{ij}\right). \end{aligned} \tag{1.27}$$

Thus, the expectation value of the operator $O_{mn}^{(\mathrm{Im})}$ on the state ρ returns the imaginary part of the off-diagonal element ρ_{ij}. However, neither $O_{mn}^{(\mathrm{Im})}$ is a projection operator. These considerations are at the basis of practical implementations of QST, which look for alternative strategies to determine the off-diagonal elements of the density matrix.

We now prove the following inequality using the properties of any density matrix:

$$|\rho_{ij}| \leq \sqrt{\rho_{ii}\rho_{jj}}, \tag{1.28}$$

where the equality holds for pure states. If the state is pure, $\rho = |\psi\rangle\langle\psi|$, then the calculation is straightforward:

$$\begin{aligned} |\rho_{ij}| &= |\langle i|\rho|j\rangle| = |\langle i|\psi\rangle\langle\psi|j\rangle| = |\langle i|\psi\rangle||\langle\psi|j\rangle| \\ &= \sqrt{(\langle i|\psi\rangle\langle\psi|i\rangle)(\langle\psi|j\rangle\langle j|\psi\rangle)} \\ &= \sqrt{(\langle i|\psi\rangle\langle\psi|i\rangle)(\langle j|\psi\rangle\langle\psi|j\rangle)} \\ &= \sqrt{\langle i|\rho|i\rangle\ \langle j|\rho|j\rangle} \\ &= \sqrt{\rho_{ii}\rho_{jj}}. \end{aligned} \tag{1.29}$$

More in general, considering a density matrix ρ, we can define the following vectors:

$$|u_i\rangle = \sqrt{\rho}\,|i\rangle \qquad\qquad |u_j\rangle = \sqrt{\rho}\,|j\rangle\,, \tag{1.30}$$

with $\sqrt{\rho}$ the unique positive square root of ρ [9, 11, 12]. Then:

$$\rho_{ij} = \langle i|\rho|j\rangle = \langle i|\sqrt{\rho}\sqrt{\rho}\,|j\rangle = \langle u_i|u_j\rangle\,. \tag{1.31}$$

Applying the Cauchy–Schwarz inequality we find:

$$|\rho_{ij}| = |\langle u_i|u_j\rangle| \leq \|u_i\| \cdot \|u_j\|. \tag{1.32}$$

But $\|u_i\|^2 = \langle i|\rho|i\rangle = \rho_{ii}$ and similarly $\|u_j\|^2 = \rho_{jj}$. Hence:

$$|\rho_{ij}| \leq \sqrt{\rho_{ii}\rho_{jj}}. \tag{1.33}$$

A second possible proof is the following one. The positive semi-definiteness of ρ implies that $\langle v|\rho|v\rangle \geq 0 \quad \forall\,|v\rangle$. Let $|v\rangle = a\,|i\rangle + b\,|j\rangle$, with $a, b \in \mathbb{C}$. Then:

$$\langle v|\rho|v\rangle = |a|^2\rho_{ii} + |b|^2\rho_{jj} + 2\mathrm{Re}\,(a^*b\rho_{ij}) \geq 0. \tag{1.34}$$

Now choose $a = 1$ and $b = -\alpha\rho_{ij}^*$, with $\alpha \in \mathbb{R}$. Then:

$$\langle v|\rho|v\rangle = \rho_{ii} + \alpha^2|\rho_{ij}|^2\rho_{jj} - 2\alpha|\rho_{ij}|^2 \tag{1.35}$$

$$= \left(|\rho_{ij}|^2\rho_{jj}\right)\alpha^2 - 2|\rho_{ij}|^2\alpha + \rho_{ii} \geq 0. \tag{1.36}$$

This is a quadratic inequality in α, and is satisfied for every value of α if and only if $(|\rho_{ij}|^2\rho_{jj}) > 0$ and the discriminant has to be negative or equal to zero. Since the first condition is always satisfied, we can compute the discriminant:

$$4|\rho_{ij}|^2\left(|\rho_{ij}|^2 - \rho_{ii}\rho_{jj}\right) \leq 0. \tag{1.37}$$

Since $4|\rho_{ij}|^2 \geq 0$, then $|\rho_{ij}|^2 - \rho_{ii}\rho_{jj} \leq 0$, hence the thesis.

A notable and immediate consequence of this result is that if one of the diagonal elements of the density matrix is zero, then all the off-diagonal elements on the corresponding row and column will be zero too. Similarly, if one of the diagonal elements is small, then the off-diagonal elements on the corresponding row and column will be small too. This will be crucial in determining the number of observables in the proposed QST protocols.

1.2.2 Purity and von Neumann entropy

Given a quantum state ρ that describes the state of a system of dimension D, how can we determine if it is pure or not? Eventually, can we quantify how far it is from a pure state? There are several ways to answer these questions. One way is to compute the spectral decomposition of ρ. If the spectral decomposition returns a single non-vanishing eigenvalue, equal to 1, then the corresponding eigenvector provides the state as $\rho = |\psi\rangle\langle\psi|$. If the spectral decomposition returns several non-vanishing eigenvalues, then we can retrieve the state via (1.20). We can also quantify "how much pure" a quantum state ρ is by defining the **purity** of ρ:

$$\mathcal{P}(\rho) = \text{Tr}\left(\rho^2\right), \tag{1.38}$$

which can be expressed as a function of the elements of the density matrix:

$$\begin{aligned}
\mathcal{P}(\rho) = \text{Tr}\left(\rho^2\right) &= \sum_{i=0}^{D-1} \left(\rho^2\right)_{ii} \\
\left(\rho^2\right)_{ii} &= \sum_{k=0}^{D-1} \rho_{ik}\rho_{ki} \\
&= \sum_{k=i} \rho_{ii}\rho_{ii} + \sum_{k\neq i} \rho_{ik}\rho_{ki} \\
&= \sum_{k=i} \left(\rho_{ii}\right)^2 + 2\sum_{k<i} |\rho_{ik}|^2 \\
\implies \mathcal{P}\left(\rho\right) &= \sum_{i=0}^{D-1} \left[\sum_{k=i} \left(\rho_{ii}\right)^2 + 2\sum_{k<i} |\rho_{ik}|^2\right]
\end{aligned} \tag{1.39}$$

The purity satisfies some properties that we express as a theorem.

1.2.1 Theorem. *Consider a Hilbert space of dimension D associated with a quantum system and let ρ be a state of the system. The purity satisfies the following properties:*

$$(i) \quad 0 \leq \frac{1}{D} \leq \mathcal{P}(\rho) \leq 1 \quad \forall \rho \ \ (\mathcal{H} = \mathbb{C}^D) \tag{1.40a}$$

$$(ii) \quad \mathcal{P}\left(U\rho U^\dagger\right) = \mathcal{P}(\rho) \tag{1.40b}$$

$$(iii) \quad \mathcal{P}(\rho) = 1 \iff \rho = |\psi\rangle\langle\psi| \tag{1.40c}$$

Proof. To prove (i) we observe that the purity is positive by definition, for it is a sum of positive terms, it is minimum when all the off-diagonal terms are vanishing and the diagonal elements are all equal to $1/D$, and it is maximum when there is only one non-vanishing diagonal element equal to 1. To prove (ii) we exploit the invariance of the trace under permutations:

$$\begin{aligned}
\mathcal{P}\left(U\rho U^\dagger\right) &= \text{Tr}\left[\left(U\rho U^\dagger\right)^2\right] \\
&= \text{Tr}\left[\left(U\rho U^\dagger\right)\left(U\rho U^\dagger\right)\right] \\
&= \text{Tr}\left[U\rho\rho U^\dagger\right] = \text{Tr}\left[U\rho^2 U^\dagger\right] \\
&= \text{Tr}\left[\rho^2 U^\dagger U\right] = \text{Tr}\left(\rho^2\right) = \mathcal{P}\left(\rho\right).
\end{aligned}$$

Finally, if $\mathcal{P}(\rho) = 1 = \text{Tr}(\rho)$, then $\rho^2 = \rho$, which is true when $\rho = |\psi\rangle\langle\psi|$, that is, the state is pure. If $\rho = |\psi\rangle\langle\psi|$, then:

$$\mathcal{P}(\rho) = \text{Tr}\left(\rho^2\right) = \text{Tr}\left(|\psi\rangle\langle\psi|\,|\psi\rangle\langle\psi|\right) = \text{Tr}\left(|\psi\rangle\langle\psi|\right) = 1.$$

□

The state achieving the lower bound of the purity is called the **maximally mixed state**:

$$\rho_{\text{maxmix}} = \frac{\mathbb{I}_D}{D}. \tag{1.41}$$

The **von Neumann entropy** quantifies the "mixedness" of a quantum state as:

$$S(\rho) = -\text{Tr}\left(\rho \ln \rho\right) = -\sum_k p_k \ln p_k, \tag{1.42}$$

where the second expression derives from the spectral decomposition of ρ. It holds $0 \leq S(\rho) \leq \ln D$, where the lower bound is attained by pure states, and the upper bound by the maximally mixed state. Logarithms with different bases can be used in the definition of the von Neumann entropy, and the upper bound changes consequently.

1.2.3 The Pauli basis and the Bloch vector

The Puali matrices $\{\sigma_x, \sigma_y, \sigma_z\}$, along with the 2×2 identity matrix $\mathbb{I}_2 \equiv \sigma_0$, form an orthogonal basis of the Hilbert space of all complex 2×2 matrices over $\mathbb{C}$, when equipped with the Hilbert-Schmidt scalar product:

$$\begin{aligned} (\cdot,\cdot) : \mathcal{B}(\mathcal{H}) \times \mathcal{T}(\mathcal{H}) &\to \mathbb{C} \\ (O, A) &= \text{Tr}\left(O^\dagger A\right). \end{aligned} \tag{1.43}$$

We can expand the density operator of a qubit on this basis:

$$\rho = \sum_{\alpha=0}^{3} \text{Tr}\left(\sigma_\alpha^\dagger \rho\right) \sigma_\alpha = \frac{1}{2}\left(\mathbb{I} + \underline{v} \cdot \underline{\sigma}\right) = \frac{1}{2}\begin{pmatrix} 1 + v_z & v_x - iv_y \\ v_x + iv_y & 1 - v_z \end{pmatrix}, \tag{1.44}$$

with $\underline{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ and $\mathbb{R}^3 \ni \underline{v} = (v_x, v_y, v_z) = \left(\text{Tr}\left(\sigma_x \rho\right), \text{Tr}\left(\sigma_y \rho\right), \text{Tr}\left(\sigma_z \rho\right)\right)$ is called the **Bloch vector**. If we compute the purity of a state we find that:

$$\mathcal{P}(\rho) = \frac{1}{2}\left(1 + |\underline{v}|^2\right) \leq 1 \implies |\underline{v}| \leq 1. \tag{1.45}$$

This constraint offers a nice geometrical interpretation, as it identifies any density matrix of a qubit with a point within the unit sphere of $\mathbb{R}^3$, called **Bloch sphere**. On its surface there are pure states ($|\underline{v}| = 1$), while mixed states are in its interior ($|\underline{v}| < 1$). In the center is located the maximally mixed state ($|\underline{v}| = 0$).

If one can measure the expectation value of the Pauli matrices, then only three observables are required to define the qubit density matrix. However, the fourth measurement is required to normalize the count statistics. The use of experimental data to calculate the expectation values to be put into (1.44) can lead to matrices with negative eigenvalues, which are non-physical. Statistical techniques such as maximum likelihood estimation can be used to retrieve physical density matrices, as we will see in the following.

1.2.4 Composite systems

We can extend the formalism to describe systems made up of two or more subsystems, generally interacting with each other. For simplicity, let us assume that there are two subsystems, and label them with A and B. Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be the corresponding Hilbert spaces of dimension D_A and D_B, respectively. Then the Hilbert space of the composite system is given by the tensor product of the two Hilbert spaces:

$$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B. \tag{1.46}$$

The dimension of $\mathcal{H}_{AB}$ is the product of the dimension of the two Hilbert spaces, that is, $D_{AB} = D_A \times D_B$. The tensor structure implies that if $\{|\psi_i\rangle\}_{i=}^{D_A}$ and $\{|\phi_j\rangle\}_{j=1}^{D_B}$ are bases of $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, then:

$$\{|\psi_i\rangle \otimes |\phi_j\rangle\}_{i,j=1}^{D_A,D_B} \tag{1.47}$$

is a basis of $\mathcal{H}_{AB}$. Consequently, any element $|\psi\rangle_{AB} \in \mathcal{H}_{AB}$ can be expanded on that basis as:

$$|\psi\rangle_{AB} = \sum_{i=1}^{D_A}\sum_{j=1}^{D_B} c_{ij} \, |\psi_i\rangle \otimes |\phi_j\rangle \, . \tag{1.48}$$

Analogously, any operator acting on the composite system can be written as:

$$O = \sum_k O_k^{(A)} \otimes O_k^{(B)}, \quad O_k^{(A,B)} \in \mathcal{B}\left(\mathcal{H}_{A,B}\right). \tag{1.49}$$

In particular, the set of statistical operators of the composite system is $\mathcal{S}\left(\mathcal{H}_A \otimes \mathcal{H}_B\right)$, and any statistical operator of the composite system can be written as:

$$\rho_{(AB)} = \sum_k p_k \, |\psi_k\rangle\langle\psi_k|_{AB} \, . \tag{1.50}$$

The simplest example is given by the product state $\rho = \rho_A \otimes \rho_B$, which physically represents an uncorrelated state that can be prepared by acting locally and independently on the two subsystems. In this case, if we measure an observable $O \otimes Q$, then the probabilities of the respective outcomes factorize as follows:

$$P\left(O = o, Q = q\right) = \mathrm{Tr}\left[\left(O \otimes Q\right)\left(\rho_A \otimes \rho_B\right)\right] = \mathrm{Tr}\left(O\rho_A\right)\mathrm{Tr}\left(Q\rho_B\right). \tag{1.51}$$

A particular, but relevant, case is when we measure an observable on one of the subsystems and ignore the other. Without loss of generality, suppose we measure an observable $O \in \mathcal{B}\left(\mathcal{H}_A\right)$. Then the expectation value of such an observable is given by:

$$\langle O_A \otimes \mathbb{I}_B\rangle_{\rho_{AB}} = \mathrm{Tr}\left[\left(O_A \otimes \mathbb{I}_B\right)\rho_{AB}\right] = \mathrm{Tr}_A\left(O_A\rho_A\right), \tag{1.52}$$

where we have introduce the **reduced density operator**:

$$\rho_A = \mathrm{Tr}_B\left(\rho_{AB}\right) = \sum_{j=1}^{D_B} \langle j|\rho_{AB}|j\rangle \, , \tag{1.53}$$

with $\{|j\rangle\}_{j=1}^{D_B}$ a basis of $\mathcal{H}_B$, and the **partial trace operation**:

$$\mathrm{Tr}_{A(B)} : \mathcal{B}\left(\mathcal{H}_A \otimes \mathcal{H}_B\right) \to \mathcal{B}\left(\mathcal{H}_{A(B)}\right) \tag{1.54}$$

$$O \to \mathrm{Tr}_{A(B)}(O). \tag{1.55}$$

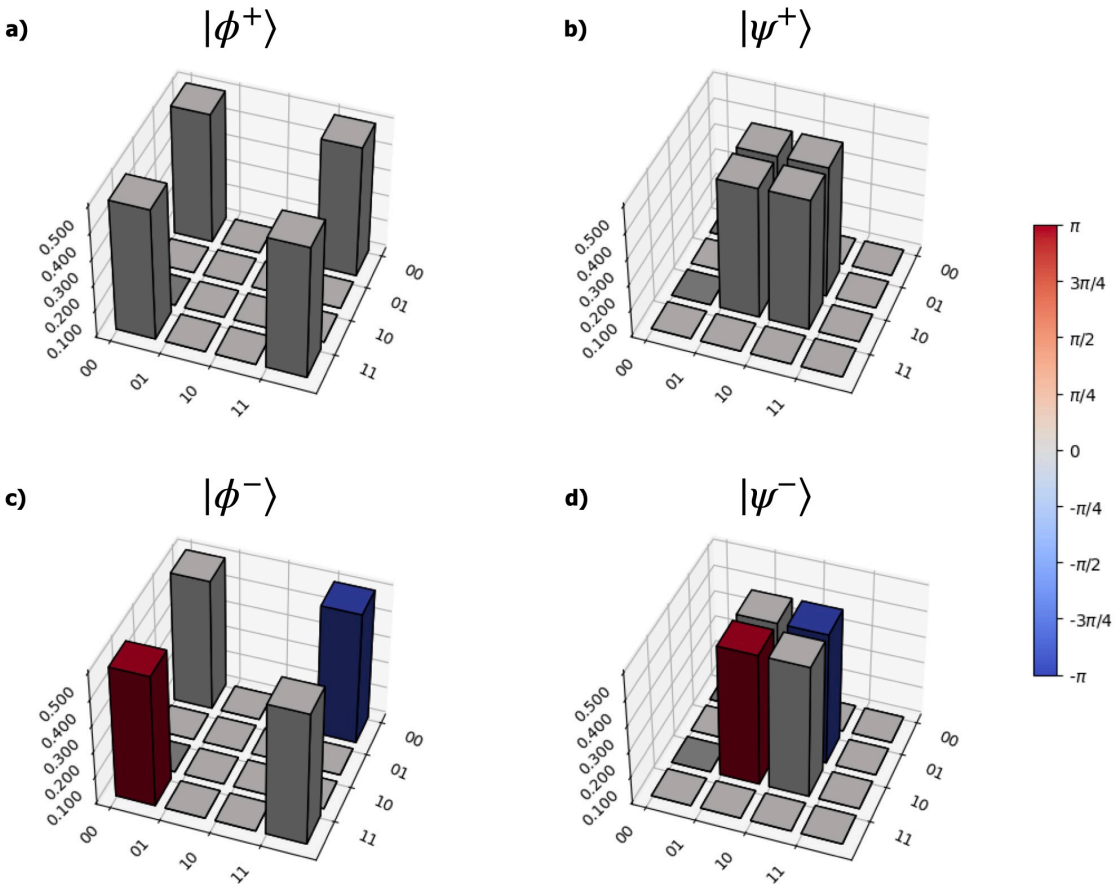


Figure 1.1: **Density matrices of Bell states.** Representation of the four Bell states **a)** $|\phi^+\rangle$, **b)** $|\psi^+\rangle$, **c)** $|\phi^-\rangle$, and **d)** $|\psi^-\rangle$ in the standard basis. The height of each bar represents the modulus of each element, while the phase is encoded in the color as shown in the colorbar.

1.2.5 Examples of quantum states

Let us now consider some relevant examples of quantum states. In general, we will consider states of systems composed of n_q qu*d*its, that is, n_q d-level carriers of information. The dimension of the associated Hilbert space is $D = d^{\,n_q}$.

The most relevant states of 2 qubits $(n_q = 2, d = 2)$ for quantum technology applications are the **Bell states**:

$$\left|\phi^{\pm}\right\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle \pm |11\rangle\right), \qquad \left|\psi^{\pm}\right\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle \pm |10\rangle\right). \tag{1.56}$$

The representation of the four Bell states in the standard basis is shown in Figure 1.1. All the Bell states are pure, with only two non-zero diagonal elements, and thus only one non-zero off-diagonal element above (or below) the diagonal. The Bell states are the maximally entangled 2-qubit states.

If we consider 3 or more qubits, we can consider other types of maximally entangled

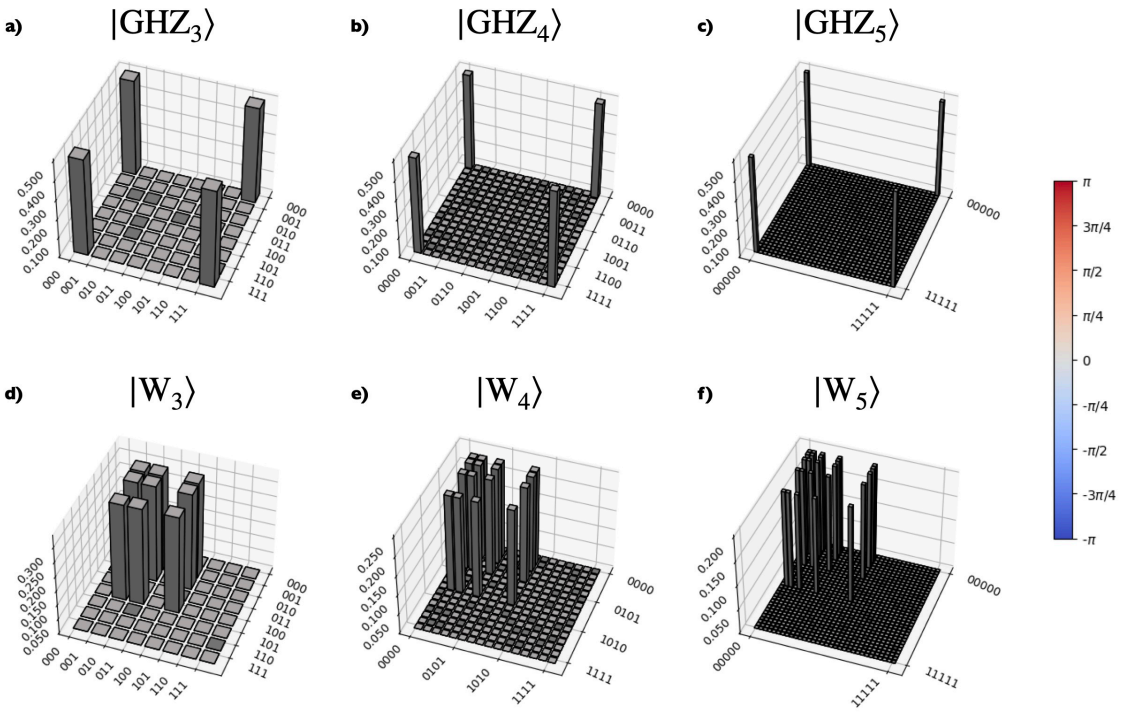


Figure 1.2: **Density matrices of GHZ and W states. a), b), c)**: Representation of density matrix of GHZ states of 3, 4, and 5 qubits in the standard basis, respectively. **d), e), f)**: Representation of density matrix of W states of 3, 4, and 5 qubits in the standard basis, respectively. The modulus and the phase are encoded as in Figure 1.1. Note that in panels **b), c), e)**, and **f)** we reported only some of the ticks on the horizontal axes.

states. Examples of paramount importance are the GHZ and W states, defined as:

$$\begin{aligned} \left|\mathrm{GHZ}_{n_q}\right\rangle &= \frac{1}{\sqrt{2}}\left(|0\rangle^{\otimes n_q} + |1\rangle^{\otimes n_q}\right) \\ \left|\mathrm{W}_{n_q}\right\rangle &= \frac{1}{\sqrt{n_q}}\left(|100\ldots0\rangle + |010\ldots0\rangle + \cdots + |000\ldots1\rangle\right). \end{aligned} \tag{1.57}$$

The GHZ and W states represent two different kinds of n_q-partite entanglement, as they cannot be converted into each other by local operations and classical communication (LOCC) [13]. In the tripartite case for example, the W state retains maximally bipartite entanglement if any of the three qubits is traced out, while the GHZ state becomes fully separable. They are exploited in several applications of quantum technologies, such as quantum computing [14, 15], communication [16], sensing [17], metrology [18], and secret sharing [19]. Finally, we note that the density matrix of a n_q-qubit GHZ state has only 2 non-vanishing off-diagonal elements, regardless of the number of qubits, while the one of a n_q-qubit W state has $2 \times \binom{n}{2} = n!/(n-2)! = n(n-1)$ non-vanishing off-diagonal elements, which scales quadratically with n_q.

Finally, we consider qudits with $d \geq 3$. Relevant examples are the maximally entangled states of two qudits, generalizations of the Bell states, defined as:

$$|\psi_d\rangle_{\text{ent}} = \frac{1}{\sqrt{d}} \sum_{m=0}^{d-1} |mm\rangle\,. \tag{1.58}$$

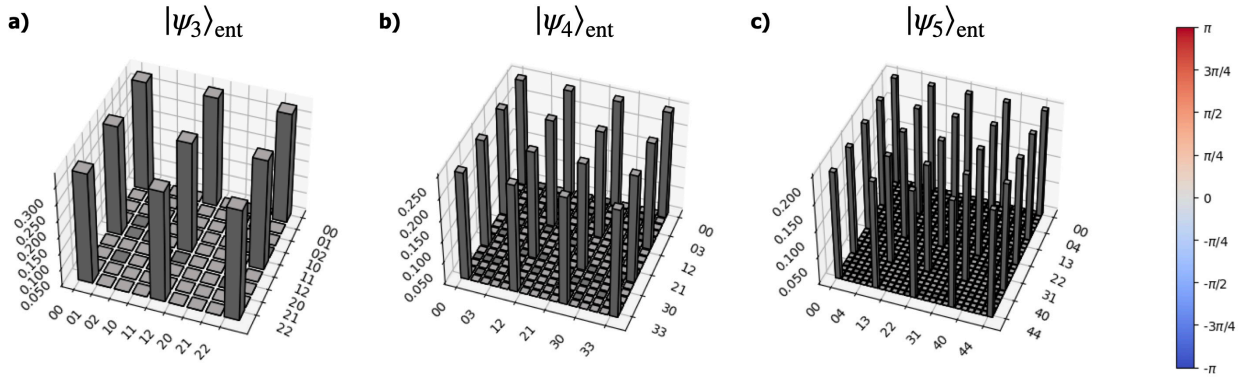


Figure 1.3: **Density matrices of maximally entangled states.** Representation of maximally entangled states of **a)** two qutrits ($d = 3$), **b)** two ququarts ($d = 4$), and **c)** two ququints ($d = 5$) in the standard basis. The encoding of the modulus and the phase is the same as of Figure 1.1.

As an example, the maximally entangled state of two qutrits is:

$$|\psi_3\rangle_{\text{ent}} = \frac{1}{\sqrt{3}} \sum_{m=0}^{2} |mm\rangle = \frac{1}{\sqrt{3}} \left(|00\rangle + |11\rangle + |22\rangle\right). \tag{1.59}$$

1.2.6 Measures of similarity between quantum states

Given two quantum states ρ and σ, we need to quantify how similar they are. A widely used figure of merit in this sense is the **fidelity**, defined as:

$$\mathcal{F}(\rho, \sigma) \equiv \mathcal{F}_\sigma^\rho = \text{Tr}\left(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right). \tag{1.60}$$

The fidelity is symmetric, $\mathcal{F}(\rho, \sigma) = \mathcal{F}(\sigma, \rho)$, is bounded between 0 and 1 and is preserved under unitary evolution of both states. Another possible figure of merit is the **trace distance** between ρ and σ, defined as:

$$D(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1, \tag{1.61}$$

with $\|\cdot\|_1$ the trace norm, $\|A\|_1 = \text{Tr}\left(\sqrt{A^\dagger A}\right)$ for a given operator A. The trace distance and the fidelity are related by the Fuchs-van de Graaf relations:

$$1 - \sqrt{F(\rho, \sigma)} \leq D(\rho, \sigma) \leq \sqrt{1 - F(\rho, \sigma)^2}. \tag{1.62}$$

1.2.7 Measurements

We already know that any physical quantity that can be measured, called an observable, is associated with an Hermitian operator, whose eigenvalues represent the possible

outcomes of a measurement of the observable. The probability distribution over the outcomes of that measurement can be computed from the density operator representing the state of the system via the Born rule:

$$P(x_i) = \text{Tr}(\Pi_i \rho), \tag{1.63}$$

with Π_i a projection operator onto the eigenvector corresponding to the measurement outcome x_i. The expectation value of an observable A is then given by:

$$\langle A \rangle = \text{Tr}(A\rho). \tag{1.64}$$

Pure states are described by rank-1 projectors, and assigning a pure state to a quantum system implies certainty about the outcome of some measurement on that system. The representation of such a state in its eigenbasis will therefore be a diagonal matrix with a single non-vanishing diagonal element, equal to 1.

This definition of measurement can be extended to the so-called generalized quantum measurement, or **positive-operator-valued measure (POVM)**. In the case of a Hilbert space with finite dimension D, a POVM is a set of a finite number $\mathcal{M}$ of positive semi-definite operators $\{M_i\}_{i=0}^{\mathcal{M}-1}$ that sum to the identity:

$$\sum_{i=0}^{\mathcal{M}-1} M_i = \mathbb{I}_D, \quad M_i \geq 0 \ \ \forall i. \tag{1.65}$$

Note that the number of elements $\mathcal{M}$ of the POVM can be different in general from the dimension D of the Hilbert space. A POVM can be seen as the most general way of assigning probabilities to a given quantum state ρ. More specifically, we can think of a POVM as a two-argument machine that takes a measurement outcome i and a quantum state ρ and returns the probability of measuring outcome i on the state ρ as $\text{Tr}(M_i\rho)$, which is another statement of the Born rule. A POVM can be interpreted as a probability distribution over the measurement outcomes in the following sense: we assign to each measurement outcome i a positive semi-definite operator M_i (which explains the name positive operator-valued measure), and then specify the quantum state and the Born rule to obtain actual probabilities. We now look at some examples of POVM.

1.2.2 Example. We can construct by hand a simple example of a POVM for a single qubit with the following operators:

$$M_0 = \lambda |0\rangle\langle 0| \tag{1.66}$$
$$M_1 = (1-\lambda)|0\rangle\langle 0| + |1\rangle\langle 1|, \tag{1.67}$$

with $\lambda \in [0, 1]$. These operators are positive and sum up to the 2-dimensional identity, and thus form a POVM. The rationale behind this construction is the following. The outcome 0 represents the system being found in the state $|0\rangle$, while outcome 1 means that it can be found in either $|0\rangle$ or $|1\rangle$ with different probabilities. For a qubit density matrix we have:

$$P_0 = \text{Tr}(M_0\rho) = \lambda\rho_{00} \tag{1.68}$$
$$P_1 = \text{Tr}(M_1\rho) = 1 - \lambda\rho_{00}. \tag{1.69}$$

Notice that even if $\rho_{00} = 1$, that is the system is surely in the state $|0\rangle$, we can obtain outcome 1 with a certain probability.
We can also construct POVM with a number of elements larger than the dimension of the Hilbert space. As an example:

$$M_0 = q\,|0\rangle\langle 0| \tag{1.70}$$
$$M_1 = q\,|+\rangle\langle +| \tag{1.71}$$
$$M_2 = 1 - M_0 - M_1. \tag{1.72}$$

Suppose you are given a state that is either $|1\rangle$ or $|-\rangle$. If you measure the system and find outcome 0, then you know that the state is $|-\rangle$, for $\langle 0|1\rangle = 0$. Similarly if you find outcome 1, then the state is $|1\rangle$, for $\langle +|-\rangle = 0$. Finally, if you find outcome 2, then you do not learn anything about the state.

1.2.3 Example. We can construct a POVM starting from one of the Platonic solids, that is, the tetrahedron, the cube, the octahedron, the dodecahedron, and the icosahedron. Let us analyze the case of the tetrahedron for simplicity, and consider the POVM corresponding to the vertices of the tetrahedron in the Bloch sphere. To find the elements of the POVM, consider the following set of vectors [20]:

$$V_4 = \left\{ \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} \alpha \\ -\beta \end{pmatrix}, \begin{pmatrix} \beta \\ i\alpha \end{pmatrix}, \begin{pmatrix} \beta \\ -i\alpha \end{pmatrix} \right\}, \quad \alpha = \sqrt{\frac{3+\sqrt{3}}{6}},\ \beta = \sqrt{\frac{3-\sqrt{3}}{6}} \tag{1.73}$$

These vectors identify the vertices of a tetrahedron as follows. Consider $|\psi\rangle \in V_4$, construct the density matrix $\rho = |\psi\rangle\langle\psi|$, and compute the corresponding Bloch vector, the components of which are the coordinate of the corresponding vertex of the tatrahedron. We find:

$$\begin{aligned}
|\psi_0\rangle &= \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \implies \underline{v}_0 = \left(\sqrt{\frac{2}{3}}, 0, \sqrt{\frac{1}{3}} \right) \\
|\psi_1\rangle &= \begin{pmatrix} \alpha \\ -\beta \end{pmatrix} \implies \underline{v}_1 = \left(-\sqrt{\frac{2}{3}}, 0, \sqrt{\frac{1}{3}} \right) \\
|\psi_2\rangle &= \begin{pmatrix} \beta \\ i\alpha \end{pmatrix} \implies \underline{v}_2 = \left(0, \sqrt{\frac{2}{3}}, -\sqrt{\frac{1}{3}} \right) \\
|\psi_3\rangle &= \begin{pmatrix} \beta \\ -i\alpha \end{pmatrix} \implies \underline{v}_3 = \left(0, -\sqrt{\frac{2}{3}}, -\sqrt{\frac{1}{3}} \right)
\end{aligned} \tag{1.74}$$

These are the vertices of a tetrahedron in the Bloch sphere. It holds:

$$\sum_{i=0}^{3} |\psi_i\rangle\langle\psi_i| = 2\mathbb{I}_2, \tag{1.75}$$

thus the elements of the tetrahedron POVM are $\left\{ M_i = \frac{1}{\sqrt{2}} |\psi_i\rangle\langle\psi_i| \right\}_{i=0}^{3}$. Similar calculations lead to the POVMs for the remaining Platonic solids.

There exist physical platforms, such as the superconducting or trapped-ion ones, where the cost of measuring the expectation value of a single projector is comparable to the cost of simultaneously measuring the expectation value of several projectors. In these cases, the most convenient observable to introduce is the **measurement setting**. If we consider a system composed of n_q qu$d$its, a measurement setting s (or measurement basis in the literature) is the tensor product of the generators of SU(d) that specifies which generator is measured on each qudit. Each setting s has d^{n_q} possible outcomes corresponding to the number of ways of picking one of the d eigenvectors of the SU(d) generators that appear in the setting. Consequently, from a practical point of view a measurement setting s can be interpreted as a set of d^{n_q} projectors $\{\Pi_k\}_{k=1}^{d^{n_q}}$ onto the eigenbasis of the observable that defines the setting. By repeatedly measuring all the qudits one retrieves information about the statistical distribution of all d^{n_q} possible outcomes associated with that setting, that is, we gain access to the probability distribution $\{p_k = \mathrm{Tr}\,(\rho\Pi_k)\}_{k=1}^{d^{n_q}}$. The values $\{p_k\}_{k=1}^{d^{n_q}}$ are equal to the diagonal elements of the density matrix ρ when represented in the eigenbasis of the observable that defines the setting.

Without loss of generality, let us now consider the case of qubits ($d = 2$), where the SU(2) generators are the Pauli matrices $\mathcal{S} = \{X, Y, Z\}$. Let then $\mathcal{S}_{n_q}$ be the Cartesian product of $\mathcal{S}$ with itself n_q times:

$$\mathcal{S}_{n_q} = \underbrace{\mathcal{S} \times \mathcal{S} \times \cdots \times \mathcal{S}}_{n_q \text{ times}} \tag{1.76}$$

Therefore $\mathcal{S}_{n_q}$ is the set of the 3^{n_q} possible settings of n_q qubits, and $\mathcal{S}_{n_q} \ni s = s_0 s_1 \ldots s_{n-1}$ is a n_q-tuple representing a specific measurement setting of n_q qubits. Let then $\mathcal{O} = \{+1, -1\}$ be the set of outcomes of (one of) the Pauli matrices, and $\mathcal{O}_{n_q}$ the Cartesian product of $\mathcal{O}$ with itself n_q times, that is, the set of outcomes of the product of n_q Pauli matrices:

$$\mathcal{O}_{n_q} = \underbrace{\mathcal{O} \times \mathcal{O} \times \cdots \times \mathcal{O}}_{n_q \text{ times}}, \tag{1.77}$$

with $\mathcal{O}_{n_q} \ni o$ representing one of the possible outcomes of the setting. The probability of outcome o of setting s is given by:

$$P_\rho\,(o|s) = \mathrm{Tr}\,(\rho\Pi_{s_o}) = \langle s_o|\rho|s_o\rangle\,, \tag{1.78}$$

with $|s_o\rangle$ the eigenvector corresponding to the outcome o. Let us now apply these results to the case of $n_q = 2$ qubits. The sets $\mathcal{S}_2$ and $\mathcal{O}_2$ are:

$$\begin{aligned} \mathcal{S}_2 &= \{XX, XY, XZ, YX, YY, YZ, ZX, ZY, ZZ\}\,, \\ \mathcal{O}_2 &= \{(+1,+1), (+1,-1), (-1,+1), (-1,-1)\}\,. \end{aligned} \tag{1.79}$$

Each setting $s \in \mathcal{S}_2$ has $2^{n_q} = 4$ possible outcomes corresponding to the number of ways of picking one of the two eigenvectors of each of the Pauli matrices appearing in the setting. As an example, the four possible outcomes and the associated eigenvectors of the setting $s = XY$ are $(+1,+1)$ with $|x_+ y_+\rangle$, $(+1,-1)$ with $|x_+ y_-\rangle$, $(-1,+1)$ with $|x_- y_+\rangle$, and

$(-1,-1)$ with $|x_-x_-\rangle$. Then the probabilities of the outcomes are:

$$o=(+1,+1) \implies |s_o\rangle = |x_+y_+\rangle \implies P_\rho\left((+1,+1)|XY\right) = \langle x_+y_+|\rho|x_+y_+\rangle \tag{1.80}$$
$$o=(+1,-1) \implies |s_o\rangle = |x_+y_-\rangle \implies P_\rho\left((+1,-1)|XY\right) = \langle x_+y_-|\rho|x_+y_-\rangle \tag{1.81}$$
$$o=(-1,+1) \implies |s_o\rangle = |x_-y_+\rangle \implies P_\rho\left((-1,+1)|XY\right) = \langle x_-y_+|\rho|x_-y_+\rangle \tag{1.82}$$
$$o=(-1,-1) \implies |s_o\rangle = |x_-y_-\rangle \implies P_\rho\left((-1,-1)|XY\right) = \langle x_-y_-|\rho|x_-y_-\rangle\,. \tag{1.83}$$

Finally, the measurement setting $s=XY$ can be interpreted as the set of 4 projectors onto its eigenbasis $\{|DR\rangle\,,|DL\rangle\,,|AR\rangle\,,|AL\rangle\}$, that is:

$$\{\Pi_{DR}=|DR\rangle\langle DR|\,,\Pi_{DL}=|DL\rangle\langle DL|\,,\Pi_{AR}=|AR\rangle\langle AR|\,,\Pi_{AL}=|AL\rangle\langle AL|\}\,. \tag{1.84}$$

The values $\{p_k\}=\{\mathrm{Tr}\,(\rho\Pi_k)\}=\{p_{DR},p_{DL},p_{AR},p_{AL}\}$ are equal to the diagonal elements of the density matrix ρ when represented in the eigenbasis of XY.

1.2.8 Quantum channels

We now characterize the transformations of the set of statistical operators, that is, transformations that map quantum states into quantum states. These are called **quantum channels**, or **quantum dynamical maps**, and are useful to characterize noise sources.

A transformation Λ of the set of statistical operators $\mathcal{S}\left(\mathcal{H}\right)$ is a quantum channel if it satisfies the following three properties of linearity, trace preservation, and complete positivity, called CPTP in general:

$$\Lambda(\alpha\rho+\beta\sigma)=\alpha\Lambda(\rho)+\beta\Lambda(\sigma),\quad \alpha,\beta\in\mathbb{C} \tag{1.85}$$
$$\mathrm{Tr}\,(\Lambda\rho)=\mathrm{Tr}\,(\rho)=1 \tag{1.86}$$
$$\left(\Lambda_S\otimes\mathbb{I}_E\right)(\rho_{SE})\geq 0 \text{ for any } \mathcal{H}_E,\rho_{SE}\in\mathcal{S}\left(\mathcal{H}_S\otimes\mathcal{H}_E\right) \tag{1.87}$$

Linearity guarantees that the convex structure of the set of quantum states is preserved. Along with trace preservation, it ensures that the probabilistic interpretation of quantum mechanics is maintained. Complete positivity is required such that the states, even when considered as parts of composite systems, correctly transform as states. A functional Λ that satisfies these three properties is a quantum dynamical map, or quantum channel. An important theorem by Kraus guarantees the equivalence of the properties CPTP with the so-called **Kraus form**.

1.2.4 Theorem. *A functional $\Lambda:\mathcal{B}\left(\mathcal{H}\right)\to\mathcal{B}\left(\mathcal{H}\right)$ is a quantum dynamical map if and only if it can be written as:*

$$\Lambda(\rho)=\sum_k \Omega_k\rho\Omega_k^\dagger, \tag{1.88}$$

with $\{\Omega_k\}$ called the Kraus operators that satisfy the completeness relation $\sum_k\Omega_k^\dagger\Omega_k=\mathbb{I}$.

It is important to note that the Kraus operators Ω_k do not need to be Hermitian or unitary. Let us now look at two relevant examples of quantum channels.

1.2.5 Example (Depolarizing channel)**.** Given a state ρ of a quantum system of dimension D, the depolarizing channel transforms it into a linear combination of itself and the maximally entangled state:

$$\Lambda_{\text{depol}}(\rho) = (1-p)\rho + \frac{p}{D}\mathbb{I}_D, \tag{1.89}$$

with $0 \leq p \leq 1 + 1/(D^2-1)$ a parameter quantifying the strength of the channel. If $p = 0$ the depolarizing channel does not modify the state, while if $p = 1$ the depolarizing channel returns the maximally mixed state. We can provide a geometrical interpretation of the action of the depolarizing channel via the Bloch vector. Indeed:

$$\rho = \frac{1}{2}(\mathbb{I} + \underline{v}\cdot\underline{\sigma}) \implies \Lambda_{\text{depol}}(\rho) = \frac{1}{2}[\mathbb{I} + (1-p)\,\underline{v}\cdot\underline{\sigma}] \tag{1.90}$$

$$(v_x, v_y, v_z) \implies ((1-p)v_x, (1-p)v_y, (1-p)v_z)\,. \tag{1.91}$$

Thus, the depolarizing channel acts as a uniform contraction of the Bloch vector towards the center of the Bloch sphere. Finally, the depolarizing channel is unital, that is, $\Lambda_{\text{depol}}(\mathbb{I}) = \mathbb{I}$, and it can never increase the purity of the state. Figure 1.4 illustrates the action of the depolarizing channel on a pure state represented in its eigenbasis. The channel gradually suppresses the coherences, that is, the off-diagonal elements of the state, as the value of the noise parameter p increases, which in turn gradually reduces its purity.

1.2.6 Example (Experimental state-error)**.** Given a state ρ of a quantum system of dimension D, the experimental state-error channel [21] maps the state into a convex combination of itself and a random state:

$$\Lambda_{\text{exp-state}}(\rho) = (1-\varepsilon)\rho + \varepsilon\rho_{\text{random}} \tag{1.92}$$

$$\rho_{\text{random}} = \frac{R^\dagger R}{\text{Tr}\,(R^\dagger R)} \tag{1.93}$$

$$R = 2\text{rand}(D) - 1 + i\,[2\text{rand}(D) - 1]\,. \tag{1.94}$$

The function $\text{rand}(D)$ generates a $D \times D$ matrix of random values sampled from a uniform distribution over the interval (0,1). This channel simulates experimental error in the preparation of the state. In fact, the actually prepared state always differs from the intended state by a small amount, here quantified by the ε parameter and implemented via a random density matrix.

1.3 Quantum State Tomography

Now we dive into QST, which is the main topic of this work. The following paragraphs is a short introduction to QST and not a systematic presentation of the methods with which QST is implemented. The goal is to clarify the specific meaning of QST in this work and to describe a paradigmatic QST experiment to establish the nomenclature.

In general, QST is the process of reconstructing the representation of a quantum state for a given basis of the Hilbert space from the measurement of a certain number of observables. Despite its simplicity, this definition raises several questions. Which measurements

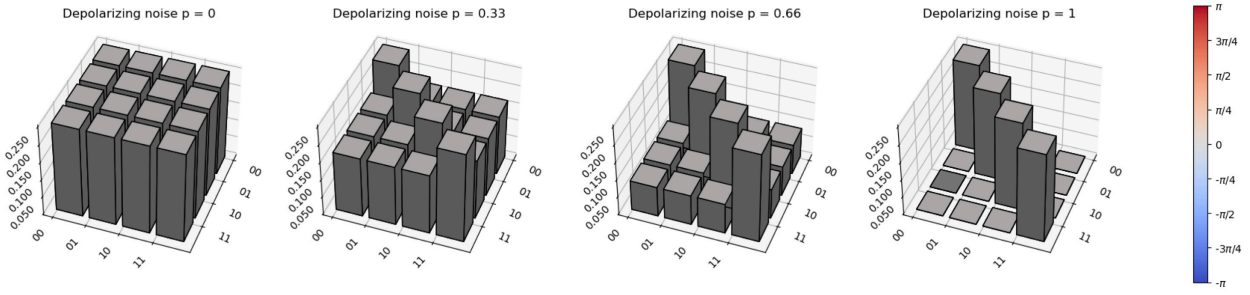


Figure 1.4: **Effect of the depolarizing channel on a density matrix.** Representation of the pure state $|\psi\rangle = (|00\rangle + |01\rangle + |10\rangle + |11\rangle)/2$ in the standard basis when the depolarizing channel is applied with different values of noise parameter p, from 0 to 1.

should one perform: projective, POVMs, or measurement settings? Do measurements involving entangled observables provide any advantage over local measurements? How many measurements should one perform? Once we have performed the measurements and recorded the outcomes, what tools and strategies should one use to reconstruct the state? Before reviewing the literature and looking for answers to these questions, it is important to clarify what we mean by QST and "measurement". Indeed, what was referred to as "traditional" or "conventional" QST changed over the years because researchers developed several different ways to perform QST. Furthermore, the term "measurement" in the literature can refer to projective measurements, measurement settings, or other kinds of measurement. However, the precise meaning of "measurement" is not always clearly specified and can consequently be misleading.

In this work, **QST is the process of reconstructing the representation of a quantum state of n qu$d$its, that is, the density matrix in a given basis of the Hilbert space, from the expectation value of *at least* d^{2n} projectors**. This definition encompasses complete QST, where the number of projective measurements is exactly d^{2n}, and over-complete QST, where it is larger, typically equal to d^{3n}. Furthermore, the way in which one performs such measurements is irrelevant to the definition, that is, the choice between projective measurements, POVMs, or measurement settings, is related to the convenience of implementing the measurements and to the physical platform on which one is working.

Let us now describe the workflow of a typical QST experiment without referring to a specific physical platform and considering a quantum system whose state is encoded in a given degree of freedom. The first operation is **state preparation**, which in turn includes two sub-steps, state initialization and state evolution. With state initialization, we prepare the system in a fiducial state, typically $|00\ldots0\rangle$, the ground state of a Hamiltonian that we can control, or the vacuum state. With state evolution, we apply one or more unitary transformations to the initial fiducial state such that it evolves to a final state of interest. The second operation is the **implementation of the projectors.** After the state preparation, a layer composed of some experimental gadgets projects the prepared state onto one or more of the basis states of the chosen degree of freedom. This setup implements one or more projectors at once, depending on if one performs single projective measurements or measurement settings. The third and final operation is the **detection**, where a series of detectors or sensors is placed after the experimental gadgets.

These three operations, called a **measurement** on the whole, represent a single step of a typical QST experiment and consume a single copy of the state. Other names used

for this series of operations are **event** or **shot**. A complete QST experiment requires running the three operations many times to determine the expectation values of each chosen observable. Consequently, we consume several copies of the state, due to the intrinsic probabilistic nature of quantum mechanics. In the following, we use the phrase "measure an observable" to mean that we measure the expectation value of that observable, that is, we perform a certain number of measurements or equivalently consume a certain number of copies of the state to determine it. For instance, when we say "measure the projector(s)", it means that we measure the expectation value(s) of the projector(s). Analogously, we use the phrase "measure the (measurement) setting(s)" to mean that we measure the expectation values of all the projectors contained in the setting.

Let us now apply the workflow described here to a typical QST experiment, following the example proposed in [22]. Here, the state preparation is implemented with a black box that generates pairs of photons in an (almost) arbitrary quantum state of their polarization degree of freedom. The implementation of the projectors takes place by putting proper gadgets (waveplates and polarizers) such that they project the photons onto a polarization state chosen by the user, thus implementing a specific projector. Finally, photons are measured by photodetectors that record coincidence counts. Suppose now that the gadgets were set to project onto the state $|HH\rangle$, we generate a pair of photons, and both the photodetectors click. Can we conclude that the state of the pair of photons was $|HH\rangle$? The answer is no. The intrinsic probabilistic nature of quantum mechanics allows one to define the probability of observing a certain outcome only upon repeating the same experiment many times, that is, consuming many copies of the state. We generate 10^4 pairs of photons to measure the projector onto $|HH\rangle$ and record 5×10^3 clicks of the photodetectors. Then the probability of finding the system in the state $|HH\rangle$ is $5 \times 10^3/10^4 = 0.5$. This is the implementation of a single projective measurement. After this, one changes the gadget setup to project onto a different polarization state and repeat the procedure just described.

1.4 Literature review: history and state of the art

In this section, we review the literature about QST starting from some seminal papers that paved the way for the study of quantum state characterization, or determination, from which QST (as meant nowadays) later derived. Then we will review the state of the art of QST by looking through the main approaches developed in the past years, and highlighting pros and cons of each method. Some reviews on QST are [23], [24], [25], and [26]. We stress that this is not an exhaustive list, but rather a guide to orient the reader to the most relevant and up-to-date QST approaches that one can find in the literature.

1.4.1 Pioneering works

In 1933 Wolfgang Pauli [27] poses the following problem, named the "Pauli problem" after him. The state of a quantum particle is characterized by a complex wave function in a suitable Hilbert space. However, one cannot directly observe a complex quantity in experiments, for observables are given by the expectation values of Hermitian operators computed on the quantum state. Examples for a particle are the probability distributions of position and momentum. Then, is it possible to determine the wavefunction from the

distributions of position and momentum? In his paper Pauli states that this remains an open question. In his 1937 textbook [28], Kemble stated that the probability distribution of position and its time derivative suffice to determine the wavefunction. However, in 1968 Gale, Guth, and Trammell [29] demonstrated that this approach is valid only for a particle moving in one dimension, while a complete characterization of the wave function of a spinless particle in three dimensions requires not only the probability distribution, but also the probability current of position. For particles with spin, the density matrix can be reconstructed by measuring the mean values of the non-vanishing multipole moments of the spin distribution. These multipole moments capture higher-order correlations between spin components, providing a more comprehensive picture of the state compared to individual distributions. In 1992 Weigert [30] proposed a simple solution to the Pauli problem for the case of a pure state of a spin-s system, with $s = n/2, n \in \mathbb{N}$, later extended to the determination of the pure state state of a particle moving in a one-dimensional potential in [31]. Here, the author shows that the measurement of two position probability densities at subsequent times, plus the expectation values of a finite number of non-local operators, is equivalent to the knowledge of the pure state. Orlowski and Paul [32] addressed the Pauli problem in quantum optics in 1994, where the position and momentum distributions can be directly measured on a single-mode radiation field by optical homodyning. In this case, we have to identify position and momentum with the quadrature components of the electric field.

In 1957 Fano [33] addressed for the first time the problem of "identification of states", that is, determining the parameters characterizing the density matrix. It is stated that in general, for a finite-dimensional system described by a density matrix with D rows and columns, the number of independent real parameters is equal to $D^2 - 1$. Thus, D^2 linearly independent operators suffice to reconstruct the density matrix. Some examples are provided: spin orientation of spin-1/2 particles, reconstructed by using the Pauli matrices; generalization to particles with different spins; polarization of electromagnetic radiation; polarization of photon pairs. In [34] Wichmann suggests that to reconstruct a quantum system when only partial information about the system is available, we have to select the "most chaotic" ensemble, that is, the one that maximizes entropy, that still aligns with the known ensemble averages of the considered set of observables.

In 1970 Band and Park [35] advocated for data-driven determination of quantum states, since any preparation scheme cannot guarantee a pure state. This information can only be revealed through the analysis of experimental data. Consequently, inferring the quantum state associated with a preparation scheme should rely on data analysis, not assumptions. They highlight the necessity of multiple measurements to accurately determine a system's state, inherent to the statistical interpretation of quantum mechanics, and propose a method using a minimal set of non-commuting observables to determine the quantum state of spin-1/2 particles, offering a practical approach. The same authors published two other works [36, 37] (actually one work divided into two parts) where they face the problem of quantum state determination. In the first part, they develop a multipole algebra to construct a quorum of observables for state determination. Given a D-dimensional Hilbert space, a **quorum** of observables can be found *mathematically* by choosing $D^2 - 1$ observables $\{A_i\}$ such that the $D^2 - 1$ corresponding expectation values, computed on the density matrix via Born rule $\langle A_i \rangle = \mathrm{Tr}\,(\rho A_i)$, result in a system of linear algebraic equations possessing a unique solution set for the $D^2 - 1$ real independent

unknown parameters required to define the density matrix[1]. In the second part, they provide explicit examples of quorums and discuss criteria for distinguishing between pure and mixed states. They also emphasize the main limitation of their work. They say that the quorum they developed is still related to "classical" concepts like position or energy, but when working with quantum systems, these may not be so easy to measure. Thus, they suggest and believe that it should be possible to carry out a program in which measurements are carried out with suitable probing devices or techniques. In a subsequent work of 1979 [39], they determined the quorum for a spinless particle in one dimension. In particular, they show that the corresponding density matrix can be determined empirically by analyzing position data associated with various instants subsequent to the state preparation.

In 1989 Royer [40] proposed the use of phase-space representations, particularly the Wigner function, for state measurement and detailed operational procedures to measure it. The theoretical gap between theoretical descriptions and experimental measurements was bridged in the same year by Vogel and Risken [41]. They demonstrate not only that the probability distribution of a rotated quadrature phase can be determined by the quasi-probability distributions, such as the P, the Q, and the Wigner function, but also that the quasi-probability distributions can be retrieved via the inverse Radon transformation if we know all marginal distributions corresponding to the various settings of the local oscillator phase (covering angles between 0 and π). By measuring specific probability distributions related to well-defined quadrature phases, we can derive the quasi-probability distributions of the corresponding state. Smithey et al. have conducted an experimental implementation of this method [42]. In this work, optical homodyne tomography, a technique that employs balanced homodyne detection to measure probability distributions of quadrature field amplitudes, is applied to characterize electromagnetic field modes. Specifically, the authors utilized optical homodyne tomography for the first time to directly determine the Wigner function and density matrix of both vacuum and quadrature-squeezed states of a light mode.

In 1991 Peres and Wootters [43] explored how to best determine the quantum state of two identical quantum systems prepared at different locations. The authors compare the effectiveness of various measurement techniques, including separate measurements on each system and global measurements that involve measuring multiple particles together. They find that a measurement technique that uses entangled operators yields more information about the state of the system than separate measurements, even when the latter are enhanced by a novel iterative strategy.

In the same year Jones [44] used Bayesian inference to define an optimal state determination protocol for arbitrary finite-dimensional Hilbert spaces through some concepts from communication theory, such as mutual information, communication channels, and the optimal receiver problem. The author extended the analysis in a later work [45] (1994), in which an information-theoretic framework is adopted. The author considers a scenario with limited resources, when only a finite number N of identical copies of a quantum state are available, and frames the problem as a communication task between two hypothetical experimenters, "Norman" and "Noeline". Norman, who acts as a transmitter, encodes a message by preparing a quantum state and sends N identical copies of

[1]Such a set of observables will be termed **informationally complete** for the first time by Prugovecki [38] in 1977.

this state to Noeline, who acts as a receiver, and performs measurements to determine the state and decode the information. In this framework, the inherent uncertainty in quantum measurements is viewed as noise affecting the transmission of information about the state. The goal of the protocol is to maximize information transfer, and the effectiveness of Noeline's measurement scheme is assessed by the amount of information she can extract about the prepared state, much like evaluating the capacity of a communication channel. The author leverages Shannon information theory to quantify the mutual information between the prepared state and the measurement outcomes to establish fundamental bounds on how much information can be extracted about a quantum state from a finite number of measurements, similar to how Shannon's theory establishes limits on classical communication channels. The upper bound for Shannon information obtainable from N identical copies of a d-level system is:

$$\{\psi, \Phi_N\} \leq \ln \left[\frac{\Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{\nu d}{2} + N\right)}{\Gamma\left(\frac{\nu d}{2}\right) \Gamma\left(\frac{\nu}{2} + N\right)} \right] + N \left\{ \Psi\left(\frac{\nu}{2} + N\right) - \Psi\left(\frac{\nu d}{2} + N\right) \right\} \tag{1.95}$$

with $\{\psi, \Phi_N\}$ the mutual information between the unknown quantum state (ψ) and the data Φ_N obtained from N measurements, which quantifies the information gain from the measurements, $\Gamma(\cdot)$ denotes the gamma function, $\Psi(\cdot)$ represents the digamma function, and ν is the type index of the Hilbert space, which takes the value 1, 2, or 4 for systems with time-reversal invariance, corresponding to real, complex, or quaternionic Hilbert spaces, respectively. As the number of copies N increases, the information gain approaches:

$$\{\psi, \Phi_N\} \leq \frac{\nu(d-1)}{2} \ln N - \frac{\nu(d-1)}{2} - \left[\frac{\Gamma\left(\frac{\nu d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \right] + O\left(\frac{1}{N}\right), \tag{1.96}$$

which shows that the information grows logarithmically with N, indicating that an infinite number of measurements would be required to completely determine the quantum state. Finally, the author relies on a Bayesian methodology for quantum state inference. This approach systematically updates the experimenter's knowledge about the quantum state based on the accumulated measurement data. It leverages prior information, which can include some prior knowledge about the state or complete ignorance, represented by a prior probability distribution over the possible states, and updates this information using the likelihood function, which quantifies the probability of observing the measured data given a particular state. This iterative process refines the experimenter's understanding of the state as more data is gathered.

In 1995 Leonhardt [46] developed the formalism of the discrete Wigner function such that the quadrature histograms, which can be observed experimentally, are projections of the Wigner function. The density matrix can then be reconstructed from the quadrature histograms by means of the discrete Fourier transform.

Given a finite ensemble of identically prepared quantum states, their state can be determined only approximately. Some questions naturally arise in this context: How much knowledge can be obtained from such a finite ensemble? How quickly does one approach exact knowledge as the ensemble becomes large? What experimental strategies furnish the maximum knowledge? Massar and Popescu addressed these questions in the case of spin-1/2 particles in 1995 [47]. They answered these questions by developing a game.

This game consists of many runs. In each run, the player receives N spin-1/2 particles, all polarized in the same directions. The player knows that all the N spins are parallel, in each run are polarized randomly, uniformly distributed in space, and the Hamiltonian is the same in every run. The player is allowed to do any measurement and finally required to guess the polarization direction, that is, to indicate a direction. The score of each run is $\cos^2(\alpha/2)$, with α the angle between the guessed and the correct direction. The final score is the average of the scores for each run. Basically, the score is between 0 and 1. If no measurement is performed, the polarization direction is just guessed and the score is 1/2. Any improvement with respect to 1/2 represents an information gain about the polarization direction. To answer the first two questions, the authors calculated the maximum score available for this game: $S_{max} = (N+1)/(N+2)$. This score tends to 1 as $N \to \infty$: this is coherent with the expected result in the case of an infinite ensemble, where the direction of the polarization can be determined exactly. The optimal experiment to achieve the maximum score would require an infinite isotropic set of projectors. Finally, to answer the third question, the authors show that there exist no optimal experiments consisting of separate measurements on each spin, that is, an optimal measurement on an ensemble of parallel spins necessarily treats the ensemble as a single composite system. This would correspond to a von Neumann measurement in an infinitely dimensional extension of the Hilbert space of the state. Thus, even if they found answers to the initial questions, the optimal procedure cannot be experimentally implemented.

The infinite number of projectors required by this strategy makes it unpractical. Derka et al. [48] addressed this issue in 1998 by proposing an algorithm to find the POVM that returns the best estimate of an unknown pure quantum state, provided that we have at our disposal only a finite (but large enough) number of copies of the state. They provide a compact set of equations that define the POVM , together with an upper bound of the mean fidelity. This result partially improves the result of [47], as there is no systematic way to determine the number of POVM elements, and the authors do not provide details on the practical implementation of the POVM.

Latorre, Pascual, and Tarrach further improved these results in [49], where they looked for the most efficient way to extract information about the unknown pure state of a qubit using a minimal number of measurements. They find optimal, *finite*, and minimal POVMs as follows. First, they constructed a set of equations that an optimal POVM must satisfy. Second, they systematically analyzed these equations for an increasing number N of copies to find solutions with the smallest possible number n of projectors. Finally, they identify the minimal n for each N and the corresponding arrangement of the projectors using mathematical techniques and numerical analysis. Interestingly, the minimal POVMs for smaller N correspond to platonic solids: N=2 a tetrahedron, N=3 an octahedron, N=5 an icosahedron. For N=4, a slightly irregular "twisted prism with pyramidal caps" emerges as the optimal solution. The authors also propose a formula to estimate the minimal number of projectors n_{min} needed for a given N. This formula, based on their analysis of the equations and observed patterns, predicts n_{min} accurately for the cases they explicitly solved (N = 2 to 7).

Finally, this line of research was also studied by Gill and Massar, who examined the problem of estimating the state of a large but finite number N of identical quantum systems [50]. We dispose of a finite number N of copies of an unknown quantum state ρ, pure or mixed, and our goal is to determine ρ as well as possible. Let us suppose that the unknown state depends on a vector of p unknown real parameters: $\rho(\theta)$, with

$\theta = (\theta_1, \ldots, \theta_p)$. After performing measurements on the N copies of ρ, one guesses what is θ: Let us call the guessed value $\hat{\theta}^N$. If the estimation strategy is good, then the mean quadratic error (MQE) $V_{ij}^n(\theta)$ decreases as $1/N$:

$$V_{ij}^n(\theta) \simeq \frac{W_{ij}(\theta)}{N}, \tag{1.97}$$

with $W_{ij}(\theta)$ the scaled MQE matrix. The main question of the paper is: What scaled MQE matrices are attainable through arbitrary measurements and estimation procedures? If the parameter is one-dimensional $p = 1$, the problem has already been solved by providing the quantum Cramer-Rao bound, and a strategy for attaining that bound in the limit of large N has been proposed. In this paper, the authors present a bound for W in the multi-parameter case. The conclusions of the work can be summarized as follows. If the state is pure, and it belongs to the Hilbert space of a qubit, the bound is sharp. It provides a necessary and sufficient condition that the scaled MQE matrix has to satisfy in order to be attainable. The bound can be attained by performing separate measurements on each particle. If the state is mixed, it belongs to the Hilbert space of a qubit, and if one restricts oneself to measurements that act separately on each particle, then the bound is sharp. If the state is pure and belongs to a Hilbert space of dimension $D > 2$, the bound applies but it is not sharp. If the state is mixed, it belongs to a Hilbert space of dimension $D > 2$, and if one restricts oneself to measurements that act separately on each particle, then our bound applies but it is not sharp. If the state is mixed and one allows collective measurements, then the bound is not necessarily satisfied. The scaling laws for MSE provided in this work have been experimentally addressed in adaptive QST experiments [51, 52, 53, 54, 55].

1.4.2 Maximum likelihood

Maximum likelihood estimation (MLE) is a frequentist statistical inference technique to find the parameter values that maximize the likelihood of observed data. It is conceptually straightforward, flexible, and computationally affordable. However, it can be biased when applied to small samples because of its frequentist nature. How does MLE relate to QST? We can start from **linear inversion** to understand why and how MLE has been introduced in the context of QST. Consider the Born rule $\langle A_i \rangle = \mathrm{Tr}(\rho A_i)$ for a quorum, or informationally complete set, of observables $\{A_i\}$. The left-hand side of the Born rule is a number representing the measured expectation value, while the right-hand side is a linear function of the elements of the density matrix. This linear system of equations has a unique solution because of the informational completeness of the set of observables. However, linear inversion has a major flaw when applied to this problem. If the expectation values are affected by noise, and this happens whenever we use experimental data, then the method is not guaranteed to return a density matrix representing a valid physical state, that is, a trace-one, Hermitian, and positive semi-definite complex matrix. In particular, the positive semi-definiteness condition is the one that is not guaranteed, while being trace-one and Hermitian can be imposed by construction. An explicit example of a non-physical density matrix obtained with linear inversion is provided by James et al. in [22]. Smolin et al. provided an algorithm to find the density matrix closest to a given trace-one, Hermitian matrix under the 2-norm, thus correcting the linear inversion flaw [56].

Hradil [57], Banaszek et al. [58], and James et al. [22] reported the failure of linear inversion to return physical results in QST experiments, while Mogilevtsev et al. pointed it out in homodyne experiments [59]. We now analyze how they contributed differently to the development of MLE for QST.

In 1997 Hradil [57] highlighted the limitations of traditional techniques, such as linear inversion, and introduced an algorithm for QST, later called $R\rho R$ algorithm, based on maximum likelihood. The author showed that a state ρ_0 that maximizes the likelihood function satisfies the following equation:

$$R\left(\rho_0\right)\rho_0 R\left(\rho_0\right) = \rho_0, \tag{1.98}$$

with R a non-negative state-dependent operator, hence the name of the algorithm. The limitations of this method are the lack of a recipe to apply in the laboratory and the numerical challenges due to the solution of the nonlinear equation for the density matrix. In [60] Hradil et al. rephrase the problem of maximizing the likelihood as a multidimensional nonlinear optimization problem, producing a system of coupled equations to be solved, which in turn can be translated into a nonlinear equation for the density matrix. Hradil et al. applied this framework to the quantum state of a spin-1/2 system [61]. They find an equation for the Bloch vector that can be solved conveniently by iteration. Rehacek et al. proposed [62] an iterative algorithm to solve the non-linear equation that results from [60]. This iterative algorithm was later applied in some experiments [63, 64, 65]. In [66] Rehacek et al. showed that the $R\rho R$ algorithm may not converge, for there is no guarantee that the likelihood function increases at every iteration step, and provide a counterexample where a cycle of length two arises. Then they modify the algorithm by introducing dilution, a control step size at each iteration with a parameter ε, and prove that the diluted $R\rho R$ algorithm converges to the MLE solution if at each iteration the optimal value of the step size is chosen. They provide strategies to choose ε at each step and analyze the initial counterexample, an experiment of homodyne tomography [65], and one with four ion qubits [67]. Later, Goncalves et al. [68] modified the diluted $R\rho R$ algorithm by using an inexact step size selection and proved it to be globally convergent.

Banaszek et al. [58] introduced an efficient parametrization for the numerical solution of the ML estimation procedure. The density matrix is parametrized as $\rho = T^\dagger T$, with T a complex lower or upper triangular matrix, with real elements on the diagonal. This parametrization takes inspiration from the Cholesky decomposition, valid for an arbitrary positive semi-definite Hermitian matrix. If we consider a D-dimensional Hilbert space, the number of independent real parameters of the matrix $T^\dagger$ is $D + 2D(D-1)/2 = D^2$, which is exactly the number of independent real parameters of a Hermitian matrix. Furthermore, this parametrization automatically guarantees that ρ is positive semi-definite and Hermitian:

$$\rho^\dagger = \left(T^\dagger T\right)^\dagger = T^\dagger \left(T^\dagger\right)^\dagger = T^\dagger T = \rho, \tag{1.99}$$

$$\langle\psi|\rho|\psi\rangle = \langle\psi|T^\dagger T|\psi\rangle = \|T\,|\psi\rangle\,\|^2 \geq 0. \tag{1.100}$$

In this work, the normalization condition $\mathrm{Tr}\left[\rho\right] = 1$ is considered in the optimization problem. The authors provide some considerations for numerical optimization. In particular, they replace the likelihood by its natural logarithm, called log-likelihood, which does not change the position of the maximum, and rely on Lagrange multipliers to impose the normalization. Finally, the authors apply MLE to quantum homodyne tomography

of a single-mode radiation field, a two-mode field, and a pair of spin-1/2 particles in the singlet state.

The paper by James, Kwiat, Munro, and White is a cornerstone in the development and practical implementation of MLE for QST [22, 69].[2] They focus on qubits, on which they perform only local projective measurements, that is, tensor products of projectors onto the eigenvectors of the Pauli matrices. They introduce the notion of *tomographically complete* set of projective measurements, and provide a way to realize it given the states on which we are projecting. Regarding the ingredients of MLE, they adopt the same parametrization of the density matrix as [58] with a slight modification to directly encode the normalization:

$$\rho = \frac{T^\dagger T}{\mathrm{Tr}\,(T^\dagger T)}, \tag{1.101}$$

and the parameters are introduced into the lower or upper triangular matrix $T(t)$:

$$T\left(\boldsymbol{t}\right) = \begin{pmatrix} t_1 & 0 & 0 & 0 \\ t_5 + it_6 & t_2 & 0 & 0 \\ t_{11} + it_{12} & t_7 + it_8 & t_3 & 0 \\ t_{15} + it_{16} & t_{13} + it_{14} & t_9 + it_{10} & t_4 \end{pmatrix}. \tag{1.102}$$

Instead of maximizing the likelihood, they minimize the negative log-likelihood, which is mathematically equivalent. Furthermore, minimization problems suffer from fewer convergence issues than the corresponding maximization versions. Finally, numerical optimization to find the optimal set of parameters is carried out using standard methods in this paper. Another relevant feature of this work is the development of a public library available on Github and an online interface that implement the procedure described above. These tools allow one to perform QST up to 2 qubits by selecting which and how many measurements have been performed, eventually taking into account experimental imperfections as well. The major limit of this approach is that the number of measurements required is exponential with the number n_q of qubits composing the system. This work was extended by Thew et al. to systems composed of many qudits with $d > 2$ [70]. This procedure assumes that one can measure the the projectors onto the eigenvectors of the generators of SU(d). In principle, it requires $d^{2n_q} - 1$ measurements, still exponential with n_q. They provide sets of measurement based on orthogonal and non-orthogonal states, and show explicit examples for one and two qutrits.

The numerical maximization of the likelihood can be expensive if the dimension of the system is large. In fact, it can happen that the size of the experiments is limited by the complexity of numerical state recovery rather than the experimental capabilities of the system as in [71], where the authors reconstructed an 8-qubit W state realized with trapped ions. This led to a whole line of research dedicated to improving the numerical optimization step. In [21] Kaznady and James analyzed the QST problem from a purely computational perspective and compared different ways of obtaining a valid density matrix from tomographic data to understand if we always need an expensive MLE routine or if and when some shortcut techniques suffice. Interestingly, the authors point out that a data structure of maximum size $2^{n_q} \times 2^{n_q}$ would allow us to go up to 15-16 qubits, a

[2] A detailed analysis of these works can be found at the following link.

point at which the density matrix itself would become a storage problem. This is very close to the maximum number of qubits on which we tested one of our proposed protocols with numerical experiments on simulated data. Furthermore, they highlight that an analytic form of the gradient is necessary to speed up the algorithm and reduce the computation time. In [72] the authors discuss the role of local solutions in MLE. They show that for convex negative log-likelihood functions, all local minima of the unconstrained optimization problem ar global. Thus, any minimizer leads to the maximum likelihood estimation of the density matrix. Baumgratz et al. [73] proposed a fixed point algorithm based on matrix product states and operators to maximize the likelihood. This algorithm is particularly advantageous when we consider one-dimensional quantum systems composed of n_q qudits and if the observables to be measured have a matrix product operator representation of low bond dimension. Goncalves et al. [74] introduced a projected gradient algorithm to optimize some functionals on the set of density matrices, or on the intersection of the same with another one defined by functional constraint. They applied the proposed algorithm to the minimization of the likelihood function in QST and compared it to existing methods, such as the $R\rho R$ algorithm. Shang et al. [75] developed a "superfast" MLE algorithm by combining a conjugate-gradient algorithm with an accelerated projected-gradient algorithm and used it to reconstruct the state of [71]. Finally, in [76] the authors highlight that the negative log-likelihood function is not smooth nor Lipschitz, thus there are no sample complexity guarantees. They developed a first-order algorithm starting from Soft Bayes and online portfolio selection to address this problem.

MLE was originally conceived as a *frequentist* statistical inference tool. However, one might not adhere to the frequentist interpretation of probability and provide sound reasons to choose another interpretation. In this respect, Blume-Kohout described the major flaws of MLE for QST, why they arise, and proposed an alternative based on Bayesian inference [77]. We now briefly report only the former, while exposing the latter in Section 1.4.3 on Bayesian approaches to QST. MLE can produce estimates with eigenvalues equal to 0, and vanishing probabilities imply absolute certainty about the outcomes of future measurements about the probability itself. This is implausible for a number of reasons. First, an absolute certainty of not observing a specific eigenvalue cannot be justified by a finite amount of data. Second, if an outcome has never been observed, it does not justify the assertion that it will never occur. Third, zero probabilities are not compatible with any error bars. MLE produces vanishing eigenvalues because it tends to produce rank-deficient state estimates. Numerical exploration showed that the MLE estimate usually has at least as many zero eigenvalues as the linear inversion estimate has negative ones. Why is this not good for QST? MLE is a frequentist method, and frequentist techniques are based on the axiom that the future will look (statistically) identical to the past. If this is true, then MLE is the best possible estimate. For classical systems, it is always possible that this axiom holds, while for quantum systems, it cannot possibly hold: Any future observer could violate it by simply making a novel measurement. In [78] Blume-Kohout suggested a modified version of MLE, called hedged MLE (HMLE), where the standard likelihood function is replaced with the product of itself and a "hedging function" that depends on a parameter β. The role of this function is to hedge against as yet unseen probabilities. HMLE is tested on single-qubit systems using quantum relative entropy, infidelity, Euclidean distance, and trace distance as measures of inaccuracy. HMLE provides improved accuracy for almost all metrics. The author concludes that the optimal value of β is not clear, in contrast to the classical case where an asymptotic value is known.

In [79] the authors develop a minimax estimator for QST show that HMLE can achieve a worst-case risk very close to the one of the minimax estimator, can achieve an accuracy comparable to the minimax estimator, and is less computationally complex than the minimax estimator.

Scholten et al. [80] analyze QST as a model selection problem. QST is a statistical inference process in which the data are used to fit the parameters (the elements of the density matrix to be reconstructed) of a statistical model (the set of all possible density matrices in a Hilbert space). However, it is not always easy to determine which model to use, for it is not always a priori known what the Hilbert space is or its dimension. Choosing a proper Hilbert space is an instance of model selection, and the authors address this problem by deriving a generalization of the Wilk's theorem to the case of QST.

1.4.3 Bayesian QST

Bayesian methods for statistical inference are rooted in a subjective interpretation of probability, centered around the use of Bayes' theorem to update prior beliefs about parameters in light of observed data. In this framework, we start with an initial understanding or belief about the parameters, represented by a prior probability distribution, and then revise this belief after observing the data to obtain a posterior probability distribution. They can explicitly incorporate prior knowledge or beliefs, provide a natural framework for quantifying uncertainty, and can exhibit good performance even with small samples. However, the choice of the prior distribution is non-trivial, and Bayesian methods can often be computationally intensive, particularly for complex models that require advanced computational techniques like Markov Chain Monte Carlo (MCMC) to approximate the posterior distribution.

As anticipated previously, in [44] and [45] the authors connected Bayesian inference to QST. In addition to these works, Derka et al. [81] linked Bayesian quantum inference with the maximum entropy (MaxEnt) or Jaynes principle, which selects the density operator that maximizes the von Neumann entropy while satisfying the constraints imposed by the measured expectation values of the chosen observables. It stems from the most likely state being the one with the highest entropy, as it assumes the least amount of information beyond what is actually measured. However, MaxEnt requires precise knowledge of the expectation values of the observables, which implies an infinite number of measurements. If only a finite number of copies is available, then Bayesian inference comes into play. They apply this scheme to a spin-1/2 system as an example. Buzek et al. [82, 83] commented on how to extend Bayesian inference to mixed states, not considered in previous works, the choice of a proper prior distribution, an integration measure for mixed states, and the required computational complexity. They also provided some examples for systems composed of one or two spin-1/2 particles.

Several papers studied Bayesian inference in the context of quantum state estimation. Schack et al. [84] derived a quantum version of the Bayes rule , while Neri [85] showed that the same rule follows from the ordinary quantum measurement theory applied to density operators that represent our a priori knowledge of a system. Tanaka et al. [86] showed that Bayesian predictive density operators are the best predictive density operators when evaluated using the average relative entropy based on a prior distribution. Finally, Baier et al. [87] compared Bayesian and least squares methods for quantum state

estimation, while Schmied et al. [88] compared quantum Bayesian inference with other methods to reconstruct single-qubit states.

The work of Blume-Kohout [77] represents a major advance in the field of Bayesian inference applied to QST. We previously analyzed the critiques about MLE reported in this work, while we now illustrate the proposed alternative, called Bayesian mean estimation (BME), based on the Bayesian framework. BME provides plausible state estimates, avoiding implausible zero eigenvalues, and assigns non-zero probabilities even to events not observed in the data. It provides error bars that identify a plausible region that contains not only the best estimate but also nearby solutions that are almost equally likely. Finally, the method can optimize the accuracy measures that can be expressed as operational divergences. The author also notices that metrics like the trace distance and the fidelity are not of this kind. The practical aspects of this method come at a price. BME deals with computing integrals, usually computationally expensive due to the large dimension of the parameter space and more difficult than the optimization involved in MLE. The choice of the prior is challenging too, for a proper prior should extend over all the possible states. This is not trivial, because there is no unique measure over mixed states. Finally, the accuracy of the best estimate is highly dependent on the choice of the prior distribution. The author suggests that future research should address how to find the prior distribution that matches the unknown distribution of quantum states, improve the numerical integration techniques, and make the approach scalable to larger Hilbert space dimensions.

The practicality of BME is evident from the number of its experimental implementations on different platforms, such as frequency bin qubits [89] and qudits [90], biphoton frequency spectra [91] (along with compressed sensing, see 1.4.4), plasmonic nanometrology [92], and homodyne and heterodyne detection (properly adapted to continuous variable states) [93]. The Bayesian approach has been used in combination with other QST approaches, such as adaptive [94, 95] and shadow-based methods [96] (see 1.4.7 for more details on shadow tomography).

Specialized work improved the main ingredients that compose the Bayesian framework for QST, such as the choice of the prior [97, 98, 99, 100, 101, 102], properly and efficiently sampling from the state space [103, 104, 105, 106], and the numerical tractability of the method [107, 100]. However, we observe that the proposed improvements are rarely tested on Hilbert spaces of large dimension, thus limiting their applicability in large-size experiments. Finally, Bayesian QST has also been studied as a model selection problem [108, 109], to understand if the description of the Gibbs state can be justified solely from tomographic data [110], and to provide bounds on the average fidelity [111]. An open source library of statistical Bayesian inference is available for applications in quantum technologies [112].

1.4.4 Compressed sensing QST

Compressed sensing (CS), also known as compressive sensing, compressive sampling, or sparse sampling, is a signal processing technique to reconstruct signals that are sparse in a known basis [113]. CS leverages this sparsity to reconstruct the signal using a significantly reduced number of measurements compared to those required by conventional sampling theory. From a mathematical point of view, taking inspiration from matrix com-

pletion, the reconstruction process can be formulated as a rank or L_1 minimization problem, which is a convex optimization problem that can be solved via linear programming or other specific algorithms. CS has been applied in various fields, such as medical imaging [114], astronomy [115], spectroscopy [116], and neural engineering [117].

Compressed sensing QST (CS QST) concentrates on states that are well approximated by density matrices of rank much smaller than the dimension of the associated Hilbert space. This means that the method can be applied in real situations where the ideal (or expected) state is (nearly) pure and the actual (noisy) state still has low entropy. However, there are some experimental constraints to be satisfied, as one usually restricts to measurements that can actually be performed in the laboratory. Finally, the procedure to find the (nearly) pure state consistent with the experimental data has to be computationally efficient.

CS QST provides a solution that meets all these criteria [118, 119]. Consider a system of n_q qubits ($D = 2^{n_q}$), and let $\mathcal{S}$ be the set of all 4^{n_q} measurement settings, that is, observables of the form $S = \sigma_1 \otimes \cdots \otimes \sigma_{n_q}$, with $\sigma_i \in \{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}$. The compressed sensing protocol proceeds as follows. We choose m measurement settings $\{S_1, S_2, \ldots, S_m\}$ by sampling independently and uniformly at random from $\mathcal{S}$ (alternatively, one can choose these Pauli operators without replacement, but independent sampling simplifies analysis [120]), and measure the expectation values $\mathrm{Tr}\,[S_i\rho]$. Then we recover the density matrix by solving the following convex optimization problem: Find a Hermitian matrix σ that fits the data, that is, $\mathrm{Tr}\,(S_i\sigma) = \mathrm{Tr}\,(S_i\rho)$, while minimizing the trace norm $\|\sigma\|_{\mathrm{Tr}} = \mathrm{Tr}\left(\sqrt{\sigma^\dagger\sigma}\right)$. The main result of Gross et al. [118] is to show that we can uniquely reconstruct a rank-r state ρ by solving the convex optimization problem using $m = cDr\log^2 D$ randomly chosen measurement settings with exponentially small failure probability in c.

Several theoretical analyses have been performed to better understand why and how CS QST works. Koltchinskii [121] studied the relation between the estimation error and the precision of the low-rank representation of the unknown state. Liu [122] showed that an unknown rank-r matrix of dimension D can be recovered from almost all sets composed of $O\left(rD\log^6 D\right)$ measurement settings via trace norm minimization. Flammia et al. [123] provided a detailed theoretical analysis of CS QST. They theoretically analyzed the method with differential geometry tools, showed that the sample complexity of tomography decreases with the rank, compared the method with MLE, and suggested a generalization to quantum process tomography of channels with small Kraus rank. Wang [124] studied the statistical relationship between matrix completion and CS QST. Kalev et al. [125] showed that the positivity constraint is what actually allows the unique identification of a low-rank quantum state among all possible quantum states, and the measurements required by this protocol are called strictly informationally complete. This work was extended to rank-bounded states by Baldwin et al. [126]. Kueng et al. [127] studied low-rank matrix recovery under the lens of linear algebra, with applications to QST. Badveli et al. [128] generalized the original method to qudits whose Hilbert space dimension is not a power of two.

CS QST has been experimentally validated in several quantum systems. For example, on different degrees of freedom of photons, such as path [129], OAM [130], and polarization [131], as well as with trapped ions [132]. CS can also be used in combination with other approaches. Smith et al. [133] combined CS with continuous measurements to perform experiments on the spins of cesium atoms. Ahn et al. developed an adaptive version

of CS QST and performed experiments and numerical studies [134, 135]. Gil-Lopez et al. made this adaptive version suitable for low-rank spectral-temporal optical signals [136]. Schreiber et al. [137] combined CS and shadow tomography to perform tomography of parametrized quantum states and channels.

As suggested by Flammia et al. [123], the compressed sensing approach can also be applied to quantum process tomography. Theoretical studies [138] have been carried out in this respect, together with experiments on quantum gates implemented with photons [139], superconductors [140], NMR [141], and trapped ions [142]. Finally, further generalizations of the original method are a non-convex algorithm for CS QST [143], the construction of confidence sets for CS QST [144], and the application to Hamiltonian learning [145].

1.4.5 Mutually unbiased bases (MUBs)

The first paper in which mutually unbiased bases (MUBs) in the context of QST was by Ivanovic [146], even if the name is due to Wootters and Fields [147]. Recall that the problem of quantum state determination consists of recovering the density matrix of an ensemble of identical systems with dimension D, which may be in a pure or mixed state. Such a density matrix is specified by $D^2 - 1$ independent real parameters. Any given measurement setting (or basis) will yield precisely $D - 1$ real numbers, namely the probabilities of all but one of the D possible outcomes, the last probability being determined by the requirement that the probabilities sum to unity. Thus, the minimum number of different bases that one needs to determine the state uniquely is $(D^2 - 1)/(D - 1)$, that is, $D + 1$. Ivanovic [146] has shown that if one can find $D + 1$ bases that are all unbiased with respect to each other, then the measurements corresponding to these bases are guaranteed to be sufficient to determine the density matrix. For $D = 2$, the sufficiency of $D + 1 = 3$ mutually unbiased bases to reconstruct the density matrix is clear. The density matrix in this case can be written in the Bloch vector form, and the measurements represented by the Pauli observables precisely determine the three components of the Bloch vector. The author also showed by explicit construction that such bases exist for each prime value of D.

The second relevant work in the context of MUBs for QST is [147] by Wootters and Fields. Two bases $\{|v_i\rangle\}$ and $\{|u_j\rangle\}$ over an D-dimensional complex space are **mutually unbiased** if the inner products between all possible pairs of vectors with one vector from each basis have the same magnitude:

$$|\langle v_i|u_j\rangle| = \frac{1}{\sqrt{D}} \quad \forall i, j \tag{1.103}$$

Alternatively, one can express the notion of mutual unbiasedness in terms of projectors. Let P_i be the projection onto the vector $|v_i\rangle$ and let Q_j project onto $|u_j\rangle$. Then mutual unbiasedness requires that for all P_i and Q_j:

$$\mathrm{Tr}\,(P_i Q_j) = \frac{1}{D}. \tag{1.104}$$

Schwinger [148] discussed pairs of bases related in this way for finite D and noted that these bases represent measurements that are "maximally non-commutative". In other

words, a measurement over one basis leaves one completely uncertain as to the outcome of a measurement over a basis unbiased with respect to the first. As an example, the eigenbases of the Pauli operators are examples of MUBs for a single qubit. In fact, a measurement of spin in the x direction for a spin-1/2 particle leaves one completely uncertain about the component of the spin in the y direction. The authors then addressed the question of the existence of $D+1$ mutually unbiased bases in a Hilbert space of dimension D. They extended the result of [146] to all values of D that are powers of primes, and pointed out that the construction of the bases is different for powers of odd primes than for powers of two. Finally, they proved that for no value of D it is possible to find more than $D+1$ MUBs.

Suppose that we have been able to construct a set of $D+1$ MUBs for a Hilbert space of dimension D. How can we use it to reconstruct an unknown state ρ? Label the $D+1$ bases as $\{|a\alpha\rangle\}$, with $a = 0, \ldots, D$ the basis number, and $\alpha = 0, \ldots, D-1$ one of the basis states belonging to the particular basis a. The mutually unbiasedness condition can be formulated as follows:

$$|\langle a\alpha|b\beta\rangle|^2 = \frac{1}{D}\left(1-\delta_{a,b}\right) + \delta_{a,b}\delta_{\alpha,\beta}. \tag{1.105}$$

This condition implies that each basis is orthonormal and that two arbitrary states belonging to different bases are equiangular, that is, $|\langle a\alpha|b\beta\rangle|^2 = \frac{1}{D}$ if $a \neq b$. Let us now consider the projectors on the states that form the MUBs, $\Pi_{a\alpha} = |a\alpha\rangle\langle a\alpha|$, which satisfy the following properties:

$$\begin{aligned} \mathrm{Tr}\left(\Pi_{a\alpha}\Pi_{b\beta}\right) &= \frac{1}{D}\left(1-\delta_{a,b}\right) + \delta_{a,b}\delta_{\alpha,\beta} \\ \sum_{\alpha=0}^{D-1}\Pi_{a\alpha} &= \mathbb{I}_D \quad \forall a \\ \sum_{a=0}^{D}\sum_{\alpha=0}^{D-1}\Pi_{a\alpha} &= (D+1)\mathbb{I}_D. \end{aligned} \tag{1.106}$$

Let us now define the probabilities of finding the state of the system in $|a\alpha\rangle$ via the usual Born rule:

$$p_{a\alpha} = \langle a\alpha|\rho|a\alpha\rangle = \mathrm{Tr}\left(\rho\Pi_{a\alpha}\right), \tag{1.107}$$

which satisfy the following properties:

$$\begin{aligned} \sum_{\alpha=0}^{D-1} p_{a\alpha} &= 1, \\ \sum_{a=0}^{D}\sum_{\alpha=0}^{D-1} p_{a\alpha} &= (D+1). \end{aligned} \tag{1.108}$$

Then any state ρ can be expressed as [149]:

$$\rho = \sum_{a=0}^{D}\sum_{\alpha=0}^{D-1} p_{a\alpha}\left(\Pi_{a\alpha} - \frac{1}{D+1}\mathbb{I}_D\right). \tag{1.109}$$

Several physical platforms can be leveraged to experimentally realize QST with MUBs, such as trapped ions [150], electrons in a double quantum dot [151], polarization [152], linear transverse [153] and orbital angular momentum [154] of photons, and driven cavity QED systems [155]. Wang et al. [156] utilized quantum circuits to construct MUBs, studied in [157, 158], as the ensemble in shadow tomography.

As anticipated, the existence of MUBs has been proved only for Hilbert spaces with dimension equal to the power of a prime number. However, how to actually construct such bases is non-trivial, and several strategies have been developed. For example, [159], [160], and [161] constructed MUBs for the dimensions where the existence is guaranteed, while the authors of [162] constructed a MUB for a Hilbert space of dimension 6 by factorizing the space into the direct product of a two-dimensional and a three-dimensional space, which both possess complete sets of MUBs. Klimov et al. [163] found a method to generate a minimal set of MUBs for permutationally invariant states of n_q qubits.

Some theoretical studies on MUBs include the review by Durt et al. [164], the analysis of MUBs within the framework of star product scheme for spin-state tomography carried out by Filippov et al. [149], the interpretation of MUBs as complex projective 2-designs [165], and a generalization of the previous results to the case of measurements not described by rank one projectors [166].

1.4.6 Symmetric Informationally Complete POMVs (SIC-POVMs)

Let us recall that, given a Hilbert space $\mathcal{H}$ of dimension D, a POVM over $\mathcal{H}$ is a set of $\mathcal{M}$ positive semi-definite operators $\{M_i\}_{i=0}^{\mathcal{M}-1}$ that sum to the identity: $\sum_{i=0}^{\mathcal{M}-1} M_i = \mathbb{I}_D$. If a POVM consists of at least D^2 operators that span the space of self-adjoint operators acting on $\mathcal{H}$, then it is an **informationally complete POVM (IC-POVM)**. IC-POVMs consisting of exactly D^2 elements are called minimal. Alternatively, a POVM is called informationally complete if its statistics completely determine the quantum state on which the measurements are carried out. Symmetric informationally complete POVMs (SIC-POVMs) are IC-POVMs with an additional property. A set of D^2 rank-1 projectors $\{\Pi_i\}_{i=0}^{D^2-1}$ that have equal pairwise Hilbert-Schmidt inner products:

$$\mathrm{Tr}\left(\Pi_i \Pi_j\right) = \frac{D\delta_{ij}+1}{D+1} = \begin{cases} 1, & \text{if } i = j \\ \frac{1}{D+1}, & \text{if } i \neq j \end{cases} \tag{1.110}$$

defines a minimal IC-POVM with elements:

$$M_i = \frac{1}{D}\Pi_i. \tag{1.111}$$

This POVM is called a **symmetric informationally complete POVM (SIC-POVM)**. The property stated above about the pairwise Hilbert-Schmidt inner products is what makes the POVM *symmetric*. Indeed, any pair of elements is equivalent to any other pair with respect to that inner product. The operators composing the SIC-POVM are linearly independent; thus, any quantum state can be reconstructed from the measurement statistics of a SIC-POVM $p_i = \mathrm{Tr}\left(\Pi_i \rho\right)$ as [167]:

$$\rho = \sum_{i=1}^{D^2} \left[(D+1)\, p_i - \frac{1}{D} \right] \Pi_i. \tag{1.112}$$

The simplest example of a SIC-POVM is for the Hilbert space of a qubit ($D = 2$). The $D^2 = 4$ rank-1 projectors can be constructed from the following set of states:

$$\begin{aligned} |\psi_1\rangle &= |0\rangle\,, \\ |\psi_2\rangle &= \frac{1}{\sqrt{3}}|0\rangle + \sqrt{\frac{2}{3}}|1\rangle\,, \\ |\psi_3\rangle &= \frac{1}{\sqrt{3}}|0\rangle + \sqrt{\frac{2}{3}}e^{i\frac{2\pi}{3}}|1\rangle\,, \\ |\psi_4\rangle &= \frac{1}{\sqrt{3}}|0\rangle + \sqrt{\frac{2}{3}}e^{i\frac{4\pi}{3}}|1\rangle\,, \end{aligned} \tag{1.113}$$

and the elements of the SIC-POVM are $F_i = \Pi_i/2$.

It would be useful to know whether SIC-POVMs could be constructed in a Hilbert space of arbitrary dimension. Zauner proposed a conjecture about the existence of SIC-POVMs in his dissertation dissertation [168], later inserted as "Problem 23" into the list of Open Problems in Quantum Information drafted by researchers from IQOQI in Vienna. Here, three different formulations of the problem are proposed, each based on slightly different viewpoints, but all boiling down to the same question: ***Do SIC-POVMs exist for Hilbert space of arbitrary dimension?*** A conclusive answer to this problem is still lacking. However, much work has been done and some partial results have been obtained. The pioneering work of Renes et al. [169], along with [170, 171], linked the search of SIC-POVMs with other algebraic objects, such as frames, frame potentials, and spherical designs. Later, several works leveraged theoretical [172, 173, 174, 175, 176, 177, 178] and numerical [179, 180, 181, 182, 183] methods to discover new SIC-POVMs. SIC-POVMs have been compared with other functional QST approaches, such as over-complete sets of observables [184, 185], and have been experimentally used for QST of photonic states [186, 167].

1.4.7 Shadow tomography

Shadow tomography has a different goal from QST. While the latter aims at reconstructing the representation of a quantum state starting from a certain number of measurement outcomes, the former aims at predicting one or more properties of a quantum state, that is, the expectation value of one or more observables, using a finite number of copies of the state, *without fully reconstructing the state*. This idea was previously studied by Ekert et al. [187], who developed a protocol to estimate the overlap $\mathrm{Tr}\,(\rho_a\rho_b)$ between any two quantum states without tomography, and Brun [188], who discovered how to determine any m-degree polynomial function of the elements of the density matrix by measuring the expectation value of an observable on m copies of the state without tomography. Flammia and Liu [189] followed a similar approach for quantum fidelity estimation, later improved in [190]. Audenaert et al. [191] and Wunderlich et al. [192] showed how to quantify the entanglement content of a quantum state without reconstructing the entire state, while Zhu et al. [193] followed a similar line of reasoning with entanglement witnesses. Zhang et al. [194] employed the Krylov subspace method to estimate the quantum Fisher information.

Let us now precisely formulate the problem. Consider a system composed of n_q qubits and a set of T observables $\{A_1, \ldots, A_T\}$. Suppose now that we are given a finite number

of copies of an unknown quantum state $|\psi\rangle$, and we want to estimate the expectation value of these observables with respect to $|\psi\rangle$ within some error bounds. In principle, this would require an exponential number of copies of the state, without any advantage with respect to QST. Shadow tomography allows one to estimate a large number of observables using a small number of copies of the unknown state. The shadow tomography protocol produces a classical shadow of the state, which turns out to be an efficient representation to estimate the expectation value of the observables of interest. For such a protocol to be useful, we would like to minimize the required number of copies of the state, the complexity of producing the classical shadow, and the complexity of estimating the expectation value of the observables given the classical shadow.

The main result of Aaronson's paper [195], somehow anticipated in [196], was to prove that there exists an algorithm allowing one to estimate the expectation value of an exponential number of observables using only a polynomial number of copies of the state. However, the protocol depends on the observables whose expectation value is being estimated, the number of copies is poly-logarithmic in the dimension of the associated Hilbert space and in the number of observables, the time complexity of the algorithm is still exponential, and the dimension of the shadow is about polynomial in the dimension of the Hilbert space. Huang, Kueng, and Preskill proposed a shadow tomography protocol that can accurately estimate M different functions of the state with high probability using only $\log M$ copies of the state [197]. This protocol does not depend on the observables of interest, and the time complexity to produce the shadow is polynomial in the number of qubits. The procedure to extract the classical shadow is quite simple: apply a random Clifford unitary U to rotate the state, and then measure it in the computational basis to get a length-n_q bitstring $|b\rangle$. The average mapping (over the choice of the unitary and the outcome distribution) from the state ρ to its classical snapshot $U^\dagger |b\rangle\langle b| U$ can be seen as a quantum channel:

$$\mathbb{E}\left(U^\dagger |b\rangle\langle b| U\right) = \mathcal{M}(\rho) \implies \rho = \mathbb{E}\left[\mathcal{M}^{-1}\left(U^\dagger |b\rangle\langle b| U\right)\right]. \tag{1.114}$$

In this way, we produce a single classical snapshot $\mathcal{M}^{-1}\left(U^\dagger |b\rangle\langle b| U\right)$ of the unknown state. Repetition of this procedure N times yields the classical shadow of ρ, that is, an array of N independent classical snapshots of ρ:

$$S = \left\{\rho_1 = \mathcal{M}^{-1}\left(U_1^\dagger |b_1\rangle\langle b_1| U_1\right), \ldots, \rho_N = \mathcal{M}^{-1}\left(U_N^\dagger |b_N\rangle\langle b_N| U_N\right)\right\}. \tag{1.115}$$

The authors show how this classical shadow can be used to estimate linear and nonlinear functions of the state, such as the fidelity, entanglement witnesses, two-point correlation functions, and entanglement Rényi entropies. The proposed approach has been experimentally validated with photonic [198, 199] and superconducting qubits [200].

Several variants and generalizations of the protocol have been proposed. In [201, 202, 203] the authors study alternative unitary ensembles to the Clifford group, while in [204] the authors generalize the protocol considering POVMS instead of unitary transformations, similarly to what was done in [205]. Akhtar et al. [206] improved the computational burden by means of tensor network methods. Zhou et al. [207] developed a quantum-classical hybrid framework based on the combined use of a generalized swap test and a shadow protocol.

The protocol has been generalized to continuous variable systems [208, 209, 210], process tomography [211], measurements that entangle a certain number of copies of the

state [212, 213], analog quantum simulators [214], and QST with guaranteed performance [215]. In recent years, reviews on learning states from classical shadows [216] and randomized measurements [217] have been published.

1.4.8 Other approaches and problems related to QST

Some QST methods can leverage on available prior information about the structure of particular classes of states to provide a more favorable scaling of the required resources with the number of qubits, or eventually qudits, than QST. Examples in this respect are permutationally invariant [218, 219, 220], matrix product [221, 222, 223, 224, 225, 226, 227, 228], and pure states [229, 230, 231, 232, 233, 234, 235] (the latter are also of particular interest for CS QST, see 1.4.4).

Adaptive QST [236, 237, 238, 239] and self-guided QST [240, 241, 242, 243, 244, 245, 246] choose an iterative strategy to reconstruct the state, that is, they produce an initial estimate of the state and subsequently refine it as long as data are acquired. Although a large enough number of iterations usually guarantees convergence to the true state, it can contrast with a reduction in the amount of resources one is looking for. Furthermore, they may require the implementation of projectors on entangled states, which is a non-trivial task from an experimental point of view.

Original approaches to QST are based on continuous variables [247, 248, 249, 250, 251], equidistant states [252, 253, 254], reduced density matrices [255], maximum entropy principle [256, 257, 258, 259], and quantum overlap [260, 261]. Some approaches were designed to be suitable for a specific experimental platform, such as stimulated emission tomography [262, 263, 264] for non-linear quantum processes, quantum-field tomography scheme for ultra-cold atoms [265], and scalable on-chip QST for multi-photon states [266].

The process of retrieving the representation of a quantum state from the measurement outcomes has been extended to gates, or processes, and thus called quantum process tomography (QPT) [267, 268]. QPT has been experimentally verified for the quantum Fourier transform [269], the CNOT [270] and Toffoli gates [271], and the decoherence and resonant coupling of atoms in an optical lattice [272]. Existing QST approaches have been adapted to quantum process tomography, such as AI models [273] and Bayesian methods [274]. Quantum detector tomography [275, 276, 277, 278, 279, 280, 281], whose objective is to reconstruct the POVM that describes a detector, along with state and process tomography, completes the triad to fully specify a quantum experiment. Finally, Hamiltonian learning was initially proposed as a protocol to verify that a given quantum simulator actually simulates the correct quantum dynamics by inferring the parameters of the Hamiltonian of the simulator itself [282, 283]. It was experimentally demonstrated with trapped ions [284], spins in circuit QED [285], photons [286], and superconducting qubits [287]. Successive work focused on local Hamiltonians [288, 289], open systems [290, 291], and placed it within the Bayesian framework [292]

Many works improved numerical algorithms or estimators employed in previously presented methods, such as linear regression estimator [293, 294, 295, 296], projective methods [297, 298], gradient descent [299, 300, 301], SPSA algorithm [302].

Another interesting problem regarding QST is **sample complexity**, the number of copies of the state required to faithfully reconstruct the state. To be more precise, it is the number of copies of the state required to achieve a target value of a specific figure of

merit that quantifies the quality of the reconstructed state. The most commonly used figures of merit for sample complexity are infidelity and trace distance [303], the Frobenius distance [304], and the Bures distance [305]. The work by Anshu et al. [306] is a recent review on the complexity of learning quantum states.

1.5 Final remarks

In this chapter, we introduced the fundamental elements of quantum mechanics and QST, the core of this work, specifying the nomenclature to clarify several definitions that can be found in the literature. We then reviewed some pioneering works on the more general problem of quantum state characterization or determination, and analyzed some of the approaches proposed over the years, different by statistical interpretation (MLE and Bayesian), reconstruction technique (compressed sensing), or measurements implemented (MUBs and SIC-POVMs). In addition, we highlighted the advantages and limitations of each method. This preparatory analysis of the literature led us to understand the desirable elements of any QST approach, presented at the beginning of the next chapter, in which we illustrate our proposed novel QST protocol.

Chapter 2

Threshold Quantum State Tomography (tQST)

In this chapter, we introduce a novel approach to QST, called threshold quantum state tomography (tQST). We present the workflow of the protocol, detail the main features, and unravel each step for some paradigmatic examples. After this, we validate its effectiveness through numerical analysis. Finally, we show experimental results on two different platforms. The main references for this chapter are [1], [4], and [307] for the public Github library.

2.1 Desiderata of a QST protocol

The literature review carried out in the previous chapter allowed us to identify the most desirable features of a QST protocol, summarized in the following points:

1. **No hypothesis on the state to reconstruct.** The efficacy of a QST protocol should not depend on any hypothesis on the structure of the state, such as for CS or permutationally invariant QST. These methods are advantageous only if the state of the system satisfies some underlying hypothesis, and this is possible only if we have somehow already characterized the experimental apparatus that generates the state. If this characterization is not possible for some reasons, such as the size or complexity of the system, then no prior information about the system is available and we can no longer leverage these methods.

2. **Reduction in the number of observables.** The major problem of QST is the so-called *curse of dimensionality*, that is, the number of observables whose expectation value is required to reconstruct the state scales exponentially with the dimension of the Hilbert space associated with the system. However, other resources necessary to complete any QST experiment, such as the time required to perform the measurements, the computational burden of state reconstruction, and the storage of the reconstructed state, follow the same scaling law. Any protocol that aims to significantly improve QST has to provide a reduction in the number of measurements.

3. **Computational efficiency.** Some of the methods mentioned in the previous chapter can be limited by the demanding computational burden, such as MCMC for

Bayesian QST. Thus, the numerical routines adopted to reconstruct the state should not require large amounts of computational power or time.

4. **Experimental feasibility.** Any QST protocol should be not only numerically validated, but also experimentally tested. To this end, the kind of measurement that it prescribes should be easy to implement in the laboratory. Separable projective measurements and measurement settings represent a standard type of measurement to characterize quantum systems, whereas projective measurements on entangled states are more challenging and can be implemented only in some special cases.

To our knowledge, there was no QST protocol that simultaneously met all these criteria. However, by carefully analyzing how the information is distributed among the elements of the density matrix and how to extract it, we developed an original QST protocol called **threshold quantum state tomography**, or **tQST** for short, which we now illustrate step by step.

2.2 The tQST protocol

We begin by noting that in QST one gest all information first, and then processes it. Indeed, the number of observables is dictated only by the dimension of the Hilbert space associated with the system, and one can decide in advance which measurements to perform (IC-POVMs, tomographically complete POVMs, overcomplete set of projectors, etc.). Thus, in a QST experiment, one performs all the required measurements in an arbitrary sequence, records the corresponding outcomes, and then processes them to reconstruct the state. However, information is not uniformly distributed among the elements of the density matrix, as some carry more information than others. In the first chapter, we proved that the three main properties of a density matrix ρ, that is, having trace equal to 1, being Hermitian and positive semi-definite, imply that any element ρ_{ij} of a physical density matrix must satisfy the bound $|\rho_{ij}| \leq \sqrt{\rho_{ii}\rho_{jj}}$. Thus, determining the diagonal elements of the density matrix (that is, measuring the projectors on the states of the chosen computational basis) provides some information on the off-diagonal terms. In fact, if one of the diagonal elements ρ_{ii} is found to be zero, then all the elements of the ith row and column of ρ can be immediately set to zero. Similarly, if ρ_{ii} and ρ_{jj} are different from zero but small compared to the other diagonal elements, then the modulus of ρ_{ij} will also be small. Finally, not all information about the system may be experimentally accessible, due to the signal-to-noise ratio. These considerations are at the core of the tQST protocol, represented pictorially in Figure 2.1, which we now illustrate.

1. Measure the diagonal elements $\{\rho_{ii}\}$ of the density matrix. This can be done by projecting onto the elements of the chosen computational basis.

2. Choose a **threshold** t, and determine the relevant off-diagonal elements, that is, those that satisfy $\sqrt{\rho_{ii}\rho_{jj}} \geq t$, and the number of observables.

3. Construct a set of local projective measurements to target the relevant off-diagonal elements and measure.

4. Use the measurement outcomes to reconstruct the state.

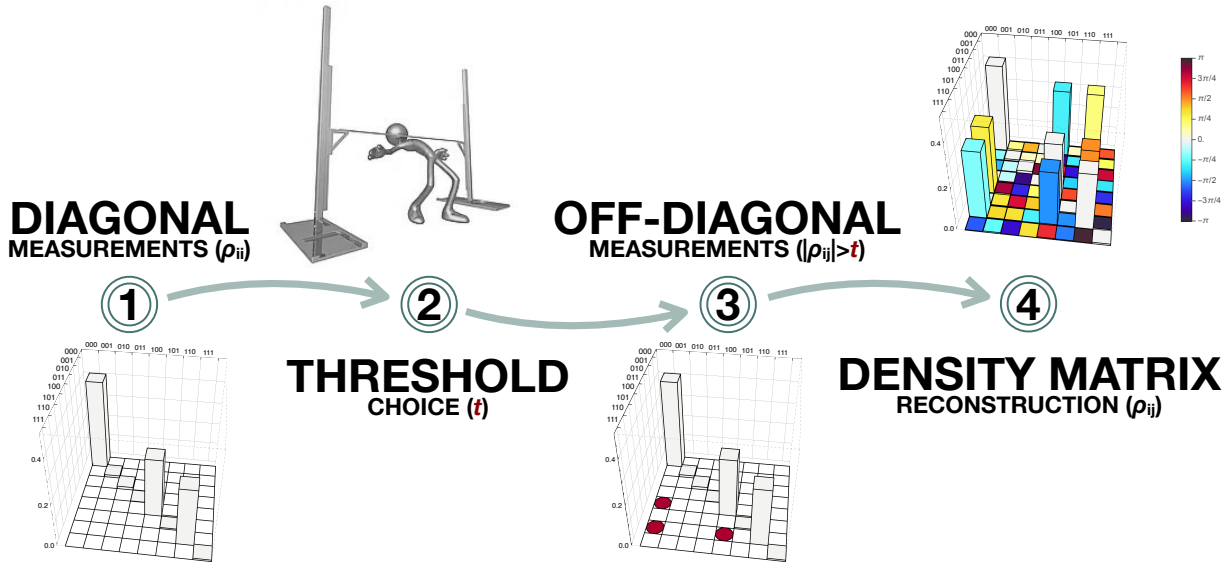


Figure 2.1: **Illustration of the tQST protocol.** (1) Measurement of the diagonal elements $\{\rho_{ii}\}$. (2) Choice of the threshold t. (3) Construction and measurement of the observables associated with the off-diagonal elements. (4) Reconstruction of the density matrix. We plot the absolute value of the density matrix elements colored by their phase as indicated by the color bar. Figure from [1].

Before proceeding, we highlight some relevant points about the tQST protocol. First, tQST does not require any hypothesis on the state to reconstruct since prior information is not required to start the protocol or to move from step to step. In tQST, the subsequent projectors can all be determined once the system is measured in the chosen computational basis, as opposed to adaptive approaches in which the necessary observables cannot be known a priori, since each one is chosen based on the outcomes of the previous. Second, the user can decide the number of projectors to measure by setting the threshold based on some mathematical or physical criteria or experimental constraints, such as the available time or resources for the experiment. In this respect, tQST can be interpreted as a tailor-made QST protocol. In tQST, the information retrieved from the measurement of the diagonal elements $\{\rho_{ii}\}$ is immediately used to decide the subsequent projectors to be measured, by choosing to neglect the off-diagonal elements of ρ whose absolute value is lower than a certain threshold. Unlike QST, the resources necessary to reconstruct the state are not uniquely determined by the dimension of the quantum system, but can be controlled with the threshold t. For example, if $t = 0$, all elements of ρ are considered and the protocol reduces to QST. In contrast, if $t > 0$, the protocol may require fewer projectors than those needed with QST and, in any case, no more than them. Finally, since the information is not uniformly distributed, even with a few projectors, one may be able to reconstruct the state with high fidelity.

At this point, one might ask several questions. How do we choose the threshold? What projectors should one perform once the relevant off-diagonal measurements have been identified? How do we reconstruct the state from the measurement outcomes? Does the protocol actually work? All these questions will be answered in the remainder of the chapter.

2.3 The threshold

In the following, we consider systems composed of n_q qubits, and $D = 2^{n_q}$ is the dimension of the Hilbert space associated with the system.

Let us now analyze in detail the threshold t, starting from the analogy between the choice of the threshold in the tQST protocol and the positioning of the bar in a limbo challenge. If the bar is very low ($t \approx 0$), a successful performance will certainly be spectacular. However, if the bar is too low, the dance can be prohibitive even for the most skilled athlete. This is the situation of QST, when the number of projective measurements to be performed is unreasonable because of the exponential scaling with the number of qubits. In contrast, for a sufficiently high bar, even a beginner may be able to dance, but the result may not be particularly noteworthy. To be more precise, the threshold t is a parameter that ranges from 0 to 1, and the higher it is, the smaller the number of projective measurements required by the protocol. If $t = 0$, we recover QST and have to perform 4^{n_q} projective measurements. For this reason, in the rest of the chapter we name QST the tQST protocol with $t = 0$, which implements all 4^{n_q} projectors unless otherwise specified. If $t > 0$, one has to perform fewer than 4^{n_q} measurements, and the magnitude of this reduction depends on the representation of the state under consideration.

In the tQST protocol, the threshold t determines the amount of information that one is willing to trade in exchange for fewer measurements. This has several consequences. First, one may be able to reconstruct the density matrix of a large state by reducing the number of projective measurements to a level compatible with the available experimental resources. In this respect, we emphasize that once a value of t is chosen, the protocol does not simply reduce to measuring a subspace of the Hilbert space of the whole system. Second, fewer projective measurements can give a significant advantage in terms of storing and handling experimental data. Third, one may be able to avoid useless projective measurements and increase the integration time for the remaining ones, leading to an improvement in the signal-to-noise ratio. The amount of information available in the characterization of a quantum state is always limited, for example, by noise or the finite precision of the experimental setup. Such experimental constraints de facto bound the accuracy with which the state can be determined, even for QST. Thus, although one may naively expect that reducing the number of projective measurements will decrease the quality of the results, we shall see in the following that a wise choice of t can still provide results compatible with the ones obtained by QST, while requiring fewer resources.

The dependency of the diagonal elements, and consequently the choice of the threshold and the number of measurements, on the representation of the state for a given basis deserves some consideration. In fact, this dependency is not a hypothesis on the state, rather than a matter of the convergence of the protocol, since different representations of the same state could require different numbers of projective measurements. In this respect, we can consider several scenarios. The simplest case is a pure state represented in its eigenbasis, with a single diagonal element equal to 1, and all the others equal to 0. This is the most favorable case for tQST, as it requires to measure only the diagonal of the density matrix, that is, only 2^{n_q} projectors instead of 4^{n_q} as required by QST. The opposite situation is a state, pure or mixed, with diagonal elements that are all equal to $1/D$, that is, a state whose diagonal is equal to the one of the maximally mixed state. In this case tQST and QST are equivalent, both require measuring 4^{n_q} projectors. If we consider states between these two limits, then a proper choice of the threshold may lead to a significant

reduction in the number of measurements without compromising the quality of the reconstructed state. In particular, the more sparse the representation of a state, the smaller the number of measurements required by tQST.

We identified two analytical criteria to set the threshold, based on the Gini index and the Shannon entropy, respectively. These criteria allowed us to define a threshold as a function of the outcomes of the initial projective measurements required by the tQST protocol, that is, the diagonal elements of the state, and to relate the threshold to quantities such as the sparsity or the information contained in the diagonal elements. Let us start with the criterion based on the Gini index. In the theory of signals, a representation of a signal is said to be sparse if a small number of its coefficients contain a large portion of the energy. This definition can be adapted to our needs as follows: a density matrix is sparse if a small number of its elements contain a large portion of the information contained in the state due to Eq. (1.28). Although there exist several measures of sparsity, the authors of [308] identified six desirable criteria that a good measure of sparsity should have, and proved that the Gini index is the only one that satisfies all of them. Let now $\underline{c} = \left(c_{(1)}, c_{(2)}, \ldots, c_{(D)}\right)$, such that $c_{(1)} \leq c_{(2)} \leq \cdots \leq c_{(D)}$ and $c_{(i)} \geq 0 \; \forall i$. The Gini index is defined as:

$$\mathrm{GI}\left(\underline{c}\right) = 1 - 2\sum_{k=1}^{D} \frac{c_{(k)}}{\|\underline{c}\|_1} \left(\frac{D-k+\frac{1}{2}}{D}\right), \tag{2.1}$$

with $\|\underline{c}\|_1 = \sum_i c_i$. One can prove that the Gini index satisfies the following bound:

$$0 \leq \mathrm{GI}\left(\underline{c}\right) \leq 1 - \frac{1}{D}, \tag{2.2}$$

with the minimum attained when all the elements of the argument vector are equal to $1/D$, and the maximum when there is a single non-vanishing element equal to 1. Thus, the Gini index increases as the sparsity of the argument vector increases. We adapted this definition to make it a suitable threshold for tQST. First, the argument vector of the Gini index is now the diagonal of the density matrix, that is, $\underline{c} \equiv \underline{\rho} = (\rho_{11}, \rho_{22}, \ldots, \rho_{DD})$. After computing $\mathrm{GI}\left(\underline{\rho}\right)$, we divide it by $1 - 1/D$ to normalize it between 0 and 1, and then divide by D again to obtain the so-called **Gini threshold**:

$$t_{\mathrm{g}} = \frac{\mathrm{GI}\left(\underline{\rho}\right)}{1-\frac{1}{D}} \cdot \frac{1}{D} = \mathrm{GI}\left(\underline{\rho}\right) \cdot \frac{D}{D-1} \cdot \frac{1}{D} = \frac{\mathrm{GI}\left(\underline{\rho}\right)}{D-1} = \frac{\mathrm{GI}\left(\underline{\rho}\right)}{2^{n_q}-1}. \tag{2.3}$$

The Gini threshold t_{g} represents the average sparsity of the diagonal of the density matrix. Thus, the criterion $\sqrt{\rho_{ii}\rho_{jj}} \geq t_{\mathrm{g}}$ requires the measurement of the observables associated with the off-diagonal elements ρ_{ij} that are above the Gini threshold.

The second criterion for setting the threshold is based on the Shannon entropy, here defined as:

$$S\left(\underline{\rho}\right) = -\sum_{i=0}^{D-1} \rho_{ii} \log_D\left(\rho_{ii}\right). \tag{2.4}$$

We adopted the base-D logarithm instead of the usual base-2 version for the Shannon entropy, normalized between 0 and 1. The **entropy threshold** is then defined as:

$$t_{\mathrm{e}} = \frac{1 - S\left(\underline{\rho}\right)}{D}. \tag{2.5}$$

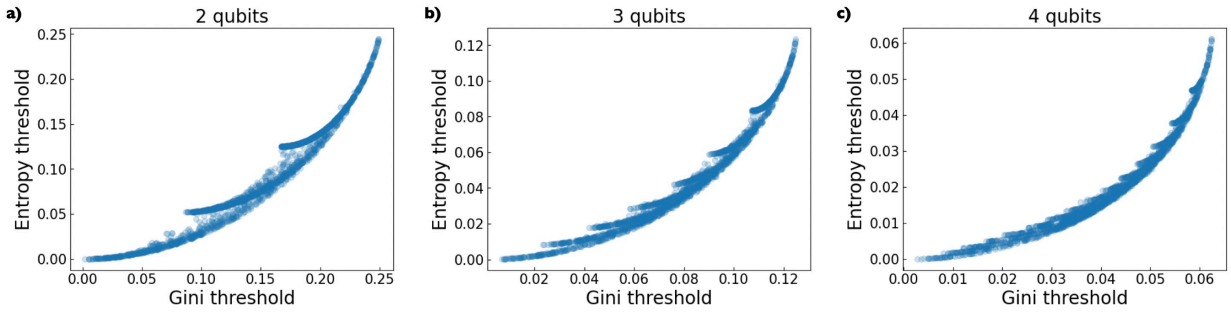


Figure 2.2: **Correlation between the Gini and the entropy threshold.** We show the correlation plots between the Gini and the entropy threshold for **a)** 2, **b)** 3, and **c)** 4 qubits, respectively. The corresponding Pearson correlation coefficient are 0.96, 0.95, and 0.93.

The entropy threshold quantifies the average information provided by the diagonal of the density matrix, and the criterion $\sqrt{\rho_{ii}\rho_{jj}} \geq t_{\mathrm{e}}$ requires the measurement the observables associated with the off-diagonal elements that contain more information than the average provided by the entropy threshold.

We performed some numerical simulations to quantify how the Gini and the entropy thresholds are related to the sparsity of the density matrix, that is, how information is distributed among its elements. We constructed a dataset of 2,000 density matrices, with an equal share of pure and mixed states. For pure states, whose rank is always equal to 1, the number of vanishing diagonal elements z was uniformly chosen between 0 and $2^{n_q} - 2$. For mixed states, z determines the possible values of the rank r of the mixed density matrix, as it varies between 2 and $2^{n_q} - z$. Thus, we first randomly chose z as for pure states and then uniformly chose r accordingly. We generated such datasets for 2, 3, and 4 qubits. For each density matrix of each dataset we computed the Gini threshold, the entropy threshold, the diagonal sparsity, and the matrix sparsity, defined as follows:

$$\text{Diagonal sparsity} \equiv \mathbf{d} = \frac{\text{\# of vanishing diagonal elements}}{D} \tag{2.6}$$

$$\text{Matrix sparsity} \equiv \mathbf{m} = \frac{\text{\# of vanishing matrix elements}}{D^2} \tag{2.7}$$

The correlation plots between Gini and entropy threshold are shown in Figure 2.2. The correlation between the two definitions of the threshold is quantified by the Pearson correlation coefficient, equal to 0.96, 0.95, and 0.93 for 2, 3, and 4 qubits, respectively. Furthermore, the two thresholds cover the same range of values and can be used interchangeably from the point of view of the numerical value of the threshold. We also quantified the correlation between the Gini and entropy thresholds with the diagonal and matrix sparsity defined in (2.6). The correlation plots are reported in Figures 2.3 and 2.4, while the corresponding Pearson correlation coefficients are reported in Table 2.1. We noted that the correlation coefficient between the thresholds and the sparsities increased with the number of qubits. This happens because the possible values that the sparsities can assume increases with the number of qubits, as is evident also from the plots, thus increasing the correlations between them and the threshold values considered here. The data allowed us to conclude that the Gini threshold is clearly correlated with diagonal and matrix sparsity. This is consistent with the definition of the Gini threshold, based on

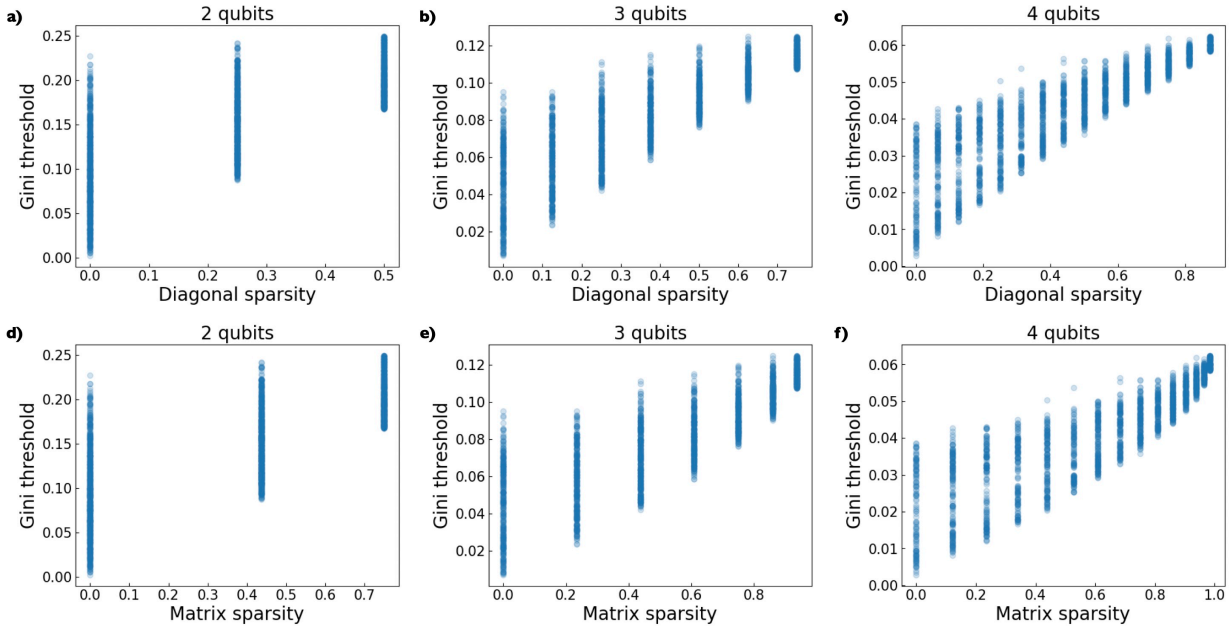


Figure 2.3: **Correlation between the Gini threshold and diagonal and matrix sparsity.** We show the correlation plots between the Gini threshold and the diagonal **(a, b, c)** and matrix **(d, e, f)** sparsity for 2, 3, and 4 qubits, respectively.

Number of qubits	$\mathrm{Corr}(t_{\mathrm{g}}, \mathbf{d})$	$\mathrm{Corr}(t_{\mathrm{g}}, \mathbf{m})$	$\mathrm{Corr}(t_{\mathrm{e}}, \mathbf{d})$	$\mathrm{Corr}(t_{\mathrm{e}}, \mathbf{m})$
2	0.76	0.76	0.83	0.82
3	0.86	0.85	0.91	0.87
4	0.88	0.87	0.93	0.86

Table 2.1: **Correlation between thresholds and sparsities.** We report the number of qubits, the Pearson correlation coefficient between: the Gini threshold and the diagonal sparsity; the Gini threshold and the matrix sparsity; the entropy threshold and the diagonal sparsity; the entropy threshold and the matrix sparsity.

the Gini index that quantifies the sparsity of the diagonal, which in turn is related to the sparsity of the entire density matrix. Similarly, we can conclude that the entropy threshold is also positively correlated with the diagonal and matrix sparsity. Furthermore, the correlation plots show that the functional relation with diagonal and matrix sparsity is different for the two definitions of the thresholds. We believe that this reflects the different quantities considered for the definition of the two thresholds.

2.4 How to construct the projective measurements

The choice of the threshold determines the most relevant off-diagonal elements. We now have to measure the expectation value of a certain number of observables to find the real and imaginary parts of those off-diagonal elements. Given a state ρ, we know that the expectation values of the observables defined in (1.24) and (1.26) allow one to retrieve

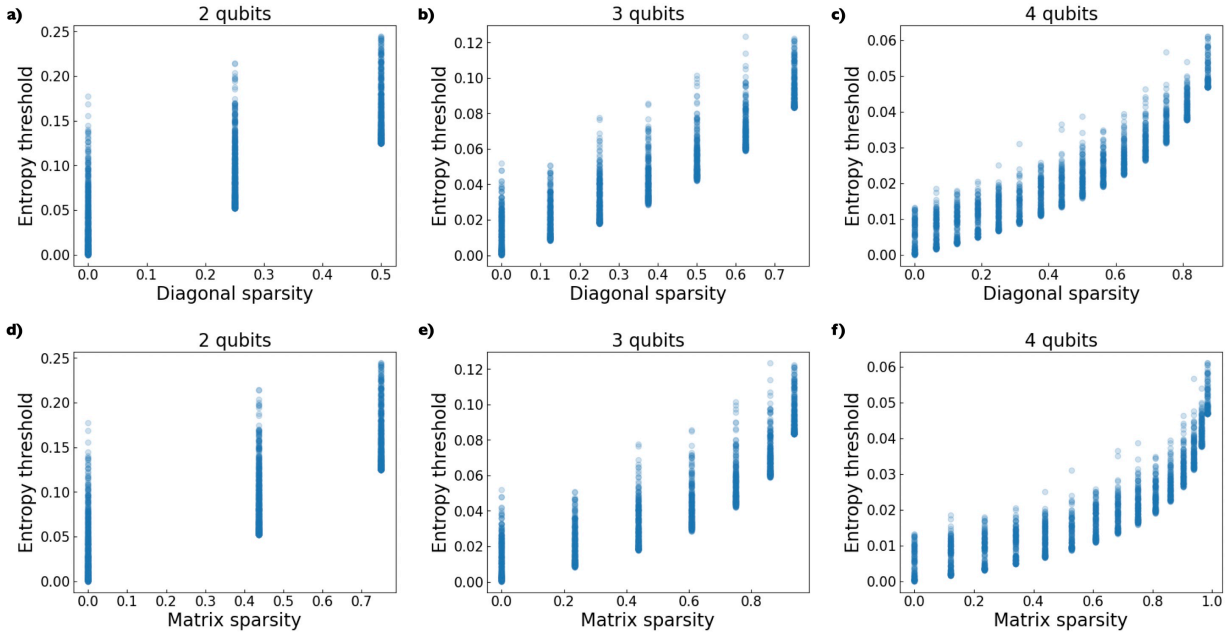


Figure 2.4: **Correlation between the entropy threshold and the diagonal and matrix sparsity.** We show the correlation plots between the entropy threshold and the diagonal **(a, b, c)** and matrix **(d, e, f)** sparsity for 2, 3, and 4 qubits, respectively.

the real and imaginary parts of ρ_{ij} through $\mathrm{Tr}\left(O_{ij}^{(\mathrm{Re,Im})}\rho\right)$, respectively. However, we already proved that these observables are not projective operators, and their experimental implementation is still unknown. In contrast, the observables provided by the tQST protocol are local projective measurements, that is, $\Pi = |\psi\rangle\langle\psi|$, with $|\psi\rangle = \otimes_{k=1}^{n_q} |\phi_k\rangle$ and $|\phi_k\rangle$ eigenstate of one of the Pauli matrices. Then, we associate two projectors with each off-diagonal element, $\Pi_{ij}^{(\mathrm{Re})}$ for the real part and $\Pi_{ij}^{(\mathrm{Im})}$ for the imaginary part. In this way, we recover the 4^{n_q} projectors of QST if we consider all off-diagonal elements. For a given off-diagonal element ρ_{ij}, the associated Π is the projector that is closest to the observables defined in (1.24) and (1.26) according to the 2-norm:

$$\Pi_{ij}^{(\mathrm{Re,\,Im})} = \arg\min_{\Pi} \|O_{ij}^{(\mathrm{Re,Im})} - \Pi\|_2. \tag{2.8}$$

This way of constructing the projectors allows us to retrieve as much information as possible about the real and imaginary parts of the off-diagonal elements they are associated with, while still being feasible to implement from the experimental point of view.

We recall the definition of a *tomographically complete* set of projectors introduced in [22]. A set of projectors is tomographically complete if there exists a unique set of scalars $a_K^{(\mathrm{Re,Im})}$, with $K = \{ij\}$ a multi-index, such that the density matrix can be uniquely reconstructed, that is:

$$\rho = \sum_K a_K^{(\mathrm{Re,Im})} \Pi_K^{(\mathrm{Re,Im})}. \tag{2.9}$$

Equivalently, suppressing the "Re" and "Im" superscript indices, we get:

$$\mathrm{Tr}\left(\Pi_L^\dagger \rho\right) = \sum_K a_K \mathrm{Tr}\left(\Pi_L^\dagger \Pi_K\right) = \sum_K M_{LK} a_K, \tag{2.10}$$

which means that the matrix $M_{LK} = \mathrm{Tr}\left(\Pi_L^\dagger \Pi_K\right) = |\langle\psi_L|\psi_K\rangle|^2$ has to be invertible and the scalars a_K are uniquely determined as $a_K = (M^{-1})_{KL}\,\mathrm{Tr}\left(\Pi_L^\dagger \rho\right)$ [22]. Notice that due to the presence of equivalent minima in (2.8), it is not guaranteed that a complete set exists. However, if a complete set exists, then the matrix M is positive definite.

We found that for systems composed of n_q qubits, a tomographically complete set of projectors associated to the elements of a n_q-qubit density matrix can be constructed from the eigenvectors of the Pauli matrices. We recall the definition of such eigenvectors for convenience here:

$$\begin{aligned} &|H\rangle\,; && |D\rangle = \frac{1}{\sqrt{2}}\left(|H\rangle + |V\rangle\right); && |R\rangle = \frac{1}{\sqrt{2}}\left(|H\rangle + i\,|V\rangle\right), \\ &|V\rangle\,; && |A\rangle = \frac{1}{\sqrt{2}}\left(|H\rangle - |V\rangle\right); && |L\rangle = \frac{1}{\sqrt{2}}\left(|H\rangle - i\,|V\rangle\right). \end{aligned} \tag{2.11}$$

If we consider a single qubit, our protocol provides the following projectors:

$$\begin{aligned} \Pi_{00}^{\mathrm{re}} &= |H\rangle\langle H|\,, \quad \Pi_{11}^{\mathrm{re}} = |V\rangle\langle V|\,, \\ \Pi_{01}^{\mathrm{re}} &= |D\rangle\langle D|\,, \quad \Pi_{01}^{\mathrm{im}} = |R\rangle\langle R|\,. \end{aligned} \tag{2.12}$$

The projectors on the states of the computational basis determine the diagonal elements of the single-qubit density matrix, while the other two are associated with the real and imaginary part of the only one off-diagonal element, respectively. We can recast these projectors in a more suggestive form, which turns out to be particularly useful in constructing the projectors for more than one qubit:

$$\pi_1 \equiv \begin{bmatrix} |H\rangle & |D\rangle + i\,|R\rangle \\ 0 & |V\rangle \end{bmatrix}. \tag{2.13}$$

This is a 2×2 table structured in such a way that, according to the tQST protocol, the real and imaginary parts of ρ_{ij} are associated with the expectation value of the projectors onto the real and imaginary part of the entry (i, j) in (2.13), respectively. For example, the imaginary part of ρ_{01} is associated with the expectation value of the projector onto the imaginary part of entry $(0, 1)$ of (2.13), that is, $\mathrm{Tr}\left(\rho\Pi_{01}^{\mathrm{im}}\right) = \mathrm{Tr}\left(\rho\,|R\rangle\langle R|\right) = \langle R|\rho|R\rangle$. Entries with a value of 0, which indeed acts as the zero element for the recursive operations below, do not need to be explicitly determined as the density matrix is Hermitian. Consequently, we will assume, without loss of generality, that $j \geq i$ whenever we select off-diagonal elements above threshold and aim to determine Π_{ij}.

A set of separable tomographically complete projectors for $n_q > 1$ qubits can be constructed using the following recursive relation:

$$\pi_{n_q} = \begin{bmatrix} |H\rangle\,\pi_{n_q-1} & |D\rangle\,\pi_{n_q-1} + i\,|R\rangle\,\overline{\pi}_{n_q-1} \\ 0 & |V\rangle\,\pi_{n_q-1} \end{bmatrix}, \tag{2.14}$$

having defined:

$$\overline{\pi}_1 \equiv \begin{bmatrix} |H\rangle & 0 \\ |D\rangle - i\,|R\rangle & |V\rangle \end{bmatrix}, \tag{2.15}$$

and analogously for all n_q. In (2.14), the notation $|X\rangle\,\pi$, with $X = \{H, V, D, R\}$, represents a table whose entries are the result of the (tensor) product of the ket $|X\rangle$ and the kets contained in π. Therefore, π_{n_q} is a table that has twice the number of rows and columns as π_{n_q-1}. As an example, tables $|X\rangle\,\pi_1$ and $|X\rangle\,\overline{\pi}_1$ read:

$$|X\rangle\,\pi_1 = \begin{bmatrix} |XH\rangle & |XD\rangle + i\,|XR\rangle \\ 0 & |XV\rangle \end{bmatrix}, \tag{2.16a}$$

$$|X\rangle\,\overline{\pi}_1 = \begin{bmatrix} |XH\rangle & 0 \\ |XD\rangle - i\,|XR\rangle & |XV\rangle \end{bmatrix}. \tag{2.16b}$$

If we now consider 2 qubits, (2.14) yields:

$$\pi_2 = \begin{bmatrix} |HH\rangle & |HD\rangle + i\,|HR\rangle & |DH\rangle + i\,|RH\rangle & |DD\rangle + i\,|DR\rangle \\ 0 & |HV\rangle & |RR\rangle + i\,|RD\rangle & |DV\rangle + i\,|RV\rangle \\ 0 & 0 & |VH\rangle & |VD\rangle + i\,|VR\rangle \\ 0 & 0 & 0 & |VV\rangle \end{bmatrix}. \tag{2.17}$$

The real and imaginary parts of ρ_{12} is associated with the expectation value of the projector onto the real and imaginary part of the entry $(1, 2)$ of π_2, that is, $\mathrm{Tr}\,(\rho\,|RR\rangle\langle RR|) = \langle RR|\rho|RR\rangle$ and $\mathrm{Tr}\,(\rho\,|RD\rangle\langle RD|) = \langle RD|\rho|RD\rangle$, respectively. The procedure described yields a total of 4^{n_q} projectors, and we have numerically verified that up to $n_q = 14$ the corresponding matrix M is invertible, that is, the set of 4^{n_q} projectors is tomographically complete.

The outlined recursive procedure removes the necessity of generating the entire set of projectors upfront. Instead, we can generate projectors on demand, specifically for the elements of the density matrix identified by the threshold. To this end, we divide the table π_{n_q} into four quadrants, with "1", "2", "3", and "4" referring to the upper left, upper right, lower left, and lower right quadrants, respectively. Each of quadrants 2 and 3 is further divided into an upper ("u") and a lower ("l") triangular part. The real part of the elements along the diagonal is assigned to the u portion, while the imaginary part is assigned to the l portion. To determine the projector corresponding to the density matrix element ρ_{ij}, we initially locate it within π_{n_q}, which immediately determines the projector associated with the first qubit, that is $|H\rangle$, $|D\rangle$, $|R\rangle$, or $|V\rangle$ according to the number of the quadrant where it is located: 1, $2u$, $2l$, or 4, respectively. Then we continue to split the quadrant where the element is found until we reach a resulting quadrant size of 2×2. At each splitting step, if the element falls into quadrants 1 or 4, the projector associated with the next qubit is $|H\rangle$ (for 1) or $|V\rangle$ (for 4). In contrast, if it falls into quadrants 2 or 3, the projector choice depends on its position in the previous splitting step. For an element in an upper quadrant, we select $|D\rangle$; and, for an element in a lower quadrant, we select $|R\rangle$ unless the previous quadrant was either $2l$ or $3l$, in which case the choice is reversed. The number of splitting steps is equal to the number of qubits, which in turn is equal to the number of single-qubit projectors to determine. Table 2.2 summarizes these steps. Figure 2.5 visualizes the proposed algorithm for 3 and 4 qubits. More specifically, in the case of 3 qubits, we consider the imaginary part of the density matrix element ρ_{35}, whose location in π_3 is described by the sequence "$2l3l4$", corresponding to the projector onto $|RDV\rangle$. In the 4-qubit case, the location of the real part of the density matrix element ρ_{49} in π_4 is represented by the sequence "$2l3u12u$" corresponding to the projector onto $|RRHD\rangle$.

Previous quadrant	Next quadrant	1-qubit projector
any	1	$\|H\rangle$
any	4	$\|V\rangle$
any except $2l$, $3l$	$2u$, $3u$	$\|D\rangle$
any except $2l$, $3l$	$2l$, $3l$	$\|R\rangle$
$2l$, $3l$	$2u$, $3u$	$\|R\rangle$
$2l$, $3l$	$2l$, $3l$	$\|D\rangle$

Table 2.2: **Recursive construction of the tQST projectors.** Summary of how to choose the single-qubit projector at each recursive step in (2.14). In the first step, "previous quadrant" is always '"any".

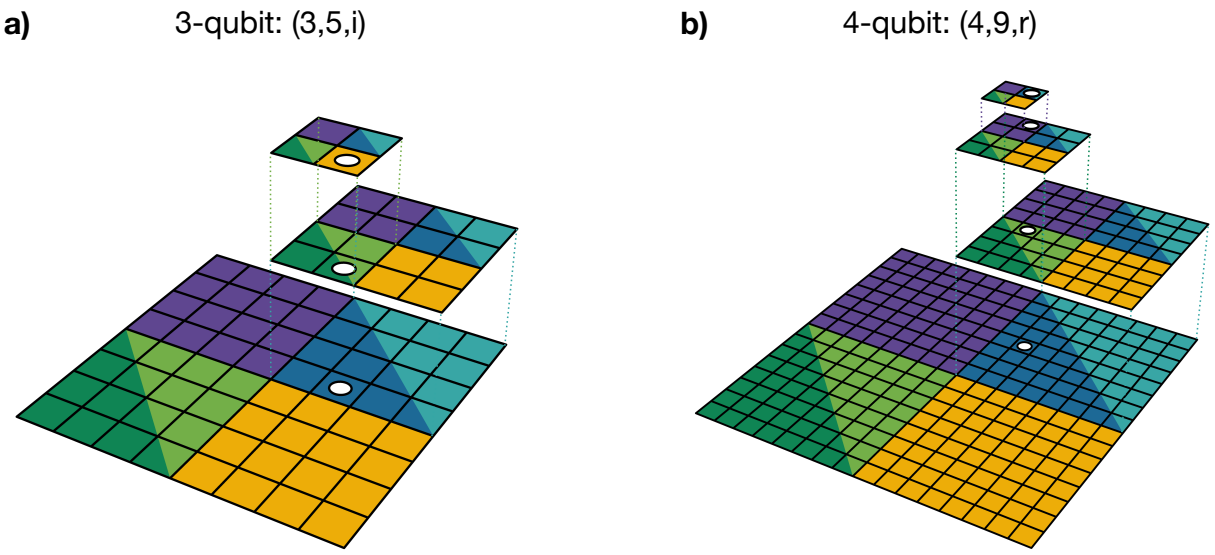


Figure 2.5: **Recursive determination of the tQST projectors associated with a density matrix element for 3 and 4 qubits.** At each step, the density matrix is divided into 4 different quadrants: quadrant "1" is colored in violet, quadrant "2" in blue, quadrant "3" in green, and quadrant "4" in golden yellow. Quadrants "2" and "3" are further subdivided into an upper (lighter) and a lower (darker) part. Each plane represents a successive iteration of the algorithm outlined in the main text, which proceeds from bottom to top. The number of planes is equal to the number of qubits, equal in turn to the number of single-qubit projectors to determine. **a)** The 3-qubit density matrix element $\rho_{35}^{(\mathrm{Im})}$ is located in quadrants "$2l3l4$" and corresponds to the projector onto $|RDV\rangle$ according to Table 2.2. **b)** The 4-qubit density matrix element $\rho_{49}^{(\mathrm{Re})}$ is located in quadrants "$2l3u12u$" therefore corresponding to the projector onto $|RRHD\rangle$ according to Table 2.2. Figure from [1].

2.5 State parametrization and reconstruction

Suppose that we have determined the diagonal elements of the density matrix, chosen a threshold that prescribes a certain number of projective measurements, and performed those measurements. We now have a set of measurement outcomes, a representation of the performed measurements, and we need to process them to reconstruct the state. How do we do this? An important point to consider in the choice of the reconstruction procedure is that the guarantee that the reconstructed state is normalized, Hermitian, and positive semi-definite. These properties can be readily embedded into maximum likelihood estimation (MLE) using a proper parametrization of the density matrix. To this end, we introduced the **triangular** and the **g5** models to parametrize the state. The triangular model is based on the Cholesky decomposition, valid for any Hermitian positive semi-definite matrix, according to which the density matrix is written as:

$$\rho = \frac{T^\dagger T}{\mathrm{Tr}\,(T^\dagger T)}, \tag{2.18}$$

with T an upper or lower triangular complex matrix. This parametrization guarantees that the reconstructed matrix is the proper representation of a quantum state, that is, a physical density matrix. The triangular model is unbiased and the number of parameters to be determined is equal to 4^{n_q} for a state of n_q qubits. However, if one has reasons to believe that ρ describes a high-purity state, then one can express it as:

$$\rho = \frac{R^\dagger R}{\mathrm{Tr}\,(R^\dagger R)} \tag{2.19}$$

with R a complex matrix with shape $(n_{\mathrm{vec}}, 2^{n_q})$, and n_{vec} is the number of eigenvectors used to approximate ρ. This parametrization is similar to the one introduced in [105]. In the case of a high purity state, n_{vec} can be much smaller than the dimension of the Hilbert space; thus the number of parameters needed for the reconstruction will scale more favorably with the number of qubits. In general, for a given value of n_{vec}, the number of parameters is equal to $2 \times n_{\mathrm{vec}} \times 2^{n_q}$. This implies a minimization of the likelihood over a parameter space with a smaller dimension, leading to shorter computational times and more accurate results. Furthermore, it eventually allows one to repeat the minimization multiple times to find results that are closer to the global minimum.

According to the best practices of computational science, we minimize the negative log-likelihood instead of directly maximizing the likelihood itself. In particular, we minimize the following function:

$$L = \sum_K \left(\frac{n_K - N_K}{2\sqrt{n_K}} \right)^2, \tag{2.20}$$

which derives from the assumption that the noise on the counts has a Gaussian probability distribution [22]. Here $n_K = \mathrm{Tr}\,(\rho P_K) = \langle \psi_K | \rho | \psi_K \rangle$ and N_K are the expected, theoretical, and the measured, experimental, outcomes of measuring the projector $P_K = |\psi_K\rangle\langle\psi_K|$ on the state ρ, respectively. Minimization is carried out using the `minimize` numerical routine provided by Scipy [309]. The solver is the L-BFGS-B [310], with a maximum number of iterations equal to 1,000 and a tolerance on the gradient equal to 10^{-4}. The starting

point for the minimization is chosen at random. The specifications of the minimization algorithm are those just presented, unless otherwise specified. The minimization solver is gradient-based; thus it requires one to compute the gradient at every step. It is possible to speed up the minimization by providing an analytical expression of the gradient of the likelihood function. To this end, consider the triangular parametrization of the density matrix, that is, $\rho = T^\dagger T$. If $n_K = \langle\psi_K|\rho|\psi_K\rangle$, then we can expand the state onto which we are projecting on the computational basis as:

$$|\psi_K\rangle = \sum_i v_i^{(K)} |i\rangle \,, \tag{2.21}$$

such that the expectation values are given by:

$$n_K = \langle\psi_K|\rho|\psi_K\rangle = \sum_{a,b,c} v_a^{(K)*} T_{ab}^\dagger T_{bc} v_c^{(K)} \tag{2.22}$$

$$= \sum_b \left(\sum_a v_a^{(K)*} T_{ab}^\dagger \right) \left(\sum_c T_{bc} v_c^{(K)} \right) \tag{2.23}$$

$$= \sum_{a,b} \left| T_{ab} v_b^{(K)} \right|^2 \tag{2.24}$$

$$= \sum_{a,b} \left| T_{ab} V_{bK} \right|^2 , \tag{2.25}$$

having introduced the matrix V with matrix elements $V_{iK} = v_i^{(K)}$. We can use the Wirtinger approach to compute the gradient. In particular, we have to derive the likelihood with respect to T_{ij}^*:

$$\frac{\partial L}{\partial T_{ij}^*} = \frac{\partial L}{\partial n_K} \frac{\partial n_K}{\partial T_{ij}^*} = \sum_K \left(\frac{n_K^2 - N_K^2}{4 n_K^2} \right) \frac{\partial n_K}{\partial T_{ij}^*}, \tag{2.26}$$

from which we find:

$$\frac{\partial n_K}{\partial T_{ij}^*} = \sum_{c,k} T_{ic} V_{ck} V_{kj}^\dagger . \tag{2.27}$$

Finally:

$$\frac{\partial L}{\partial T_{ij}^*} = \sum_{c,k,K} T_{ic} V_{ck} \left(\frac{n_K^2 - N_K^2}{4 n_K^2} \right) V_{kj}^\dagger = \left(TVDV^\dagger \right)_{ij} , \tag{2.28}$$

having defined the matrix D via its matrix element $D_{ik} = \frac{n_K^2 - N_K^2}{4 n_K^2} \delta_{ik}$.

We quantitatively compared the triangular and g5 parametrizations. In Figure 2.6 we show the scaling of the number of variables required by the two parametrizations as a function of the number of qubits. A reduction of up to 3 orders of magnitude can be achieved using the g5 parametrization with $n_{\text{vec}} = 1$, allowing faster minimization. To substantiate and quantify this statement, we considered GHZ and W states from 2 to 10

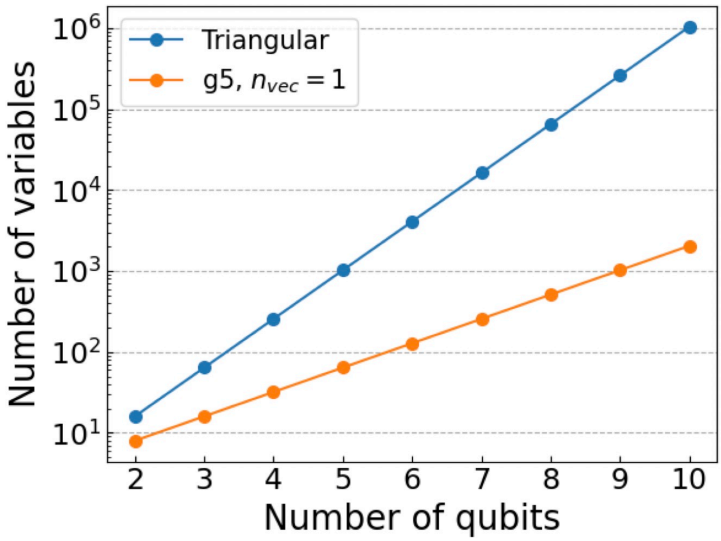


Figure 2.6: **Number of variables required by the triangular and the g5 parametrization with $n_{\text{vec}} = 1$.** We show the number of variables required by the triangular and the g5 parametrizations ($n_{\text{vec}} = 1$) with respect to the number of qubits. Note the semi-logarithmic scale on the vertical axis.

qubits and compared the minimization time and the fidelity between the reconstructed and the target states for the two models. The g5 parametrization was implemented with $n_{\text{vec}} = 1$, because in these specific cases we know in advance that the states are pure. The minimization time for the GHZ and W states is measured using the `timeit` function of the `time` Python library, which provides an average and a standard deviation on the measured computational times. We adopted the Gini threshold to determine the number of measurements for each state. All numerical tests were carried out on a standard desktop computer with an Intel Core i7-4790 processor, 3.60 GHz CPU, and 32 GB system memory, without GPUs. The fidelities achieved with the two parametrizations for the GHZ and W states are reported in Tables 2.3 and 2.4, respectively, while the minimization times are shown in Figure 2.7. The g5 parametrization requires minimization times that are up to 4 orders of magnitude smaller than the ones required with the triangular parametrization. At the same time, the quality of the reconstruction does not degrade, as proved by the fidelity close to unity in all the cases for both the GHZ and the W states. We believe that the decrease in the fidelity for the 10-qubit GHZ state is due to the likelihood reaching a local minimum that is not close to the global one, which in turn is due to the huge dimension of the parameter space where the minimization is carried out.

Finally, we considered an ensemble of 10^3 random pure states of 2, 3 and 4 qubits, and compared the average minimization time and average fidelity of the two parametrizations. We measured the minimization time for every state with the `time` function of the `time` Python library and then computed the average and the standard deviation. We set $n_{\text{vec}} = 2$ for these simulations, for it is possible that the number of variables required by $n_{\text{vec}} = 1$ is too small to achieve high enough fidelities, even if we know in advance that all the states are pure. The results reported in Table 2.5 confirm the conclusions obtained with the GHZ and W states, that is, the g5 parametrization allows a faster minimization of the negative log-likelihood without compromising the quality of the reconstruction. Fur-

Number of qubits	Triangular fidelity	g5 fidelity
2	0.99954	0.99999
3	0.99875	0.99999
4	0.99735	0.99999
5	0.99661	0.99999
6	0.98980	0.99999
7	0.97353	0.99999
8	0.98286	0.99999
9	0.95528	0.99969
10	0.90419	0.99999

Table 2.3: **Fidelity of the triangular and g5 parametrization for GHZ states.** We report the fidelity between the target state $\left|\mathrm{GHZ}_{n_q}\right\rangle$ and the tQST reconstructed state with the triangular and g5 parametrization, respectively, for $2 \leq n_q \leq 10$.

Number of qubits	Triangular fidelity	g5 fidelity
2	0.99977	0.99999
3	0.99622	0.99999
4	0.99470	0.99999
5	0.99340	0.99999
6	0.99233	0.99999
7	0.98892	0.99999
8	0.98940	0.99999
9	0.98852	0.99999
10	0.98010	0.99999

Table 2.4: **Fidelity of the triangular and g5 parametrization for W states.** We report the fidelity between the target state $\left|\mathrm{W}_{n_q}\right\rangle$ and the tQST reconstructed state with the triangular and g5 parametrization, respectively, for $2 \leq n_q \leq 10$.

thermore, we notice that the results for the two parametrizations are compatible within the errors.

Before showing some examples of application of tQST, we highlight that an implementation of the protocol using the Python language is available online on Github [307]. We developed this library and made it public so that other researchers and experimental collaborators could apply and implement tQST independently without the need for specific training from our side.

Number of qubits	Time triangular (s)	Fidelity triangular	Time g5 (s)	Fidelity g5
2	0.007(3)	0.991(24)	0.004(1)	0.991(28)
3	0.014(8)	0.998(5)	0.007(2)	0.999(5)
4	0.050(35)	0.999(2)	0.014(6)	0.999(3)

Table 2.5: **Fidelity and minimization time of triangular and g5 parametrization for random pure states.** We report the average minimization time and the average fidelity between the target state and the tQST reconstruction with the triangular and g5 parametrization, respectively. The errors represent one standard deviation from the average.

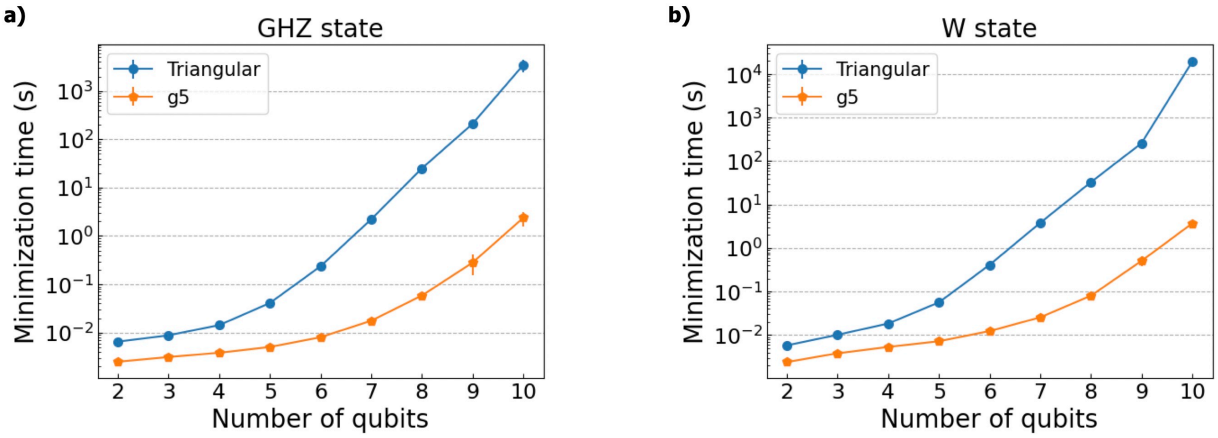


Figure 2.7: **Minimization time of the triangular and the g5 parametrization.** We show the minimization time for **a)** GHZ and **b)** W states with respect to the number of qubits. Note the semi-logarithmic scale on the vertical axis of both panels.

2.6 Complete examples of the tQST protocol

We now illustrate the tQST protocol step by step when applied to different 3-qubit states. Let us start with the following random mixed state:

$$\mathrm{Re}\,(\rho) = \begin{pmatrix} 0.239 & -0.087 & 0 & -0.058 & 0 & 0.085 & 0 & -0.047 \\ -0.087 & 0.274 & 0 & 0.085 & 0 & 0.046 & 0 & 0.129 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.058 & 0.085 & 0 & 0.224 & 0 & 0.124 & 0 & 0.082 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.085 & 0.046 & 0 & 0.124 & 0 & 0.167 & 0 & 0.051 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.047 & 0.129 & 0 & 0.082 & 0 & 0.051 & 0 & 0.096 \end{pmatrix} \tag{2.29}$$

$$\mathrm{Im}\,(\rho) = \begin{pmatrix} 0 & -0.052 & 0 & 0.056 & 0 & 0.086 & 0 & -0.04 \\ 0.052 & 0 & 0 & -0.184 & 0 & -0.102 & 0 & -0.016 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.056 & 0.184 & 0 & 0 & 0 & -0.038 & 0 & 0.05 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.086 & 0.102 & 0 & 0.038 & 0 & 0 & 0 & 0.032 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.04 & 0.016 & 0 & -0.05 & 0 & -0.032 & 0 & 0 \end{pmatrix} \tag{2.30}$$

We first measured the expectation value of the 8 projectors onto the states of the computational basis $\{|HHH\rangle, |HHV\rangle, \ldots, |VVV\rangle\}$ to determine the diagonal elements of the density matrix. We then calculated the Gini threshold for this diagonal, $t_{\mathrm{g}} = 0.069$, from which the tQST protocol identifies 10 off-diagonal elements, listed below along with the corresponding projectors:

$$\begin{aligned}
(0,1,\mathrm{Re}) &\leftrightarrow |HHD\rangle\langle HHD| & (0,1,\mathrm{Im}) &\leftrightarrow |HHR\rangle\langle HHR| \\
(0,3,\mathrm{Re}) &\leftrightarrow |HDD\rangle\langle HDD| & (0,3,\mathrm{Im}) &\leftrightarrow |HDR\rangle\langle HDR| \\
(0,5,\mathrm{Re}) &\leftrightarrow |DHD\rangle\langle DHD| & (0,5,\mathrm{Im}) &\leftrightarrow |DHR\rangle\langle DHR| \\
(0,7,\mathrm{Re}) &\leftrightarrow |DDD\rangle\langle DDD| & (0,7,\mathrm{Im}) &\leftrightarrow |DDR\rangle\langle DDR| \\
(1,3,\mathrm{Re}) &\leftrightarrow |HDV\rangle\langle HDV| & (1,3,\mathrm{Im}) &\leftrightarrow |HRV\rangle\langle HRV| \\
(1,5,\mathrm{Re}) &\leftrightarrow |DHV\rangle\langle DHV| & (1,5,\mathrm{Im}) &\leftrightarrow |RHV\rangle\langle RHV| \\
(1,7,\mathrm{Re}) &\leftrightarrow |DDV\rangle\langle DDV| & (1,7,\mathrm{Im}) &\leftrightarrow |DRV\rangle\langle DRV| \\
(3,5,\mathrm{Re}) &\leftrightarrow |RRV\rangle\langle RRV| & (3,5,\mathrm{Im}) &\leftrightarrow |RDV\rangle\langle RDV| \\
(3,7,\mathrm{Re}) &\leftrightarrow |DVV\rangle\langle DVV| & (3,7,\mathrm{Im}) &\leftrightarrow |RVV\rangle\langle RVV| \\
(5,7,\mathrm{Re}) &\leftrightarrow |VDV\rangle\langle VDV| & (5,7,\mathrm{Im}) &\leftrightarrow |VRV\rangle\langle VRV|
\end{aligned}$$

The total number of required projectors is thus 28, less than half of the 64 required by QST. By performing these measurements, we finally reconstruct the state using the triangular model and the maximum likelihood, achieving a fidelity between the target and the reconstructed state of 0.99905.

We then examined the following random pure state:

$$\mathrm{Re}\,(\rho) = \begin{pmatrix} 0.028 & 0 & 0 & 0 & 0.044 & 0.026 & 0.092 & 0.115 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.044 & 0 & 0 & 0 & 0.08 & 0.054 & 0.171 & 0.19 \\ 0.026 & 0 & 0 & 0 & 0.054 & 0.04 & 0.117 & 0.117 \\ 0.092 & 0 & 0 & 0 & 0.171 & 0.117 & 0.369 & 0.403 \\ 0.115 & 0 & 0 & 0 & 0.19 & 0.117 & 0.403 & 0.483 \end{pmatrix} \tag{2.31}$$

$$\mathrm{Im}\,(\rho) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0.017 & 0.021 & 0.042 & 0.014 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.017 & 0 & 0 & 0 & 0 & 0.017 & 0.009 & -0.05 \\ -0.021 & 0 & 0 & 0 & -0.017 & 0 & -0.031 & -0.075 \\ -0.042 & 0 & 0 & 0 & -0.009 & 0.031 & 0 & -0.127 \\ -0.014 & 0 & 0 & 0 & 0.05 & 0.075 & 0.127 & 0 \end{pmatrix} \tag{2.32}$$

By measuring the projectors onto the elements of the computational basis, we determined the diagonal elements and used them to compute the Gini threshold, equal to $t_{\mathrm{g}} = 0.098$. The off-diagonal elements above threshold and the corresponding projectors are:

$$\begin{aligned} &(0,6,\mathrm{Re}) \leftrightarrow |DDH\rangle\langle DDH| && (0,6,\mathrm{Im}) \leftrightarrow |DRH\rangle\langle DRH| \\ &(0,7,\mathrm{Re}) \leftrightarrow |DDD\rangle\langle DDD| && (0,7,\mathrm{Im}) \leftrightarrow |DDR\rangle\langle DDR| \\ &(4,6,\mathrm{Re}) \leftrightarrow |VDH\rangle\langle VDH| && (4,6,\mathrm{Im}) \leftrightarrow |VRH\rangle\langle VRH| \\ &(4,7,\mathrm{Re}) \leftrightarrow |VDD\rangle\langle VDD| && (4,7,\mathrm{Im}) \leftrightarrow |VDR\rangle\langle VDR| \\ &(5,6,\mathrm{Re}) \leftrightarrow |VRR\rangle\langle VRR| && (5,6,\mathrm{Im}) \leftrightarrow |VRD\rangle\langle VRD| \\ &(5,7,\mathrm{Re}) \leftrightarrow |VDV\rangle\langle VDV| && (5,7,\mathrm{Im}) \leftrightarrow |VRV\rangle\langle VRV| \\ &(6,7,\mathrm{Re}) \leftrightarrow |VVD\rangle\langle VVD| && (6,7,\mathrm{Im}) \leftrightarrow |VVR\rangle\langle VVR| \end{aligned}$$

The fidelity between the reconstructed state and the target state using the measurement outcomes corresponding to the listed projectors is 0.99873.

The W state is one of the states for which tQST is most favorable, as it has only $n_q\,(n_q - 1)$ off-diagonal elements, as can be seen from the density matrix shown in Figure 1.2d). Consequently, the number of projectors required by tQST for W states scales quadratically with the number of qubits, which is more feasible than the exponential scaling of QST. If $n_q = 3$, then tQST requires one to measure only 6 additional projectors after the diagonal ones, correctly identified by the Gini threshold $t_{\mathrm{g}} = 0.089$. In this case, the off-diagonal elements and the corresponding projectors are:

$$\begin{aligned} &(1,2,\mathrm{Re}) \leftrightarrow |HRR\rangle\langle HRR| && (1,2,\mathrm{Im}) \leftrightarrow |HRD\rangle\langle HRD| \\ &(1,4,\mathrm{Re}) \leftrightarrow |RHR\rangle\langle RHR| && (1,4,\mathrm{Im}) \leftrightarrow |RHD\rangle\langle RHD| \\ &(2,4,\mathrm{Re}) \leftrightarrow |RRH\rangle\langle RRH| && (2,4,\mathrm{Im}) \leftrightarrow |RDH\rangle\langle RDH| \end{aligned}$$

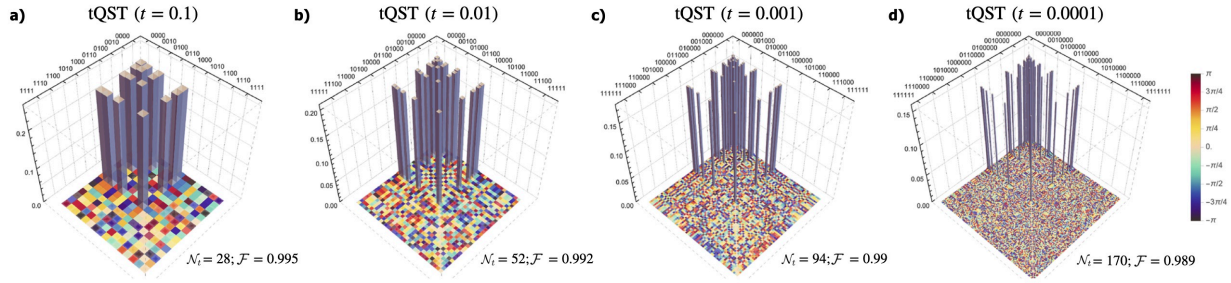


Figure 2.8: **W states and tQST.** We show the tQST reconstructions of W states for different numbers of qubits and threshold values: **a)** 4 qubits, $t = 0.1$, **b)** 5 qubits, $t = 0.01$, **c)** 6 qubits, $t = 0.001$, and **d)** 7 qubits, $t = 0.0001$. For each number of qubits we report the number $\mathcal{N}_t$ of projectors used and the fidelity $\mathcal{F}$ with the corresponding target state. Figure adapted from [1].

The tQST reconstruction with these 14 projective measurements achieves a fidelity equal to 0.99793 with the target state. In Figure 2.8 we show the numerical reconstruction of the density matrix of W states for numbers of qubits ranging from 4 to 7 using tQST. Even for the case of 7 qubits, in which QST would have required 16,384 measurements, we could reconstruct the density matrix achieving a fidelity of about 0.99 with the target state by implementing only 170 projective measurements. Furthermore, the number of projective measurements is exactly equal to the number of diagonal elements (2^{n_q}) plus the number of non-vanishing off-diagonal elements of a W state $[n_q(n_q - 1)]$. We stress that the dramatic reduction in the number of projective measurements is not simply given by the choice of a particular state but rather by the employed computational basis, which affects the state representation.

Finally, we consider the GHZ state, which is the non-trivial state exhibiting the best advantage when applying tQST, for it has only 2 non-vanishing off-diagonal elements, regardless of the number of qubits that make up the system, as shown in Figure 1.2a). The Gini threshold t_g identifies only two off-diagonal elements, whose projectors are:

$$(0, 7, \mathrm{Re}) \leftrightarrow |DDD\rangle\langle DDD| \qquad (0, 7, \mathrm{Im}) \leftrightarrow |DDR\rangle\langle DDR| \,.$$

The fidelity between the tQST reconstruction and the target state is 0.99912.

2.7 Numerical analysis

The previous examples referred to states that are random or relevant for technological applications and for which tQST is particularly advantageous. However, this is not sufficient to conclude that the protocol is generally effective. To this end, we performed numerical simulations to analyze the effectiveness of the tQST protocol. For all of the following analyses, we used the same datasets of quantum states as in Section 2.3.

We first examined whether there was a relation between the number of measurements required by the protocol, the purity, and the sparsity (quantified by the Gini threshold) of the considered density matrices. The results, shown in Figure 2.9, suggest that the number of projective measurements required by tQST is inversely proportional to the sparsity of the state, regardless of whether the state is pure or mixed. This result is in

agreement with the analyses conducted so far, as tQST can greatly reduce the number of measurements for states that have many vanishing or small elements by a proper choice of the threshold.

Then, we reconstructed each state of the 2-, 3-, and 4-qubit datasets using the Gini threshold and the triangular parametrization. Figure 2.10 a), d), g) show the distribution of the number of projective measurements for the three datasets considered. The Gini threshold allowed us to use a number of projective measurements that is actually different from the one of QST in most of the cases. In particular, we performed fewer projective measurements than QST for 79%, 93%, and 97% of the 2-, 3-, and 4-qubit states, with an average number of measurements of $9(3)$, $28(16)$, $91(64)$, respectively. Here, the uncertainty is the standard deviation from the average. However, the distribution of the number of measurements is not enough to prove that tQST works effectively, that is, the states are reconstructed with high fidelity. We achieved an average fidelity of $0.994(19)$, $0.998(5)$, and $0.998(2)$ for 2, 3, and 4 qubits, respectively. For comparison, QST achieved average fidelities of $0.99977(294)$, $0.99947(201)$, and $0.99935(169)$ for the same number of qubits. Figure 2.10b), e), h) shows the fidelity between the reconstructed and target states with respect to the number of projective measurements required by tQST for each state, while Figure 2.10c), f), i) shows a zoomed-in version of the same plot, along with the average fidelity for each number of projective measurements. The data clearly show that tQST can reconstruct the states with high fidelity, even when we perform much fewer projective measurements than QST. Furthermore, the average fidelity is consistently high for each number of projective measurements, allowing us to conclude that the global average is not biased by the state for which many projectors are required.

2.8 Fidelity lower bound

The previous results show a notable feature of tQST, that is, a significant reduction in the number of projective measurements does not necessarily lead to large errors in the reconstruction of the density matrix. However, this feature is strongly dependent on the state representation. In general, one may be interested in using tQST because the available resources are simply not enough to implement the QST. In this case, it is important to have an estimate of the largest error associated with the threshold choice. Specifically, given the diagonal elements $\{\rho_{ii}\}$ and a corresponding threshold t, one can estimate a lower bound for the fidelity achievable through a tQST reconstruction.

To determine the fidelity bound, we start from Fuchs–van de Graaf inequalities:

$$1-\sqrt{F\left(\rho_1,\rho_2\right)} \leq \frac{1}{2}\|\rho_1-\rho_2\|_1 \leq \sqrt{1-F\left(\rho_1,\rho_2\right)^2}, \tag{2.33}$$

with $\|\rho_1-\rho_2\|_1$ the trace norm between ρ_1 and ρ_2, with $\|A\|_1 = \mathrm{Tr}\left(\sqrt{A^\dagger A}\right)$. A second useful result is (see the last equation of section IIA in [311]):

$$\|\rho_1-\rho_2\|_1 \leq 2\sqrt{\min\left[\mathrm{rank}\left(\rho_1\right),\mathrm{rank}\left(\rho_2\right)\right]}\|\rho_1-\rho_2\|_2, \tag{2.34}$$

with $\|A\|_2 = \sqrt{\mathrm{Tr}\left(A^\dagger A\right)}$ the Frobenius norm of A [312].

Let us define $\mathcal{C}(t)$ as the set of pairs of indexes (i,j) such that $\sqrt{\rho_{ii}\rho_{jj}} \leq t$, consider the case $\rho_1=\rho$ and $\rho_2=\rho_t$, with $\rho_t = \{0\ \forall(i,j)\in\mathcal{C}(t);\rho_{ij}\ \text{otherwise}\}$ (note that $\rho=\rho_t$ if

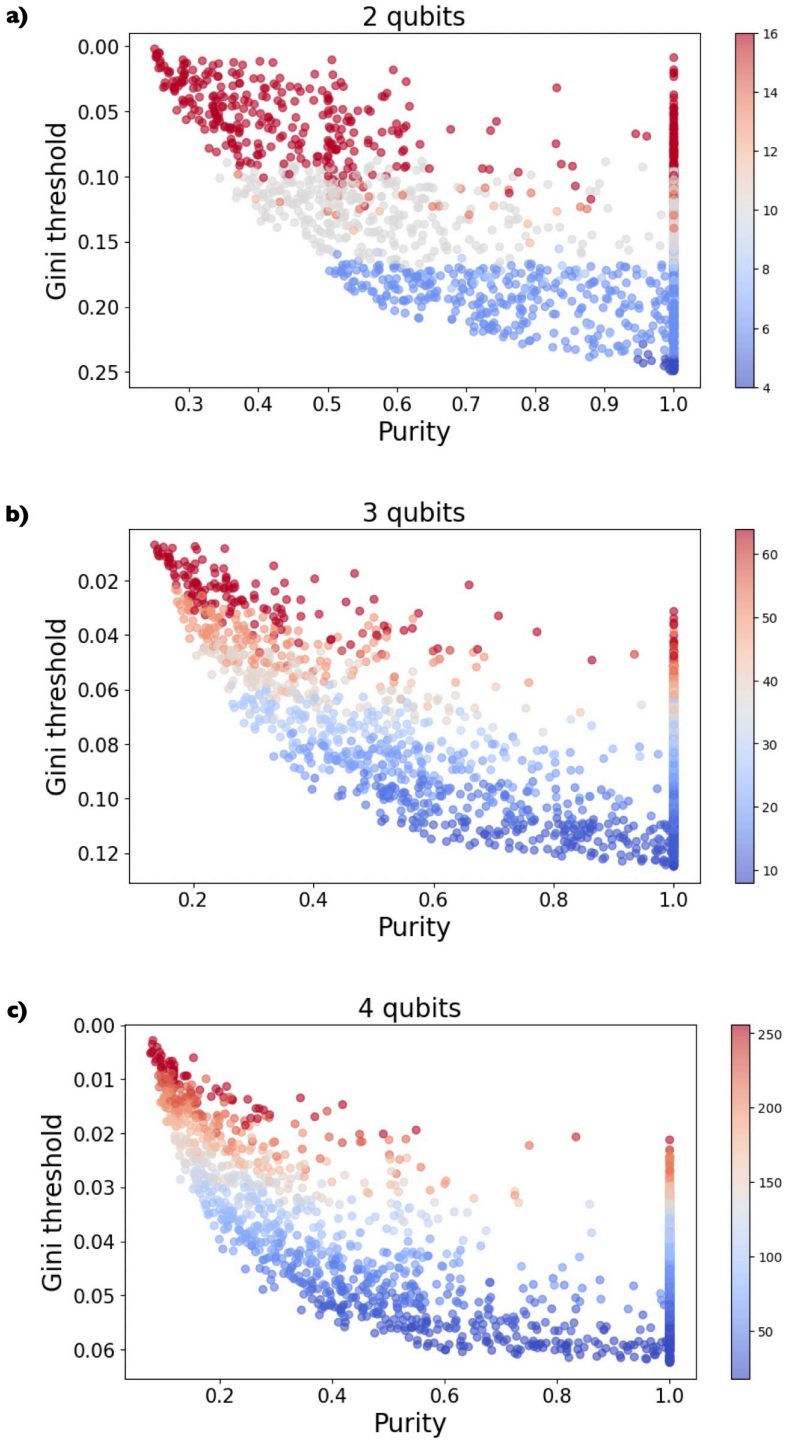


Figure 2.9: **Relation between sparsity, purity, and number of projective measurements.** We show the correlation plot between sparsity (Gini threshold), purity, and number of projective measurement of each state in the dataset for **a)** 2, **b)** 3, and **c)** 4 qubits. The number of projective measurements is encoded in the colorbar and ranges from 2^{n_q} to 4^{n_q}. Each point represents a different quantum state of the dataset. Note that reverse vertical axis, increasing from top to bottom.

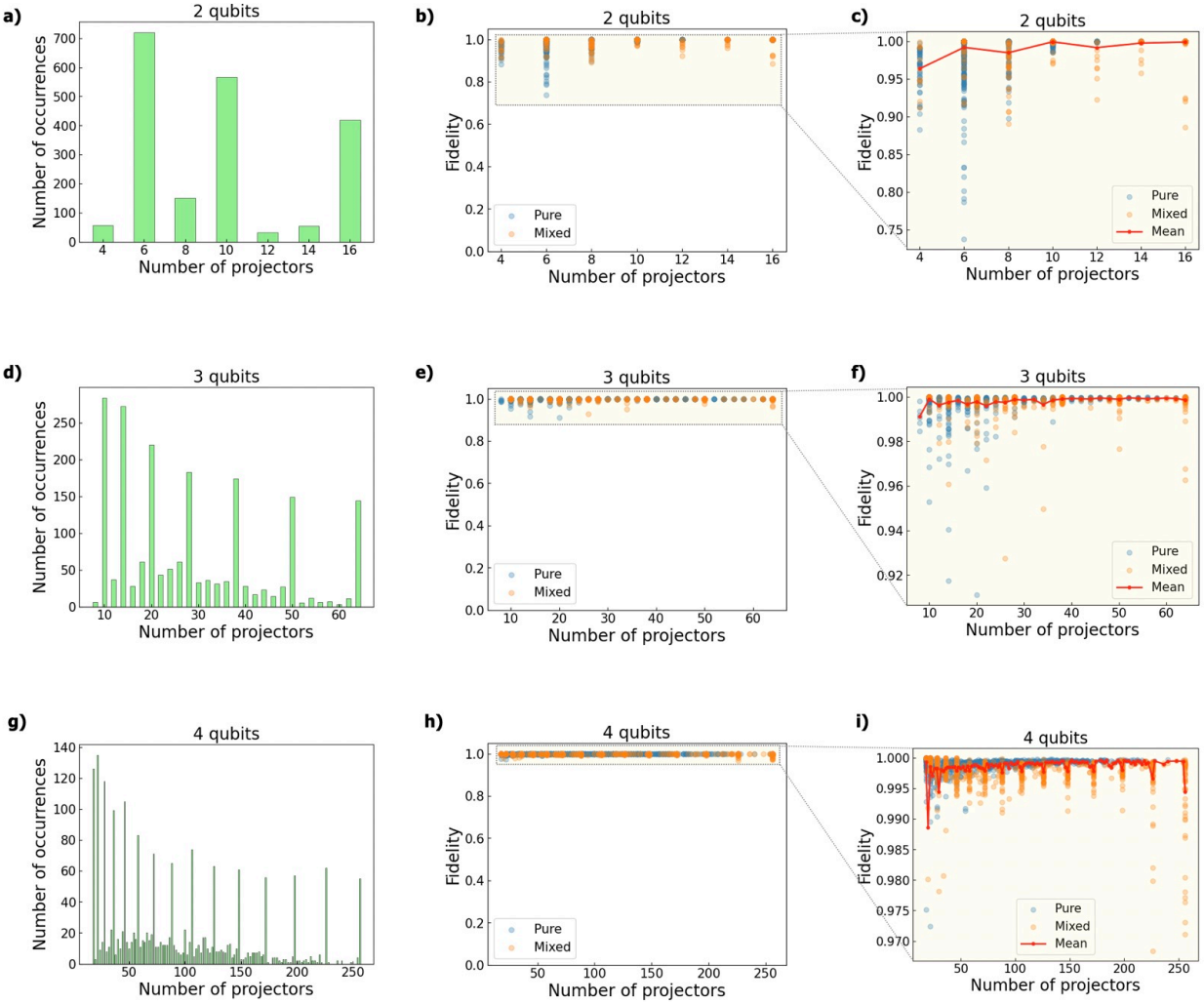


Figure 2.10: **Numerical validation of tQST. a)-d)-g)**: Distribution of the number of projective measurements used to reconstruct the states. **b)-e)-h)**: Plot of the fidelity between the reconstructed and the target state with respect to the number of projective measurements for 2, 3, and 4 qubits, respectively. Pure and mixed states are plotted with different colors, shown in the legend. **c)-f)-i)**: Detail of the vertical axis of the plots of the central column. We show in red the trend of the mean computed for each number of projective measurements.

$t = 0$), and define $\Delta = \rho - \rho_t$. Let's briefly examine the properties of Δ. We can prove that $\mathrm{Tr}\left[\Delta\right] = 0$:

$$\mathrm{Tr}\left[\Delta\right] = \mathrm{Tr}\left[\rho - \rho_t\right] = \mathrm{Tr}\left[\rho\right] - \mathrm{Tr}\left[\rho_t\right] = 1 - 1 = 0, \tag{2.35}$$

since the diagonal elements of ρ are unchanged after applying the threshold t. Furthermore, Δ is Hermitian:

$$\Delta^\dagger = (\rho - \rho_t)^\dagger = \rho^\dagger - \rho_t^\dagger = \rho - \rho_t = \Delta. \tag{2.36}$$

By definition $\Delta_{ii} = 0 \ \forall i$, $\Delta_{ij} \neq 0$ for $(i,j) \in \mathcal{C}\left(t\right)$. Furthermore, $|\Delta_{ij}|^2 \leq \rho_{ii}\rho_{jj}$. From the first part of Eq. (2.33):

$$\begin{aligned} 1 - \sqrt{F\left(\rho, \rho_t\right)} &\leq \frac{1}{2}\|\rho - \rho_t\|_{\mathrm{Tr}} \\ &\leq \frac{1}{2} 2\sqrt{\min\left(\mathrm{rank}\left(\rho\right)\mathrm{rank}\left(\rho_t\right)\right)}\|\rho - \rho_t\|_2 \\ &\leq \sqrt{\min\left(\mathrm{rank}\left(\rho\right)\mathrm{rank}\left(\rho_t\right)\right)}\|\Delta\|_2. \end{aligned} \tag{2.37}$$

We can observe that $\mathrm{rank}(\rho) \leq \mathrm{rank}(\rho_t)$, for considering $t > 0$ leads, in general, to a decreasing of the state purity. Then:

$$\sqrt{\min\left(\mathrm{rank}\left(\rho\right)\mathrm{rank}\left(\rho_t\right)\right)} = \sqrt{\mathrm{rank}\left(\rho\right)}. \tag{2.38}$$

By using the properties of Δ we find:

$$\|\Delta\|_2 = \sqrt{\mathrm{Tr}\left[\Delta^\dagger\Delta\right]} = \sqrt{\mathrm{Tr}\left[\Delta^2\right]}. \tag{2.39}$$

To compute $\mathrm{Tr}\left[\Delta^2\right]$ we can start from:

$$\Delta_{ij}^2 = \sum_m \Delta_{im}\Delta_{mj} = \sum_m \Delta_{im}\Delta_{jm}^*. \tag{2.40}$$

Hence:

$$\Delta_{ii}^2 = \sum_m \Delta_{im}\Delta_{im}^* = \sum_m |\Delta_{im}|^2. \tag{2.41}$$

Finally:

$$\mathrm{Tr}\left[\Delta^2\right] = \sum_i \Delta_{ii}^2 = \sum_i \sum_m |\Delta_{im}|^2 \leq \sum_{(i,j)\in\mathcal{C}(t)} \rho_{ii}\rho_{jj}. \tag{2.42}$$

We can then conclude that:

$$\sqrt{\mathrm{Tr}\left[\Delta^2\right]} \leq \sqrt{\sum_{(i,j)\in\mathcal{C}(t)} \rho_{ii}\rho_{jj}}. \tag{2.43}$$

With this result at hand, by combining (2.33) and (2.34) we find:

$$F \geq F_{\mathrm{bound}}\left(t\right) = \left(1 - \sqrt{\mathrm{rank}(\rho)}\sqrt{\sum_{(i,j)\in\mathcal{C}(t)} \rho_{ii}\rho_{jj}}\right)^2. \tag{2.44}$$

This result is very general and may be helpful in choosing an appropriate threshold. However, it should be noted that the fidelity with which the state is reconstructed could be in general much greater than the lower bound given by (2.44).

2.9 Experimental results on an IBMQ processor

In this section, we show the experimental results of using tQST to reconstruct quantum states generated on a superconducting platform provided by IBM. In particular, we experimentally demonstrate tQST using the IBMQ processor `Lagos`, which allows one to prepare states of up to 7 superconducting qubits (see Appendix A.1 for the platform specifications and performance metrics). In our implementation, we program the system to generate a target quantum state by constructing the corresponding quantum circuit. We first reconstruct the corresponding density matrix by performing QST as implemented in `qiskit` [313, 314], which we call IBMQ-QST. This protocol reconstructs the density matrix using linear inversion on the outcomes of an over-complete set of 6^{n_q} observables. In fact, the `qiskit` package implements measurement settings by default, the total number of which is 3^{n_q}, each corresponding to 2^{n_q} projectors, thus the total number of observables required by IBMQ-QST. This set comprises all possible projectors that can be constructed as tensor products of the projectors onto the eigenvectors of the Pauli matrices. Reconstruction via linear inversion implemented by `qiskit` might yield a normalized, Hermitian, but non-positive semi-definite matrix, which is subsequently rescaled to be positive semi-definite using the method outlined in [56]. We then reconstruct the same state using tQST, where the threshold t is chosen considering the typical signal-to-noise ratio (SNR) of the system to avoid unnecessary measurements. In this case, we leveraged maximum likelihood estimation to obtain the reconstructed density matrix from the measured expectation values, as explained in the previous sections. To compare the results achieved by the two protocols, we compute the fidelities between the target state and the reconstructed ones, and also the fidelity between the two reconstructed states.

In Section 2.3, we anticipated that the choice of an appropriate threshold t depends on several factors, such as the specific physical system used to implement the qubits, which determines the kind and level of noise, the amount of available resources, such as the usable time for the experiment, or the target quantum state to be generated. In the case of IBMQ quantum processors, it is possible to evaluate a circuit-specific threshold using the IBMQ simulator available in the `qiskit` package. First, we simulate the unitary evolution of a ground-state initialized quantum register according to the circuit itself, project it onto the computational basis, and determine the diagonal counts of the target state using $n_{\text{shots}} = 10,000$ shots, according to `qiskit` nomenclature. These counts are generated without any kind of experimental error, such as the ones due to the finite sample size or decoherence of the system, and are separated into zero and non-zero counts. Second, we use the IBMQ simulator, which includes noise effects close to the experimental ones, to run the circuit a certain number of times, equal to 100 in our case. For every run, we record the maximum value of the counts among the expected zero elements of the diagonal, $c_0^{\max}$, and the minimum value of the counts for the smallest expected non-zero diagonal element, $c_{>0}^{\min}$. Third, we define the *noise threshold* as $t_0 = c_0^{\max} + n_q\sqrt{c_0^{\max}}$, and the *signal threshold* as $t_{>0} = c_{>0}^{\min} - n_q\sqrt{c_{>0}^{\min}}$. The square root terms consider the variability of the counts $c_0^{\max}$ and $c_{>0}^{\min}$ in every run of the circuit, while the factor n_q takes into account that noise increases with the number of qubits n_q for the quantum processor considered. Finally, we define a circuit-specific threshold as:

$$t_n = \frac{\max(t_0, t_{>0})}{n_{\text{shots}}}. \tag{2.45}$$

This circuit-specific threshold basically does not consider the diagonal entries that are most affected by noise. For the depth-3 quantum circuits analyzed here, we generally have $t_{>0} \gtrsim t_0$. In particular, tQST works best whenever $t_{>0} \gg t_0$, while $t_{>0} \sim t_0$ indicates an unfavorable SNR.

In our analysis, we generated 300 random states for each number of qubits from 4 to 7 and sorted them according to the diagonal filling, that is, the percentage of expected non-vanishing diagonal elements. Random states were generated from random circuits with circuit depth equal to 3. We performed 10 independent runs for each circuit and in each run we reconstructed the state with the IBMQ-QST protocol. Then, for each of the considered circuits, we set the circuit-specific threshold t_n as discussed above and reconstructed the state with the tQST protocol. Due to the presence of noise, different runs may result in a different number of projective measurements to be performed because some of the diagonal elements might have slightly different values. Thus, we also performed 10 independent runs for each circuit with the tQST protocol. In Figure 2.11, we show representative results for diagonal filling equal to 25% [a), (d)], 50% [b), (e)], and 75% [c), f)] for the cases of 4 and 7 qubits. Each panel of the figure shows the density matrix reconstructed with IBMQ-QST and the one obtained with tQST. In the latter case, we also report the value of the threshold t_n and the corresponding number $\mathcal{N}_t$ of projective measurements used in the tQST reconstruction. Other representative results for 5 and 6 qubits are shown in Figure 2.12, where we also display the circuits used to generate the states. The representative states shown in Figures 2.11 and 2.12 correspond to the runs where the minimum number of measurements was performed. For the cases of 4, 5, and 6 qubits, we observe up to a 100-fold reduction in the number of required measurements compared to IBMQ-QST. For the 7-qubit state shown in Figure 2.11d), the reduction factor is ≈ 300.

As is evident from the state representation, the states reconstructed with IBMQ-QST and tQST are very similar, despite the substantial difference in the number of projective measurements. We properly quantify this similarity by calculating the fidelity between the two reconstructed states, which in all cases is about 0.90 or higher, as shown in Figure 2.13. The same figure also reports the fidelity between the tQST-reconstructed state and the target state. Remarkably, this is always comparable, within error bars, to the fidelity between the IBMQ-QST reconstruction and the target state, as reported in Table 2.6, where a detailed analysis of the states shown in Figure 2.12 is presented. No advantage is obtained with IBMQ-QST by performing more measurements than those set by the circuit-specific threshold in tQST. This suggests that when the threshold is determined by the SNR of the system, the tQST protocol can extract all the information accessible with IBMQ-QST, yet with a smaller number of measurements.

The IBM quantum processor `Lagos` limited our analysis to 7 qubits. Nevertheless, we further investigated the performance of tQST by extrapolating the analysis to a larger number of qubits. To this end, we considered simulated data with an SNR analogous to that of the IBMQ system and followed the same strategy as before to choose the threshold. We considered W states up to 14 qubits, now limited only by our hardware. In all cases, we obtained fidelities that exceeded 0.90, as reported in Table 2.7. We stress that in the cases of 14 qubits, the 16,556 projective measurements required by tQST make the reconstruction in principle experimentally accessible today. In contrast, QST would require at least $4^{14} = 268,435,456$ projective measurements that, at the moment, appear as a prohibitive number.

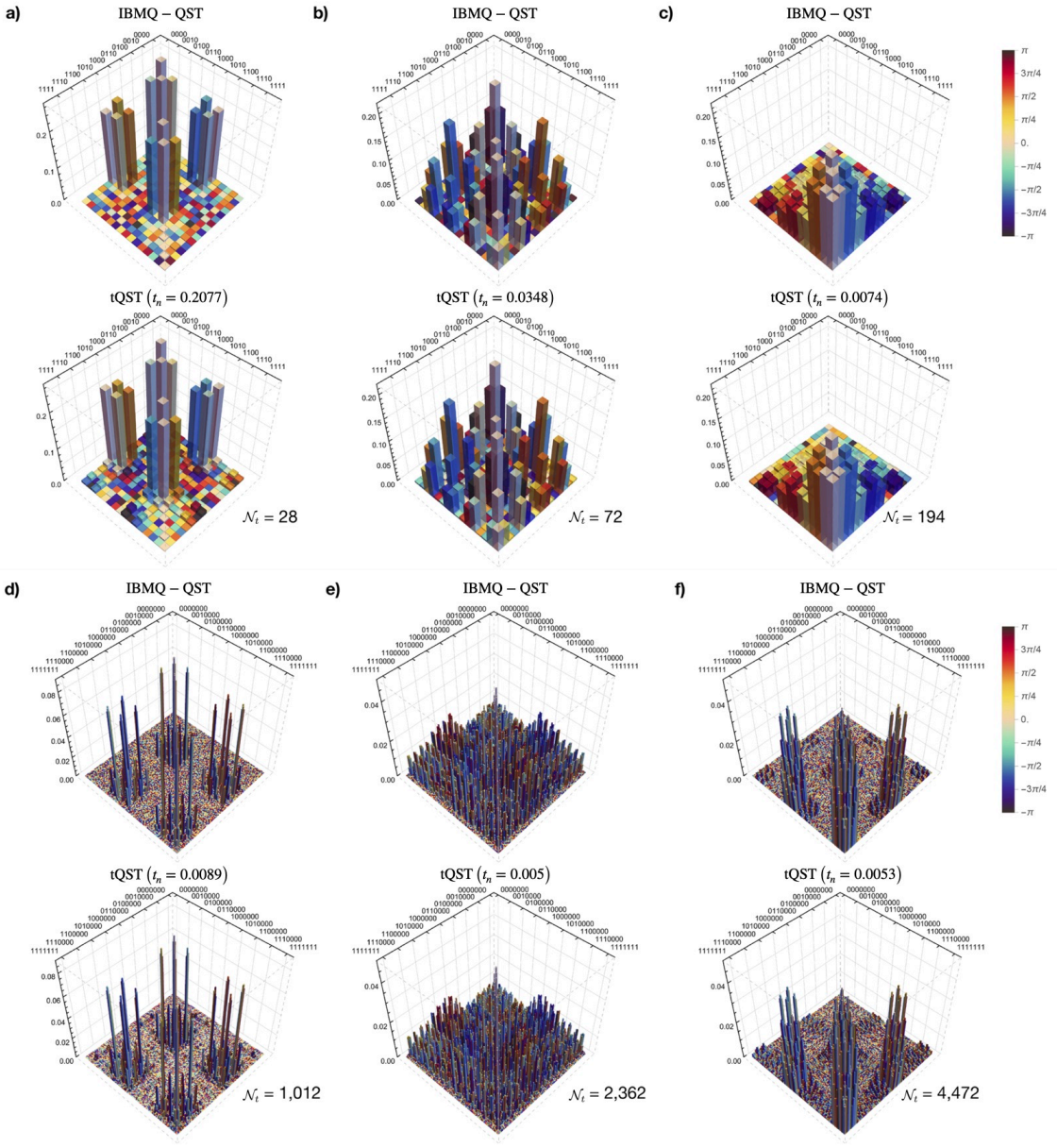


Figure 2.11: **Reconstruction of random states with tQST on an IBMQ processor.** We show random states reconstructed with IBMQ-QST and tQST on an IBMQ processor. In each panel, the top and bottom sub-panels show the IBMQ-QST and the tQST reconstructions, respectively. In the tQST sub-panel we report the threshold value t_n and the number $\mathcal{N}_t$ of projectors. **a)**, **b)**, **c)**: Reconstructions of 4-qubit random states with diagonal filling equal to 25%, 50%, and 75%, respectively. **d)**, **e)**, **f)**: Reconstructions of 7-qubit random states with diagonal filling equal to 25%, 50%, and 75%, respectively. Figure adapted from [1].

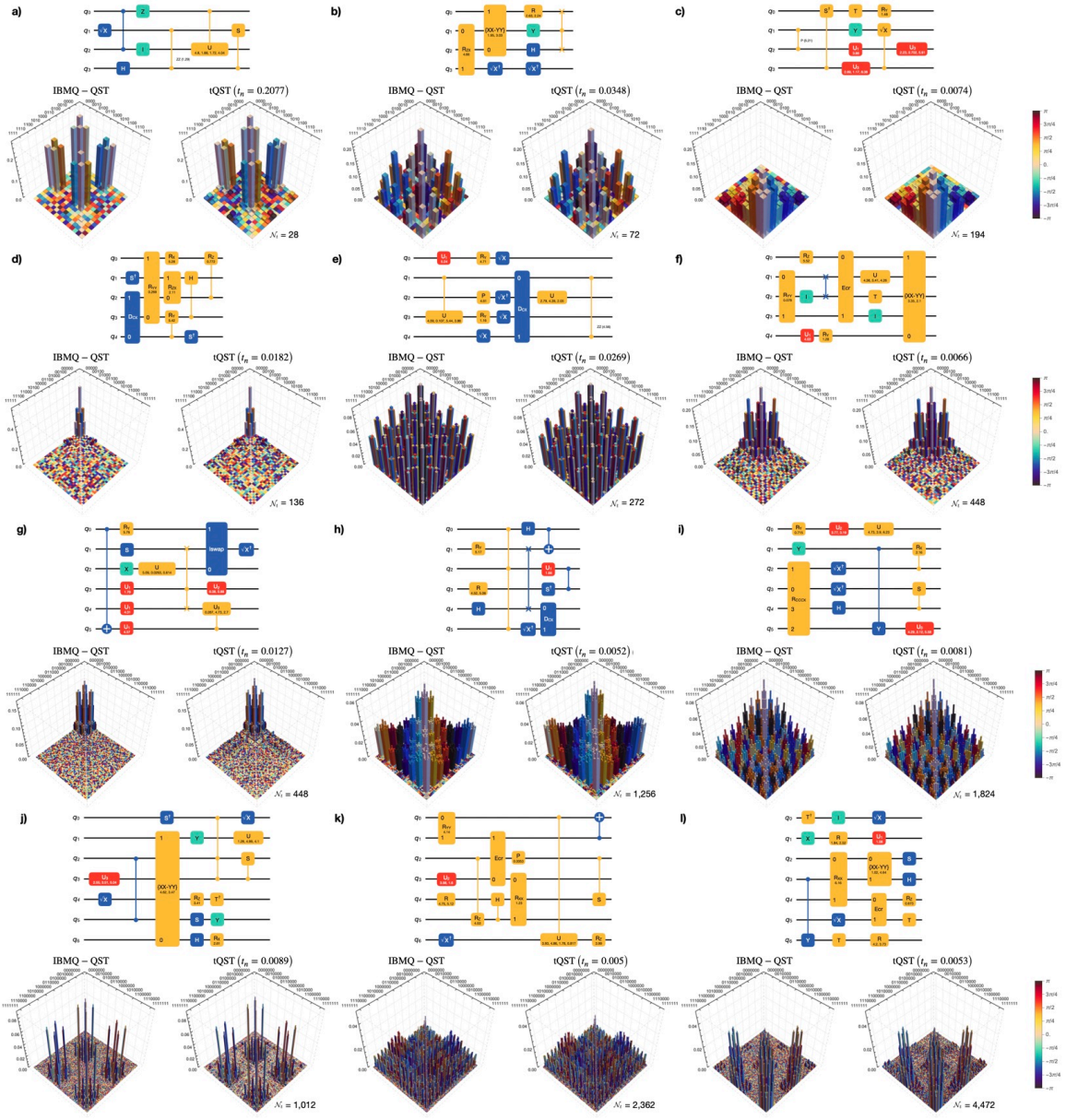


Figure 2.12: **Representative results of tQST applied to random circuits implemented in a IBMQ processor.** We show the reconstructions with IBMQ-QST and tQST of random states from 4 to 7 qubits on an IBMQ processor. In each panel, the top sub-panel shows the circuit constructed to generate the state, while the bottom left and bottom right sub-panels show the IBMQ-QST and the tQST reconstructions, respectively. In the tQST sub-panel we report the threshold value t_n and the number $\mathcal{N}_t$ of projectors. **a) - c)**: 4-qubit random states generated with the shown circuits with diagonal filling equal to of 25% , 50%, and 75%, respectively. **d) - f), g) - i), j) - l)**: same as in **a)-c)** with 5, 6, and 7 qubits, respectively. Figure adapted from [1].

		IBMQ-QST			tQST							
n_q	Circuit	$\overline{\mathcal{F}}^{\mathrm{IBMQ-QST}}_{\mathrm{target}}$	$\mathcal{N}_{\mathrm{IBMQ-QST}}$	3^{n_q}	t_n	$\overline{\mathcal{F}}_{\mathrm{bound}}$	$\overline{\mathcal{F}}^{\mathrm{tQST}}_{\mathrm{target}}$	$\overline{\mathcal{F}}^{\mathrm{tQST}}_{\mathrm{IBMQ-QST}}$	$\min \mathcal{N}_t$	$\max \mathcal{N}_t$	$\min \ell$	$\max \ell$
4	a)	0.934(1)	1,296	81	0.2077	0.47(1)	0.935(4)	0.969(3)	28	28	9	9
	b)	0.894(3)	1,296	81	0.0348	0.56(3)	0.89(1)	0.952(9)	72	134	27	65
	c)	0.960(1)	1,296	81	0.0074	0.934(5)	0.954(3)	0.980(2)	190	210	79	81
5	d)	0.935(1)	7,776	243	0.0182	0.708(6)	0.934(3)	0.948(4)	130	150	69	87
	e)	0.941(1)	7,776	243	0.0269	0.574(7)	0.940(3)	0.959(3)	272	272	81	81
	f)	0.927(1)	7,776	243	0.0066	0.857(3)	0.925(6)	0.963(3)	456	584	187	239
6	g)	0.873(2)	46,656	729	0.0127	0.59(2)	0.869(5)	0.900(6)	556	636	229	277
	h)	0.851(1)	46,656	729	0.0052	0.76(1)	0.844(3)	0.929(2)	1,404	2,322	433	637
	i)	0.868(1)	46,656	729	0.0081	0.606(6)	0.851(4)	0.919(2)	1,804	1,986	563	585
7	l)	0.890(1)	279,936	2,187	0.0089	0.57(2)	0.885(4)	0.902(6)	894	1,084	295	343
	m)	0.802(1)	279,936	2,187	0.0050	0.64(1)	0.792(3)	0.877(2)	4,696	5,160	1,003	1,221
	n)	0.926(1)	279,936	2,187	0.0053	0.612(6)	0.922(2)	0.926(2)	2,450	2,608	963	1,067

Table 2.6: **Reconstruction of of random states generated by depth-3 random circuits on an IBM quantum processor.** We grouped circuits according to their diagonal fillings, that is, the percentage of non-zero elements of the corresponding density matrix diagonal, for each number of qubits. With reference to Figure 2.12, the diagonal fillings are 25% [a), d), g), j)] 50% [b), e), h), k)], and 75% [c), f), i), l)] for 4, 5, 6 and 7 qubits, respectively. For each circuit, which we simulated 10 times, we report, for the IBMQ-QST protocol, the average fidelity $\overline{\mathcal{F}}^{\mathrm{IBMQ-QST}}_{\mathrm{target}}$ between the target state and the reconstruction with IBMQ-QST, the number $\mathcal{N}_{\mathrm{IBMQ-QST}}$ of projective measurements, and the number of measurement settings, equal to 3^{n_q}. For the tQST protocol, we report the circuit-specific threshold t_n, the fidelity lower bound $\overline{\mathcal{F}}_{\mathrm{bound}}$, the average fidelity $\overline{\mathcal{F}}^{\mathrm{tQST}}_{\mathrm{target}}$ between the target state and the reconstruction with tQST, the average fidelity $\overline{\mathcal{F}}^{\mathrm{tQST}}_{\mathrm{IBMQ-QST}}$ between the reconstructed state with the two protocols, the minimum $\min \mathcal{N}_t$ and maximum $\max \mathcal{N}_t$ number of projective measurements required by the tQST protocol, and the minimum $\min \ell$ and maximum $\max \ell$ number of settings required by the tQST protocol. All the quoted uncertainties are one standard deviation from the average.

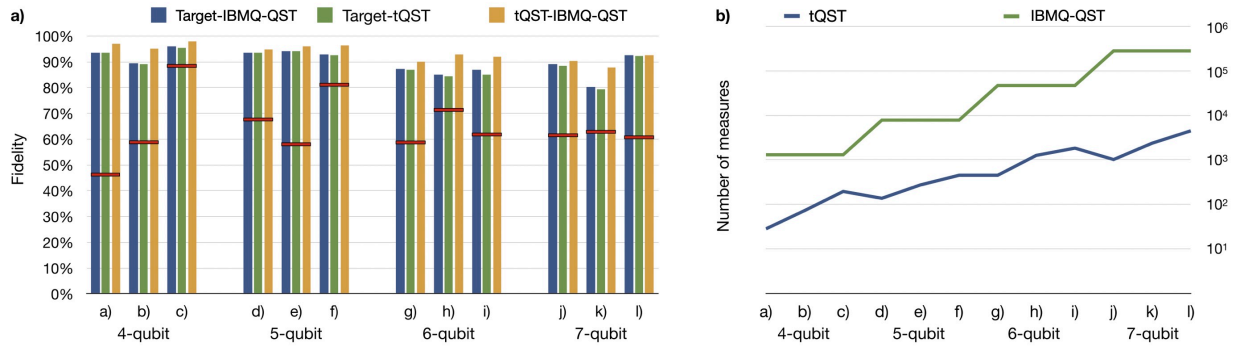


Figure 2.13: **Quantitative comparison between IBMQ-QST and tQST on the random states illustrated in Figure 2.12. a)**: Comparison of IBMQ-QST and tQST fidelities. For each state we show the fidelity between IBMQ-QST reconstruction and the target state, between tQST reconstruction and the target state, and between the IBMQ-QST and tQST reconstructions. The red line represents the tQST fidelity bound corresponding to the circuit-specific threshold. **b)**: Scaling of the number of projective measurements with the number of qubits for both QST and tQST reconstructions. Figure adapted from [1].

Qubits	Threshold (t_n)	$\mathcal{N}_t$	Fidelity
8	0.053	312	91.5%
9	0.047	584	91.9%
10	0.042	1,114	91.2%
11	0.038	2,158	91.4%
12	0.035	4,228	91.4%
13	0.032	8,348	91.3%
14	0.030	16,556	91.3%

Table 2.7: **Numerical reconstruction of W state with tQST up to 14 qubits.** We introduced an ad-hoc noise similar to the one observed in [71]. We report the number of qubits, the threshold t_n, the number of projectors $\mathcal{N}_t$, and the fidelity between the tQST reconstruction and the target state.

2.10 Experimental results on a photonic integrated circuit

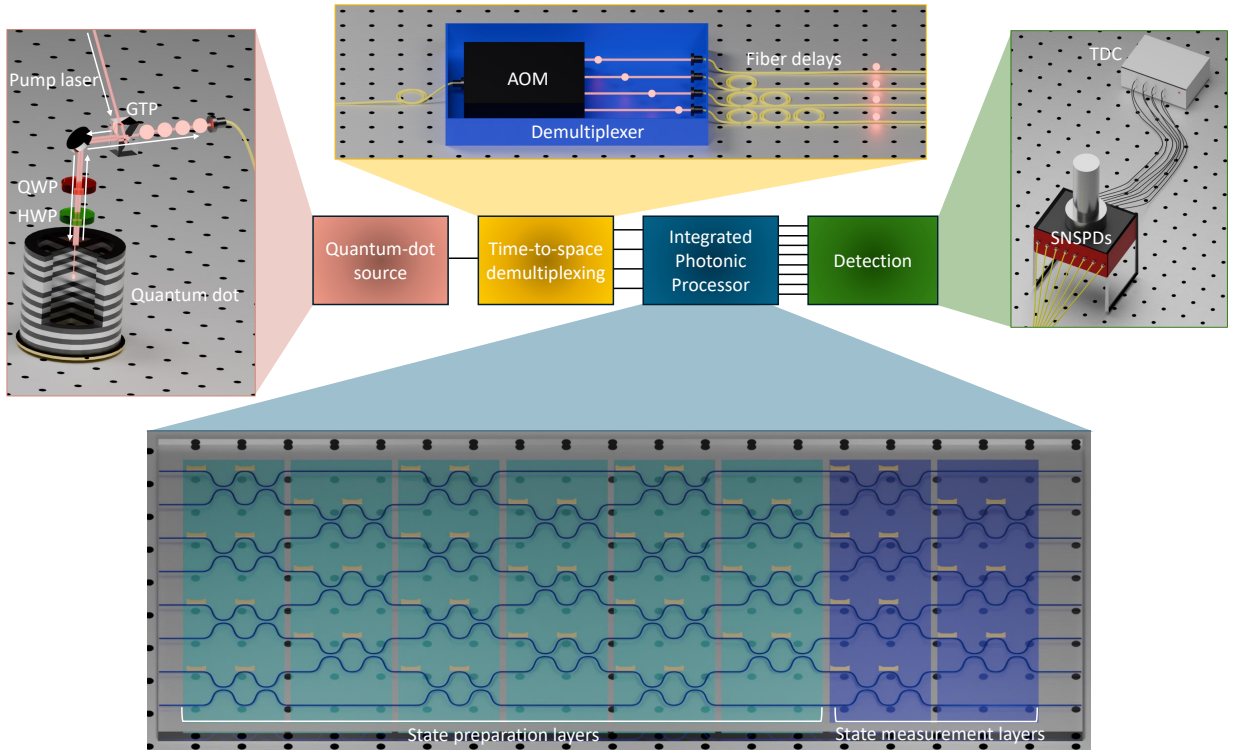


Figure 2.14: **Architecture of the hybrid-photonic platform employed in the experiment.** Pictorial representation of the experimental apparatus, based on a hybrid approach, used to generated multi-qubit states, and to implement the tQST protocol to reconstruct them. A pulsed laser, manipulated with a pulse-shaping stage, excites a QD single-photon source employing a RF optical excitation technique, comprising a Glan-Thompson polarizer (GTP), a half wave plate (HWP) and a quarter wave plate (QWP). The single-photon stream emitted by the QD source is then interfaced with the DMX module, based on an AOM, to obtain the required multi-photon resource. The output resource is then interfaced with an 8-mode fully-reconfigurable integrated photonic processor, fabricated via the femtosecond-laser-writing technology, with internal structure allowing for universal operation. The first six layers (cyan) of RBSs are used to perform the state preparation task with the dual rail encoding, while the remaining two layers (blue) are used to measure the qubits in different bases. At the end of the setup, photon detection is carried out via SNSPDs. Photon detection events, corresponding to n_q-fold coincidences, are recorded via a time-to-digital TDC. Figure from [4].

We now analyze the experimental results of the implementation of tQST on a reconfigurable photonic integrated circuit. Let us start by describing the experimental setup, comprising a state-of-the-art hybrid photonic architecture, interconnecting a diverse array of technologies tailored for multi-photon experiments. Specifically, it comprises three subsequent stages, as depicted in Figure 2.14. The photons are generated via a single-photon source based on quantum dot (QD) technology excited at a repetition rate of 158 MHz in the resonance fluorescence (RF) regime. We measure both the second-order auto-

correlation $g^{(2)}(0)$ and the Hong-Ou-Mandel interference visibility V_{HOM} through standard correlation experiments to quantify the quality of the single photons emitted by the QD. We typically register values of $g^{(2)}(0) \sim 0.01$ and $V_{\mathrm{HOM}} \sim 0.90$ with the QD source used in the experiment. Then, a time-to-space demultiplexing module (DMX) based on an acousto-optical modulator (AOM) converts the initial train of single photons, emitted at a fixed time interval, into sets of photons distributed in several spatial modes. This procedure allowed us to generate a multi-photon resource, which can be injected into the input ports of the experiment. The pairwise indistinguishability between the photons comprising the generated multi-photon state is achieved by finely tuning free-space delay lines and polarization paddle controllers. In particular, the temporal indistinguishability is ensured by fine-tuning, on the millimeter scale, the position of the internal couplers of the DMX device while tracking the visibility of standard HOM interference experiments carried out on the integrated device. The multi-photon resources generated with such an approach are then injected into a 8-mode fully reconfigurable integrated photonic processor. The first six layers of the processor are earmarked for state preparation, while the remaining two are used to implement projective measurements. The present device was fabricated (see Appendix B.1 for more details) using the femtosecond laser-writing technique, and its internal structure is based on an interferometric mesh of 28 Mach-Zehnder unit cells, each of which acts as a Reconfigurable Beam Splitter (RBS), arranged according to the universal rectangular geometry described in [315]. Moreover, the integrated processor used here features polarization-independent operation [316]. The reconfigurability of the device is induced by thermal phase shifters and is suitably controlled via current controllers that can be set to implement a given chosen unitary transformation U [316]. This requires knowledge of the current-optical phase response of the device, which can be done via suitable calibration procedures. As reported in Appendix B.1, we achieved an average moduli fidelity of 0.991 with our calibrated device, computed over randomly drawn Haar random matrices. Finally, at the output of the setup, the photons are collected and sent to a suitable detection system comprising high-efficiency superconducting nanowire single-photon detectors (SNSPDs), paired with a time-to-digital converter (TDC) used to discriminate and record n_q-fold coincidence events.

Quantum states are generated in the photon path degree of freedom by harnessing linear optics elements, such as beam splitters and phase shifters, and considering n_q-fold coincidence events measured in post-selection on a subset of all possible combinations of n_q photons in M modes. More specifically, the states are encoded via the *dual rail logic* in the photon spatial modes [317]. Thus, we can non-deterministically generate a n_q-qubit state by considering an optical interferometer with an even number $M = 2n_q$ of modes split into n_q dual rails, that is, adjacent pairs of spatial modes labeled with states $|0\rangle$ and $|1\rangle$ from top to bottom. Overall, when n_q photons are injected in the circuit, we consider valid for the state generation only the output combinations of photons distributed over the M modes, such that each dual rail is occupied by one and only one photon, while the others are discarded via a post-selection process. The controlled generation of the n_q-qubit state and its subsequent reconstruction via tomographic measurements take place in the aforementioned reconfigurable integrated photonic processor. As shown in Figure 2.14 and in Appendix B.2, the first six layers of the interferometer are configured appropriately to prepare a target state. The generation of a n_q-qubit state is obtained by mapping an initial n_q-photon state into a set of $2n_q$ modes via a suitable linear optical transformation and by applying a suitable post-selection procedure upon the detection of an n_q-photon

coincidence event in mode combinations satisfying the dual rail logic. In particular, if we inject n_q photons into the circuit, $N_{\text{RBS}} = M(M-1)/2$ RBSs constitute a M-mode universal multiport interferometer and are exploited in the state preparation stage (see Appendix B.2). The remaining two layers of the interferometer are dedicated to implementing projective measurements. More specifically, this relies on the possibility of implementing any arbitrary single-qubit transformation via a single elementary cell. Thus, by appropriately programming the transformation of an RBS acting on the two modes of a qubit, it is possible to implement any projective measurement on the output state. Thus, we can measure the observables needed to retrieve the target quantum state by applying the appropriate sequence of measurements according to the chosen reconstruction protocol, either QST or tQST in this experiment.

We now present the results of the tQST protocol implemented on the illustrated photonic platform. We analyze different classes of states, corresponding to different levels of sparsity of the density matrices. In the remainder of the section, we call QST the tQST with threshold $t = 0$. The main figure of merit to quantify the efficacy of the tQST approach is given by the fidelity between the density matrix derived using tQST and the one obtained through the QST procedure ($\mathcal{F}_{0,t}$). A second set of relevant parameters are the purities $\mathcal{P}_0$ and $\mathcal{P}_t$ of the states reconstructed with QST and tQST, respectively. Comparing $\mathcal{P}_{\text{t}}$ with $\mathcal{P}_0$ let us understand how much the purity of state reconstructed with tQST is affected by the reduced number of projectors. Finally, the remaining relevant parameter is the fidelity between the density matrix derived using the QST procedure and the expected state from the apparatus ($\mathcal{F}_{0,\text{m}}$) (see B.3 for details on the model that takes into account the experimental imperfections of the apparatus). This expected state is obtained by numerical simulations that consider the main noise effects in the apparatus, due to non-ideal single-photon emission from the QD source [318, 319, 320, 321], to losses, and to an imperfect dialing of the unitary matrix on the integrated photonic processor. The noise effects that we consider are related to (i) a non-zero multi-photon emission probability from the QD source, (ii) a degree of partial distinguishability between the photons, and (iii) directional couplers of the elementary cell with a splitting ratio different from $50{:}50$ (see Appendix B.3). The parameters related to the different effects have been estimated by independent measurements. More specifically, noise source i) is characterized by a standard measurement of the second-order autocorrelation via a Hanbury Brown and Twiss interferometer. In contrast, partial distinguishability (ii) can be estimated by measuring the Hong-Ou-Mandel visibilities V_{HOM} between photons emitted at different times. Finally, the splitting ratios of the directional couplers (iii) were determined during circuit calibration.

We first experimentally implemented the tQST protocol on a set of random states of 2 and 3 qubits. More precisely, we selected a set of 40 2-qubit states and 10 3-qubit states, each set characterized by equally spaced values of the Gini index, from the minimum to the maximum value, calculated on the diagonal elements of the ideal density matrices. The values of the Gini index are reported in Tables 2.8 and 2.9. This approach allowed us to test the performance of tQST by considering states that uniformly cover the full range of possible degrees of sparsity on the diagonal. The results of QST and tQST are shown in Figure 2.15. More specifically, for each state we show the fidelity $\mathcal{F}_{0,\text{t}}^{(n_q)}$ between the density matrices obtained with the two protocols and $\mathcal{F}_{0,\text{m}}^{(n_q)}$ between the QST reconstruction and the theoretical model, and the ratio $\mathcal{N}_{\text{t}}/\mathcal{N}_0$ between the number of projectors

State number	Gini index	State number	Gini index
1	0.000944	21	0.12627012
2	0.00663219	22	0.13204937
3	0.01249306	23	0.13790891
4	0.01814765	24	0.14387539
5	0.02421449	25	0.15024978
6	0.03227125	26	0.1567441
7	0.03700212	27	0.1639128
8	0.04446993	28	0.16907807
9	0.05089243	29	0.17615138
10	0.05707232	30	0.1824406
11	0.06313556	31	0.18807192
12	0.06867141	32	0.19487812
13	0.07593002	33	0.20180418
14	0.0811412	34	0.20715015
15	0.08702726	35	0.21351848
16	0.09500147	36	0.22003493
17	0.09970305	37	0.22618367
18	0.10596755	38	0.23185384
19	0.11257372	39	0.23896472
20	0.11851598	40	0.24496395

Table 2.8: **Gini index of the 2-qubit random states.** Values of the Gini index for the 40 2-qubit states chosen in the reported experiment. The reported Gini index is computed on the diagonal of the target state, in absence of experimental imperfections.

State number	Gini index	State number	Gini index
1	0.0078638	6	0.06720037
2	0.02123207	7	0.08189606
3	0.03362007	8	0.09318023
4	0.04380996	9	0.10441946
5	0.05467637	10	0.11615141

Table 2.9: **Gini index of the 3-qubit random states.** Values of the Gini index for the 10 3-qubit states chosen in the reported experiment. The reported Gini index is computed on the diagonal of the target state, in absence of experimental imperfections.

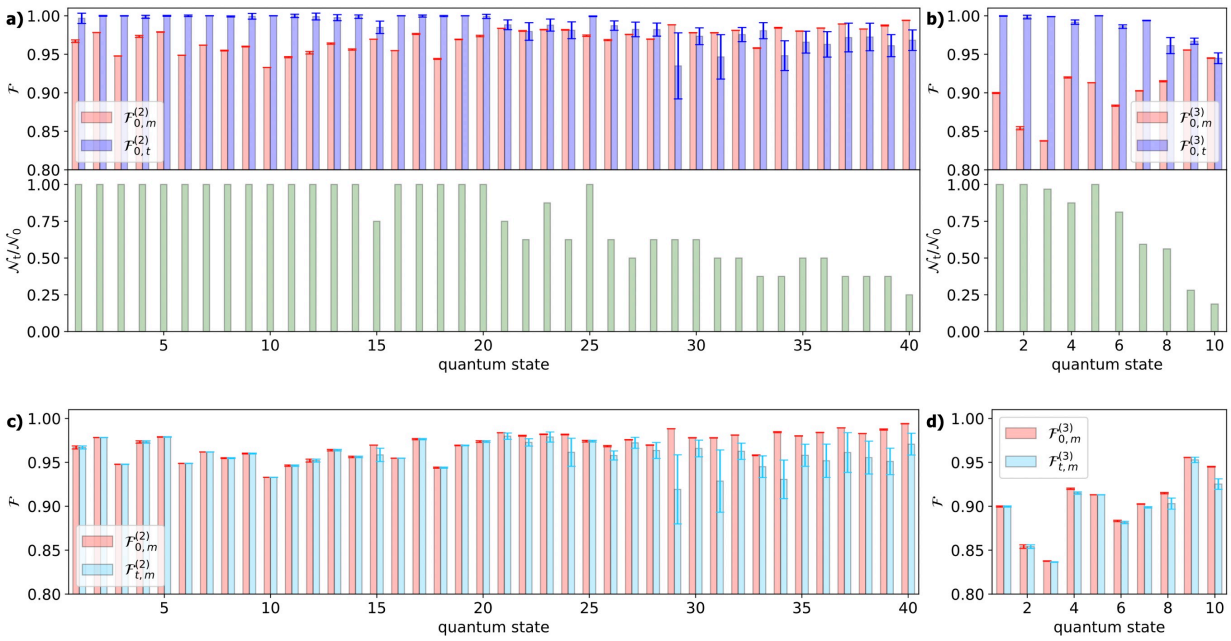


Figure 2.15: **Benchmarking tQST on 2- and 3-qubit random states.** The tQST protocol is tested by generating and analyzing a set of 40 2-qubit and 10 3-qubit random states with equally spaced Gini index. **a), b)**: In the upper panels, we show the fidelity $\mathcal{F}_{0,\mathrm{m}}^{(n_q)}$, with $n_q = 2, 3$, between the state reconstructed with QST and the theoretical model that takes into account experimental imperfections (red bars), and the fidelity $\mathcal{F}_{0,t}^{(n_q)}$, with $n_q = 2, 3$, between the states reconstructed with QST and tQST with the Gini threshold (blue bars); in the lower panels, we show the ratio $\mathcal{N}_t/\mathcal{N}_0$ (green bars) between the number of projectors $\mathcal{N}_t$ required by tQST with the Gini threshold, and the number of projectors $\mathcal{N}_0$ required by QST. **c), d)**: We show the fidelity $\mathcal{F}_{0,\mathrm{m}}^{(n_q)}$ as in panels **a)** and **b)**, and the fidelity $\mathcal{F}_{t,\mathrm{m}}^{(n_q)}$, with $n_q = 2, 3$, between the state reconstructed with tQST with the Gini threshold and the theoretical model taking into account experimental imperfections. In all panels the states on the horizontal axis are sorted by increasing Gini index. Error bars are computed by standard bootstrapping techniques, considering the standard deviation of the fidelities computed from a set of independent solutions to the state reconstruction problem. Each solution is obtained by initializing the optimization algorithm with a different seeds, that is, by choosing a different guess for the starting point of the maximum likelihood algorithm. Figure from [4].

used by tQST ($\mathcal{N}_\mathrm{t}$) and those used by QST ($\mathcal{N}_0$) in the upper and lower panels of Figure 2.15a) and b), respectively. In Figure 2.15c) and d) we also show the fidelity $\mathcal{F}_{\mathrm{t,m}}^{(n_q)}$ between the reconstructed states with tQST and the expected state calculated with a theoretical model that takes into account experimental imperfections. Here, the threshold value in the tQST protocol was chosen as the Gini threshold. We observe that for states with the lowest sparsity values, the two techniques coincide given that the tQST approach requires measuring a tomographically complete set of 4^n projectors. When measuring states with increasing sparsity in the diagonal elements, the tQST starts to be advantageous due to the measurement of a progressively lower number of projectors, as shown in Figure 2.15. The quality of all the reconstructed states shows that tQST is effective in reducing the number of projectors while having only a very limited loss of information with respect to the QST approach, since all the fidelities are found to be $\mathcal{F}_{\mathrm{0,t}}^{(n_q)} > 0.935$. The reduction in the number of projective measurements becomes more pronounced when the state dimensionality increases, given the exponential increase in the number of measurements required by QST. Finally, we observe that the reconstructed states show good agreement with the theoretical expectations from the model, as quantified by the average fidelities $\overline{\mathcal{F}}_{\mathrm{0,m}}^{(2)} = 0.969(14)$ and $\overline{\mathcal{F}}_{\mathrm{t,m}}^{(2)} = 0.959(18)$ for the 2-qubit scenario, and $\overline{\mathcal{F}}_{\mathrm{0,m}}^{(3)} = 0.903(35)$ and $\overline{\mathcal{F}}_{\mathrm{t,m}}^{(3)} = 0.899(33)$, for the 3-qubit case.

We then applied tQST to reconstruct specific maximally entangled states characterized by a sparse density matrix in the standard basis. In this case, tQST is expected to provide a significant advantage over QST. Indeed, in a noiseless scenario, tQST would allow us not to measure the projectors associated with the vanishing elements of the density matrix, while in a noisy scenario we can achieve the same result by properly choosing the threshold. We tested tQST on several maximally entangled states, such as $|\Psi^+\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$ for 2 qubits, $|\mathrm{GHZ}_3\rangle = (|010\rangle + |101\rangle)/\sqrt{2}$ and $|\mathrm{W}_3\rangle = (|100\rangle + |010\rangle + |001\rangle)/\sqrt{3}$ for 3 qubits, and $|\mathrm{GHZ}_4\rangle = (|0101\rangle + |1010\rangle)/\sqrt{2}$ for 4 qubits. The Bell states and the GHZ states are generated by following the post-selected approach of [318], while the W states are generated by exploiting the full reprogrammability of the device (see Appendix B.2 for details on the configuration for the generation layout.).

The results of the reconstructed density matrices with the two protocols, and their comparison with the theoretical model of the apparatus, are shown in Figure 2.16, while additional analyses are reported in Appendix B.4. As a general comment, we observe that the states reconstructed with tQST are close to those reconstructed with QST. We quantified this conclusion by a more detailed analysis, reported in Table 2.10 and Figure 2.17, where we analyzed tQST using a different number of projectors $\mathcal{N}$ compared to the one indicated by the Gini threshold.

We start by analyzing the fidelity between the QST and tQST reconstructions $\mathcal{F}_{\mathrm{0,t}}^{(\alpha)}$ as a function of the number of projectors $\mathcal{N}$, with $\alpha = \{\Psi^+, \mathrm{GHZ}_3, \mathrm{W}_3, \mathrm{GHZ}_4\}$. We observe that for all states, the initial reduction of the number of projectors in a regime above the value identified by the Gini index is accompanied by a small loss of information, that is, a small decrease in the fidelity. We observe that the Gini threshold identifies the correct boundary where the minimum number of projectors is measured for the $|\Psi^+\rangle$ state. For the $|\mathrm{GHZ}_3\rangle$ state, our approach sets the threshold for the number of projectors very close to the boundary, and the obtained value of the fidelity can be increased up to 0.95 by adding a very limited number of additional projectors (≤ 10). Conversely, for the $|\mathrm{W}_3\rangle$ and $|\mathrm{GHZ}_4\rangle$ states, we observe that the Gini threshold corresponds to a slightly conserva-

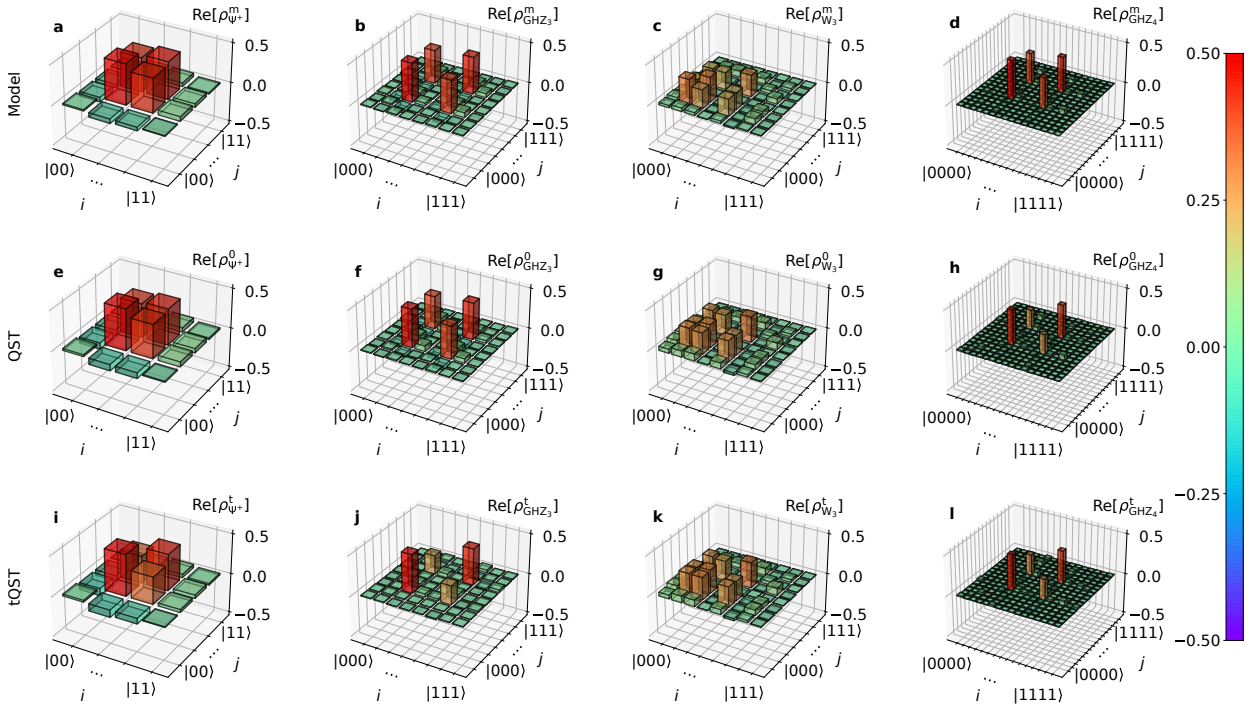


Figure 2.16: **Real part of the density matrices of maximally entangled states reconstructed with tQST.** We show the real part of the density matrices of the different maximally entangled states considered, namely (from left to right) $|\Psi^+\rangle$, $|\mathrm{GHZ}_3\rangle$, $|\mathrm{W}_3\rangle$, and $|\mathrm{GHZ}_4\rangle$. In the first row (panels **a-d**) we show the expected density matrices $(\rho_\alpha^{\mathrm{m}})$ estimated from the theoretical model taking into account experimental imperfections. The second row (panels **e-h**) shows the density matrices (ρ_α^0) reconstructed with QST. Finally, the third row (panels **i-l**) shows the density matrices (ρ_α^t) reconstructed with tQST with the Gini threshold. The superscript index α labels the states as $\alpha = \Psi^+, \mathrm{GHZ}_3, \mathrm{W}_3, \mathrm{GHZ}_4$. On the right part of the figure, we report the colormap for the density matrix bars, equal for all panels **a-l**. Figure from [4].

State (α)	$\mathcal{N}_0$	$\mathcal{N}_{\mathrm{t}}$	$\mathcal{F}_{0,\mathrm{t}}^{(\alpha)}$
$\|\Psi^+\rangle$	16	6	0.94(3)
$\|\mathrm{GHZ}_3\rangle$	64	10	0.84(5)
$\|\mathrm{W}_3\rangle$	64	32	0.9833(7)
$\|\mathrm{GHZ}_4\rangle$	256	66	0.948(5)

Table 2.10: **Reconstruction of maximally entangled states with tQST.** We applied tQST to reconstruct some exemplary maximally entangled states. We report the state to be reconstructed, the number of projectors $\mathcal{N}_0$ required by QST, the number of projectors $\mathcal{N}_{\mathrm{t}}$ required by tQST with the Gini threshold, and fidelity $\mathcal{F}_{0,\mathrm{t}}^{(\alpha)}$ between the reconstructions with QST and tQST. The uncertainty represents the standard deviation from the average.

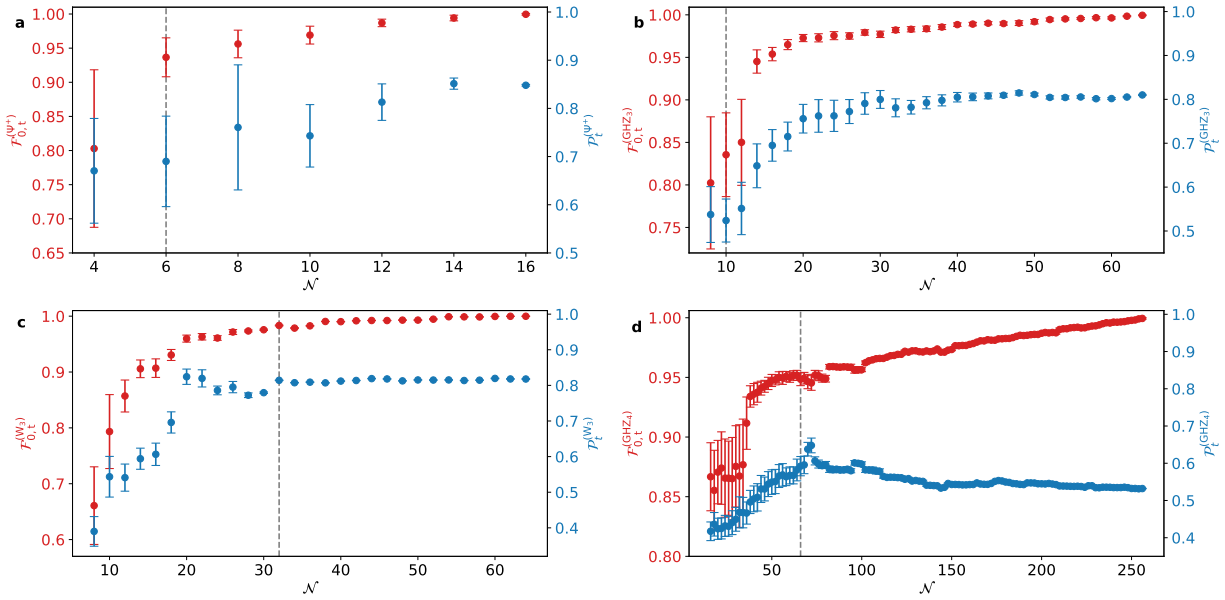


Figure 2.17: **Reconstruction of maximally entangled states varying the number of measured projectors.** In each panel, the red points represent the fidelity $\mathcal{F}_{0,t}^{(\alpha)}$ for different numbers of projectors $\mathcal{N}$, obtained by progressively varying the threshold t. The blue points represent the purity $\mathcal{P}_{\mathrm{t}}^{(\alpha)}$ of the states reconstructed via tQST as a function of $\mathcal{N}$. The superscript index α labels the states as $\alpha = \{\Psi^+, \mathrm{GHZ}_3, \mathrm{W}_3, \mathrm{GHZ}_4\}$. We show these plots for the state $|\Psi^+\rangle$, $|\mathrm{GHZ}_3\rangle$, $|\mathrm{W}_3\rangle$, and $|\mathrm{GHZ}_4\rangle$ in panels **a**, **b**, **c**, and **d**, respectively. In each plot, the vertical dashed line corresponds to the number of projectors required by the Gini threshold. Error bars are computed by standard bootstrapping techniques, considering the standard deviation of the fidelities computed from a set of independent solutions to the state reconstruction problem. Each solution is obtained by initializing the optimization algorithm with a different seeds, that is, by choosing a different guess for the starting point of the maximum likelihood algorithm. Figure from [4].

tive choice, since it is still possible to further reduce the number of projectors without a relevant decrease in fidelity. The same trend is confirmed by observing the purity of the reconstructed states with tQST. Removing an exceeding amount of projectors with respect to those indicated by the Gini threshold implies a significant reduction of the purity of the reconstructed density matrix, for not enough information was collected to properly estimate this quantity. We also observe that the purity of the reconstructed states decreases with the size of the system. This is not due to the reconstruction method, but it depends on the increased impact of the experimental imperfections of the state generation process as the number of photons increases. Finally, we observed that in our experiments the tQST approach with Gini threshold requires a number of projectors different from the one required for the corresponding noiseless states. Indeed, the presence of experimental noise adds other non-zero values in the diagonal elements, which in turn implies to measure additional projectors than those required by the corresponding noiseless state.

2.11 Final remarks

In this chapter, we introduced a novel QST protocol, called tQST, and provided a detailed demonstration that it fulfills the four desiderata outlined at the beginning of the chapter. Before discussing these criteria, we briefly summarize the tQST workflow. The protocol begins with the determination of the diagonal elements of the density matrix, which are obtained by measuring the expectation values of the projectors onto the selected computational basis. In the second step, a threshold parameter t is chosen to identify the off-diagonal elements that satisfy $\sqrt{\rho_{ii}\rho_{jj}} \geq t$ and to determine the number of observables necessary to complete the tomography. Then, we construct local projective measurements to target the relevant off-diagonal elements and measure their expectation value. Finally, the measurement outcomes are processed to reconstruct the quantum state.

The structure of the protocol inherently satisfies the first desideratum, namely the absence of any need for prior knowledge about the state. The procedure is entirely data-driven, enabling its execution from beginning to end without assumptions or initial hypotheses concerning the nature of the state. The second desideratum, on the reduction in the number of required observables, is directly addressed by the role of the threshold t. In QST, the number of observables scales exponentially with the dimension of the Hilbert space. In contrast, tQST exploits the sparsity of the representation of the state in a given basis to reduce the number of observables up to two orders of magnitude for some states. Notably, this dependency on the state representation is not an assumption, but rather a feature that affects only the convergence of the method.

We propose two quantitative criteria to determine the threshold t using only the diagonal elements, the only ones available at that stage of the protocol. The first is based on the Gini index, capturing the sparsity of the diagonal; the second employs the Shannon entropy, reflecting the average information content. Despite their conceptual differences, both definitions of the threshold exhibit similar quantitative behavior and correlate positively with matrix sparsity. Once the number of relevant observables is established, we identify the specific measurements required. We restrict our focus to local projective measurements, that is, the tensor products of projectors onto the eigenstates of Pauli operators. For each significant off-diagonal element, a recursive algorithm identifies two such projectors that optimally extract information regarding its real and imaginary com-

ponents.

After the measurement of the observables, reconstruction of the quantum state is carried out using MLE. We provide two parametrizations for the numerical implementation of this reconstruction: the triangular parametrization, which is unbiased, and the g5 parametrization, which is advantageous in cases where the state is known or assumed to be (approximately) pure. Furthermore, we derive an analytical expression for the gradient of the likelihood function to enhance optimization efficiency. We compare these parametrizations in terms of minimization time and reconstruction fidelity on GHZ and W states (up to 10 qubits) and random pure states (up to 4 qubits).

To further validate the protocol, we presented comprehensive, step-by-step applications of tQST to various quantum states, specifically, a random mixed state, a random pure state, a W state, and a GHZ state, each composed of three qubits. A subsequent numerical analysis for systems of 2, 3, and 4 qubits revealed that tQST often requires substantially fewer projective measurements than QST while preserving high reconstruction fidelity. These results highlight the distinctive nature of tQST and its clear advantages in efficiency.

Addressing the fourth desideratum, experimental feasibility, we implemented the tQST protocol on two distinct quantum platforms: a superconducting quantum processor provided by IBMQ and a fully reprogrammable integrated photonic chip. On the IBMQ platform, we performed a comparative analysis between tQST and IBMQ-QST, the default tomography protocol available on the IBMQ interface. We conducted experiments on quantum states of up to seven qubits. The key figures of merit were the fidelity between the tQST and IBMQ-QST reconstructions, and the fidelity between each reconstruction and the theoretical target state. In most cases, the protocols yielded compatible results within the associated error bars, despite significant differences in the number of measurements required. Then, we experimentally tested tQST on a hybrid photonic platform on quantum states of up to four qubits. The results demonstrated accurate reconstruction across states with varying levels of sparsity, as quantified by average fidelity. The tQST protocol was particularly favorable for entangled states, such as the 4-qubit GHZ state, which are foundational in many quantum information protocols. Taking into account the detected coincidence rates and the fact that multiple projectors within the same measurement basis can be implemented simultaneously in our setup, the measurement time for the 4-qubit GHZ state using tQST was reduced to approximately 7.5 hours, compared to 40.5 hours required by QST.

The findings of this chapter motivate several questions. Is MLE the only viable method for state reconstruction within this framework? Is it possible to generalize the protocol to qudits or to alternative measurement strategies? Can tQST be adapted for quantum characterization tasks beyond state tomography? Moreover, are there practical applications where the benefits of tQST can be effectively leveraged? These questions will be addressed in the subsequent chapters, beginning with the exploration of alternatives to maximum likelihood estimation.

Chapter 3

Deep Learning and tQST

In this chapter, we study whether it is possible to use an artificial intelligence (AI) model to characterize quantum states. We start with some general considerations about machine learning and deep learning, and then review the literature on AI-based methods for QST. We introduce the deep learning models that we embed into the tQST protocol to carry out two tasks, state reconstruction and purity estimation. We show numerical results up to 4 qubits with noiseless and noisy data. Possible references for the first section are [322, 323], while the main reference for the rest of the chapter is [3].

3.1 Machine Learning and Deep Learning

3.1.1 General considerations about machine learning

Machine learning is a subfield of artificial intelligence (AI) defined as the field of study that gives computers the ability to learn without being explicitly programmed. We can introduce a **model**, or hypothesis, which represents a relationship between the input data x and the output y to clarify what we mean by "learning". In the simplest case, a model is a parametric function:

$$h_\theta(x) \to y. \tag{3.1}$$

The goal of learning, or **training**, is to find opportune values of the parameters θ, generally called **weights and biases**, by analyzing a data set, called the **training set**. The training set consists of **examples**, each of which consists of one or more **features**, and possibly a label. Typically, learning proceeds with a learning algorithm and is achieved by minimizing the difference between the output provided by the model and the required output. Depending on the source from which the data come, it is possible to distinguish three main possibilities. In **supervised learning** the model learns from labeled data, that is, the training data are annotated with the correct output, and training a model means to allow it to predict the right output for unknown (unlabeled) data. If the labels come from a finite set of options, called classes, then we have a classification problem, and the model is called a classifier. If the labels are real numbers, then we have a regression problem, and the model is called a regressor. In **unsupervised learning** the model learns from unlabeled data, that is, the training data are not annotated with the correct output, and it is expected to find common structures useful to understand the phenomenon under study.

Finally, in **reinforcement learning** the model is represented by an agent that learns from the interaction with an environment, real or simulated, that provides feedback about the actions performed by the agent and is used to fix the model.

Model selection involves the choice of the model and of its **hyperparameters**, specific of each model, to distinguish them from those learned during training. Examples of hyperparameters are the number of iterations or epochs, the loss function, the optimizer, and the learning rate. Let us briefly analyze each of them. The learning algorithm progresses through several **iterations**, that is, several updates of the model parameters. At every iteration, the model processes a set, or **batch**, of examples, the number of which is called **batch size**. After a certain number of iterations, the entire training set will have been fed to the model, with each example processed once: this is an **epoch**. Thus, an epoch represents $N_{\text{train}}/(\text{batch size})$ iterations, with N_{train} the total number of training examples. The **loss** measures how far the prediction of a model is from its label during the training of a supervised model.

The loss is calculated using a **loss function**, which is a mathematical function that calculates the loss on a batch of examples. A loss function returns a lower loss for models that make good predictions than for models that make bad predictions. The goal of training is typically to minimize the loss that a loss function returns. Several loss functions exist, and the choice depends on the kind of problem we are facing. The L_2 or mean squared error (MSE) loss and the logistic or cross-entropy loss are common choices in regression and classification problems, respectively. The minimization of the loss is achieved by some proper optimization algorithm. The most widely used is **gradient descent**, an iterative algorithm that gradually finds the best combination of parameters that minimize the loss. An **optimizer** is a specific implementation of the gradient descent algorithm. Common choices are AdaGrad (ADAptive GRADient descent) and Adam (ADAptive with Momentum). Finally, the **learning rate** is a floating-point number that tells the gradient descent how strongly to adjust the parameters (weights and biases) on each iteration. If we set it too low, the training will take too long. If we set it too high, the gradient descent often has trouble reaching convergence.

Typically, the values of the hyperparameters cannot be chosen in advance, and we have to try different combinations and select the best one. However, it is not a good idea to take decisions and evaluate their effects on the same data set. We can avoid this problem by using another set of data, called the **validation set**. After training, we need a measure to quantify the performance of our model. In classification problems, the standard score is the accuracy, the fraction of samples correctly classified, while in regression problems a common score is the coefficient of determination, defined as:

$$R^2 = 1 - \frac{\sum \left(y_{\text{true}} - y_{\text{pred}}\right)^2}{\sum \left(y_{\text{true}} - \mu_{\text{true}}\right)^2}, \tag{3.2}$$

with μ_{true} the mean of the true values. The ability of the model to make good predictions on unknown data is called **generalization**. To quantify this capability, we have to consider another data set, usually called the **test set**, with data that are not contained in the training set, thus never seen by the model, and then compare the performances of the model on the training and test set. The goal of a learning algorithm is that the model performs well on the training set and generalizes well to new data, tuning the "complexity" of the model (a rough estimate of which is the number of parameters of the model). Another way to

limit the complexity of a classifier is to constrain the values of its parameters, and this is usually done with a technique called regularization, which consists of introducing an additional term in the loss function optimized by the model. The training-validation-test split is usually achieved by splitting the whole data set at our disposal into three subsets. The common ratios for this splitting are 90-5-5% and 80-10-10% for train, validation, and test, respectively.

Machine learning encompasses a wide variety of algorithms for both classification and regression problems. Other very popular models, such as neural networks, belong to the realm of deep learning, which is the subject of the next subsection.

3.1.2 From artificial neurons to artificial neural networks

Artificial neural networks, simply called **neural networks** (NNs), are computational models that draw inspiration from the complex structure and functions of biological neural networks. They are modeled as collections of interconnected units, called (artificial) neurons, that communicate through a network of connections, whose biological counterpart is called the synapse. Each neuron receives inputs, processes them through a simple computation, and produces a single output that can be transmitted to multiple other neurons. The combination of the processing of multiple neurons produces complex behaviors. Thus, it is opportune to start describing the model of a single neuron.

We can model a single neuron as a unit with an activation state that depends on neighboring neurons that send their own activation state through the connection. The sum of the activations of the neighbors, weighted by the strength of the connections, measures the excitation of the neuron. In the original model [324] of the **perceptron**, proposed to solve image recognition problems, a binary threshold was proposed: if the measure of the excitation of the neuron exceeds a threshold, then the neuron is activated. However, the discontinuity of this activation function prevents the application of learning algorithms that use gradient-based optimization strategies, such as gradient descent. A neuron model with better mathematical behavior can be obtained by replacing the binary activation function with a real-valued activation that is the output of a suitable activation function $a(\cdot)$, a crucial component within each artificial neuron. This mathematical function determines whether a neuron should activate based on the weighted sum of its inputs and introduces non-linearity into the model. Moreover, the threshold is replaced by a bias b, resulting in the following mathematical formulation:

$$\text{activation} = a(z) = a\left(b + \sum_{j=0}^{n-1} x_j w_j\right), \tag{3.3}$$

with $\{x_0, x_1, \ldots, x_{n-1}\}$ the activations of the neighbors, and $w_0, w_1, \ldots, w_{n-1}$ the weights of the corresponding connections. A schematic representation of this artificial neuron model is shown in Figure 3.1. The sigmoid, the arctangent, the hyperbolic tangent, and the

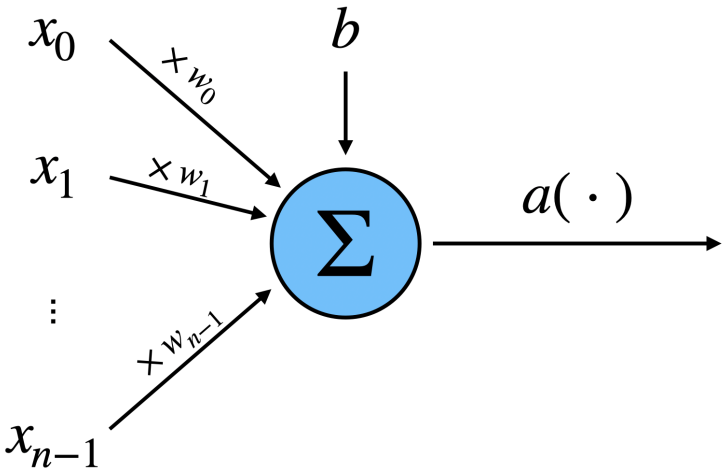


Figure 3.1: A schematic representation of the perceptron.

Rectified Linear Unit (ReLU) activation are common choices for the activation function:

$$\begin{aligned} a(z) &= \frac{1}{1+e^{-z}} && \text{sigmoid} \\ a(z) &= \tan^{-1}(z) && \text{arctangent} \\ a(z) &= \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}} && \text{hyperbolictangent} \\ a(z) &= \max(z,0) && \text{ReLU} \end{aligned} \tag{3.4}$$

A single neuron can be used for supervised learning problems, where the inputs correspond to the features of the samples. If we want to solve more complicated problems, we have to combine multiple neurons in complex ways and train them simultaneously.

The general structure of a neural network is that of a graph in which the vertices are the neurons and the edges are the connections between them. A relevant class of NNs is that of feedforward networks, where we can always arrange the neurons into a multilayer structure such that information flows in one direction: each layer transforms its inputs to outputs that feed the next layer, with no cycles or memory loops. We can divide the neurons of a feedforward NN into three groups. Input neurons compose the first layer of the network, called the input layer, have no incoming connections, and their activation is taken directly from the input to the network. Output neurons belonging to the last layer of the network, called the output layer, have no outgoing connections: Their activations form the output computed by the network. Between them, the hidden neurons of the intermediate layers, called the hidden layers, have incoming and outgoing connections: they take information from other neurons, process it, and pass the result to the other neurons. Furthermore, the layers of these networks typically are fully connected: Each neuron in a layer is connected to all neurons in the next layer, and no other connections are allowed. This type of feedforward network, a stack of fully connected layers with non-linear activations, is called **Multi-Layer Perceptron (MLP)** [325]. The number of input and output neurons is dictated by the dimension of the problem. For a binary classification problem, there would be an input neuron for each component of the feature vector and an output neuron for each class. The class predicted by the network would be

the one corresponding to the output neuron with the highest activation. The number and size of the hidden layers determine the complexity of the network.

From a mathematical point of view, MLP is basically a combination of many single neurons. Consider an MLP of $D+1$ layers, where layer 0 is the input layer, layers through 1 to $D-1$ are hidden layers, and layer D is the output layer, with D called the depth of the network. Each layer l contains a given number n_l of neurons. The activations of the input neurons are taken from the input features. Then the activation $x_j^{(l)}$ of the j-th neuron of layer $l > 0$ is:

$$x_j^{(l)} = a\left(z_j^{(l)}\right), \tag{3.5}$$

$$z_j^{(l)} = b_j^{(l)} + \sum_{r=0}^{n_l-1} W_{jr}^{(l)} x_r^{(l-1)}, \tag{3.6}$$

with $z_j^{(l)}$ the total input to the neuron, $b_j^{(l)}$ the bias, and $W_{jr}^{(l)}$ the weight of the connection with the r-th neuron of layer $l-1$. Here, the activation function is assumed to be the same for all neurons. We can gather the activations of layer l into a vector $\boldsymbol{x}^{(l)}$ and write the above equation in a more compact form:

$$\boldsymbol{x}^{(l)} = \boldsymbol{a}\left(\boldsymbol{z}^{(l)}\right), \tag{3.7}$$

$$\boldsymbol{z}^{(l)} = \boldsymbol{b}^{(l)} + W^{(l)}\boldsymbol{x}^{(l-1)}. \tag{3.8}$$

These equations represent the forward propagation and we can use them to compute the output of the network: We take the input data as the activations $\boldsymbol{x}^0$ of the input neurons; from $\boldsymbol{x}^{(0)}$ compute the activations $\boldsymbol{x}^{(1)}$ of the first hidden layer; repeat until all the activations have been computed. Finally, we take the activations $\boldsymbol{x}^{(D)}$ of the output layer and use them as predictions computed by the neural network. However, training the network to make it produce the correct outputs is not as simple as it sounds.

We recall that training an MLP consists in finding the weights and biases that minimize a suitable loss function. The minimization of the loss can be achieved with gradient-based optimization, such as gradient descent. To this end, we need to compute the derivatives of the loss with respect to the parameters. The **backpropagation** algorithm provides a convenient way to compute these derivatives, working backwards on the operations performed in the forward pass. A single step of the backpropagation algorithm for a classification problem, composed of a forward and a backward pass, can be qualitatively described as follows. The forward pass starts by computing the output of the network as described above. The goal of minimizing the loss function is to make the value of the output neuron O_c corresponding to the correct classification as close as possible to 1. However, this value depends on the many nested activation functions of the entire neural network. Thus, minimizing the loss function would entail making adjustments throughout the network that bring the value of O_c closer to 1. To do so, we need to know how any change in the previous layers will change the output of O_c, that is, we need the partial derivatives of the loss function with respect to the activations of neurons. The output of O_c activation depends on the contributions it receives from neurons in the penultimate layer. One way to change the output of O_c is to change the weights between neurons in the layer $D-1$ and O_c. However, $D-1$ neurons' output values, in turn, are influenced by weights applied to inputs they receive from $D-2$. So we can differentiate the activation

functions in $D-1$ to find the partial derivatives of the weights applied to the contributions of $D-2$. These partial derivatives show us how any change in a weight of $D-2$ will affect the outputs of $D-1$, which would subsequently affect the output value of O_c and therefore affect the loss function. Thus, in the backward pass we can recursively repeat this process until we have reached the input layer and have the gradient of the loss function: a vector of its partial derivative for each weight and bias parameter in the network. We have now completed a forward and backward pass for a single training example. We can formalize the backpropagation algorithm for a classification problem as follows:

1. Forward pass: compute the output of the network.
   (a) Set the activation of the input neurons $\boldsymbol{x}^{(0)}$;
   (b) For every layer $l = 1, 2, \ldots, D$:
      i. Compute the input:

$$\boldsymbol{z}^{(l)} = \boldsymbol{b}^{(l)} + W^{(l)}\boldsymbol{x}^{(l-1)}; \tag{3.9}$$

      ii. Compute the activations:

$$\boldsymbol{x}^{(l)} = \boldsymbol{a}\left(\boldsymbol{z}^{(l)}\right) \tag{3.10}$$

      iii. Compute the output probabilities.
2. Backward pass: compute the derivatives of the loss function L with respect to the parameters.
   (a) Compute the output derivatives, that is, the derivatives $\boldsymbol{\delta}^{(D)}$ of the loss function with respect to the activations of the output neurons, with:

$$\delta_j^{(D)} = \frac{\partial L}{\partial x_j^{(D)}}. \tag{3.11}$$

   .
   (b) For every layer $l = D-1, D-2, \ldots, 1$:
      i. Compute the derivative of the loss function with respect to the activations of the neurons of layer l:

$$\boldsymbol{\delta}^{(l)} = \left(W^{(l+1)}\right)^T \left(\boldsymbol{\delta}^{(l+1)} \odot \boldsymbol{a}'\left(\boldsymbol{z}^{(l+1)}\right)\right), \tag{3.12}$$

      which depends on the weights W^{l+1} between layer ℓ and $\ell+1$, the derivative $\boldsymbol{\delta}^{(l)}$, and the derivative of the activation function, $a'\left(z_j^{(l)}\right) = \partial x_j^{(l)}/\partial z_j^{(l)}$. Here, $\odot$ is the element-wise product.
      ii. Compute the derivatives of the loss function with respect to weights and biases as:

$$\nabla_{W^{(l)}} L = \left(\boldsymbol{\delta}^{(l)} \odot \boldsymbol{a}'\left(\boldsymbol{z}^{(l)}\right)\right)\left(\boldsymbol{x}^{(l-1)}\right)^T + \lambda W^{(l)} \tag{3.13}$$

$$\nabla_{\boldsymbol{b}^{(l)}} L = \boldsymbol{\delta}^{(l)} \odot \boldsymbol{a}'\left(\boldsymbol{z}^{(l)}\right) \tag{3.14}$$

This algorithm allows us to compute the gradients for the parameters corresponding to a single training example. Given a training set of more than one example, the gradients of the average loss are the average of the single gradients. Finally, we update the parameters by gradient descent with a given learning rate.

Let us mention the so-called *universal approximation theorem*, one of the reasons why neural networks have been so widely used. According to this result, for (almost) any problem we may want to solve, there is a neural network that performs arbitrarily well [326, 327]. However, this theorem does not tell the parameters and hyperparameters of such a network. In practice, the choice of the number and size of the hidden layers depends on our capability to make the network learn to solve the problem. We know from empirical experiments that deep neural networks, that is, neural networks with many hidden layers, tend to achieve greater expressivity than shallow neural networks with the same number of parameters. However, training deep neural networks can be numerically challenging.

3.1.3 The Neural Network Ecosystem

As the field of deep learning matured, a need arose for architectures specifically tailored to handle complex data types and solve highly specialized problems, each tailored to different data modalities and tasks. This led to the proliferation of advanced neural network designs, covering a diverse “ecosystem” of neural network architectures. These models have achieved state-of-the-art results in many fields, from computer vision and speech recognition to natural language processing (NLP), bioinformatics, and game playing. In this section, we survey some major classes of architecture, outlining the core design and how they differ from the "vanilla" MLP.

Convolutional Neural Networks (CNNs) [328, 329] are a specialized class of feedforward networks that are particularly suitable for processing grid-like data, most notably images, but also audio spectrograms and video frames, for instance. Their strength lies in their ability to recognize patterns and extract hierarchical features from visual data through a structured learning process. In practice, a CNN learns a hierarchy of feature detectors, from edges and textures in early layers to object parts and whole objects in deeper layers. To this end, the unique architecture of CNNs is built around two primary types of hidden layers: convolutional and pooling layers. The convolutional layers apply a set of learnable filters, or kernels, that slide across the input data (typically an image) to automatically extract various features such as edges, textures, and more complex shapes. From a mathematical point of view, this operation is implemented by a convolution. A key feature of CNNs is that its nodes share weights and bias values across different spatial locations, which is highly efficient because the same feature detector can be applied across different areas of the input, making the network translationally invariant. For example, an edge detector can identify edges in an image regardless of where they are located. The pooling layers typically follow the convolutional layers and serve to reduce the dimensionality of the feature maps, that is, the spatial resolution, while preserving the most essential information. A typical CNN structure involves a series of alternating convolutional and pooling layers, often culminating in one or more fully connected layers that perform the final classification or regression.

Recurrent Neural Networks (RNNs) [325] are specifically designed to process sequen-

tial data characterized by input of variable length and temporal dependencies, such as time series and sentences, where the timing and order in which the input is structured are relevant. The RNN architecture prescribes that each layer's output feeds back into itself for the next input, forming loops that allow data to circle back through the hidden layers before an output is produced, providing them with a form of "memory" that persists across different time steps. In this way, RNNs maintain a hidden state that continuously captures and updates information about previous inputs, allowing them to find structure in time-dependent data. However, RNNs suffer from vanishing or exploding gradients when learning long-range dependencies, that is, during training, gradients can become extremely small (vanishing) or excessively large (exploding), causing information to be lost or amplified over time, which complicates the training over many time steps. To address this, RNN variants were developed, such as Long Short-Term Memory (LSTM) [330] networks and Gated Recurrent Unit (GRU) [331]. Both LSTM and GRU mitigate the vanishing gradient problem by enhancing the "memory" capabilities of recurrent networks. They allow important signals to pass through many steps with minimal attenuation, enabling them to effectively learn and remember long-term dependencies in sequential data.

Transformer networks [332, 333, 334] have profoundly revolutionized the field of NLP and beyond, fundamentally altering how sequential data is processed. Unlike RNNs and LSTM networks, which process sequences step by step, transformers enable parallel processing of sequential data. Their core principle is to learn context and meaning by effectively tracking relationships between elements within a sequence, regardless of their distance from each other. The cornerstone of the transformer architecture is the self-attention mechanism that quantifies how much attention each word should pay to every other word in the sequence. In this way, the model can dynamically weigh the importance of different elements in the input and adjust their influence on the output. By stacking several such mechanisms, we obtain multi-head attention, allowing for parallel processing of the data and much faster training on specialized hardware. Originally devised for NLP tasks, such as translation and language modeling, transformers have been extended to text generation, image recognition and generation, and multi-modal tasks. This leads us to the framework of generative models, some of which we now present.

Generative Adversarial Networks (GANs) [335] represent one of the most famous generative models, characterized by a unique competitive training process involving two distinct neural networks. A generator $\mathcal{G}$ takes random noise and produces synthetic data, and a discriminator $\mathcal{D}$ tries to distinguish real training data from the output of the generator. $\mathcal{G}$ and $\mathcal{D}$ are trained simultaneously: $\mathcal{D}$ is trained to minimize its classification error between real and fake data, while $\mathcal{G}$ is trained to "fool" $\mathcal{D}$ by maximizing the classification error of $\mathcal{D}$. This mini-max "game" drives both networks to improve, with $\mathcal{G}$ ultimately learning to create outputs that are indistinguishable from real data, so realistic that $\mathcal{D}$ can no longer reliably detect it as fake.

Autoencoders [336, 337] are a fundamental class of unsupervised learning models primarily designed for data compression, dimensionality reduction, and feature extraction. An autoencoder is a type of neural network architecture designed to efficiently compress (encode) input data into a compact (latent) representation and then reconstruct (decode) the original input from this compressed representation. In doing so, an autoencoder learns salient features and can perform dimensionality reduction similar to principal component analysis, but with nonlinear transformations. Many variants exist, such

as denoising, sparse, and variational autoencoders (VAEs).

Boltzmann Machines [338, 339, 340] are a type of recurrent neural network in which each neuron is connected to all other neurons. Restricted Boltzmann Machines (RBMs) simplify this by restricting connectivity: only some neurons act as input neurons, while others are hidden, and there is no full connectivity between neurons within the same layer. Deep Belief Networks (DBNs) [341] are a class of generative models constructed by stacking multiple layers of RBMs. They learn by clamping the value of input neurons to a desired pattern, after which the weights are adjusted, allowing the network to converge to the most similar state compared to their input.

Although the analysis of the ecosystem of neural networks and AI models could be much broader, our attention turns to the specific tasks of QST and quantum state characterization. We now illustrate the application of some of the previously mentioned models to these challenges.

3.2 AI models for QST

Various models in machine learning and deep learning have been developed to characterize quantum states. At its core, machine learning (or deep learning) is a form of statistical inference. Generally, we suspect that a relationship exists between the input features (the data we gathered) and the output (what we aim to classify or predict). Machine learning techniques become valuable when no analytical formula can connect these elements. In QST, the linear inversion procedure can produce non-physical states with noisy data, necessitating the use of statistical inference tools such as MLE or Bayesian inference. Thus, machine learning and deep learning models offer a viable option for reconstructing quantum states from measurement outcomes.

In this respect, different types of neural network have been employed in quantum state characterization. They encompass feedforward [342, 343, 344, 345, 346, 347], convolutional [348, 349], and recurrent neural networks [350], autoencoders [351], models within the generative framework [352, 353, 354], and attention-based ones [355, 356, 357, 358, 359]. In [360] Wu et al. applied techniques from multitask learning to predict global properties of many-body quantum states from the measurement of observables that involve only a small number of neighboring sites, transferring information to systems of higher dimension and generalizing well to unseen Hamiltonians. Models have been combined together, such as RBMs and feedforward NNs [361], or combined with other QST approaches, such as shadow tomography [362] and adaptive QST [350, 239], to improve performance. Models from alternative frameworks, such as reinforcement learning [363, 364, 365, 366] and genetic algorithms [367], have also been considered.

Machine and deep learning models have been adopted for many applications within the framework of quantum state characterization beyond QST, such as Hamiltonian learning [368], direct fidelity estimation [369], quantum process tomography [370], and classification of quantum states [371]. For instance, Ahmed et al. [372] leveraged CNNs and conditional GANs to classify or reconstruct optical quantum states from the measurements of quasi-probability distributions, such as the Wigner or Husimi ones, or the histograms of the measurement statistics, such as the photon number distributions. Furthermore, hardware tailored to QST can be designed, such as an optical neural network specific for QST [373] and a FPGA optimized for machine learning QST [374]. The review

by Krenn et al. [375] illustrates several use cases of artificial intelligence tools for quantum technologies.

Carleo and Troyer in [376] (2017) introduced a novel strategy to solve problems involving quantum many-body states. They interpreted the quantum wave-function as a computational black box that, given an input configuration of the degrees of freedom composing the system, returns a phase and an amplitude according to the probability distribution induced by the wave-function itself, and approximated the wave-function with an RBM. The reason behind this choice is that RBMs offer a compact variational representation of many-body quantum states, where the parameters of the network represent the amplitude and the phase of the state under study, capable of sustaining non-trivial correlations. This idea, originally applied to find the ground state or describe the time evolution of transverse-field Ising and antiferromagnetic Heisenberg models, has been extended in several ways, such as the connection of RBMs to tensor networks [377], the construction of deep BMs [378], and the purification of mixed states [379]. It has been applied to numerical QST [380, 381], implemented in a two-photon experiment [382], and integrated with a quantum simulator [383]. Finally, the reviews [384, 385, 386, 387] present these and other numerous results of this approach.

To our knowledge, few of these methods really tackle the problem of reducing the number of observables required for QST without any hypothesis on the state to reconstruct, or can do it only by adopting strategies such as shadow tomography that do not aim to reconstruct the density matrix. However, in the previous chapter we showed that tQST is a successful approach to solve this problem. Thus, we now study whether it is possible to embed a proper AI model in the protocol and extend tQST to a different quantum state characterization task.

3.3 Deep learning models for quantum state characterization

In the previous chapter, we adopted maximum likelihood estimation (MLE) for the reconstruction of the density matrix with the tQST protocol. However, we just showed that different tools can be used to carry out this task, such as machine learning or deep learning models. In this respect, we develop and use a physics-inspired deep learning model to retrieve the density matrix from the measurement outcomes suggested by the tQST protocol. Our goal is to explore the feasibility of deep learning-based data processing as an alternative to MLE, without further reducing the number of projective measurements. Furthermore, we use the same model and measurement outcomes to estimate the purity of quantum states, thus extending the applicability of the tQST protocol to other tasks beyond only density matrix reconstruction. We now present the two models we developed for our purposes.

The first model we considered was the MLP, as the simplest deep learning architecture to implement, with high flexibility, and due to its behavior as a universal approximator. The MLP architecture comprises a sequence of Linear layers (according to the PyTorch nomenclature [388]), each followed by an activation function. However, relying on the universal approximation property may impose a very high cost in terms of the complexity of the model, the computational burden of the training procedure, and the number

of training samples. To control these factors, we can imbue the architecture with prior knowledge of the mathematical structure of the optimal solution. To this end, many recent studies explored the concept of equivariance in neural networks [389, 390, 391]. A network (or one of its components) is equivariant to a group of transformations if applying a transformation in the group to the input causes a predictable change in the output. For example, CNNs are equivariant to translations in the plane, and graph neural networks (GNNs) are equivariant to permutations in the order used to label the vertices in the graphs [392]. If we consider density matrices, a permutation of rows and columns corresponds to a permutation of the elements of the basis chosen for the representation, which leads to a different representation, but does not modify the possible outcomes or the expectation value of a given observable. More precisely, consider a density matrix ρ and an observable O represented in a given basis $\mathfrak{B}$. The possible outcomes of a measurement of O are the corresponding eigenvalues, and the expectation value of O is given by $\langle O \rangle = \mathrm{Tr}\,(\rho O)$. Let us now permute, that is, reorder the elements of $\mathfrak{B}$ to obtain a permuted basis $\mathfrak{B}'$, and accordingly find the permuted representations of the density matrix ρ' and the observable O'. A permutation of the basis does not change the eigenvalues of O and ρO, thus the physical results as the outcomes of a measurement or the expectation value of the observable are unchanged.

We then decided to take advantage of permutation equivariance to develop the second model for our work. If we consider the case of neural networks applied to density matrices, the permutation of rows and columns should result in a similarly permuted output from the neural network, analogously as the case of adjacency matrices processed by GNNs. In this context, a neural network $F : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ is **permutation equivariant** if for any input matrix ρ and any permutation matrix π, we have:

$$F(\pi^T \rho \pi) = \pi^T F(\rho) \pi. \tag{3.15}$$

A network of this kind can be built as a sequence of permutation equivariant layers, each of which enjoys the permutation equivariance property (3.15). The general form for permutation equivariant linear (PELinear) layers was derived by Thiede et al. in [393]. The authors showed that any permutation equivariant linear operator $\rho \to \rho'$ can be expressed as follows:

$$\begin{aligned}
\rho'_{ij} = {} & w_0 \rho_{ij} + w_1 \rho_{ji} + w_2 \sum_v \rho_{iv} + w_3 \sum_v \rho_{vi} + w_4 \sum_u \rho_{uj} \\
& + w_5 \sum_u \rho_{ju} + w_6 \sum_{u,v} \rho_{uv} + w_7 \sum_u \rho_{uu} + w_8 \rho_{ii} + w_9 \rho_{jj} \\
& + \delta_{ij} \left(w_{10} \rho_{ij} + w_{11} \sum_u \rho_{uu} + w_{12} \sum_{u,v} \rho_{uv} \right. \\
& \left. + w_{13} \sum_v \rho_{iv} + w_{14} \sum_v \rho_{vi} \right),
\end{aligned} \tag{3.16}$$

with $w_0, w_1, \ldots, w_{14}$ the learnable parameters of the operator and $\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ if $i \neq j$. In practice, this formulation is extended to the case where matrix elements are vectors of real numbers: $\rho_{ij} \in \mathbb{R}^c$, $\rho'_{ij} \in \mathbb{R}^d$, and all coefficients are matrices of real-valued learnable parameters, $w_k \in \mathbb{R}^{c\times d}$. However, we need more than just linear layers

to build a neural network. Fortunately, permutation equivariance is closed under functional composition, and most auxiliary layers in neural networks are already permutation equivariant (activation functions, normalization, and pooling, for instance). This means that a complete permutation equivariant network, called permutation equivariant MLP (PEMLP) in this work, can be obtained as a sequence of PELinear layers and other auxiliary layers. The networks we experimented with received complex numbers as input, which are encoded as two-dimensional vectors before processing. Now that we have the architectures at hand, we need to create the training and test sets, which we detail in the following section.

3.4 Datasets, workflow, and training details

We generated separated datasets for the density matrices, the corresponding purities, and the measurement outcomes according to the tQST protocol. We first generated a dataset of density matrices of n_q qubits with the same share of pure and mixed states. In the case of mixed density matrices, we first uniformly randomly choose the number of vanishing diagonal elements z between 0 and $2^{n_q} - 2$. This determines the possible values of the rank r of the mixed density matrix, which varies between 2 and $2^{n_q} - z$. The number of possible pairs (z, r) is $\sum_{k=1}^{2^{n_q}-1} k = 2^{n_q-1}(2^{n_q} - 1)$. For each pair (z, r), we generated M random density matrices. Thus, the number of mixed density matrices is $M \times 2^{n_q-1}(2^{n_q} - 1)$. Finally, we generate an equal number of pure density matrices fixing the rank $r = 1$ and choosing z as before. The total number of density matrices in the dataset is then $2 \times M \times 2^{n_q-1}(2^{n_q} - 1)$. The 2-qubit dataset consisted of 24,000 density matrices $(n_q = 2, M = 2,000)$, while the 4-qubit dataset of 120,000 density matrices $(n_q = 4, M = 500)$.

We then constructed the dataset containing the measurement outcomes, which are the inputs of our models. For each density matrix, we computed the measurement outcomes according to the tQST protocol. More specifically, if $\sqrt{\rho_{ii}\rho_{jj}} \geq t_g$, with t_g the Gini threshold, the projective measurements associated with ρ_{ij} are performed and the outcomes are recorded; otherwise no measurement is performed, and we record a mock value 2. We chose to use 2 to represent the non-performed measurements as it is out of the range of the valid measurement outcomes, which are positive real numbers between 0 and 1. Finally, for each density matrix of the dataset we computed and stored the corresponding purity $\mathcal{P} = \mathrm{Tr}[\rho^2]$, thus constructing the datasets for the purity estimation.

We also generated noisy datasets to study the impact of noise on the performance of our models. In particular, we studied whether the models could reconstruct the noiseless density matrix, or estimate the corresponding purity, from noisy measurement outcomes. To generate the required datasets, we applied a noisy channel to the noiseless density matrix and then computed the noisy measurement outcomes. Applying a noisy channel results in general in a different number of required projective measurements with respect to the noiseless case, for the diagonal elements of the density matrix change after having applied the noisy channel. In our analysis, we considered the same two noisy channels presented in Section 1.2.8, which we report here for convenience. The first is the depolarizing channel (1.89), which maps a density matrix ρ into a linear combination of itself and

the maximally entangled state:

$$\mathcal{E}_{\text{depol}}(\rho) = (1-p)\,\rho + \frac{p}{2^{n_q}}\mathbb{I}_{2^{n_q}}. \tag{3.17}$$

with $0 \leq p \leq 1 + 1/(D^2-1)$ a parameter quantifying the strength of the noise, and $\mathbb{I}_{2^{n_q}}$ the 2^{n_q}-dimensional identity matrix. The second channel was the experimental state-error (1.92), which maps a density matrix ρ into a convex combination of itself and a random density matrix:

$$\begin{aligned} \mathcal{E}_{\text{exp-state}}(\rho) &= (1-\varepsilon)\,\rho + \varepsilon\rho_{\text{random}} \\ \rho_{\text{random}} &= \frac{R^\dagger R}{\text{Tr}\left[R^\dagger R\right]} \\ R &= 2\text{rand}\left(2^n\right) - 1 + i\left[2\text{rand}\left(2^n\right) - 1\right]. \end{aligned} \tag{3.18}$$

The rand (2^n) function generates a $2^n \times 2^n$ matrix of random values sampled from a uniform distribution over the interval (0,1). This channel simulates experimental error in the preparation of the state. Indeed, the prepared state always differs from the intended state by some small amount, here quantified by the ε parameter and implemented via a random density matrix.

The workflow of the proposed method for quantum state characterization based on tQST and deep learning is the following. We generate a (noiseless or noisy) density matrix and measure the observables specified by the tQST protocol. The measurement outcomes are recorded as explained above, stored in a array of size $(1, 4^{n_q})$ and fed into the model to estimate the purity or reconstruct the density matrix depending on the task. The architectures of the models are similar for both tasks and consist of an input layer of 4^{n_q} neurons, some hidden layers (whose number and kind depend on the size of the problem) and a final output layer with 4^{n_q} neurons for tQST, while a single neuron is necessary for the purity estimation. As anticipated, the output of the model has a different size depending on the performed task. On the one hand, for tQST, the values of the output neurons represent the 4^{n_q} real parameters that characterize the density matrix, ordered first with the (real) diagonal elements and then the off-diagonal elements, with real and imaginary parts interleaved. These parameters are then converted into a Hermitian, trace-one matrix μ. However, this reconstruction process does not guarantee μ to be positive semi-definite. We handle this problem by numerically implementing the solution presented in [56]. On the other hand, for the purity estimation, there is a single output output value. The model can estimate a purity value slightly higher than 1 or lower than the minimum $1/2^{n_q}$. In this case, the estimated values are rescaled to the maximum or minimum, respectively.

Before showing and commenting on the results, we provide some details about the training procedure. The code for the training of the models was written in the Python language using the PyTorch library [388]. All datasets were split into training, validation, and test sets with shares equal to 90%, 5%, and 5%, respectively, and loaded with a batch size equal to 32. We adopted MSELoss as the objective function to be minimized, since both tQST and purity estimation are regression tasks. We used the Adam optimizer with a learning rate equal to 10^{-4}. The number of epochs was 100. The 2-qubit MLP models for tQST and purity estimation have 2,128 and 1,633 parameters, respectively, and required a training time of ≈ 2 minutes on the Google Colab platform without GPUs. The PEMLP models have 3,972 and 3,989 parameters for the same tasks and required ≈ 30 minutes for

training on the same platform. The 4-qubit model for tQST has 295,748 parameters, while the one for purity estimation has 164,933 parameters, and both required ≈ 4 hours for training on a standard desktop computer with an Intel Core i7-4790 processor, 3.60 GHz CPU, and 32 GB system memory, without GPUs. Finally, data supporting the findings of this study are openly available on Zenodo at https://doi.org/10.5281/zenodo.14887723.

3.5 Training results

3.5.1 2 qubits

We show the results of our method for the two-qubit case. We consider this as a benchmark for the validation of the method. We tested the MLP and PEMLP models for tQST and purity estimation. For tQST, the MLP consists of two hidden Linear layers, each containing 32 neurons, while the PEMLP has a single hidden PELinear layer with 64 internal features. In both models, the activation function between the hidden layers is the ReLU function [394, 395]. We evaluated the performance on the test set using average fidelity as a figure of merit, defined as:

$$F = \frac{1}{N_{test}} \sum_i \mathrm{Tr}\left(\sqrt{\sqrt{\rho_{o,i}}\rho_{p,i}\sqrt{\rho_{o,i}}}\right), \tag{3.19}$$

with ρ_o (ρ_p) the observed (predicted) density matrix and N_{test} the number of samples in the test set. The MLP achieved an average fidelity $\overline{\mathcal{F}} = 0.9555(411)$ on the noiseless test set, while the PEMLP achieved $\overline{\mathcal{F}} = 0.8893(919)$. These results allowed us to conclude that both models can reconstruct density matrices of two qubits from the measurement outcomes collected according to the tQST protocol. The results are compatible with those obtained with maximum likelihood, which achieved an average fidelity of $\overline{\mathcal{F}} = 0.9941(174)$ on the same test set.

We then investigated the impact of noise on the accuracy with which the state can be reconstructed to assess the practical applicability of these models in an experimental scenario. The strengths p and ε of the noisy channels were set at 0.01, 0.05, and 0.1 in our experiments. The results, reported in Table 3.1, indicate that both the MLP and the PEMLP are robust against noise, with slightly modified performance as the noise strength increases.

The sparsity of the density matrix plays a crucial role in the tQST protocol. On the one hand, the larger the number of vanishing diagonal elements, the smaller the number of projective measurements we have to perform according to the tQST protocol. On the other hand, many of the states used in quantum communication or computation protocols are sparse in the computational basis. Thus, we further analyzed the models' performance based on the number of zeros on the diagonal of the noiseless density matrix. For each model, we computed the average fidelity for density matrices with 0, 1, and 2 vanishing diagonal elements separately. The MLP performed consistently well across these subsets, whereas the PEMLP showed slightly better performance on density matrices with two vanishing diagonal elements. These findings, which we report in Figure 3.2, suggest that both models are reliable in reconstructing sparse states, such as Bell states. We substantiate this conclusion by reconstructing the Bell state $|\phi^-\rangle = 1/\sqrt{2}(|00\rangle - |11\rangle)$ using high-quality experimental data from [396], where the authors reconstruct single-

MLP - tQST		
Noise strength	Depolarizing	State-error
0	0.9555(411)	0.9555(411)
0.01	0.9501(505)	0.9624(353)
0.05	0.9561(405)	0.9448(498)
0.1	0.9578(436)	0.9366(591)
PEMLP - tQST		
Noise strength	Depolarizing	State-error
0	0.8893(919)	0.8893(919)
0.01	0.8918(907)	0.8838(916)
0.05	0.8864(915)	0.8901(907)
0.1	0.8883(900)	0.8858(912)

Table 3.1: **Results for the MLP and PEMLP on the 2-qubit noiseless and noisy test sets for tQST.** For each value of the noise strength we reported the average fidelity and the standard deviation on the test set.

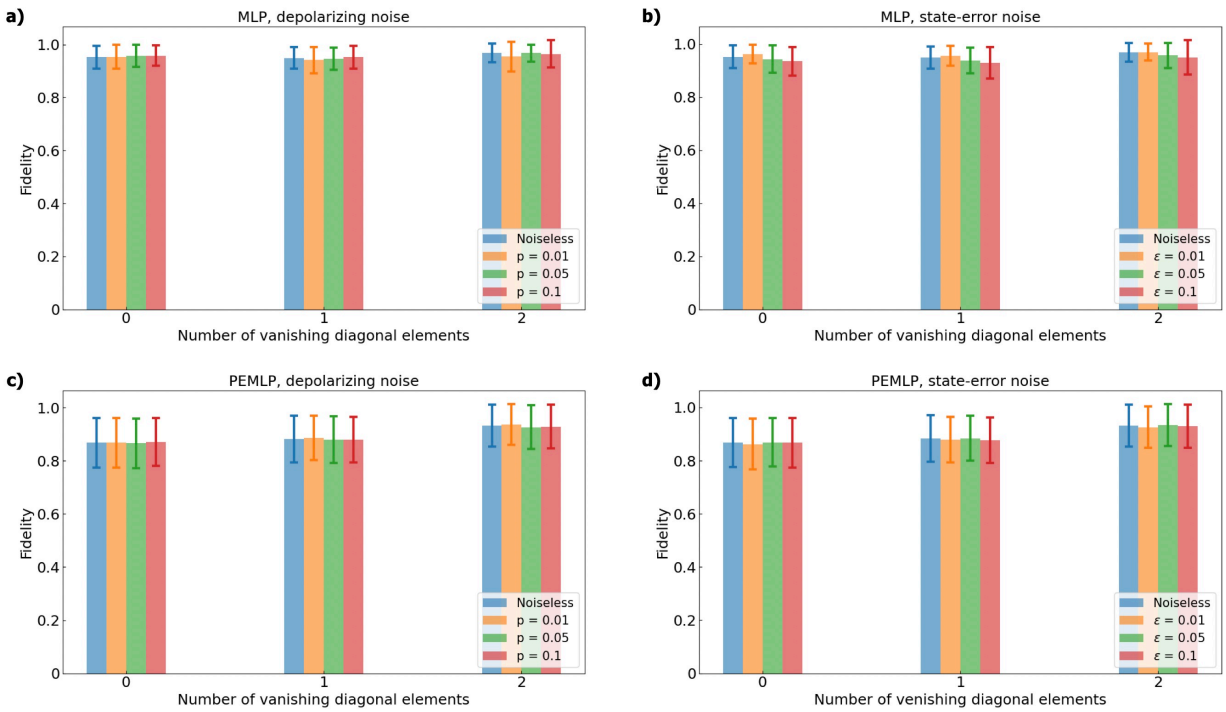


Figure 3.2: **Performance of the 2-qubit models on states with different numbers of vanishing diagonal elements.** We report the average fidelity of the 2-qubit models on states with different number of vanishing diagonal elements and different values of the noisy channels applied. The height of each bar represents the average value, while error bars represent the standard deviation from the average. We show the results for the MLP **a),b)** and the PEMLP **c),d)** under the action of the depolaziring and the state-error channel, respectively.

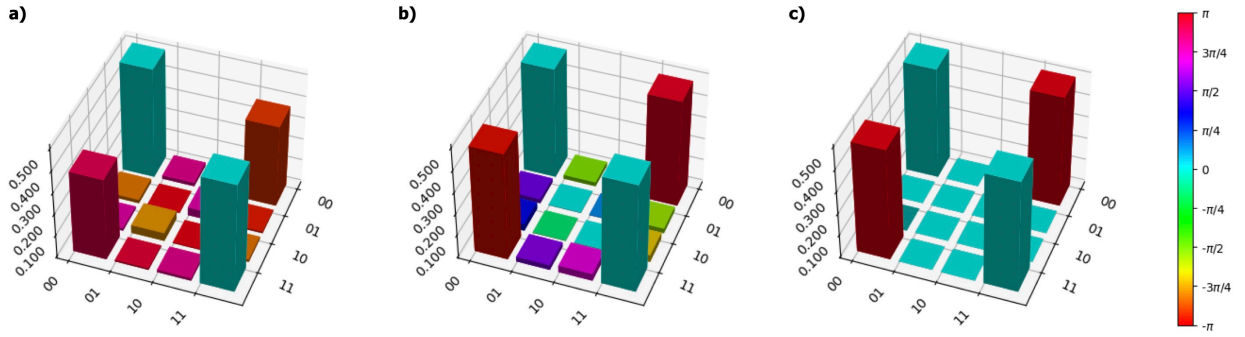


Figure 3.3: **Reconstruction of the density matrix of the Bell state $|\phi^-\rangle$.** The height of the bar represents the modulus, while the phase is encoded in the color as shown in the colorbar. Measurement outcomes are experimental data from [396]. **a)** tQST reconstruction with the MLP model trained on the depolarizing noise dataset with $p = 0.01$. The fidelity with the noiseless target state is 0.93. **b)** tQST reconstruction using maximum likelihood. The fidelity with the noiseless target state is 0.99. **c)** Target density matrix. Figure from [3].

and two-qubit states encoded in the polarization of photons generated via a non-linear optical process. The result is shown in Figure 3.3. Reconstruction of this state with our model trained on the depolarizing noise dataset with $p = 0.01$ required only 6 projective measurements instead of 16 and achieved a fidelity with the noiseless target state of $\mathcal{F} = 0.93$.

We then applied the same models to the purity estimation and evaluated their efficacy using two figures of merit. The first is the mean squared error (MSE) on the test set, defined as:

$$\text{MSE} = \frac{1}{N_{test}} \sum_i |y_{p,i} - y_{o,i}|^2, \tag{3.20}$$

with $y_{p,i}$ ($y_{o,i}$) the predicted (observed) purity, and N_{test} the number of samples of the test set as above. The second is the R^2 coefficient, or coefficient of determination, defined as:

$$R^2 = 1 - \frac{\sum_i (y_{o,i} - y_{p,i})^2}{\sum_i (y_{o,i} - \overline{y}_o)^2}, \tag{3.21}$$

with $\overline{y}_o$ the average observed purity computed on the test set. The MSE achieved by the MLP and PEMLP models was 0.0144 and 0.0097, respectively, while the corresponding R^2 coefficients were 0.7808 and 0.8528. These results demonstrate that both the MLP and the PEMLP can estimate the state purity from a limited number of projective measurements selected according to the tQST protocol. This is significant because, in principle, one would first need to reconstruct the density matrix via quantum state tomography and then compute the purity. If we reconstruct the density matrices corresponding to the test set with maximum likelihood and then estimate the purity, we achieve MSE = 0.0006 and R^2 = 0.9902.

We studied the performance of the models under the effect of the same noisy channels as in tQST. The results, reported in Table 3.2, confirmed that both models are robust against noise, as their performance is only slightly affected by increasing the noise level.

MLP - Purity estimation		
Noise strength	Depolarizing	State-error
0	0.0144 / 0.7808	0.0144 / 0.7808
0.01	0.0163 / 0.7523	0.0148 / 0.7739
0.05	0.0157 / 0.7615	0.0160 / 0.7562
0.1	0.0162 / 0.7536	0.0139 / 0.7884
PEMLP - Purity estimation		
Noise strength	Depolarizing	State-error
0	0.0097 / 0.8528	0.0097 / 0.8528
0.01	0.0101 / 0.8462	0.0123 / 0.8123
0.05	0.0104 / 0.8420	0.0118 / 0.8208
0.1	0.0106 / 0.8382	0.0149 / 0.7734

Table 3.2: **Results for the MLP and PEMLP on the 2-qubit noiseless and noisy datasets on the purity estimation.** For each value of the noise strength we reported the MSE and the R^2 on the test set.

It is interesting to note that the MLP works better in tQST, while the PEMLP gives better results in purity estimation. Maximum likelihood provides more accurate results when applied to both tasks. Nevertheless, the results on the 2-qubit datasets are a good benchmark that assesses the capability of the models to characterize quantum states. Thus, we now study the scalability of our approach to the 4-qubit case.

3.5.2 4 qubits

Based on the results achieved in the 2-qubit case, we decided to combine the MLP and PEMLP models to improve performance and conducted an ablation study [397] considering different numbers of hidden layers and neurons/internal features per hidden layer for each model separately. The activation function between the layers was the ReLU function. We trained each model on the noiseless dataset. The results we report in Table 3.3 clearly indicate that neither model, in isolation, can extract the relevant information from the input data to reliably reconstruct the density matrix.

The model with the best trade-off between the number of parameters, the training time, and the performance for tQST is composed of three hidden PELinear layers, with 32 internal features each, followed by one hidden Linear layer with 512 neurons. The activation function between each pair of hidden layers is the ReLU function. As in the 2-qubit case, we evaluated the model using the average fidelity and standard deviation on the test set. The average fidelity of the combined model on the noiseless test set is $\overline{\mathcal{F}} = 0.8811(540)$. The result shows that the combined model can carry out tQST and reconstruct 4-qubit density matrices, in contrast to isolated models. The MLE approach achieved an average fidelity of $\overline{\mathcal{F}} = 0.9982(25)$ on the same test set. In Figure 3.4 we show an example of a 4-qubit density matrix reconstructed with our model, the corresponding MLE reconstruction, and the target state. Notably, we achieved fidelity equal to 0.95 by performing only 46 projective measurements instead of 256 as in QST.

MLP					
Neurons per hidden layer	64	128	256	512	1024
1 hidden layer	0.4282(1951)	0.5094(1829)	0.5926(1842)	0.6098(1840)	0.6745(1505)
2 hidden layers	0.6072(1861)	0.6093(1853)	0.6109(1842)	0.6414(1672)	0.6424(1668)
3 hidden layers	0.6185(1783)	0.6441(1638)	0.6775(1645)	0.7102(1254)	0.6585(1586)
PEMLP					
Internal features per hidden layer	8	16	32	64	128
1 hidden layer	0.6105(1857)	0.6132(1846)	0.6067(1846)	0.6164(1837)	0.6207(1812)
2 hidden layers	0.6064(1855)	0.6043(1845)	0.6033(1837)	0.5985(1800)	0.5985(1766)
3 hidden layers	0.4114(1994)	0.6169(1807)	0.6011(1819)	0.6239(1766)	0.6338(1726)

Table 3.3: **Results of the ablation study performed on the isolated MLP and PEMLP on the 4-qubit noiseless dataset.** For each configuration, we reported the average fidelity and the standard deviation on the test set.

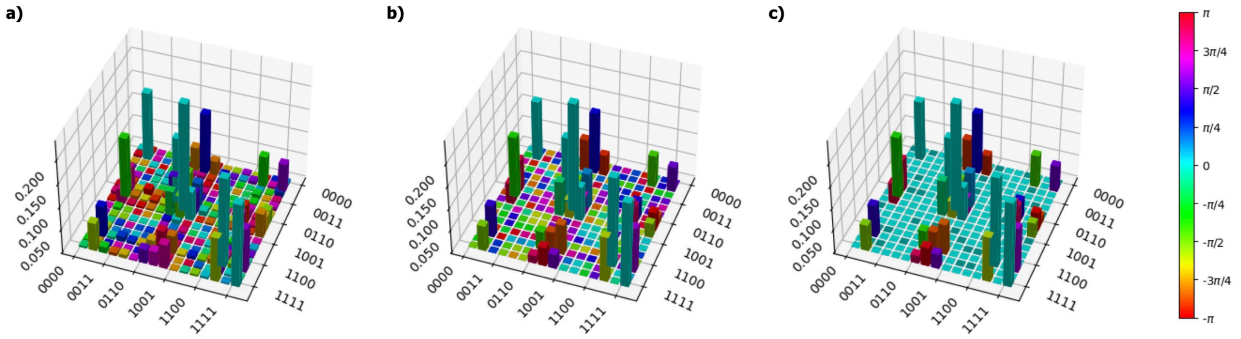


Figure 3.4: **Reconstruction of a 4-qubit density matrix.** The encoding of modulus and phase is the same as Figure 3.3. **a)** tQST reconstruction with the combined model trained on the noiseless dataset. The fidelity with the noiseless target state is 0.95. **b)** tQST reconstruction using maximum likelihood. The fidelity with the target density matrix is 0.99. **c)** Target density matrix. Figure from [3].

We believe that the success of the combined model can be explained as follows. Thanks to permutation equivariance, the PELinear layers exploit our prior knowledge about the symmetric structure of density matrices. Unlike general linear layers, PELinear layers do not need to learn the relationship among the elements of the density matrix for each possible permutation. This implies that for n_q qubits, PELinear layers are $(2^{n_q})!$ times more sample-efficient than general Linear layers. This fact is reflected in the reduction of the number of learnable parameters (which does not depend on n_q) and in the better generalization accuracy. The PELinear layers essentially elaborate "local information", as the extracted feature vectors are "located" in the position occupied by the elements they refer to. Finally, the Linear layer combines all the features holistically, providing a more reliable final prediction.

For the purity estimation, the model is the same as for tQST, with the addition of a dropout layer before the final Linear layer with probability equal to 0.5. We included this to avoid overfitting that came to light in the initial numerical experiments, as illustrated in Figure 3.5. On the noiseless dataset, the combined model achieved MSE = 0.0023 and $R^2 = 0.9837$. These results demonstrate that the model can effectively estimate the purity of a quantum state using only the measurement outcomes required by the tQST protocol without reconstructing the entire state. Furthermore, they confirm that the selected projective measurements provide sufficient information to accurately estimate a quantity that is a function of the density matrix. If we leverage maximum likelihood to reconstruct the density matrices and then estimate the state purity, we achieve MSE = 2.4×10^{-5} and $R^2 = 0.9998$.

Finally, we studied the impact of noise on the model performance. In our simulations, we set the strengths p and ε of the noisy channels to 0.01, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, and 1. Figures 3.6a) and 3.7 illustrate the performance of the combined model trained on noisy data to reconstruct the noiseless density matrix or predict the noiseless state purity, depending on the task. We can identify four regions, covering different ranges of noise strength. In the first region, from 0 to 0.1, the effect of noise is small and comparable to statistical fluctuations. In the second, ranging from 0.1 to 0.4, the noise provides a clear improvement of the results with respect to the noiseless case. Although counterintuitive,

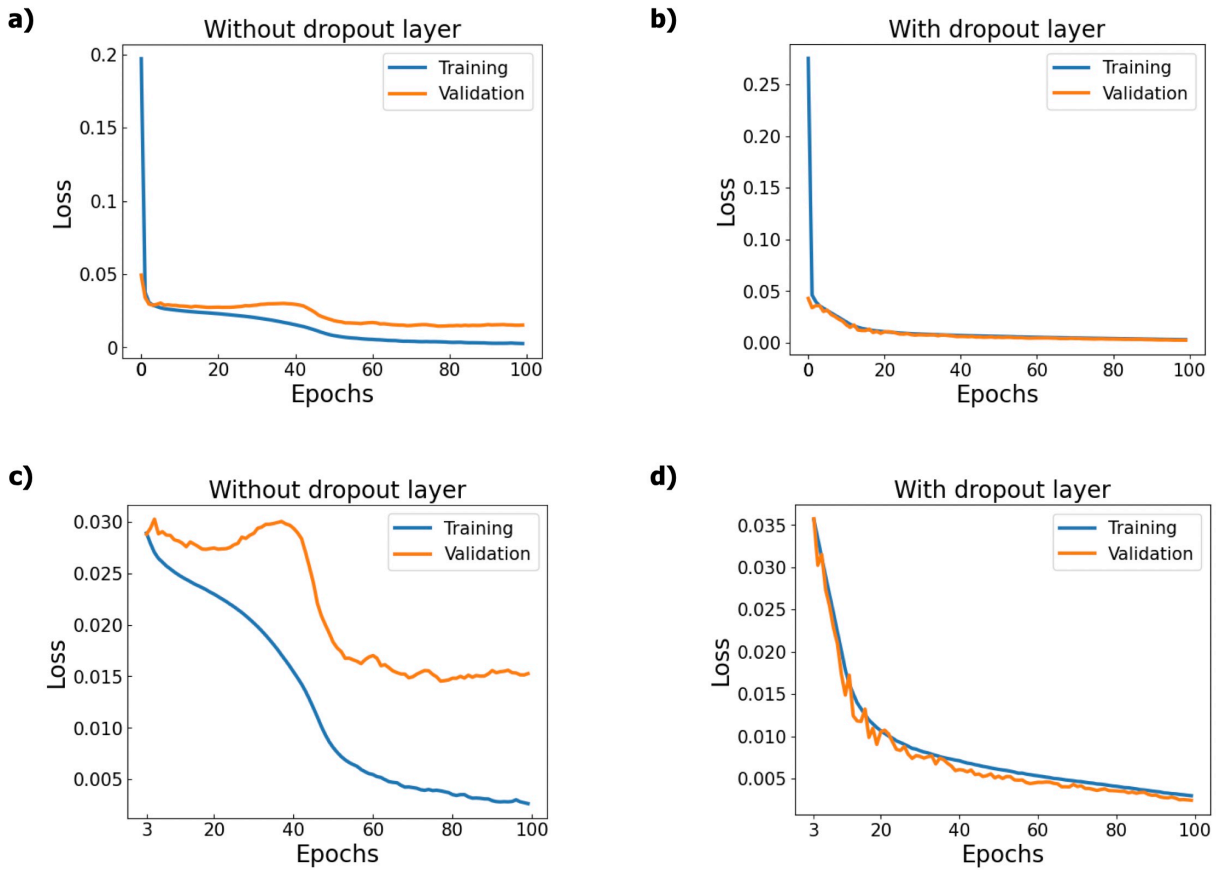


Figure 3.5: **Effect of the dropout layer on the learning procedure.** In panels **a)** and **b)** we show the training and validation loss curves for the purity estimation on the noiseless dataset without and with the dropout layer, respectively. The dropout layer removes the overfitting effect, represented by the gap between the training and validation curves. In panels **c)** and **d)** we show the same curves starting from epoch 3 to highlight the amplitude of the gap.

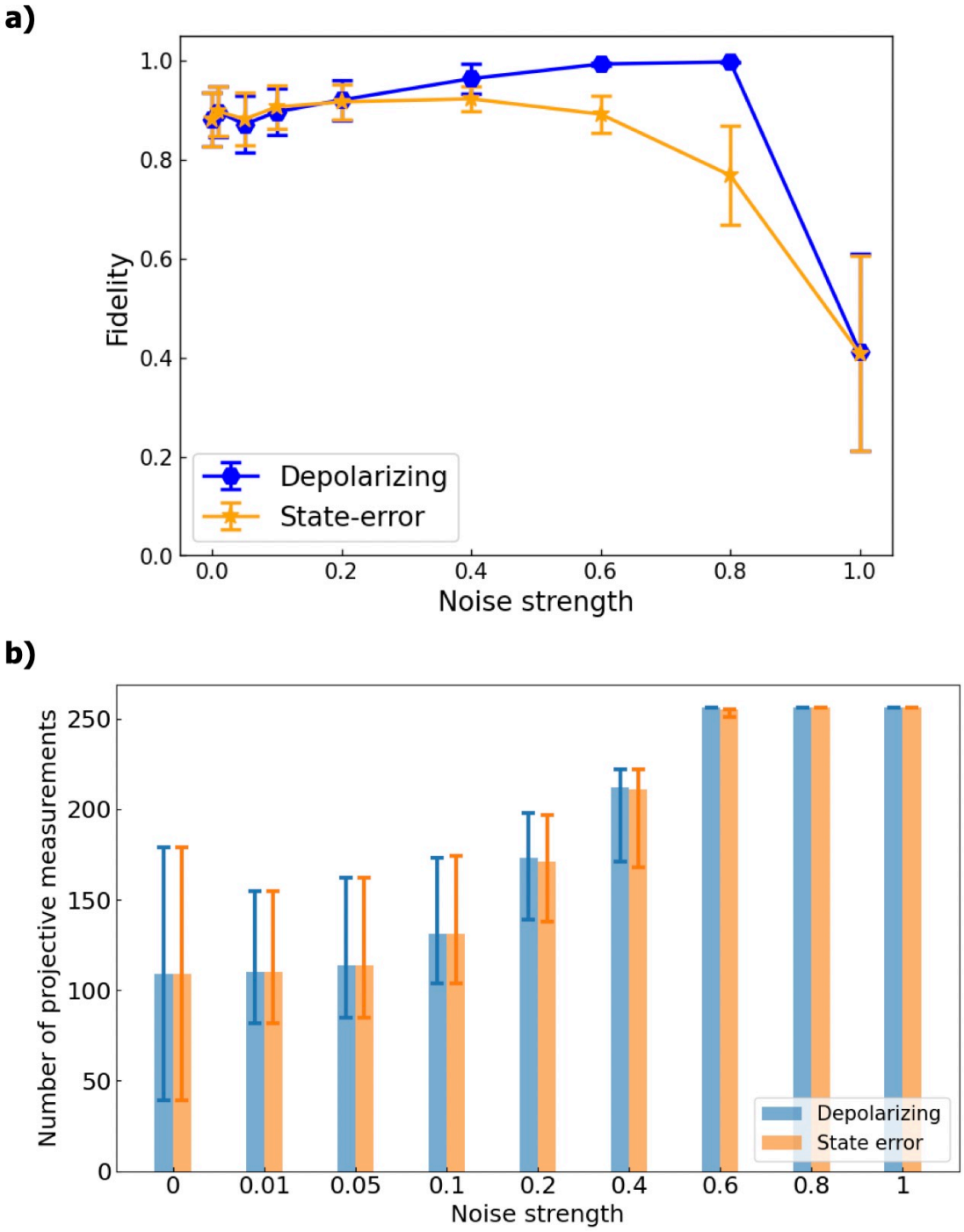


Figure 3.6: **Performance of the combined model on tQST as a function of the noise strength.** In panel **a)** and **b)** we show the average fidelity on the test set and the average number of measurements with respect to noise strength for 4-qubit states, respectively. In panel **b)**, the height of each bar represents the average value, while error bars represent the standard deviation from the average. In panel **b)**, error bars were computed asymmetrically, that is, the upper (lower) error bar quantifies the standard deviation of the data that are above (below) the average, respectively. Figure adapted from [3].

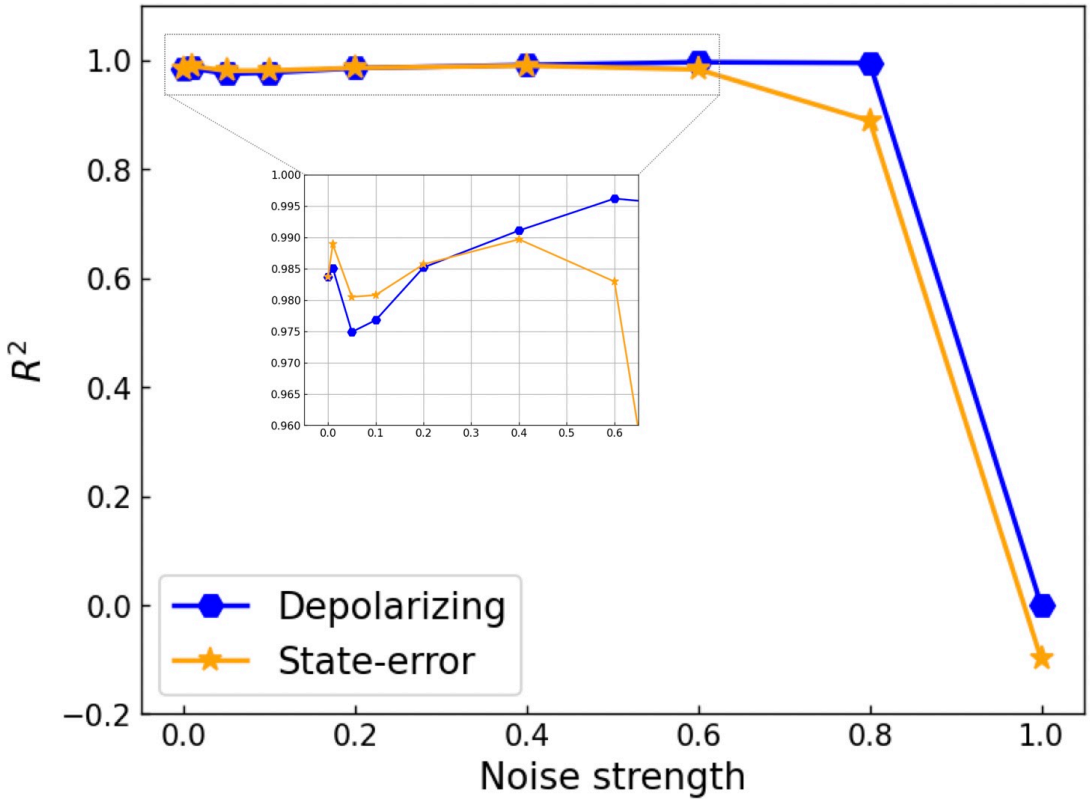


Figure 3.7: **Performance of the combined model on purity estimation as a function of the noise strength.** We show the R^2 coefficient on the test set with respect to noise strength for 4-qubit states. The inset shows a zoomed-in version of the values for noise strength between 0 (noiseless) and 0.6. Figure adapted from [3].

this is a well-documented phenomenon in machine learning. Adding controlled noise can act as a form of regularization, preventing overfitting and improving generalization by encouraging the model to learn more robust features [398, 399, 400, 322, 401]. The performance of the model in the first two regions is similar for either the depolarizing or the state-error noise channel. In contrast, in the third region, between 0.4 and 0.8, the model works much better in the case of a depolarizing channel. When noise increases, the number of projective measurements increases and approaches 256 for basically every test sample, as shown in Figure 3.6b). However, the model can counteract depolarizing noise more effectively because of its deterministic nature. In other words, unlike the case of state-error noise, which is random, the model is able to learn the correlations that characterize the depolarizing noise. Finally, in the last region, where $p = 1$ and $\varepsilon = 1$, the measurement outcomes do not provide sufficient information to complete the tasks successfully. The generalization capability of our model when trained and tested on different levels of noise depends on how different they are. If they are similar, then the performance is expected to be slightly modified, while if they are very different, then the model is likely not to generalize well. As an example, our model trained on the dataset with depolarizing noise $p = 0.4$ achieves an average fidelity $\overline{\mathcal{F}} = 0.80(13)$ when tested on data with depolarizing noise $p = 0.1$ and $\overline{\mathcal{F}} = 0.94(5)$ when tested on data with $p = 0.2$.

As we did in the 2-qubit case, we analyzed the performance of the model as a function of the number of zeros on the diagonal of the density matrix and as a function of the noise strength. We separately computed the average fidelity for states with a small (0-4), intermediate (5-9), and large (10-14) number of zeros on the diagonal. The average fidelity was 0.87, 0.87, and 0.91, respectively. We report these results in Figure 3.8. The data show that in many cases the combined model better reconstructs states that are more sparse and agrees with more general results on the tQST protocol.

Finally, we highlight a key advantage that emerges in the presence of noise. Our model can reconstruct the noiseless state from noisy measurement outcomes, whereas MLE can only retrieve the noisy state. As a result, MLE performance degrades as noise increases, while our model exhibits robustness in tQST, as explained above. For example, if we consider depolarizing noise with $p = 0.4$, our model achieves an average fidelity of 0.96(3), while MLE obtains 0.83(5). Furthermore, the inference times of our model after training and MLE are comparable, about a few seconds.

3.6 Final remarks

In this chapter, we introduced some fundamental machine learning concepts and analyzed the perceptron, a model of artificial neuron. This allowed us to study some representative members of the ecosystem of neural networks, from the multi-layer perceptron to CNNs and Boltzmann machines. We then reviewed the literature on AI models in the context of quantum state characterization and QST.

Motivated by these considerations, we showed the feasibility of embedding permutation-equivariant deep learning models into the tQST protocol. We tested our model on the state reconstruction task for 2- and 4-qubit datasets, where it successfully reconstructed density matrices demonstrating robustness against depolarizing and state-error imperfections. Our proposed model can compete with MLE, achieving compatible results in many cases. Then, we extended tQST to quantum state characterization beyond QST and

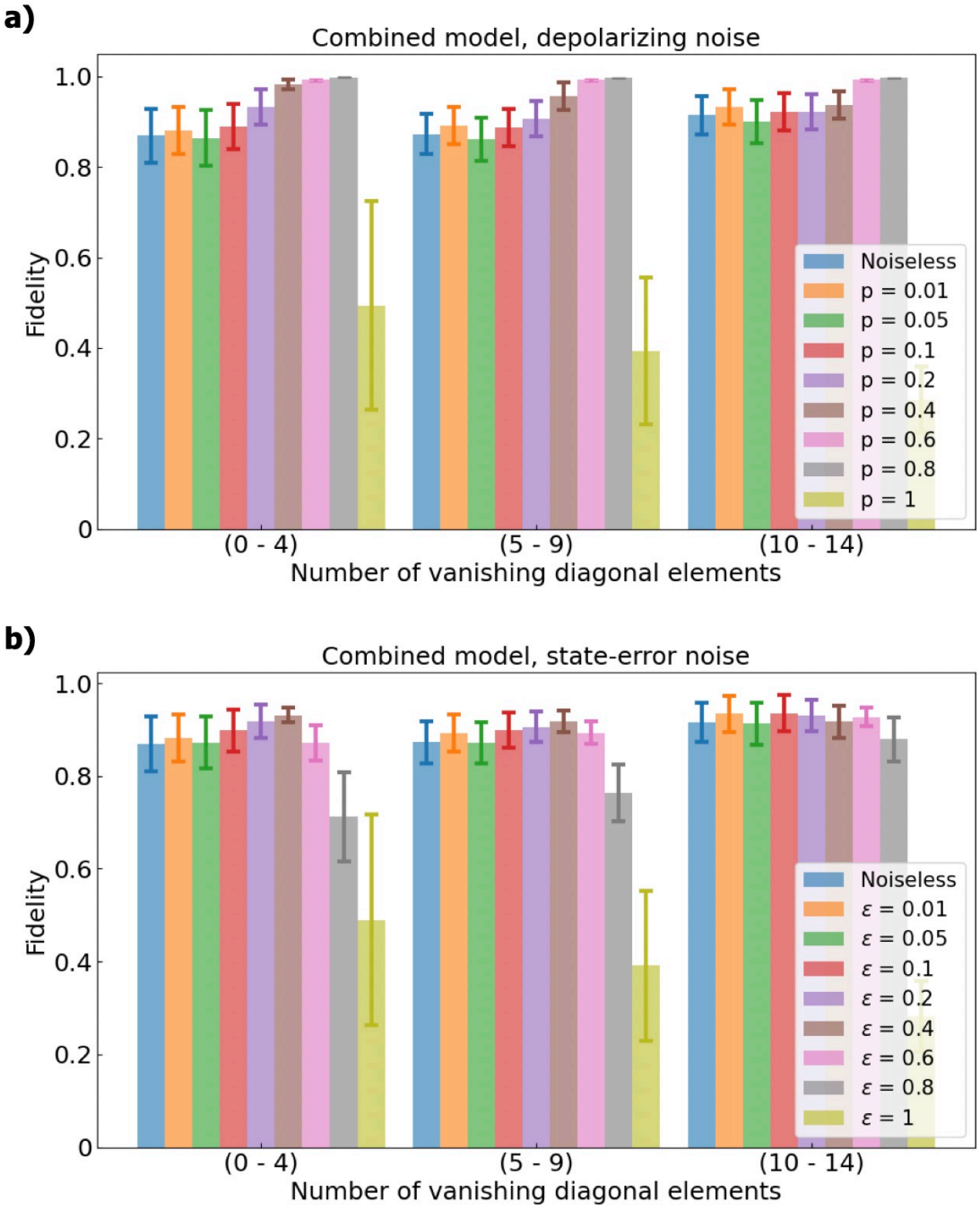


Figure 3.8: **Performance of the 4-qubit model on states with different numbers of vanishing diagonal elements.** We report the average fidelity of the 4-qubit combined model on states with different number of vanishing diagonal elements and different values of the noisy channels applied. We show the results for the depolarizing **a)** and the state-error **b)** channel. The height of each bar represents the average value, while error bars represent the standard deviation from the average.

tested our physics-inspired models to estimate the state purity from the measurement outcomes according to tQST without reconstructing the density matrix. This is a valuable result of our work for at least two reasons. First, we showed that the tQST protocol provides meaningful measurements not only to reconstruct the density matrix but also to estimate the state purity. Second, having an approach to directly estimate the purity can avoid the computational cost of MLE to reconstruct the density matrix for a large number of qubits. In both tasks, our model could manage noisy measurement outcomes to reconstruct the noiseless quantum state or predict the noiseless state purity, ensuring both robustness and accuracy despite experimental imperfections.

Although our machine learning technique does not yet outperform MLE, the results are promising enough to warrant further exploration in this area. The findings imply that incorporating physics-based architectures alongside enhanced computational methods can significantly lower resource demands and enhance the accuracy of quantum state reconstruction. This progress not only offers a crucial tool for existing quantum technology applications, but also establishes a foundation for delving into larger and more complex quantum systems. Further research is necessary to optimize these models, determine ideal thresholds, discover more efficient methods for processing input data, and extend their application to many-qubit systems, preparing for the needs of practical quantum computing and communication technologies.

Chapter 4

Enhanced Compressive Threshold Quantum State Tomography (ECT-QST)

In this chapter, we present a generalization of tQST to systems composed of qudits and measurement settings inspired by compressed sensing, called Enhanced Compressive Threshold Quantum State Tomography (ECT-QST). First we illustrate the protocol and its most relevant features, then analyze the results of the numerical simulations, and show the experimental results on superconducting and photonic platforms. The main reference for this chapter is [2].

4.1 Observables of a single qudit

In the following, we consider systems composed of n_q qu$d$its, that is, n_q d-level carriers of information. The dimension of the Hilbert space associated with these systems is $D = d^{n_q}$. We assume that for each qudit, we can measure the expectation value of the projectors onto the eigenvectors of the $d^2 - 1$ operators $\sigma^{(k)}$, corresponding to the generators of SU(d). Among these operators, $d - 1$ span the Maximal Abelian Subgroup (MAS) and all have the same eigenvectors, which define the computational basis. Thus, we only need to retain a single representative element from the MAS, or, without loss of generality, replace it with the d-dimensional identity $\mathbb{I}_d$, for it has the same eigenvectors as the generators of the MAS. This flexibility does not affect the quality of the reconstruction. In either case, we will denote the MAS representative as $\sigma^{(0)}$. The remaining $d(d-1)$ operators are Hermitian operators whose representation in the computational basis has only two non-vanishing off-diagonal elements. Half of these matrices are real, half are purely imaginary. They can be indexed by the position of their non-zero upper triangular element, starting with the real-valued matrices followed by the imaginary ones.

As an example, if we consider qubits, where $d = 2$, then there are $d^2 - 1 = 2^2 - 1 = 3$ single-qubit observables, which correspond to the usual SU(2) Pauli matrices: $\sigma^{(0)} = \sigma_z$ (the MAS generator), $\sigma^{(1)} = \sigma_x$ and $\sigma^{(2)} = \sigma_y$. If we consider qutrits, where $d = 3$, then there are $d^2 - 1 = 3^2 - 1 = 8$ single-qutrit observables, corresponding to the 8 Gell-Mann matrices, which are the SU(3) analogous of the Pauli matrices. The real observables are

the following:

$$\sigma^{(1)} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \sigma^{(2)} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \sigma^{(3)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \tag{4.1}$$

while the purely imaginary observables are:

$$\sigma^{(4)} = \begin{pmatrix} 0 & i & 0 \\ -i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \sigma^{(5)} = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \qquad \sigma^{(6)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & i \\ 0 & -i & 0 \end{pmatrix}. \tag{4.2}$$

Finally, SU(3) MAS comprises the following two generators:

$$\sigma^{(7)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \sigma^{(8)} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \tag{4.3}$$

As explained above, in our protocol we choose the d-dimensional identity $\mathbb{I}_d$ as the representative element of the MAS defining the computational basis, since it has the same eigenvectors as the generators of the MAS.

4.2 How to construct the set of measurement settings

4.2.1 Density matrix elements and measurement settings

The proposed ECT-QST protocol starts with the measurement of the D diagonal elements $\{\rho_{ii}\}_{i=1}^{D}$ of the density matrix. We can do this by measuring the setting $s^{(00\cdots0)} = \otimes_{r=1}^{n_q} \sigma^{(0)}$, which corresponds to measuring the expectation value of the projectors onto the D states constructed as tensor products of the eigenvectors of $\sigma^{(0)}$, that is, the computational basis. After this first measurement, we choose a suitable threshold t to select the matrix elements ρ_{ij} for which $\mathfrak{r}_{ij} = \sqrt{\rho_{ii}\rho_{jj}} \geq t$. As explained in Section 2.3, the choice of threshold is a crucial step of our protocol, for it determines the amount of resources required to complete the tomography experiment. Ideally, one would like to choose a threshold that is small enough to include the most relevant elements of the density matrix, yet large enough to exclude those elements that are most likely affected by noise and do not provide meaningful information about the quantum state being measured. However, the optimal value of t depends on several factors, such as the specific physical system used to implement the qubits, which in turn determines the noise level, the amount of available resources, or the target quantum state to be generated.

Once we have chosen the threshold, we look for a set S_t of measurement settings of the form $s^{(K)} = s^{(k_1 k_2 \cdots k_{n_q})} = \otimes_{r=1}^{n_q} \sigma^{(k_r)}$ that provides as much information as possible on the density matrix elements selected by the threshold. Here, $K = \left(k_1, k_2, \ldots, k_{n_q}\right)$ is a multi-index where each index k_r ranges from 0 to $d(d-1)$ and describes which observable is measured on the r-th qudit. Since the expectation value of $s^{(K)}$ on ρ is given by:

$$\mathrm{Tr}\left(s^{(K)}\rho\right) = \sum_i s_{ii}^{(K)} \rho_{ii} + 2 \sum_{i<j} \left[\mathrm{Re}\left(s_{ij}^{(K)}\right) \, \mathrm{Re}\left(\rho_{ij}\right) + \mathrm{Im}\left(s_{ij}^{(K)}\right) \, \mathrm{Im}\left(\rho_{ij}\right) \right], \tag{4.4}$$

a sufficient condition for $s^{(K)}$ to provide information on the real (imaginary) part of ρ_{ij} is that the real (imaginary) part of $s_{ij}^{(K)}$ is non-zero. From the definition of $s^{(K)}$ one has:

$$s_{ij}^{(K)} = s^{(K)}_{i_1^d i_2^d \cdots i_{n_q}^d, j_1^d j_2^d \cdots j_{n_q}^d} = \sigma^{(k_1)}_{i_1^d, j_1^d} \sigma^{(k_2)}_{i_2^d, j_2^d} \cdots \sigma^{(k_N)}_{i_{n_q}^d, j_{n_q}^d}, \tag{4.5}$$

with i_r^d the r-th digit of the base-d representation of i (and similarly for j_r^d). Equation (4.5) shows that *all* the matrix elements of the single qudit observables on the right-hand side of (4.5) must be different from zero so that $s_{ij}^{(K)}$ is different from zero.

We now need to find a setting that provides information about the real part of the density matrix element ρ_{ij}. To this end, we digit-wise compare the base-d representations of i and j and choose: $\sigma^{(0)}$ if $i_r^d = j_r^d$, and the unique real generator with a non-zero element at (i_r^d, j_r^d), if $i_r^d \neq j_r^d$. It is then sufficient to replace an odd number of single qudit observables where $i_r^d \neq j_r^d$ with the corresponding imaginary ones to construct a setting that provides information about the imaginary part of the same density matrix element. We choose to change only one single qudit observable, the one for which $i_r \neq j_r$ for the smallest value of r. Thus, to gain information on the imaginary part of ρ_{ij}, we replace the first real single qudit operator $\sigma^{(k_r)}$ different from $\sigma^{(0)}$ with the corresponding $\sigma^{(k_r + d(d-1)/2)}$. We repeat this procedure for every off-diagonal element above threshold, and construct the set S_t of *unique* measurement settings by removing duplicates. Thus, if E is the number of off-diagonal elements selected by the threshold, we have $|S_t| \leq 2E$.

As an example, let us consider a system composed of $n_q = 4$ qubits ($d = 2$), and construct the measurement settings associated with the density matrix element $\rho_{4,13}$. We first write the base-2 (binary) representation of i and j, which is composed of 4 digits, obtaining $i = 4 = 0100^2$ and $j = 13 = 1101^2$. Then we bit-wise compare the two binary representations to construct the setting for the real part of the density matrix element: if two bits are equal, take σ_z, while if they are different, take σ_x. In this way, we find $\sigma_x \otimes \sigma_z \otimes \sigma_z \otimes \sigma_x$ as the setting associated with the real part. Finally, if we replace the first occurrence of σ_x with σ_y we construct the setting for the imaginary part, which is $\sigma_y \otimes \sigma_z \otimes \sigma_z \otimes \sigma_x$. If we now consider a system composed of $n_q = 3$ qutrits ($d = 3$), to determine the measurement settings associated with the density matrix element $\rho_{12,23}$, we first calculate the base-3 representations as $i = 12 = 110^3$ and $j = 23 = 212^3$, and then determine the setting $s^{(302)} = \sigma^{(3)} \otimes \sigma^{(0)} \otimes \sigma^{(2)}$ as the one associated with the real part and $s^{(602)} = \sigma^{(6)} \otimes \sigma^{(0)} \otimes \sigma^{(2)}$ the one associated with the imaginary part.

We highlight that the rule just presented does not recover all the 3^{n_q} settings required in QST even in the limit of $t \to 0$. For example, the settings $\sigma_x \otimes \sigma_y$ and $\sigma_y \otimes \sigma_y$ cannot belong to the set S_t by construction in the case of 2 qubits. In Table 4.1 we report the maximum number $(\max|S_t|)$ of settings that our protocol prescribes and the maximum number of settings required by QST, which scales as 3^{n_q} if we consider qubits. The data allowed us to conclude that the maximum amount of experimental resources required by the proposed protocol scales more favorably with the number of qubits than QST. As a final comment, we note that for sparse density matrices with E off-diagonal elements to be determined, our protocol requires $\mathcal{O}(E)$ settings, corresponding to $\mathcal{O}(E\, d^{n_q})$ projective measurements. Consequently, it is highly likely that some settings provide information on multiple matrix elements, making them redundant. In the next section, we show how to further reduce the required number of measurement settings.

n_q	$\max\|S_t\|$	3^{n_q}
2	6	9
3	14	27
4	30	81
5	62	243
6	126	729
7	254	2,187
8	510	6,561
9	1,022	19,683
10	2,046	59,049

Table 4.1: **Maximum number of settings of ECT-QST and QST.** We report the number of qubits, the maximum number $\max|S_t|$ of *unique* measurement settings required by ECT-QST, and the maximum number of settings required by QST if we consider qubits.

4.2.2 Pruning the measurement settings

We now recall the definitions (1.24) and (1.26) of the observables that provide sharp information on the real or imaginary part of a specific off-diagonal element of the density matrix ρ. This will be useful to find the minimum number of settings that still allows one to faithfully reconstruct the state under consideration. Let m be a multi-index in the form $(i, j, \mathcal{R})$ or $(i, j, \mathcal{I})$ for the real or imaginary part of the considered off-diagonal elements, respectively. Let O_m be observables such that their only non-zero elements are (i, j) and (j, i), with their values being both $1/2$ if $m = (i, j, \mathcal{R})$ or $i/2$ and $-i/2$ if $m = (i, j, \mathcal{I})$, respectively. The following relations hold for these observables:

$$\mathrm{Tr}(O_m \rho) = \begin{cases} \mathrm{Re}(\rho_{ij}) & \text{if } m = (i, j, \mathcal{R}) \\ \mathrm{Im}(\rho_{ij}) & \text{if } m = (i, j, \mathcal{I}) \end{cases} \tag{4.6}$$

Each setting $s = s^{(K)}$ is a sequence of n_q observables $\sigma^{(k_r)}$, and corresponds to a set of D projectors $\left\{\Pi_n^{(s)}\right\}_{n=1}^{D}$. These projectors are separable, that is, they project onto states that are the tensor product of eigenstates of the observables $\sigma^{(k_r)}$. More precisely, $\Pi_n^{(s)} = |\phi_n^{(s)}\rangle\langle\phi_n^{(s)}|$, with $|\phi_n^{(s)}\rangle = \otimes_{r=1}^{n_q} |\varphi_{n_r^d}^{(k_r)}\rangle$, $\left|\varphi_{n_r^d}^{(k_r)}\right\rangle$ eigenstate of $\sigma^{(k_r)}$, and n_r^d the rth digit of the base-d representation of n. We construct the rank-3 tensor $\boldsymbol{A}$ with shape $(2E,\ D,\ |S_t|)$ and elements defined as:

$$A_{mn}^{(s)} = \langle\phi_n^{(s)}|O_m|\phi_n^{(s)}\rangle = \mathrm{Tr}\left(O_m \Pi_n^{(s)}\right). \tag{4.7}$$

The element $A_{mn}^{(s)}$ can be interpreted as the overlap between the projector $\Pi_n^{(s)}$ of the setting s and the observable returning the real or imaginary part of the matrix element contained in m. In this respect, it quantifies how much information the measurement of the projector $\Pi_n^{(s)}$ provides about the real or imaginary part of the matrix element of m. If $A_{mn}^{(s)} = 0$, then measuring the expectation value of $\Pi_n^{(s)}$ does not provide any information on the real or imaginary part of the matrix element of m. Then we construct a matrix C with shape

$(|S_t|, 2E)$ and elements defined as:

$$C_{sm} = \sum_n \left|A_{mn}^{(s)}\right|^2 \quad \forall\, s \in S_t. \tag{4.8}$$

The matrix element C_{sm} quantifies how much information one can retrieve about the matrix element of m by measuring the projectors of the setting s, that is, by measuring the setting s. Thus, C_{sm} represents the overlap between the setting s and the matrix element m. Consider then the following vector β, whose $2E$ components are defined as the maximum overlap on the matrix element of m among the settings $s \in S_t$:

$$\beta_m = \max_s C_{sm}. \tag{4.9}$$

This maximum surely exists, but may not correspond to a unique setting s. In general, $\sum_s C_{sm} \gg \beta_m$, that is, several projectors $\Pi_n^{(s)}$ in different settings $s \in S_t$ provide information on the same matrix element m. Thus, we look for a subset $S_t' \subset S_t$ such that $\sum_{s\in S_t'} C_{sm} \approx \beta_m$ in order to reduce the number of settings while keeping the total overlap with the matrix element m as large as the maximum overlap among the settings identified in S_t.

The simplest way to find such a subset is through a greedy algorithm, which attempts to find a globally optimal solution by making locally optimal choices. We start by selecting the setting s_1 such that $C_{s_1 m}$ has the fewest elements equal to zero, add it to S_t' and remove it from S_t. This ensures that s_1 provides information on the maximum number of matrix elements of interest. Then we find the setting $s_2 \in S_t$ with the next fewest zeros in $C_{s_2 m}$, excluding the columns m' for which the target value has already been achieved, that is, the columns m' such that $\sum_{s'\in S_t'} C_{s'm'} \geq \beta_{m'}$. We continue this process until $\sum_{s'\in S_t'} C_{s'm} \geq \beta_m \ \forall m$. At the end of this algorithm we have a smaller set S_t' of measurement settings whose elements have an overlap with the matrix element m ($\sum_{s'\in S_t'} C_{s'm}$) as large as the maximum overlap among the original settings (β_m). We also note that the pruning operation might or might not reduce the number of settings to be measured, depending on the specific state under consideration, as shown in the examples of Section 4.5.1. In the former case $|S_t'| < |S_t|$, while in the latter case $|S_t'| = |S_t|$.

Finally, the settings in S_t' are not equally important for the reconstruction of ρ. Thus, we assign a weight $w_s = \sum_m C_{sm}\mathfrak{r}_m$ to each setting, reflecting the total overlap of s on the selected density matrix elements. Finally, we sort the settings in descending order according to their weight. This allows one to progressively measure the settings, prioritizing the most important ones.

4.3 State reconstruction

We reconstruct the density matrix ρ using a maximum likelihood approach starting from the measurement of the settings in S_t'. The reconstruction process requires finding the density matrix ρ that minimizes the following (negative log-) likelihood function:

$$L(\rho) = \sum_{s,n} \left(\frac{\mathcal{N}_n^{(s)} - N_n^{(s)}}{2\sqrt{\mathcal{N}_n^{(s)}}} \right)^2, \tag{4.10}$$

with $\mathcal{N}_n^{(s)} = \langle \phi_n^{(s)} | \rho | \phi_n^{(s)} \rangle$ the expected (theoretical) measurement outcomes, and $N_n^{(s)}$ the measured (experimental) outcomes. In (4.10), the index n takes D values for n_q qudits, leading to an exponential increase in the number of terms that contribute to the likelihood.

We need a model, or parametrization, of the density matrix ρ to carry out the minimization. As explained in Section 2.5, the most general and unbiased model is based on the Cholesky decomposition and writes the density matrix as $\rho = T^\dagger T$, with T a lower or upper triangular matrix of size $D \times D$. However, as n_q increases, the evaluation of the likelihood and its gradient for minimization becomes computationally demanding. Furthermore, we experimented only with pure states on IBMQ. For these reasons, we found it more efficient to use the g5 parametrization of ρ, that is, $\rho = R^\dagger R$, with R a $n_{\text{vec}} \times D$ matrix, and n_{vec} is kept relatively small. In our numerical experiments, we fixed $n_{\text{vec}} = n_q$. To verify the soundness of this approximation, we verify that the number of eigenvectors n_{vec} used to approximate the state is greater than the rank, estimated as the inverse of the purity of the reconstructed state ρ_{rec}, that is, $n_{\text{vec}} > 1/\text{Tr}\left(\rho_{\text{rec}}^2\right)$. If this condition is not satisfied, we increase n_{vec} and repeat the minimization.

4.4 Connection with tQST

The ECT-QST protocol can be viewed as an extension of the original tQST approach [1], tailored for multiplexing platforms where one performs projective measurements on the eigenbasis of a given setting. This is the case for the two platforms that we used in our experiments. The connection between the tQST and ECT algorithms arises considering the rank-3 tensor $\boldsymbol{A}$ defined in (4.7) and noting that the projector $\Pi_n^{(s)}$ that provides the most information about the matrix element m is the one with the maximum overlap, that is, it is the projector identified by the index $n_{\max}$ given by:

$$n_{\max} = \text{argmax}\left(|A_{mn}^{(s)}|^2\right). \tag{4.11}$$

In this way, we can recover the original tQST protocol in two steps. First, we identify the setting s that provides the most information about the matrix element m of interest, and construct the matrix corresponding to that specific setting from (4.7). Second, we determine the most relevant projector from (4.11). In general, there may not be a unique solution to (4.11), which means that multiple projectors $\Pi_n^{(s)}$ may have the same maximum overlap. Here, we select the smallest value of n that satisfies (4.11).

These points clarify the relationship between tQST and ECT-QST. The former focuses on reducing the number of projective measurements, while the latter focuses on reducing the number of measurement settings. tQST is more suitable for platforms where one can perform only a single projective measurement at once, whereas ECT-QST is tailored for multiplexed platforms where one can simultaneously perform several projective measurements. As a final comment, we note that the number of projective measurements used in ECT-QST is $\mathcal{O}(d^{n_q})$ times greater than in tQST, introducing a significant amount of redundancy. However, the experimental results show that this redundancy makes ECT-QST more robust in noisy environments compared to tQST.

Before showing some examples of application of ECT-QST, we highlight that an implementation of ECT-QST using the Python language is available online on Github [402].

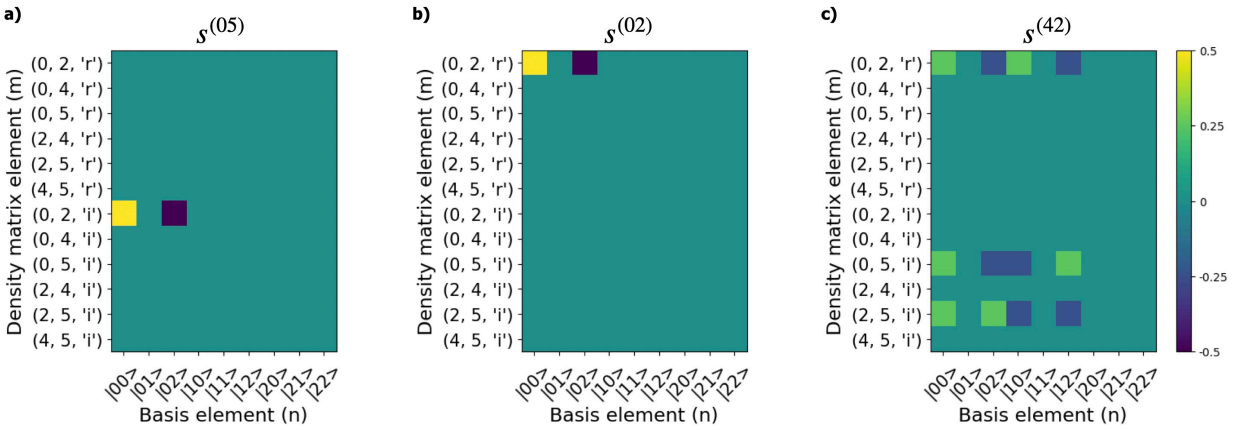


Figure 4.1: **Three components of the rank-3 tensor $\boldsymbol{A}$ of the state** (4.12)**.** We show the first three components **a)** $A^{(s^{(05)})}_{mn}$, **b)** $A^{(s^{(02)})}_{mn}$, and **c)** $A^{(s^{(42)})}_{mn}$ of the rank-3 tensor $\boldsymbol{A}$ of the state (4.12), corresponding to the first three settings of (4.13). The name of the setting is also reported in the title of each panel. We show only the real part since the imaginary part of all components is vanishing.

4.5 Examples of ECT-QST

4.5.1 Examples of the pruning procedure

We now illustrate in detail how the pruning procedure works for two different states. Let us start by considering a system composed of two qutrits $(n_q = 2, d = 3)$, whose Hilbert space has dimension $D = 9$, in the following state:

$$|\Psi\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{3}}|02\rangle + \frac{1}{\sqrt{12}}|11\rangle + \frac{i}{\sqrt{12}}|12\rangle\,. \tag{4.12}$$

We first retrieve the diagonal elements of the density matrix $\rho_\Psi = |\Psi\rangle\langle\Psi|$. The only non-vanishing elements are the ones indexed by 0, 2, 4, and 5, with values $1/2$, $1/3$, $1/12$, and $1/12$, respectively. By setting the threshold $t = 1/12$ the ECT-QST protocol identifies $E = 6$ off-diagonal elements above threshold and requires the following $|S_t| = 12$ unique settings:

$$\left\{s^{(05)}, s^{(02)}, s^{(42)}, s^{(11)}, s^{(41)}, s^{(12)}, s^{(13)}, s^{(10)}, s^{(43)}, s^{(40)}, s^{(06)}, s^{(03)}\right\}. \tag{4.13}$$

We recall that the single qutrit observables used in the tQST protocol are the ones of (4.1), (4.2), and (4.3), and the notation $s^{(05)} = \sigma^{(0)} \otimes \sigma^{(5)}$ means that we measure $\sigma^{(0)}$ on the first qubit and $\sigma^{(5)}$ on the second qubit. Equivalently, it means that we measure the expectation value of the projectors onto the eigenbasis of $\sigma^{(0)} \otimes \sigma^{(5)}$.

We now construct the rank-3 tensor $\boldsymbol{A}$ with shape $(2E, D, |S_t|) = (12, 9, 12)$ according to (4.7). In Figure 4.1 we show the components of $\boldsymbol{A}$ for the first three settings of (4.13), and report all the components in Figure C.1 of Appendix C.1. Then we compute the

matrix C according to (4.8) and obtain:

$$C = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0 & 0.25 & 0 \\ 0 & 0.25 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0 & 0 & 0.5 & 0 \\ 0.25 & 0 & 0.25 & 0 & 0.25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.25 & 0 & 0 & 0 & 0.25 & 0.25 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{4.14}$$

Finally, we compute the vector β from C according to (4.9):

$$\beta = \begin{pmatrix} 0.5 & 0.25 & 0.25 & 0.25 & 0.5 & 0.5 & 0.5 & 0.25 & 0.25 & 0.25 & 0.5 & 0.5 \end{pmatrix}. \tag{4.15}$$

The pruning procedure tells us that we can remove the settings $s^{(10)}$ and $s^{(40)}$, corresponding to the eighth and tenth row of C. The reason behind this result is the following. Let C' be the matrix obtained from C by removing the eighth and tenth rows, and let β' be the vector computed according to (4.9) from C'. We find that β' is component-wise greater than or equal to β, which is precisely the condition described above for pruning to occur. Thus, those two rows can be removed. If we remove any other different pair of rows or more rows than the two specified, we always find that at least one of the components of β' is smaller than the corresponding component in β. Thus, no more rows can be removed and the pruning procedure stops here, resulting in $|S'_t| = 10$.

Let us now consider again a system composed of two qutrits in the following state:

$$|\Phi\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{3}}|02\rangle + \frac{1}{\sqrt{12}}|10\rangle + \frac{i}{\sqrt{12}}|12\rangle. \tag{4.16}$$

The only non-vanishing diagonal elements of the corresponding density matrix $\rho_\Phi = |\Phi\rangle\langle\Phi|$ are those indexed by 0, 2, 3, and 5, with values 1/2, 1/3, 1/12, and 1/12, respectively. By setting $t = 1/12$ we find again $E = 6$ off-diagonal elements above threshold, as in the previous case. However, now we find only $|S_t| = 6$ unique settings to be measured:

$$\left\{ s^{(05)}, s^{(02)}, s^{(10)}, s^{(42)}, s^{(40)}, s^{(12)} \right\}. \tag{4.17}$$

The rank-3 tensor $\boldsymbol{A}$ now has shape $(2E, D, |S_t|) = (12, 9, 6)$, and we show the components corresponding to the first three settings of (4.17) in Figure 4.2, and show all the components in Figure C.2 of Appendix C.1. The matrix C in this case is:

$$C = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.25 & 0 & 0 & 0 & 0 & 0.25 & 0 & 0.25 & 0.25 & 0.25 & 0.25 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0.5 & 0 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{4.18}$$

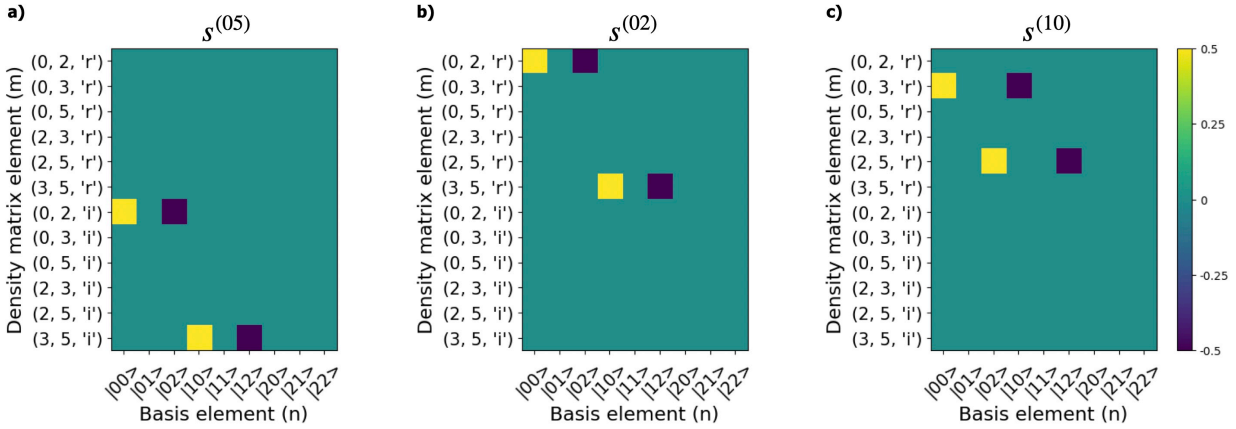


Figure 4.2: **Three components of the rank-3 tensor $\boldsymbol{A}$ of the state** (4.16)**.** We show the first three components **a)** $A^{(s^{(05)})}_{mn}$, **b)** $A^{(s^{(02)})}_{mn}$, and **c)** $A^{(s^{(10)})}_{mn}$ of the rank-3 tensor $\boldsymbol{A}$ of the state (4.16), corresponding to the first three settings of (4.17). The name of the setting is also reported in the title of each panel. We show only the real part for all the elements of the imaginary part are vanishing.

The corresponding β vector is:

$$\beta = \begin{pmatrix} 0.5 & 0.5 & 0.25 & 0.25 & 0.5 & 0.5 & 0.5 & 0.5 & 0.25 & 0.25 & 0.5 & 0.5 \end{pmatrix}. \tag{4.19}$$

For this state, the pruning procedure tells us that we cannot remove any setting from the 6 identified above. The reason is that if we construct C' by removing any row from C, and compute β' from C', we always find that at least one of the components of β' is smaller than the corresponding component β. Thus, no pruning is allowed, and $|S'_t| = |S_t| = 6$.

4.5.2 Complete examples of ECT-QST

We now show the ECT-QST protocol applied to the same states analyzed in Section 2.6. Let us start from the random mixed state (2.29). The diagonal elements, the Gini threshold, and the off-diagonal elements selected by ECT-QST are the same as of tQST, for the two protocols share these first steps. However, now ECT-QST requires the following $|S_t| = 14$ *unique* measurement settings:

$$\begin{aligned} &\{ZZX, ZXZ, ZXX, XZZ, XZX, XXZ, XXX, \\ &ZZY, ZYZ, ZYX, YZZ, YZX, YXZ, YXX\}. \end{aligned} \tag{4.20}$$

We highlight that we quote the settings and we do not report the corresponding off-diagonal elements, for a single setting typically extracts information about many off-diagonal elements. The pruning procedure reduced the number of settings to $|S'_t| = 9$, for a total of 10 measurement settings, including the first to determine the diagonal elements. Measurement of these settings corresponds to measuring the expectation value of 80 projectors, given by $(|S'_t| + 1) \times D = 10 \times 8 = 80$. The fidelity between the reconstructed and the target state is 0.9923.

If we now consider the random pure state (2.31), the $|S_t| = 10$ unique settings prescribed by ECT-QST are:

$$\begin{aligned}&\{ZZX, ZXX, ZXZ, XXX, XXZ,\\ &ZZY, ZYX, ZYZ, YXX, YXZ\}.\end{aligned} \tag{4.21}$$

The number of settings after pruning is $|S'_t| = 8$, for a total of 9. The number of measured projectors is 72, and the fidelity between the reconstructed state and the target state is 0.99987.

The ECT-QST protocol applied to reconstruct a $|W_3\rangle$ state requires the following $|S_t| = n_q\,(n_q - 1) = 6$ measurement settings:

$$\begin{aligned}&\{ZXX, XZX, XXZ\\ &ZYX, YZX, YXZ\}.\end{aligned} \tag{4.22}$$

The pruning operation does not further reduce the number of settings, therefore $|S'_t| = |S_t|$, and the corresponding number of measured projectors is 56. The fidelity between the target and the reconstructed state is 0.99999.

Finally, GHZ states require only $|S_t| = 3$ measurement settings, regardless of the number of qubits. If $n_q = 3$, the settings related to the off-diagonal elements are $\{XXX, YXX\}$ and cannot be further reduced by pruning. The number of measured projectors is 24, and the fidelity between the target and the reconstructed state is 0.99999.

4.6 Numerical analysis

After we have illustrated step by step the ECT-QST protocol for some specific states, we now conduct a more general analysis of the method by performing numerical simulations to analyze its efficacy. We first examine some relevant pure states, such as GHZ and W states, realized via proper quantum circuits. We chose the threshold t as the value of the smallest non-vanishing diagonal element of the target state density matrix.

A GHZ state of n_q qu*d*its has $d(d-1)/2$ relevant off-diagonal elements, resulting in $d(d-1) \sim \mathcal{O}(d^2)$ settings to be measured regardless of n_q. In the case of n_q qubits, only 3 settings are required to reconstruct the corresponding GHZ state. In addition to the initial setting $s^{(00\ldots0)}$ to determine the diagonal elements, the set S_t includes only two further settings, $s^{(11\cdots1)} = \sigma_x \otimes \sigma_x \otimes \cdots \otimes \sigma_x$ associated with the real part of the element at $i = 0 = \underbrace{00\cdots0^2}_{n_q \text{ times}}$ and $j = 2^{n_q} - 1 = \underbrace{11\cdots1^2}_{n_q \text{ times}}$, and $s^{(21\cdots1)} = \sigma_y \otimes \sigma_x \otimes \cdots \otimes \sigma_x$ associated with the imaginary part of the same element. Without noise, the ECT-QST protocol reconstructs the considered GHZ states with fidelity equal to 1.

The ECT-QST protocol identifies $n_q(n_q-1) \sim \mathcal{O}(n_q^2)$ settings to reconstruct an n_q-qu*d*it W state, regardless of the value of d. In the case of n_q qubits, in Figure 4.3a) and b) we show: the fidelity $\mathcal{F}^{\rho_\ell}_{\rho_{\text{target}}}$ between the ECT-QST reconstruction ρ_ℓ with ℓ settings and the target density matrix ρ_{target}; the fidelity $\mathcal{F}^{\rho_\ell}_{\rho_{\ell-1}}$ between the ECT-QST reconstruction with ℓ settings and the ECT-QST reconstruction with $\ell-1$ settings, respectively. In the latter case, we progressively measure the settings by decreasing weights as explained in Section 4.2.2. We obtained fidelities $\mathcal{F}^{\rho_{\ell_{n_q}}}_{\rho_{\text{target}}} > 0.999$ with a number of settings equal to $\ell_{n_q} = 5, 7, 5, 10$ for $n_q = 4, 5, 6, 7$, respectively, that is, with much fewer settings than those in S'_t.

Then we considered pure random states generated by 7-qubit depth-3 random circuits. For each number of qubits, we generated 100 depth-3 random circuits and extracted some representatives such that the corresponding output states have different numbers of non-vanishing diagonal elements. We selected 8 circuits of 4, 5, and 6 qubits, and 13 circuits of 7 qubits. In Figure 4.3c) and d) we show the fidelities $\mathcal{F}^{\rho_\ell}_{\rho_{\text{target}}}$ and $\mathcal{F}^{\rho_\ell}_{\rho_{\ell-1}}$ for the random states under study, whose number of non-vanishing diagonal elements is between 16 and 96, and the number of settings identified by the ECT-QST protocol is from 17 to 67. We achieved an average fidelity $\overline{\mathcal{F}}^{\rho_\ell}_{\rho_{\text{target}}} > 0.999$ with $\ell_7 = 15(7)$ settings. The fast saturation of the fidelity observed for the W and random states is driven by the progressive measurement of the settings according to their weights, demonstrating that low-weight settings minimally affect the quality of the maximum likelihood reconstruction.

We then considered a more general situation and validated the ECT-QST protocol to reconstruct random pure and mixed states. We used the same datasets of density matrices as in Section 2.7 for all the following analyses and analyzed whether there was a relation between the number of measurement settings required by ECT-QST, the purity, and the sparsity (quantified by the Gini threshold) of the density matrices under consideration. From the plots in Figure 4.4 we conclude that the number of measurement settings required by the protocol is inversely proportional to the sparsity of the states, regardless of purity, that is, for a fixed value of the purity, the sparser the state, the fewer settings required. This is consistent with the core idea of the proposed threshold protocols and with the results of Section 2.7.

We then reconstructed each of the 2,000 2-, 3-, and 4-qubit density matrices using ECT-QST with the Gini threshold and triangular parametrization. In Figure 4.5a), d), and g) we show the distribution of the number of measurement settings suggested by ECT-QST. This data confirms that the proposed protocol allows one to actually perform fewer settings than QST. The protocol requires its maximum number of settings for about half of the states of each dataset. We believe that this is due to the choice of the threshold. In fact, the Gini threshold can be a conservative choice, as discussed in Section 4.7, that is, it is likely that there exists a higher value of the threshold that requires fewer measurements than the Gini threshold without compromising the quality of the reconstruction. Furthermore, we recall that the maximum number of settings required by ECT-QST is well below the 3^{n_q} settings of QST, as reported in Table 4.1. The average number of settings was $5(1)$, $11(4)$, and $24(9)$ for 2, 3, and 4 qubits, respectively. The uncertainty is the standard deviation from the average. As mentioned in Section 2.7, the distribution of the number of settings certifies that we are implementing a protocol different from QST, but is not enough to prove that the proposed protocol works. To this end, we need to show that we can reconstruct the states with high fidelity using the settings provided by the protocol. Thus, in Figure 4.5b), e), and h) we show the fidelity between the reconstructed and the target state with respect to the number of measurements, while in Figure 4.5c), f), and i) we show a zoomed-in view of the same plots, together with the average fidelity for each number of measurement settings. The data allowed us to conclude that we can reconstruct the states with high fidelity using the settings suggested by ECT-QST, the number of which is much lower than 3^{n_q}. In particular, we achieved average fidelities of $0.992(26)$, $0.989(29)$, and $0.985(36)$ for 2-, 3-, and 4-qubit states, respectively. For comparison, QST achieved average fidelities of $0.991(27)$, $0.989(29)$, and $0.984(38)$ for the same number of qubits, which are compatible within the errors with the results of ECT-QST.

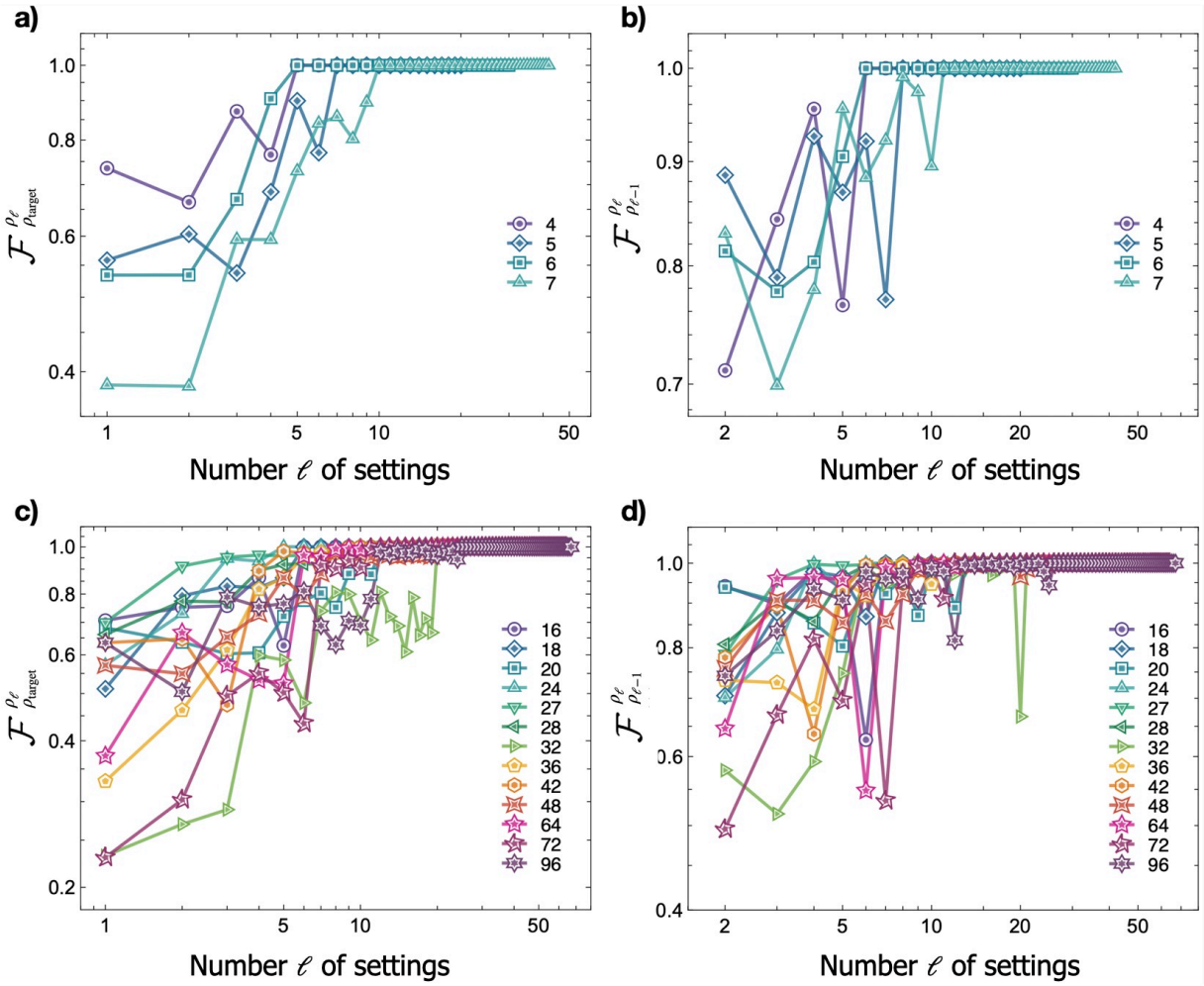


Figure 4.3: **Reconstruction of noiseless W and random states generated with noiseless quantum circuits. a)**: Fidelity $\mathcal{F}^{\rho_\ell}_{\rho_{\text{target}}}$ between the target density matrix ρ_{target} of a n_q-qubit W state and the ECT-QST reconstruction ρ_ℓ with ℓ settings as the identified settings $s \in S'_t$ are progressively measured. **b)**: Fidelity $\mathcal{F}^{\rho_\ell}_{\rho_{\ell-1}}$ between ECT-QST reconstructions with ℓ and $\ell - 1$ settings for n_q-qubit W states. In panels **a)** and **b)** each curve represents a different number of qubits, from 4 to 7. **c), d)**: Same fidelities as in **a), b)** for noiseless 7-qubit depth-3 random circuits. In panels **c)** and **d)**, each curve represents a random state with a different number of non-vanishing diagonal elements, from 16 to 96, whose symbol is reported in the legend. Figure adapted from [2].

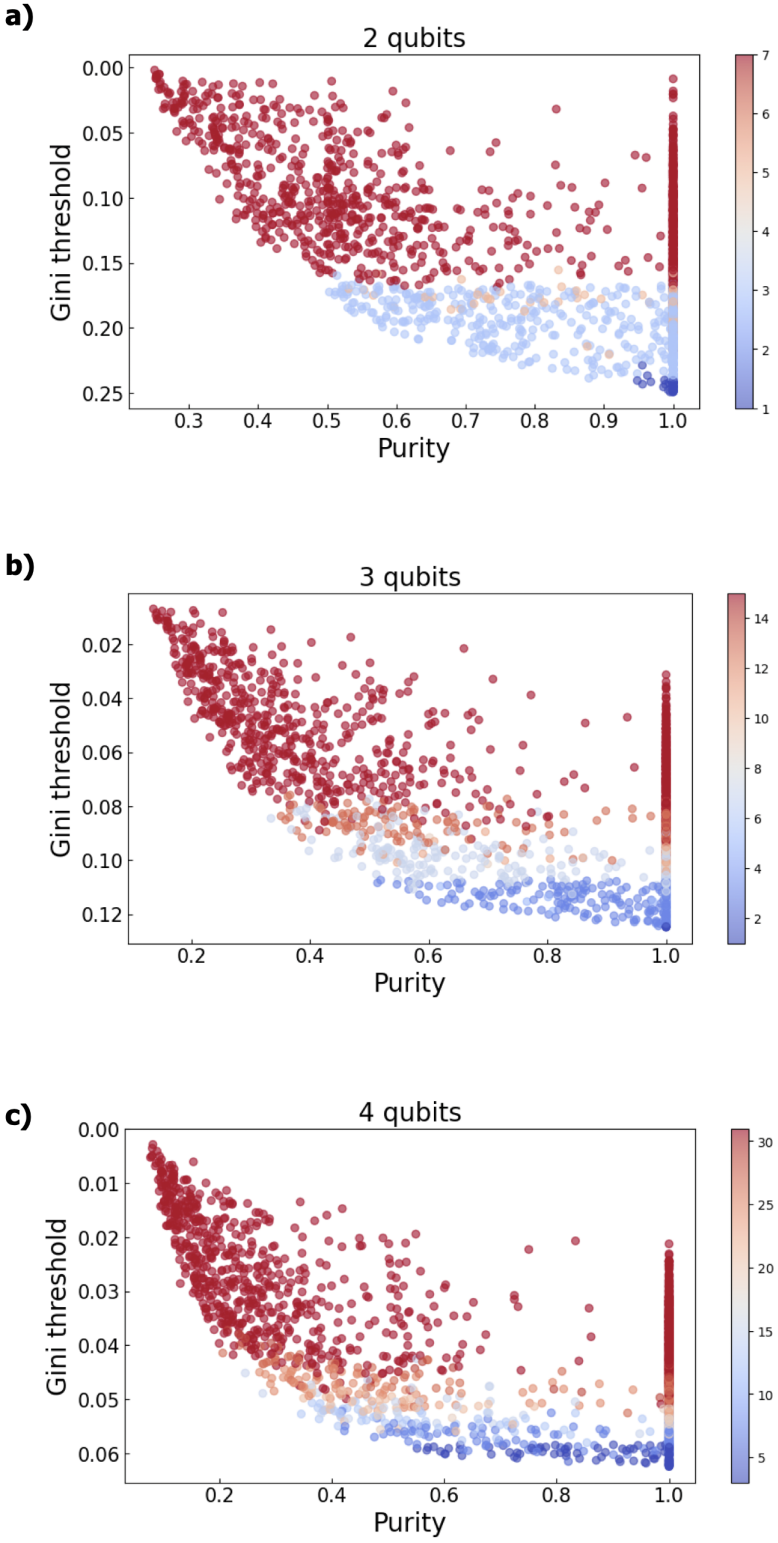


Figure 4.4: **Relation between number of settings, purity, and sparsity.** Correlation plot between the sparsity (Gini threshold), purity, and number of measurement settings of eahc state in the dataset of **a)** 2, **b)** 3, and **c)** 4 qubits. The number of settings is encoded in the colorbar. Each point represents a different state of the dataset. Note the inverted vertical axis.

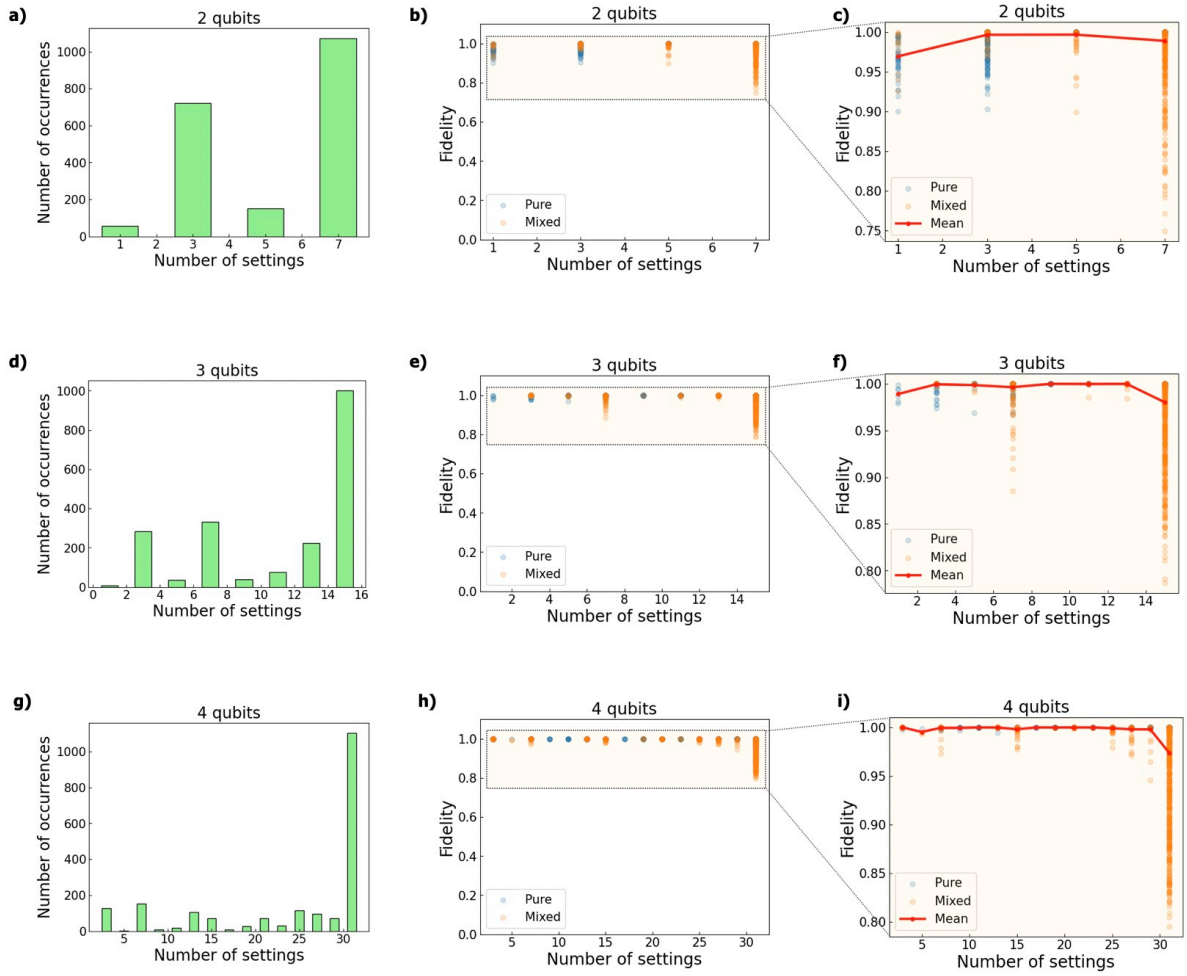


Figure 4.5: **Numerical validation of ECT-QST. a), d), g)**: Distribution of the number of measurement settings used to reconstruct the states for 2, 3, and 4 qubits, respectively. **b), e), h)**: Fidelity between the reconstructed and the target state with respect to the number of settings for 2, 3, and 4 qubits, respectively. **c), f), i)**: Detail of the vertical axis of the plot of the fidelity with respect to the number of measurements. The encoding of the color for pure and mixed states is reported in the legend. We show in red the trend of the average computed for each number of measurement settings.

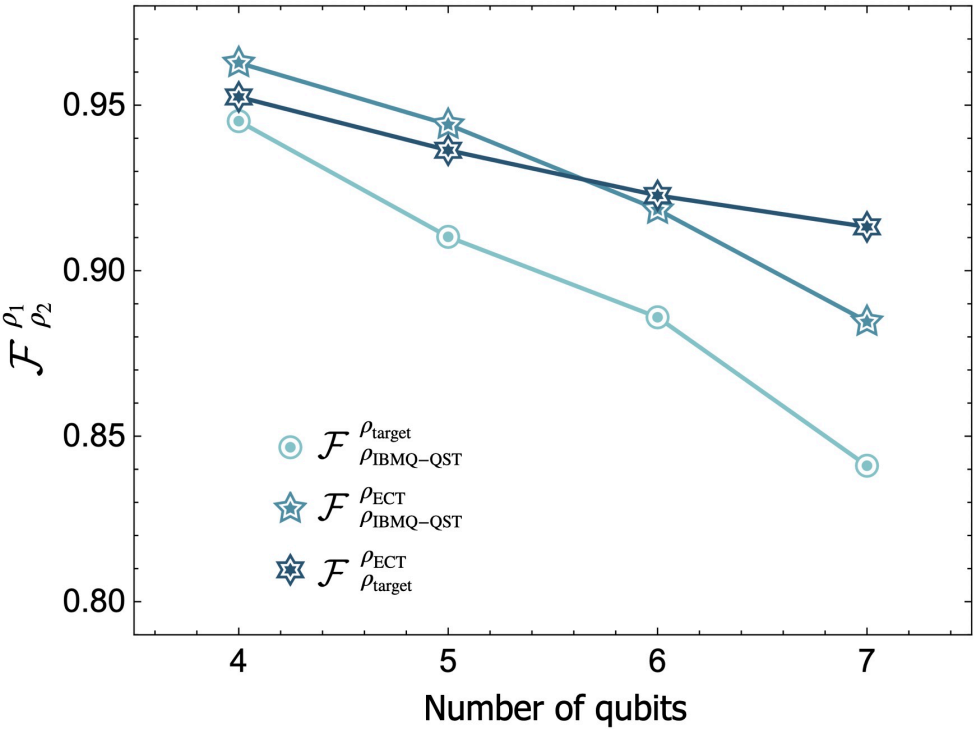


Figure 4.6: **Reconstruction of GHZ states on IBMQ with ECT-QST.** Fidelities $\mathcal{F}^{\rho_{\text{target}}}_{\rho_{\text{IBMQ-QST}}}$, $\mathcal{F}^{\rho_{\text{ECT}}}_{\rho_{\text{target}}}$, and $\mathcal{F}^{\rho_{\text{ECT}}}_{\rho_{\text{IBMQ-QST}}}$ for n_q-qubit GHZ states, with $n_q = 4, 5, 6, 7$. We define in the main text all the fidelities appearing here. Figure adapted from [2].

4.7 Experimental results on an IBMQ processor

We now show the experimental results achieved by ECT-QST to reconstruct quantum states generated on a superconducting processor provided by IBMQ. Specifically, we implemented circuits to generate GHZ, W, and random states on the IBMQ superconducting platform (see the Appendix A.1 for platform specifications and performance metrics). We chose the threshold t using the circuit-specific definition (2.45) that takes into account the noise level of the platform as we did in Section 2.9. This allowed us to make a fair comparison with the results reported in [1]. Furthermore, we compared our results with those achieved using IBMQ-QST, the default QST protocol implemented in `qiskit`. We recall that this protocol reconstructs the density matrix from linear inversion on the outcomes of an over-complete set of 6^{n_q} observables, resulting from the implementation of 3^{n_q} measurement settings, each corresponding to 2^{n_q} projectors. All the results of our experiments were obtained using $n_{\text{shots}} = 10^4$ shots.

We first implemented the circuit described in [403] to generate the n_q-qubit GHZ states, as this circuit has logarithmic time complexity and minimizes the required depth. In Figure 4.6 we show three fidelities as a function of the number of qubits: $\mathcal{F}^{\rho_{\text{target}}}_{\rho_{\text{IBMQ-QST}}}$ between the target density matrix and the one reconstructed with the IBMQ-QST protocol; $\mathcal{F}^{\rho_{\text{ECT}}}_{\rho_{\text{target}}}$ between the target density matrix and the one with the ECT-QST protocol; $\mathcal{F}^{\rho_{\text{ECT}}}_{\rho_{\text{IBMQ-QST}}}$ between the density matrix reconstructed with ECT-QST and the one reconstructed with IBMQ-QST. We note that $\mathcal{F}^{\rho_{\text{target}}}_{\rho_{\text{IBMQ-QST}}}$ decreases below 0.9 for $n_q \geq 6$, while $\mathcal{F}^{\rho_{\text{ECT}}}_{\rho_{\text{target}}}$ is above 0.9 $\forall n_q$, which means that the ECT-QST algorithm faithfully reconstructs

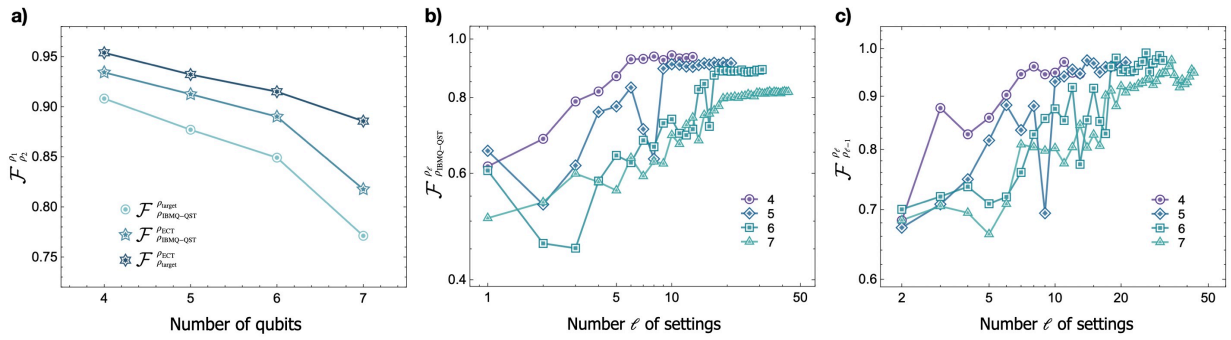


Figure 4.7: **Reconstruction of W states on IBMQ with ECT-QST. a)**: Fidelities $\mathcal{F}^{\rho_{\text{target}}}_{\rho_{\text{IBMQ}-\text{QST}}}$, $\mathcal{F}^{\rho_{\text{ECT}}}_{\rho_{\text{target}}}$, and $\mathcal{F}^{\rho_{\text{ECT}}}_{\rho_{\text{IBMQ}-\text{QST}}}$ for n_q-qubit W states, with $n_q = 4, 5, 6, 7$. The decrease of the fidelity as n_q increases is due to relaxation processes that populate the ground state component, which in the noiseless case is vanishing (see the discussion in Appendix A.2 and Figure A.3). **b), c)**: Fidelities $\mathcal{F}^{\rho_\ell}_{\rho_{\text{IBMQ}-\text{QST}}}$ and $\mathcal{F}^{\rho_\ell}_{\rho_{\ell-1}}$ for n_q-qubit W states as the settings identified by ECT-QST are progressively measured by decreasing weight, respectively. We define in the main text all the fidelities appearing here. Figure adapted from [2].

the density matrix of GHZ states. Furthermore, the fidelity between the two reconstructions is always above 0.88, indicating that the two reconstructions are quite similar despite the very different number of required settings.

Next, we leveraged the circuit with logarithmic-time complexity proposed in [403] to generate n_q-qubit W states. In Figure 4.7a), we show the fidelity $\mathcal{F}^{\rho_{\text{ECT}}}_{\rho_{\text{fQST}}}$ as in Figure 4.6, noticing that here we achieve lower values. This decrease in fidelity is due to experimental imperfections in the IBMQ processor. Although the implemented circuits generate reasonably trusting W states on the IBMQ platform, the population of the ground state $|0\rangle^{\otimes n_q}$ increases significantly as n_q increases and leads to lower fidelity values (see Appendix A.2 and Figure A.3). In Figure 4.7b) and c) we show the fidelity $\mathcal{F}^{\rho_\ell}_{\rho_{\text{IBMQ}-\text{QST}}}$ between the ECT-QST reconstruction ρ_ℓ with ℓ settings and the reconstruction with the IBMQ-QST protocol $\rho_{\text{IBMQ}-\text{QST}}$, as we progressively measure the settings identified by ECT-QST by decreasing weight, and the fidelity $F^{\rho_\ell}_{\rho_{\ell-1}}$ between ECT-QST reconstructions with ℓ and $\ell - 1$ settings, respectively. We highlight that last-measured settings least affect the fidelity, and are associated to the imaginary part of the density matrix elements identified by the threshold. This occurs because the off-diagonal elements of the target W states have vanishing imaginary parts. Consequently, the most relevant settings are those that provide information about the real part of these elements.

Finally, we implemented on IBMQ the same states generated by n_q-qubit depth-3 random circuits considered in Section 4.6, with $n_q = 4, 5, 6, 7$. We show the results for the 7-qubit random states in Figure 4.8, where we show the same fidelities as in Figure 4.7, while more details on the cases with $n_q = 4, 5$, and 6 are reported in Table 4.2 and in Appendix A.3. As a general comment, we found that ECT-QST can reduce the number of required settings by one or two orders of magnitudes with respect to IBMQ-QST, depending on the states under consideration, without compromising the quality of the reconstruction.

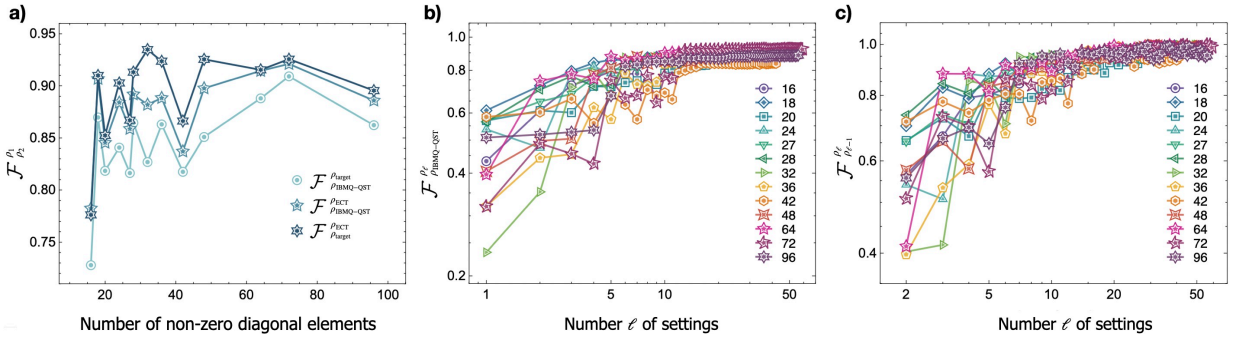


Figure 4.8: **Reconstruction of random states on IBMQ with ECT-QST. a)** Fidelities $\mathcal{F}^{\rho_{\text{target}}}_{\rho_{\text{IBMQ}-\text{QST}}}$, $\mathcal{F}^{\rho_{\text{ECT}}}_{\rho_{\text{target}}}$, and $\mathcal{F}^{\rho_{\text{ECT}}}_{\rho_{\text{IBMQ}-\text{QST}}}$ for 7-qubit random states, implemented on IBMQ depth-3 random circuits, as a function of the number of non-vanishing diagonal elements. **b)-c)**: Fidelities $\mathcal{F}^{\rho_\ell}_{\rho_{\text{IBMQ}-\text{QST}}}$ and $\mathcal{F}^{\rho_\ell}_{\rho_{\ell-1}}$ for 7-qubit pure random states as the settings identified by ECT-QST are progressively measured by decreasing weight, respectively. We define in the main text all the fidelities appearing here. Each curve represents a random state with a different number of non-vanishing diagonal elements, whose symbol is reported in the legend. Figure adapted from [2].

n_q	$\lvert S'_t\rvert$	$\overline{\mathcal{F}}^{\rho_{\text{ECT}}}_{\rho_{\text{IBMQ}-\text{QST}}}$	$\overline{\mathcal{F}}^{\rho_{\text{ECT}}}_{\rho_{\text{target}}}$	$\lvert S'_{t,\ell^*}\rvert$	$\overline{\mathcal{F}}^{\rho_{\ell^*}}_{\rho_{\text{IBMQ}-\text{QST}}}$	$\overline{\mathcal{F}}^{\rho_{\ell^*}}_{\rho_{\text{target}}}$	$\sim r\times n_q$
4	13(5)	0.97(1)	0.96(1)	7(3)	0.96(2)	0.95(3)	5
5	25(5)	0.95(1)	0.95(1)	10(3)	0.94(1)	0.94(2)	6
6	35(7)	0.92(3)	0.94(2)	12(5)	0.91(3)	0.93(3)	8
7	40(14)	0.87(4)	0.89(5)	18(6)	0.87(2)	0.89(4)	11

Table 4.2: **Results of ECT-QST for n_q-qubit depth-3 random circuits on IBMQ, with $n_q = 4, 5, 6, 7$.** $|S'_t|$ is the average number of settings determined by the pruning procedure. $\overline{\mathcal{F}}^{\rho_{\text{ECT}}}_{\rho_{\text{fQST}}}$ is the average fidelity between the ECT and IBMQ-QST reconstructions. $\overline{\mathcal{F}}^{\rho_{\text{ECT}}}_{\rho_{\text{target}}}$ is the average fidelity between the ECT-QST reconstruction and the target state. $|S'_{t,\ell^*}|$ is the average number of settings used when progressively measuring the settings in S'_t ordered by their decreasing weight and terminating when $\mathcal{F}^{\rho_{\ell^*}}_{\rho_{\ell^*-1}} > 95\%$, achieved with ℓ^* settings. $\overline{\mathcal{F}}^{\rho_{\ell^*}}_{\rho_{\text{IBMQ}-\text{QST}}}$ and $\overline{\mathcal{F}}^{\rho_{\ell^*}}_{\rho_{\text{target}}}$ are the average fidelities between the ECT-QST reconstruction with ℓ^* settings and the IBMQ-QST reconstruction and the target state, respectively. All the averages are computed over the random states selected for our analyses. The last column reports an estimate of $r \times n_q$, with r calculated as the average of the inverse purity of the IBMQ-QST reconstructions. This value corresponds to the number of settings expected by the Adaptive Compressed Sensing approach to QST proposed in [134, 135]. In all the columns, number in parentheses denote the standard deviation in the last digit(s).

4.8 A comparison with adaptive and compressed sensing QST

The ECT-QST protocol takes inspiration from CS QST. Thus, it is relevant to understand how our protocol compares with CS QST, where reconstruction of a rank-r density matrix of n_q qubits can be carried out with $\mathcal{O}(r2^{n_q}\log^2 2^{n_q})$ randomly chosen measurement settings [118]. In particular, we compare ECT-QST with an experimental implementation of CS-QST on a 7-qubit system implemented with trapped ions [132], where the authors prepare states of a topological color code [404]. The 7-qubit states we analyze are given by:

$$\begin{aligned}\left|\overline{0}\right\rangle = \frac{1}{\sqrt{8}}\left(|1010101\rangle + |1100011\rangle + |0101101\rangle + |0011011\rangle\right.\\ \left. + |1001110\rangle + |0110110\rangle + |1111000\rangle + |0000000\rangle\right),\end{aligned} \tag{4.23}$$

and

$$\begin{aligned}\left|\overline{1}\right\rangle = \frac{1}{\sqrt{8}}\left(|0101010\rangle + |1010010\rangle + |0011100\rangle + |1100100\rangle\right.\\ \left. + |0110001\rangle + |1001001\rangle + |0000111\rangle + |1111111\rangle\right).\end{aligned} \tag{4.24}$$

The states (4.23) and (4.24) are eigenstates of the operator $\mathcal{Z} = Z_1Z_2Z_3Z_4Z_5Z_6Z_7$, with $\mathcal{Z}\left|\overline{0}\right\rangle = \left|\overline{0}\right\rangle$ and $\mathcal{Z}\left|\overline{1}\right\rangle = -\left|\overline{1}\right\rangle$. At the same time, they are also eigenstates that correspond to the positive eigenvalue of all six stabilizer operators $X_1X_2X_3X_4$, $X_2X_3X_5X_6$, $X_3X_4X_6X_7$, $Z_1Z_2Z_3Z_4$, $Z_2Z_3Z_5Z_6$, and $Z_3Z_4Z_6Z_7$. As a result, such a system represents the smallest possible instance of a topological color code, which encodes a single logical qubit, resilient to one bit flip, phase flip, or a combined error of both, as an entangled state distributed over 7 physical qubits [405].

Following [132], we implement CS QST with a set of 127 measurement settings (called 'Pauli basis measurements' (PBM) settings in [132]) as a product of seven Pauli matrices:

$$s^{(j)} = V_1V_2V_3V_4V_5V_6V_7; \qquad j = 1, \ldots, 127. \tag{4.25}$$

Importantly, in CS QST, the $V_i \in \{X, Y, Z\}$ are chosen at random. In [132] the measurement was repeated 100 times for each setting , that is, the number of shots for each measurement setting was equal to 100. Thus, one ends up performing $2^7 = 128$ projective measurements for each setting, which in turn corresponds to a total amount of $127 \times 2^7 = 16,256$ projective measurements. Furthermore, the measurement outcomes collected for these measurement settings are undersampled, for the number of shots is smaller than the number of possible outcomes of each setting, resulting in a quite poor signal to noise ratio. ECT-QST with the Gini threshold identifies $8 \times 7 = 56$ off-diagonal matrix elements to reconstruct each of the states $\left|\overline{0}\right\rangle$ and $\left|\overline{1}\right\rangle$, and 14 unique measurement settings, one order of magnitude fewer than the 127 (random) settings used in [132]. We stress that, at variance with CS QST where the settings are sampled randomly, the ECT-QST protocol suggests the optimal settings that provide enough information to reconstruct the quantum state.

In Table 4.3 we report the results of the reconstruction of the states $\left|\overline{0}\right\rangle$ and $\left|\overline{1}\right\rangle$ with 127 random settings, with 14 random settings, and the 14 settings suggested by ECT-QST,

State	Number of settings	$\overline{\mathcal{F}}$
$\left\|\overline{0}\right\rangle$	127 (random)	$0.981(3)$
	14 (random)	$0.938(11)$
	14 (ECT-QST)	$0.979(2)$
$\left\|\overline{1}\right\rangle$	127 (random)	$0.981(4)$
	14 (random)	$0.929(12)$
	14 (ECT-QST)	$0.979(1)$

Table 4.3: **Comparison between CS QST and ECT-QST.** We compare CS QST and ECT-QST on reconstructing the states $\left|\overline{0}\right\rangle$ and $\left|\overline{1}\right\rangle$ with 127 randomly chosen settings, as originally done in [132], 14 randomly chosen settings, and the 14 settings required by ECT-QST. In each case we use 100 shots. The average $\overline{\mathcal{F}}$ fidelity and the standard deviation are computed over 10 independent reconstructions.

with the number of shots equal to 100 as in [132]. We note that the average fidelities obtained with 127 random settings and with the 14 settings suggested by ECT-QST are always compatible within the error, meaning that the two reconstructions are similar despite the differences in choosing the settings. Furthermore, the lower fidelity achieved with 14 random settings confirms that ECT-QST can actually identify the settings that provide information on the elements above threshold.

Finally, we can compare ECT-QST with another version of CS QST, called adaptive compressive tomography (ACT), introduced in [134, 135]. ACT is an adaptive protocol that starts from the outcomes of an initial random measurement setting to identify the next best setting (according to a proper figure of merit) using rank-deficient estimates of the state to be reconstructed. We can compare this approach with ECT-QST in several respects. For example, regarding the number of measurement settings, we achieved similar results in our IBMQ implementation. In fact, in the last column of Table 4.2 we report the quantity $r \times n_q$, with r calculated as the average of the inverse of the purity of the states reconstructed with IBMQ-QST. This value represents the number of settings required by ACT, and is compatible within the error bars with the number of settings $|S'_{t,\ell^*}|$ reported in Table 4.2, except for $n_q = 5, 7$ where it is beyond the error bar by a single measurement setting. However, there are major differences between the two approaches. ACT is an adaptive protocol, meaning that one cannot definitely know the amount of resources (number of observables, time to be spent in the laboratory, etc.) necessary to complete the experiment. In contrast to this, our protocol completely determines the amount of resources after having determined the diagonal elements and chosen the threshold. Furthermore, our protocol does not rely on any hypothesis on the state, while ACT leverages rank-deficient estimates of the state to find the next measurement setting.

4.9 Experimental results on a photonic integrated circuit

We now present the results of ECT-QST implemented on an integrated reprogrammable photonic processor. The device is the same as in Section 2.10, and we refer to there and to Appendix B.1 for further details. However, we remind the reader that the device is partic-

ularly suitable for the experimental implementation of ECT-QST, as the MZI architecture allows the natural implementation of the measurement settings, as shown in Figure 4.9. Furthermore, in this section, we name "QST" the protocol of ECT-QST with threshold $t = 0$, with the number of settings determined after the pruning procedure. The key figures of merit to quantify the efficacy of the proposed protocol are still the fidelity, between two reconstructed states or between the target and the reconstructed state, and the purity of the reconstructed state.

We performed a comprehensive analysis of ECT-QST applied to maximally entangled states of different numbers of qubits from 2 to 4, namely:

$$\begin{aligned}
\left|\psi^+\right\rangle &= \frac{1}{\sqrt{2}}\left(|01\rangle + |10\rangle\right) & \left|\psi^-\right\rangle &= \frac{1}{\sqrt{2}}\left(|01\rangle - |10\rangle\right) \\
\left|\mathrm{GHZ}_3^+\right\rangle &= \frac{1}{\sqrt{2}}\left(|010\rangle + |101\rangle\right) & \left|\mathrm{GHZ}_3^-\right\rangle &= \frac{1}{\sqrt{2}}\left(|010\rangle - |101\rangle\right) \\
|\mathrm{W}_3\rangle &= \frac{1}{\sqrt{3}}\left(|100\rangle + |010\rangle + |001\rangle\right) & |\mathrm{GHZ}_4\rangle &= \frac{1}{\sqrt{2}}\left(|0101\rangle + |1010\rangle\right)
\end{aligned} \tag{4.26}$$

Figure 4.10 examines how the fidelity $\mathcal{F}_{0,t}$ between the QST and ECT-QST reconstructions (upper panels), and the purity $\mathcal{P}$ of the ECT-QST reconstruction (lower panels), change as a function of the threshold t for the six maximally entangled states of (4.26). A key observation from all the upper panels (a–f) is that as the threshold t decreases and, consequently, the number ℓ of measurement settings increases, the fidelity between the two methods approaches 1. This convergence indicates that ECT-QST can approximate QST with high accuracy when enough settings are included. However, an important advantage of the ECT-QST method is evident: fidelity quickly saturates, that is, high fidelity is achieved even with a significantly smaller number of measurement settings compared to QST. This efficiency gain confirms one of the cornerstones of the proposed threshold protocols, both tQST and ECT-QST: The information content is not uniformly distributed among the elements of the density matrix. In fact, only a subset of the measurement settings contributes most significantly to the state reconstruction. Another noteworthy point is that not only the average fidelity increases with the number of settings, but the standard deviation also decreases. This trend reflects greater stability and reliability in the reconstruction process as more measurement settings are included. Finally, the purity trends mirror the ones of the fidelity, for the purity increases as more settings are included. The maximum value of the purity decreases as the number of qubits increases as a result of the greater impact of experimental imperfections on the generation of the states.

The results also confirm the dependence on the specific representation and the entanglement class of the quantum state under study. For instance, Figure 4.10a) and b) show similar behavior for the Bell states $|\psi^+\rangle$ and $|\psi^-\rangle$, which is expected due to their identical entanglement structure and symmetry under local unitary operations. In contrast, the differences between the $|\mathrm{W}_3\rangle$ and $|\mathrm{GHZ}_3\rangle$ states are more pronounced, which is consistent with the fact that the W and GHZ states belong to distinct entanglement classes. These differences manifest themselves in the fidelity trends and the number of settings required for convergence. Furthermore, $\left|\mathrm{GHZ}_3^+\right\rangle$,$\left|\mathrm{GHZ}_3^-\right\rangle$ and $|\mathrm{GHZ}_4\rangle$ achieved average fidelity $\mathcal{F}_{0,t} > 0.9$ with only three measurement settings, regardless of the number of qubits. This confirms that the structural sparsity of GHZ states makes them particularly attractive for ECT-QST, which can properly identify the minimal core set of settings that capture the essential information and result in a high-quality reconstruction.

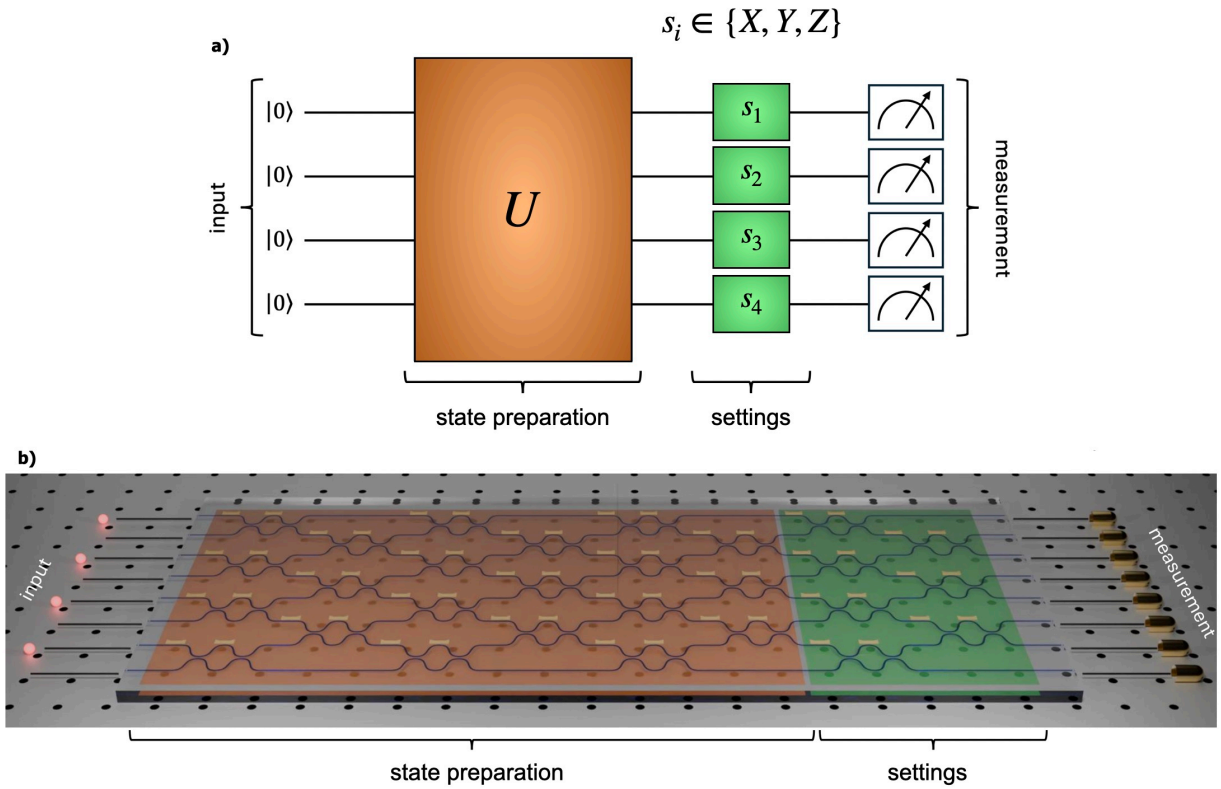


Figure 4.9: **Illustration of the hybrid photonic chip for the implementation of measurement settings. a)** After injecting photons in the mode corresponding to the $|0\rangle$ state of each qubit, the photonic processor is programmed to implement a unitary transformation U that generates the target quantum state and the observables $\{s_i\}_{i=1}^{n_q}$ of the chosen measurement setting. Finally, a series of photodetectors record the counts for each projector in the setting. In panel **b)** we show that the first six layers of the interferometer (state preparation) are properly configured to prepare a target state, while the remaining two layers of the interferometer (settings) are dedicated to the implementation of the measurement settings. Thus, by appropriately programming the transformation of an RBS acting on the two modes of a qubit, it is possible to implement any measurement setting on the output state. Thus, we can measure the observables needed to reconstruct the target quantum state by applying the appropriate sequence of settings according to the chosen reconstruction protocol, either QST or ECT-QST in this experiment. Figure to appear in [5].

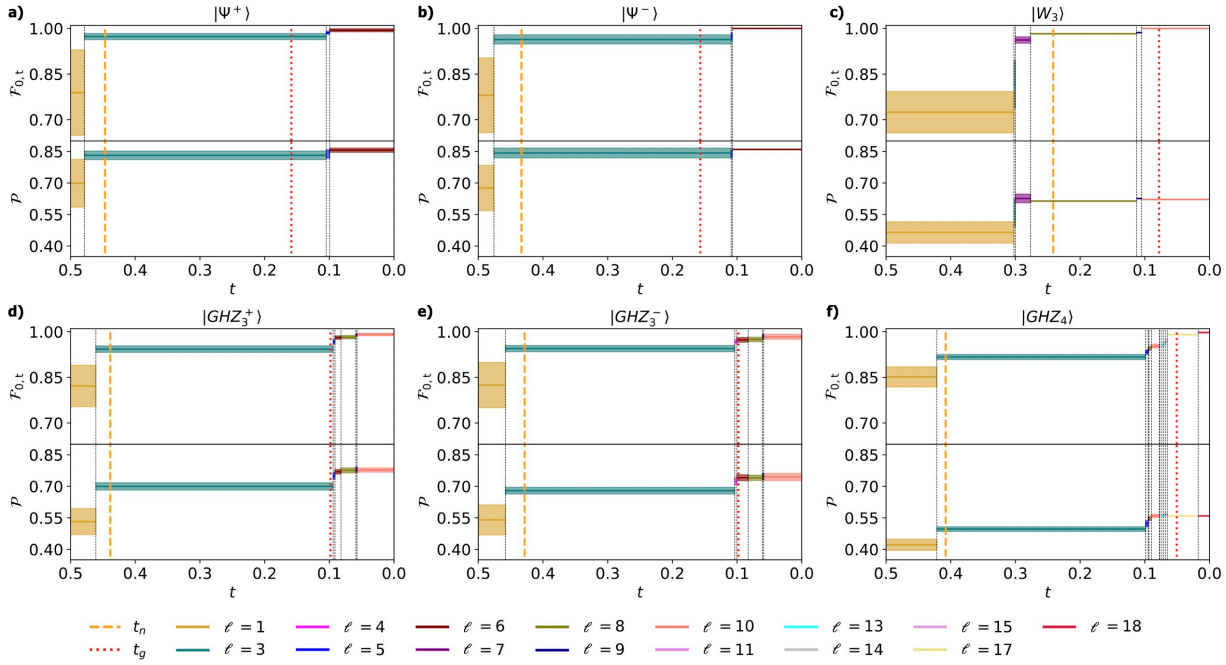


Figure 4.10: **Figures of merit of ECT-QST applied to maximally entangled states as a function of the threshold.** We show the $\mathcal{F}_{0,t}$ (upper panel) and purity $\mathcal{P}$ (lower panel) as a function the threshold t for six maximally-entangled states $|\psi^+\rangle$ **a)**, $|\psi^-\rangle$ **b)**, $|\mathrm{W}_3\rangle$ **c)**, $\left|\mathrm{GHZ}_3^+\right\rangle$ **d)**, $\left|\mathrm{GHZ}_3^-\right\rangle$ **e)**, and $|\mathrm{GHZ}_4\rangle$ **f)**. Vertical black dashed lines separate regions corresponding to different numbers ℓ of measurement settings, the value of which is reported in the color code. Horizontal lines and the associated shaded regions represent the average fidelity/purity and the corresponding standard deviation, respectively, computed with a standard bootstrap approach over 100 samples. More specifically, for each sample, the initialization parameters are chosen through a random draw of the seed. Vertical orange and red dashed lines represent the values of the Gini threshold t_g and noise-aware threshold t_n, respectively. Figure to appear in [5].

The comparison between the Gini threshold t_g and the noise-aware threshold t_n is significant as well. In general, the Gini threshold represents a more conservative choice that provides a number of settings greater than or equal to the one provided by the noise-aware threshold. Specifically, for $|\psi^+\rangle$, $|\psi^-\rangle$, and $\left|\mathrm{GHZ}_3^+\right\rangle$ the two thresholds provide the same number of settings, while for the states $|\mathrm{W}_3\rangle$, $\left|\mathrm{GHZ}_3^-\right\rangle$, and $|\mathrm{GHZ}_4\rangle$ the Gini threshold suggests 2, 3, and 14 more settings than the noise-aware threshold, respectively. However, the fidelity achieved using the noise-aware threshold is above 0.9 despite the reduced number of settings, emphasizing the advantage of choosing the threshold by taking into account the experimental noise.

Figure 4.11 complements the findings of Figure 4.10 by presenting the behavior of the fidelity $\mathcal{F}_{0,t}$ and $\mathcal{F}_{t,m}$ when the measurement settings are progressively measured according to their weight. The trends observed here mirror those of Figure 4.10: the fidelity converges to 1 as ℓ increases. The convergence is evident across all six maximally entangled states, confirming that only a subset of the measurement settings is crucial for high-fidelity reconstruction. The data also validate the effectiveness of the ranking system introduced in Section 4.2.2 to arrange the measurement settings into a hierarchy and reinforce the fact that some measurements are more informative than others. By assigning decreasing weights to the settings based on their relevance, the method correctly prioritizes those that carry the most significant information for state reconstruction. Selecting the most informative settings first allows ECT-QST to achieve high fidelity with fewer resources. The values of t_g and t_n are the same as in Figure 4.10.

The previous results allowed us to conclude that, on average, ECT-QST can achieve results that are comparable with, or better than, QST with fewer measurement settings. This is possible thanks to several factors, such as how we select and construct the settings and the weight system. However, other resources can be optimized, other than measurement settings. For instance, also the number n_c of copies of the state, or counts of the detectors, available to compute the expectation value of all observables is a resource, finite by definition in experiments. The question now is the following: Does the number of copies affect the results found so far? If so, to what extent?

To this end, we studied how the fidelities $\mathcal{F}_{0,m}$ and $\mathcal{F}_{t,m}$ vary with the number n_c of copies used for all measurement settings required by the protocol. We focus on the $|\mathrm{GHZ}_4\rangle$ state, for which QST and ECT-QST require $\ell = 18$ and $\ell = 3$ settings, respectively. Thus, for a fixed value of n_c, QST consumes $n_c/18$ copies per setting, while ECT-QST consumes $n_c/3$ copies per setting. In Figure 4.12a) we show the trend of the two fidelities as a function of n_c, and three points are highlighted. Point A identifies the low statistics regime, where the number of copies available is relatively small and ECT-QST performs better than QST, for $\overline{\mathcal{F}}_{t,m} > \overline{\mathcal{F}}_{0,m}$, without overlapping the error bars. By increasing the number of copies, we reach an intermediate regime where the fidelities reach a crossing point, namely point B, where the results of ECT-QST and QST are compatible, $\overline{\mathcal{F}}_{t,m} \sim \overline{\mathcal{F}}_{0,m}$. Finally, if we move to the high statistics regime, identified by point C, the results are opposite to the low statistics regime: QST performs better than ECT-QST, with $\overline{\mathcal{F}}_{t,m} < \overline{\mathcal{F}}_{0,m}$ and a smaller uncertainty for QST. These results can be interpreted as follows. On the one hand, ECT-QST performs better when few copies are available, since it properly identifies the most relevant measurement settings on which we can spend the available (scarce) resources, while QST, which consumes fewer copies per setting than ECT-QST, cannot collect enough information for a high-fidelity reconstruction. On the other hand, when many copies of the state are available, QST performs better than ECT-QST despite the

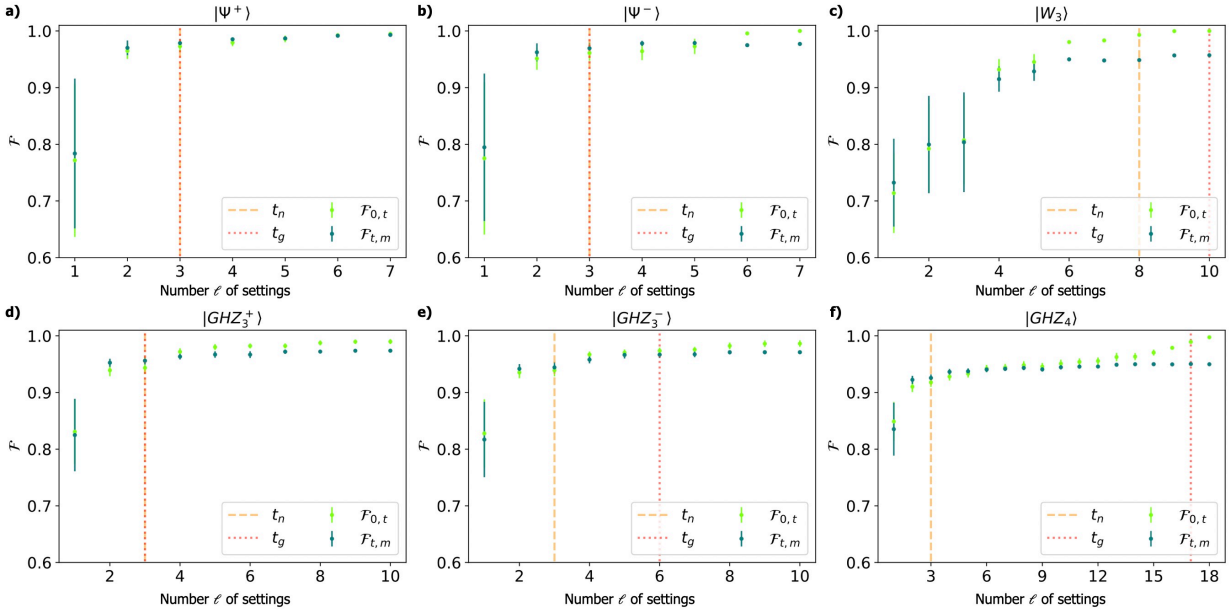


Figure 4.11: **Fidelity of ECT-QST applied to maximally entangled states as a function of the number of measurement settings.** We show the fidelities $\mathcal{F}_{0,t}$ (light green) and $\mathcal{F}_{t,m}$ (azure) by progressively increasing the number ℓ of measurement settings used to reconstruct the density matrix of the six maximally-entangled $|\psi^+\rangle$ **a)**, $|\psi^-\rangle$ **b)**, $|\mathrm{W}_3\rangle$ **c)**, $\left|\mathrm{GHZ}_3^+\right\rangle$ **d)**, $\left|\mathrm{GHZ}_3^-\right\rangle$ **e)**, and $|\mathrm{GHZ}_4\rangle$ **f)**. More specifically, the threshold t is set equal to 0, and thus the number of settings to measure $|S_t'|$, which is state-dependent and equal to the largest value on the horizontal axis of the corresponding state. The method then identifies a hierarchy for the settings through a weight system, as explained in Section 4.2.2. Each point of the plots represent the average fidelity achieved using the first ℓ measurement settings according to the weight hierarchy. The fidelity $\mathcal{F}_{0,m}$ between QST and imperfection model is computed from the reconstruction with $\ell = |S_t'|$ and corresponds to the rightmost point of $\mathcal{F}_{t,m}$. Error bars represent the standard deviation and is computed, along the average, using a standard bootstrap approach over 100 samples. More specifically, for each sample, the initialization parameters are chosen through a random draw of the seed. The number of measurement settings obtained according to the noise-aware threshold t_n, derived from the imperfection model of the experimental apparatus, and to the Gini threshold t_g are highlighted with an orange and red dashed vertical line, respectively. Figure to appear in [5].

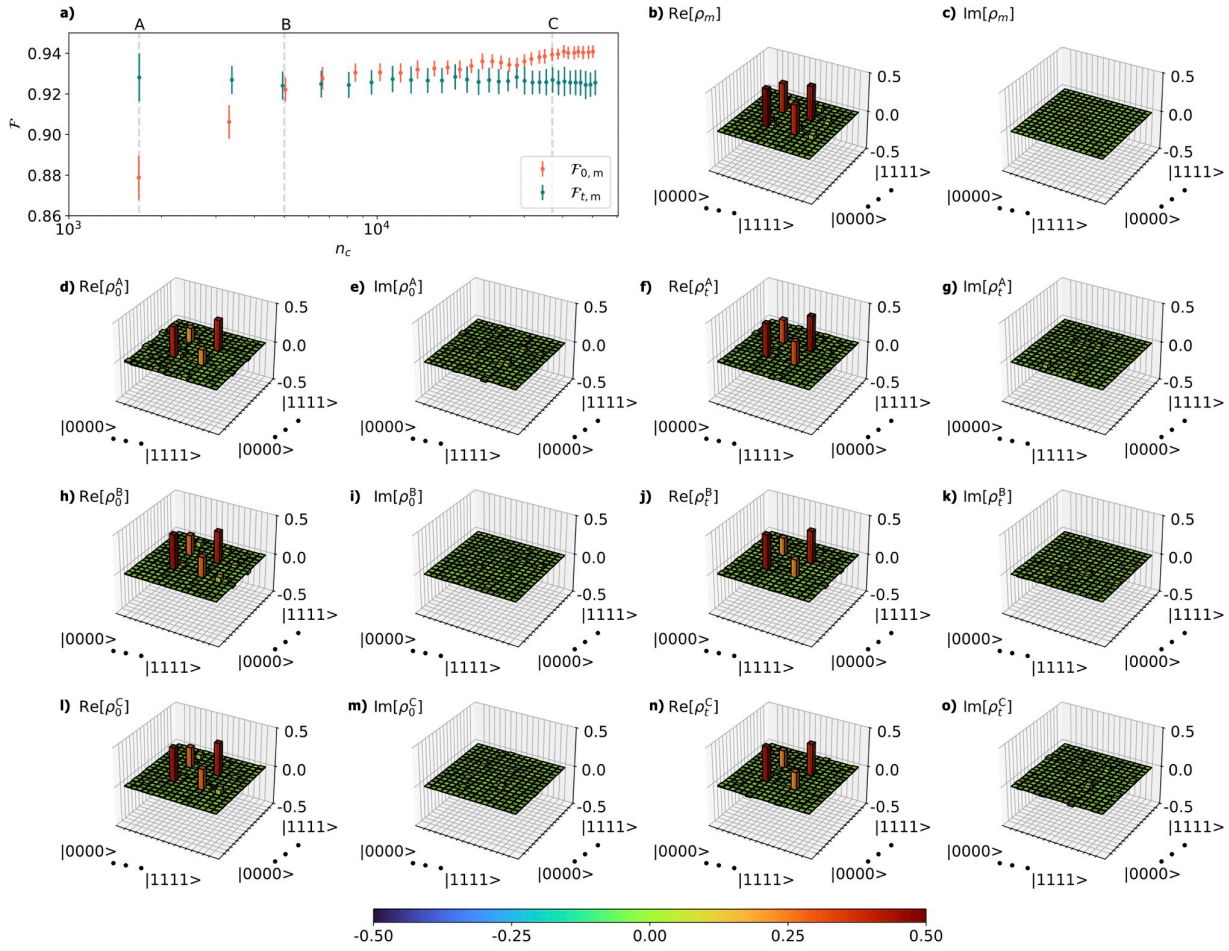


Figure 4.12: **Comparison of QST and ECT-QST for finite amount of resources. a)**: Fidelities $\mathcal{F}_{0,m}$ (red) and $\mathcal{F}_{t,m}$ (azure) as a function of the number of counts n_c used to reconstruct the density matrix of the state $|\mathrm{GHZ}_4\rangle$. Points and error bars represent the average fidelity and standard deviation, respectively, computed with a standard bootstrap approach over 100 samples. More specifically, for each sample, shot noise is introduced by Poissonian statistic on the measured counts, and jointly the initialization parameters are chosen through a random draw of the seed. In particular, n_c corresponds to the overall number of counts used by QST ($\ell = 18$) and ECT-QST ($\ell = 3$). **b)**, **c)**: Real and imaginary part of the density matrix according to the theoretical model that takes into account experimental imperfections, $\mathrm{Re}\left[\rho_m\right]$ and $\mathrm{Im}\left[\rho_m\right]$. **d)**, **e)**: Real and imaginary part of the QST reconstruction, $\mathrm{Re}\left[\rho_0^{\mathrm{A}}\right]$ and $\mathrm{Im}\left[\rho_0^{\mathrm{A}}\right]$, for $n_c \approx 1700$ (A). **f)**, **g)**: Real and imaginary part of the ECT-QST reconstruction, $\mathrm{Re}\left[\rho_t^{\mathrm{A}}\right]$ and $\mathrm{Im}\left[\rho_t^{\mathrm{A}}\right]$, for $n_c \approx 1700$ (A). **h)**, **i)**: Real and imaginary part of the QST reconstruction, $\mathrm{Re}\left[\rho_0^{\mathrm{B}}\right]$ and $\mathrm{Im}\left[\rho_0^{\mathrm{B}}\right]$, for $n_c \approx 5000$ (B). **j)**, **k)**: Real and imaginary part of the ECT-QST reconstruction, $\mathrm{Re}\left[\rho_t^{\mathrm{B}}\right]$ and $\mathrm{Im}\left[\rho_t^{\mathrm{B}}\right]$, for $n_c \approx 5000$ (B). **l)**, **m)**, Real and imaginary part of the QST reconstruction, $\mathrm{Re}\left[\rho_0^{\mathrm{C}}\right]$ and $\mathrm{Im}\left[\rho_0^{\mathrm{C}}\right]$, for $n_c \approx 37000$ (C). **n)**, **o)**: Real and imaginary part of the ECT-QST reconstruction, $\mathrm{Re}\left[\rho_t^{\mathrm{C}}\right]$ and $\mathrm{Im}\left[\rho_t^{\mathrm{C}}\right]$, for $n_c \approx 37000$ (C). The colorbar for the representation of the density matrices is shown at the bottom. Figure to appear in [5].

smaller number of copies per setting, for the use of more settings allows to gather high-quality information and achieve a better fidelity in the reconstruction. Nevertheless, both regimes confirm that the ECT-QST protocol selects the most informative measurement settings, for the fidelity is consistently above 0.92.

4.10 Final remarks

In this chapter, we introduced an original QST protocol, called ECT-QST, inspired by compressed sensing that extends tQST to systems composed of qudits where measurement settings can be implemented. The initial steps of the workflow follow those of tQST: we first determine the diagonal elements and then choose a threshold according to one of the previously described criteria. The main differences arise in the construction of the observables and in how they are associated with the off-diagonal elements above the threshold.

To address this, we showed how to construct a set of unique measurement settings that provides information on the selected off-diagonal elements. Importantly, the maximum number of settings required by ECT-QST is always less than the 3^{n_q} settings needed by standard QST for qubit systems. Furthermore, we introduced a pruning procedure that can further reduce the number of settings depending on the specific state under consideration. We also highlighted the connection between ECT-QST and tQST, showing that the latter can be recovered as a particular case of the former. As with tQST, we presented step-by-step applications of ECT-QST to the same states discussed in Section 2.6. We then moved to more general situations and performed numerical analyses on several states of 2, 3, and 4 qubits. These results demonstrated that ECT-QST can reconstruct states with high fidelity while requiring significantly fewer resources.

We also carried out experimental tests of ECT-QST on the same platforms used for tQST. On the superconducting platform provided by IBMQ, we reconstructed states such as W and GHZ states, as well as random states from 2 to 7 qubits. On the photonic platform, we reconstructed maximally entangled states of 2, 3, and 4 qubits. In both cases, the experimental results agree with the numerical analysis, confirming that ECT-QST achieves accurate reconstructions with far fewer settings than standard QST.

Finally, we compared the performance of QST and ECT-QST by evaluating the fidelities between the reconstructed states and the theoretical model as a function of the number of counts used to calculate the expectation value of the observables. This comparison revealed three regimes: ECT-QST outperforms QST when resources are scarce, then a crossing point occurs, after which QST performs better in the high-statistics regime. The advantage of ECT-QST in the low-statistics regime stems from the careful selection of measurement settings, which allows more copies of the state to be allocated to the relevant observables.

This result is particularly relevant for quantum technological applications in large Hilbert spaces. As systems grow in size, their Hilbert spaces expand, requiring more and more copies of the state to fully explore them. However, the efficient generation of large quantum states remains a challenging problem, and until more effective strategies are developed, we must contend with the low-statistics regime and rely on tools such as ECT-QST that are well suited to it.

A natural question then arises: is this the only scenario in which ECT-QST outper-

forms QST in the low-statistics regime, or are there other problems, perhaps of a more fundamental nature, where similar results hold? This question will be the focus of the next chapter.

Chapter 5

Quantum Maxwell's Demon and Finite Resources

In this chapter, we apply the ECT-QST protocol to the problem of quantum work extraction and study how finite resources affect the quality of the tomography reconstruction and the amount of work extractable from a system.

5.1 Maxwell's demon and quantum mechanics

One of the paramount results of physics is the second law of thermodynamics, according to which the total entropy of an isolated system can never decrease over time. In 1867 James Clerk Maxwell proposed a provocative thought experiment, today named the **Maxwell's demon** problem, where an imaginary being acquires microscopic information about a system and exploits this information to extract work (apparently) without consuming energy, thus seemingly violating the Second Law of Thermodynamics [406]. More precisely, a vessel filled with gas is divided into two chambers, labeled A and B, by an isolated wall with a hole with a trapdoor such that only one particle can pass through it. Both chambers are filled with the same gas as the original vessel and are initially at thermal equilibrium. An imaginary intelligent being, called a demon, controls a trapdoor that can open or close selectively without friction. The goal of the demon is to generate a temperature difference in the gas without performing any work on it. To this end, we recall that the kinetic theory of gases tells us that temperature is related to the average kinetic energy of the particles, and according to the Maxwell-Boltzmann distribution, some particles are faster or slower than the average. Thus, the demon strategy is to allow only fast particles to pass one way and slow particles the other, resulting in one chamber warming up while the other chamber cools down. This reduces the overall entropy of the system without requiring any work, thus violating the second law of thermodynamics. Equivalently, the result of the thought experiment is a heat flow from a lower to a higher temperature without any other effect, which contradicts Clausius's formulation of the second law.

This (apparent) paradox ignited several efforts to understand how and why the second law of thermodynamics was not actually violated, such as the ones by Smoluchovsky [407] and Szilard [408]. In particular, the foundational works by Landauer [409] and Bennett [410, 411] clarified why the second law cannot be violated: Any irreversible manip-

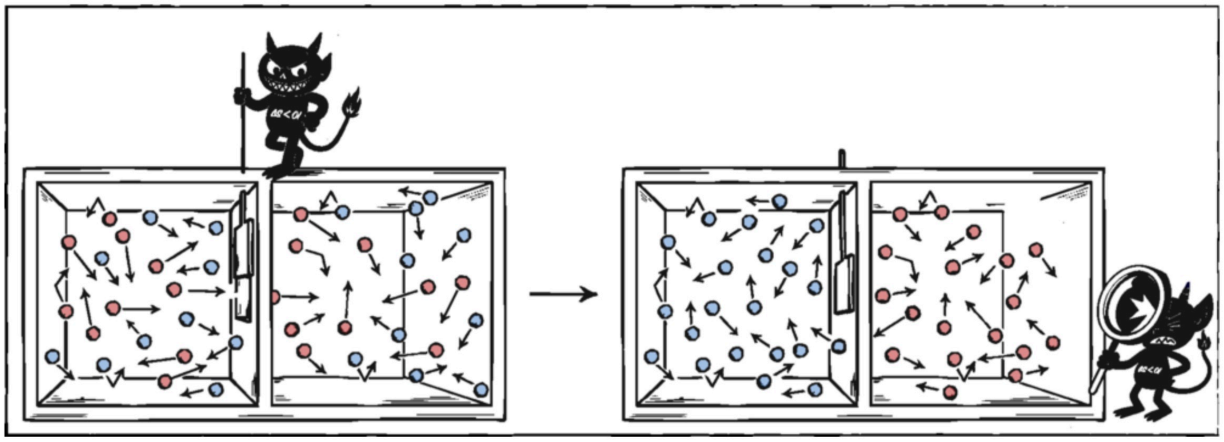


Figure 5.1: **Conceptual illustration of the Maxwell demon thought experiment.** On the left, the initial state with the two chambers at thermal equilibrium. On the right, the final, macroscopically ordered state, achieved by the demon exploiting microscopic information obtained without performing work, thus violating the second law. Red and blue particles are faster and slower than the average, respectively. Figure from [413].

ulation of information involves the consumption of energy. In Maxwell's thought experiment, the demon must be able to distinguish between the velocities of the particles that make up the gas. To do this, it measures each particle, remembers what it measured, and opens or closes the trapdoor accordingly. This implies that the information collected from the measurement is recorded in some sort of device, called a memory. However, the demon has to forget the results of the previous observations before observing another particle. Forgetting previous results is equivalent to discarding information or erasing bits of information. Landauer's principle states that the bit erasure operation has a minimum cost per bit equal to $k_B T \ln 2$, with k_B the Boltzmann constant and T the temperature at which the system is operating. This restores the validity of the second law of thermodynamics. We can then conclude that the discussion and solution of Maxwell's demon problem, highlighting the energetic cost of information processing, established a profound and fruitful connection between information theory and thermodynamics, summarized in the famous statement that "*information is physical*" [412]. The acquisition and manipulation of information require to spend energy and information about a system can be leveraged to extract work from it.

If we consider a quantum version of the problem, then the role of information becomes even more subtle and profound. The intrinsic probabilistic nature of quantum mechanics, quantified by the Born rule, allows us to determine the probability distribution of the outcomes of a given observable only by performing the same measurement on a certain number N, eventually large, of copies of the system. The measured relative frequencies will fluctuate around the probabilities computed from the Born rule, which can be recovered only in the ideal limit of infinite copies of the system. In any realistic scenario, we have access to a **finite** number of **resources**, that is, copies of the system, and thus only **partial information** about the system is accessible.

We now consider this realistic scenario and study the problem of the maximum extractable work from N copies of a quantum system by introducing a **"partial-information and finite-resources" (PIFR) quantum Maxwell demon**. While the classical version of

the demon is idealized and somehow omniscient, the quantum counterpart is a realistic agent that retrieves only partial information about the system by consuming a finite amount of resources. More specifically, it estimates the state of the system by performing a limited number of measurements on a fraction of the N copies. Then, according to its available partial information on the state, it estimates the optimal unitary transformation to extract work and implements it onto the remaining copies. This framework marks a substantial shift from traditional analyses of Maxwell's demon, for it presents a trade-off between information acquisition, accuracy of state reconstruction, and the total amount of work extracted, all under the constraint of finite resources. On the one hand, allocating more copies to gather information improves the accuracy of both the reconstructed state and the optimal work-extraction unitary, but reduces the number of copies available for actual work extraction. On the other hand, accepting a less accurate estimate may, in some cases, allow greater total work extraction. At the same time, there are analogies between the classical and the PIFR Maxwell demon. Essentially, in the classical version the demon first measures the particle (it determines its position and velocity), and then performs an action based on the acquired information (opens or closes the trapdoor). In the PIFR quantum case, the demon first reconstructs the state (the density matrix) and then performs an action based on this estimate (apply the estimated unitary transformation).

Within this framework, choosing an optimal protocol to reconstruct the state is fundamental. We know that thermodynamics inherently selects a preferred basis, the energy eigenbasis, and emphasizes the uneven compression of information into a small set of parameters, such as the temperature, the chemical potential, and others. Following this line of reasoning, we question whether QST, which considers all the possible observables equally important, is the optimal reconstruction protocol for the PIFR quantum Maxwell demon, or tQST/ECT-QST is a more efficient approach, thanks to the identification of a core set of most informative measurement.

5.2 PIFR Quantum Maxwell's demon and ergotropy

Let us start by defining the framework for the PIFR quantum Maxwell demon. We assume that a quantum system is described by a state ρ_{true}, and N i.i.d. copies of the system are available. A "perfect" determination of ρ_{true} could ideally be achieved in the limit of $N \to +\infty$, or by a large enough number of copies N such that the statistical errors due to the finite number of copies are negligible. We stress that performing a measurement on a copy and recording the outcome implies consuming that copy, that is, the copy cannot be used anymore for the subsequent work extraction task. Furthermore, we assume that the demon knows the Hamiltonian H of the system, for one typically starts from a specific model, such as the Heisenberg model, the Luttinger or the Klein-Gordon model, where the operatorial structure is known and only a few number of parameters need to be fixed, or engineered, to fully determine it.

Given the "perfect" knowledge about the state ρ_{true} and the Hamiltonian H of the system, then the maximum work that can be extracted from the system via unitary operations, that is, ideal quantum gates, is given by its **ergotropy** [414]:

$$\mathcal{W}(\rho_{\text{true}}, H) \equiv \mathcal{W} = \max_{U} \left\{ \text{Tr}\left[\left(H - U^{\dagger} H U \right) \rho_{\text{true}} \right] \right\}. \tag{5.1}$$

It is worth noting that the explicit form of the optimal ergotropic unitary that maximizes the extracted work (5.1) is given by:

$$U_{\mathcal{W}}(\rho_{\text{true}}) \equiv U_{\mathcal{W}} = \sum_k |\varepsilon_k\rangle\langle r_k| \,, \tag{5.2}$$

with $H = \sum_k \varepsilon_k \, |\varepsilon_k\rangle\langle\varepsilon_k|$ with eigenvalues sorted in ascending order $\varepsilon_k \leq \varepsilon_{k+1}$, and $\rho_{\text{true}} = \sum_j r_j \, |r_j\rangle\langle r_j|$ with eigenvalues sorted in descending order $r_k \geq r_{k+1}$. The application of (5.2) to a certain number of copies of the state produces an amount $\mathcal{W}$ of work per copy.

Let us now move to a realistic scenario in which one considers that the demon has only access to a finite number of copies N of the system. In this case, the distribution of such resources between M copies for data acquisition and state reconstruction, and $(N - M)$ copies to extract work from becomes fundamental. We start by defining $I \equiv I\,(M, \Theta)$ as the partial information retrieved from the measurements performed on M copies of the system. In general, different measurement protocols, denoted by Θ, affect the partial information retrieved, as will become clear later on. Conditional on such partial information, the demon reconstructs an estimate of the state, which we denote with $\rho_{|I}$, which in principle could be very different from the actual true state ρ_{true}. Based on this reconstruction, that is, according to its available partial information on the state, it then computes the optimal ergotropic unitary transformation:

$$U_{\mathcal{W}}(\rho_{|I}) \equiv U_{\mathcal{W},I} = \sum_k |\epsilon_k\rangle \, \langle r_{k,I}| \,, \tag{5.3}$$

and finally implements it on the remaining $(N - M)$ copies of the system, thus extracting a total work given by:

$$\mathcal{W}_I^{\text{tot}} = (N - M)\text{Tr}\left[\left(H - U_{\mathcal{W},I}^{\dagger} \, H \, U_{\mathcal{W},I}\right) \rho_{\text{true}}\right] = (N - M)\mathcal{W}_I, \tag{5.4}$$

with $\mathcal{W}_I = \text{Tr}\left[\left(H - U_{\mathcal{W},I}^{\dagger} \, H \, U_{\mathcal{W},I}\right) \rho_{\text{true}}\right]$ the work extracted by applying the unitary transformation $U_{\mathcal{W},I}$ to one copy of the state ρ_{true}. Clearly, since $U_{\mathcal{W},I}$ may differ from the optimal unitary $U_{\mathcal{W}}$, the actual amount of extracted work is necessarily smaller than or equal to $(N-M)\mathcal{W}$, which is achievable only in an ideal scenario with an infinite number of copies. We then introduce the following two figures of merit for our protocol. The first is the **reconstruction efficiency**:

$$\eta_R = \frac{\mathcal{W}_I}{\mathcal{W}}, \tag{5.5}$$

which quantifies the fraction of work that is actually extracted using $U_{\mathcal{W},I}$ compared to the maximal amount $\mathcal{W}$ when M copies are used to reconstruct the state. Alternatively, it quantifies the quality of the reconstructed state $\rho_{|I}$ when used to estimate the ergotropic unitary $U_{\mathcal{W},I}$. The second figure of merit is the **ergotropic efficiency**:

$$\eta_E = \frac{\mathcal{W}_I^{\text{tot}}}{N\mathcal{W}} = \frac{(N - M)\mathcal{W}_I}{N\mathcal{W}} = \left(1 - \frac{M}{N}\right) \frac{\mathcal{W}_I}{\mathcal{W}} = \left(1 - \frac{M}{N}\right) \eta_R, \tag{5.6}$$

which quantifies the trade-off between the quality of the state reconstructed with M copies and the amount of work extracted from $(N - M)$ copies. On the one hand, if

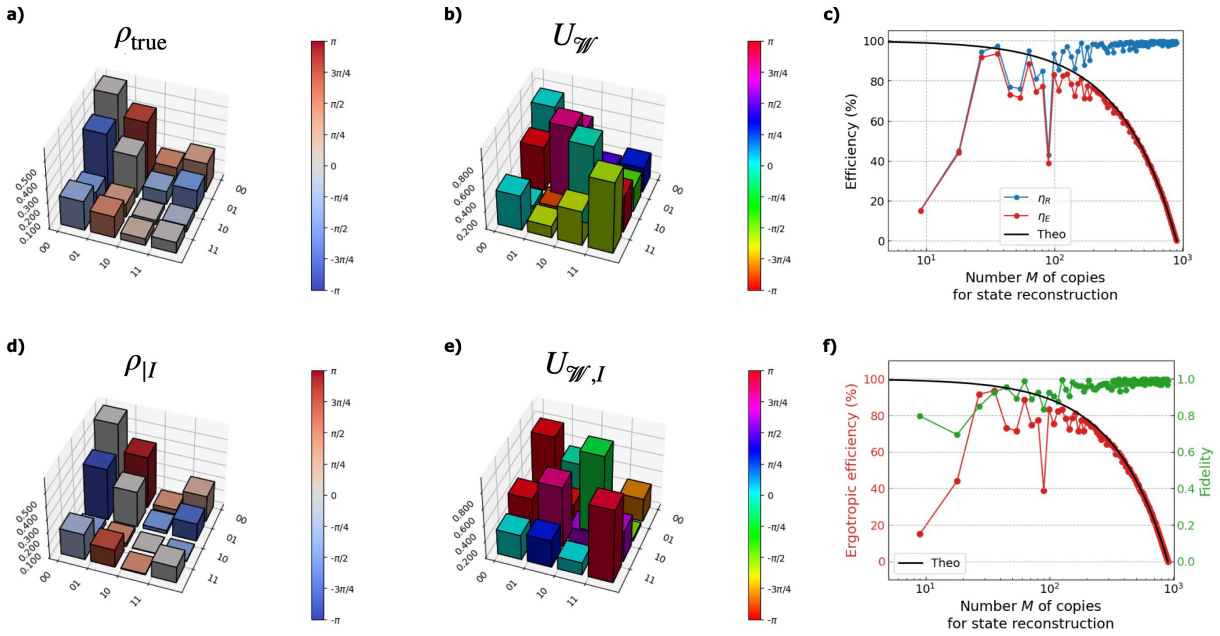


Figure 5.2: **Quantum work extraction protocol applied to a single density matrix. a), d)**: True density matrix ρ_{true} and the QST reconstruction $\rho_{|I}$ with 10 copies per setting, respectively. The fidelity between the two density matrices is $\mathcal{F} = 0.933$. **b), e)**: Ergotropic unitary $U_{\mathcal{W}}$ and the unitary transformation $U_{\mathcal{W},I}$ estimated from $\rho_{|I}$, respectively. The ergotropy is $\mathcal{W} = 8.86$J, while estimated ergotropy is $\mathcal{W}_I = 7.37$J, resulting in $\eta_R = 0.832$. **c)**: Reconstruction efficiency η_R and ergotropic efficiency η_E for the state shown in **a)** as a function of the number of copies used for state reconstruction. **f)**: Ergotropic efficiency η_E and fidelity between the ECT-QST reconstruction and the target state as a function of the number of copies used for state reconstruction, with $N = 10^3$. The black solid line is the theoretical curve $(1 - M/N)$ that represents the trend of η_E with perfect reconstruction efficiency $\eta_R = 1$. Figure to appear in [6].

we use all copies for state reconstruction $(M = N)$, then there are no copies to extract work from and $\eta_E = 0$. On the other hand, if by some means we can achieve a "perfect" reconstruction of the state $\left(\rho_{|I} = \rho_{\text{true}}\right)$ without consuming copies $(M = 0)$, then $\mathcal{W}_I = \mathcal{W}$ and $\eta_E = 1$. However, if the state reconstruction task is performed by consuming M copies, with M small but non-vanishing, then η_E is likely to be low. Indeed, if we spend few copies to reconstruct the state, we might obtain a low-fidelity reconstruction that in turn results in a sub-optimal ergotropic unitary transformation and small extracted work. Thus, a maximum of η_E must be achieved for a (state-dependent) optimal allocation of resources between state reconstruction and work extraction.

In Figure 5.2 we show an example of our protocol applied to a single density matrix for which we compute the quantities introduced above. In Figure 5.2c) and f) we show that the ergotropic efficiency η_E decreases to 0 as the number M of copies for state reconstruction approaches $N = 10^3$ and closely follows the theoretical curve $(1 - M/N)$ for large values of M. The deviations from the curve for small values of M are due to low-fidelity reconstructions of the density matrix. Furthermore, we show that both the reconstruction efficiency η_R and the fidelity $\mathcal{F}$ between the QST reconstruction and the target state approach 100% and 1 as the number of copies M allocated for state reconstruction increases,

respectively.

5.3 Numerical results

In this section, we show the results achieved by the PIFR quantum Maxwell demon. We considered systems composed of $n_q = 2, 3, 4$ qubits and 2,000 different density matrices for each number of qubits. For each density matrix, we fixed the number N of i.i.d copies of the system, M of which are available for the state reconstruction task. The two protocols available for this task were QST and ECT-QST. Here, with QST we refer to the tomography protocol that requires 3^{n_q} measurement settings. In this case, the M copies of the state were uniformly distributed among the settings, that is, we allocated $M/3^{n_q}$ copies per setting. If the demon adopts ECT-QST instead, we first consume 20% of the available copies to determine the diagonal elements of the density matrix, and then uniformly distribute the remaining 80% of the M copies among the settings identified by ECT-QST, the number $|S_t|$ of which depends on the state under consideration. As a consequence, the number of copies per setting is state-dependent too. The outcomes of each measurement setting were obtained by sampling $M/3^{n_q}$ events or $M/|S_t|$ depending on the chosen protocol, from a multinomial distribution with probabilities given by $\{p_k = \mathrm{Tr}\,(\rho\Pi_k)\}_{k=1}^{2^n}$, with $\{\Pi_k\}_{k=1}^{2^n}$ the set of projectors onto the eigenbasis of the observable that defines the setting. Finally, the Hamiltonian considered here is a Ising Hamiltonian:

$$H = -J\sum_{i=1}^{n_q} Z_i Z_{i+1} - h\sum_{i=1}^{n_q} Z_i, \tag{5.7}$$

with $J = 5$ and $h = 2$ in our numerical experiments. However, the spectrum of this Hamiltonian is degenerate, and thus we also studied the same problem with a similar, but non-degenerate Hamiltonian.

Let us start by analyzing the results regarding the reconstruction efficiency. In this case, M copies were used to reconstruct the state without considering the extraction of work, and for each state, we calculated the reconstruction efficiency according to (5.5). In Figure 5.3 we show the reconstruction efficiency as a function of the number of copies M for states of 2, 3, and 4 qubits. The trends of QST and ECT-QST are similar to those observed in Figure 4.12, although they refer to quantities with a very different physical meaning. Both QST and ECT-QST average fidelity converges to 1 as the number of copies increases, as expected, and we can identify three different regions in their behavior. In the low statistics regime, when the number of available copies is relatively small, ECT-QST achieves a higher average reconstruction efficiency than QST. Then, a crossing point takes place when the number of copies increases up to a few hundreds. Here, the average reconstruction efficiencies of the two methods are very close. Finally, after the crossing point there is a region of high statistics, where QST performs better than ECT-QST regarding the average reconstruction efficiency. Notice that in the 4-qubit case we could study the case of $M = 75$ with ECT-QST, not suitable with QST for the number of settings was larger than the number of available copies of the states, and would have resulted in less than one copy per setting. For a given number of copies in the low statistics regime, the average reconstruction efficiency decreases as the number of qubits increases, for the same amount of (scarce) resources does not allow us to properly explore Hilbert spaces

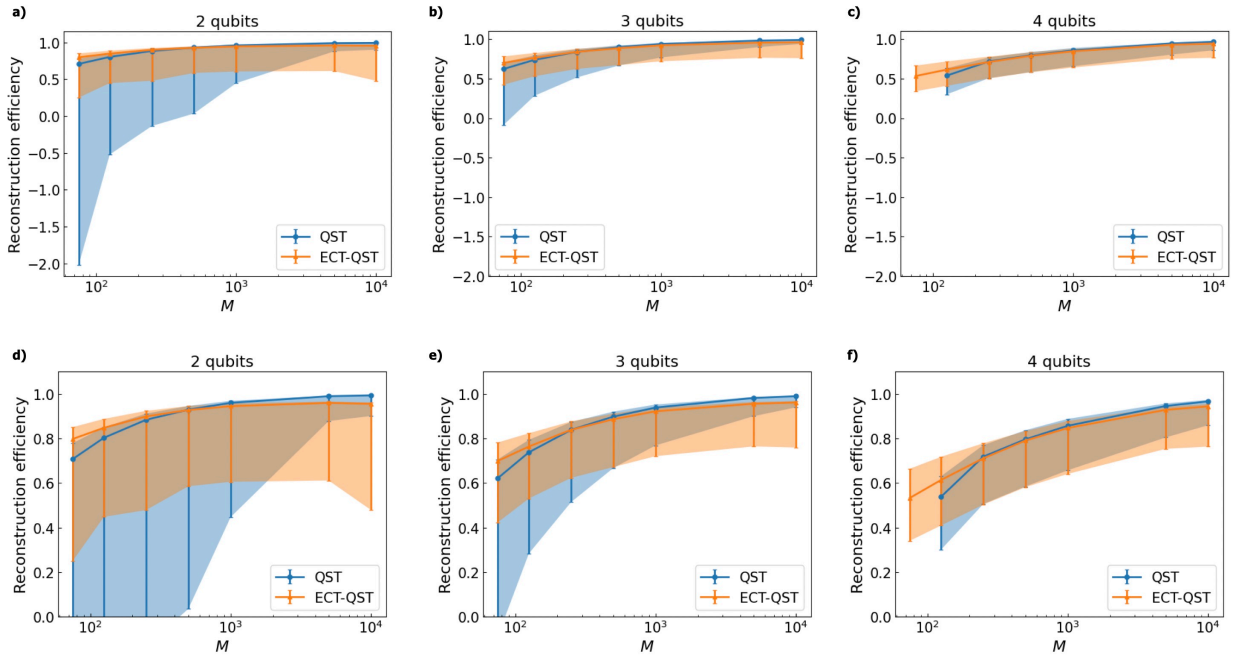


Figure 5.3: **Reconstruction efficiency of QST and ECT-QST.** We show the reconstruction efficiency of QST (blue line) and ECT-QST (orange line) for **a), d)** 2, **b), e)** 3, and **c), f)** 4 qubits. For each density matrix, we used M copies to reconstruct the state and then estimate $U_{\mathcal{W},I}$ and $\mathcal{W}_I$. The average was computed over the 2,000 density matrices composing the dataset of each number of qubits. The error bars represent one standard deviation from the average, and were computed asymmetrically, that is, the upper (lower) error bar quantifies the standard deviation of the data that are above (below) the average, respectively. Plots on the same row have the same scale on the vertical axis. Figure to appear in [6].

of increasing dimensions. Another noteworthy point is that it is possible for η_R to become negative. Although counterintuitive, we believe that there is a mathematical and a physical explanation of this. Reconstruction efficiency may be negative as a consequence of poor-quality reconstruction of the state that results in an ergotropic unitary $U_{\mathcal{W},I}$ very different from the optimal one. If we then use this low-quality unitary to estimate the ergotropy according to (5.5), it can happen that $\mathcal{W}_I$ is negative, and so η_R. This is related to the physical meaning of ergotropy. Indeed, the application of the optimal ergotropic unitary $U_{\mathcal{W}}$ to a state drives it toward a passive state and allows us to extract work. If the applied unitary is very far from being optimal, we may drive it toward a state that is very different from a passive one, such as an excited state of the Hamiltonian, which means that we are supplying work to the system instead of extracting it.

In addition to the average, the behavior of the standard deviation is interesting as well. In fact, the standard deviation of ECT-QST varies little, staying almost constant through all the values of M, while that of QST is greatly reduced when M increases. In the 2- and 3-qubit cases, in the low- and intermediate-statistics regimes, ECT-QST results are much less spread than those of QST. If we look at the high-statistics regime instead, we notice that the QST results are much less spread than the ECT-QST. In this regime QST can obtain much more precise results as it is possible to spend a sufficiently large number of copies on the maximum number of measurement settings, thus ensuring high-quality reconstructions of the state and estimates of the ergotropic unitary transformation. In the 4-qubit case the results are slightly different. While in the high-statistics regime the conclusions are the same as of the other number of qubits, in the low- and intermediate-statistics regimes the standard deviations of the two methods are comparable. This may be a peculiar feature of the protocol, or may be due to the limited number of samples analyzed in our experiments. A deeper analysis of this fact and an answer to this issue will be the subject of future studies. However, we can conclude that if enough copies of the state are available, QST is the best solution. Otherwise, ECT-QST allows for a more precise and meaningful allocation of the resources by identifying the most informative settings on which to spend the copies of the state.

Let us now consider the results for the ergotropic efficiency. We set $N = 10^3$, and considered different numbers M of copies used for state tomography. All other settings of our experiments remain unchanged. In Figure 5.4 we show the ergotropic efficiency of QST and ECT-QST for 2, 3, and 4 qubits. We also show the theoretical curve given by $1-M/10^3$ that represents the trend of the ergotropic efficiency with the perfect reconstruction efficiency, that is, $\eta_R = 1$. Consequently, the gray region above the theoretical curve is unreachable for any reconstruction protocol. If $M = N$, no copies are available for work extraction, for they have been used to reconstruct the state, and thus the ergotropic efficiency is 0. If $M \lesssim N$, QST performs slightly better than ECT-QST, achieving similar average and standard deviation values of ergotropic efficiency. Decreasing the number of copies earmarked for state reconstruction means increasing the amount for work extraction, allowing a maximum value of the ergotropic efficiency. In the low-statistics regime, the average ergotropic efficiency of ECT-QST is higher than that of QST, with smaller error bars for 2 and 3 qubits, while for 4 qubits the standard deviations are comparable. The trends of the error bars, the use of 75 copies in the 4-qubit case, and the comment on the average ergotropic efficiency for a given number of copies M mirror those observed in Figure 5.3. Finally, we observe that the closer a curve is to the theoretical one, the better the corresponding protocol is, for it achieves results that are close to optimal.

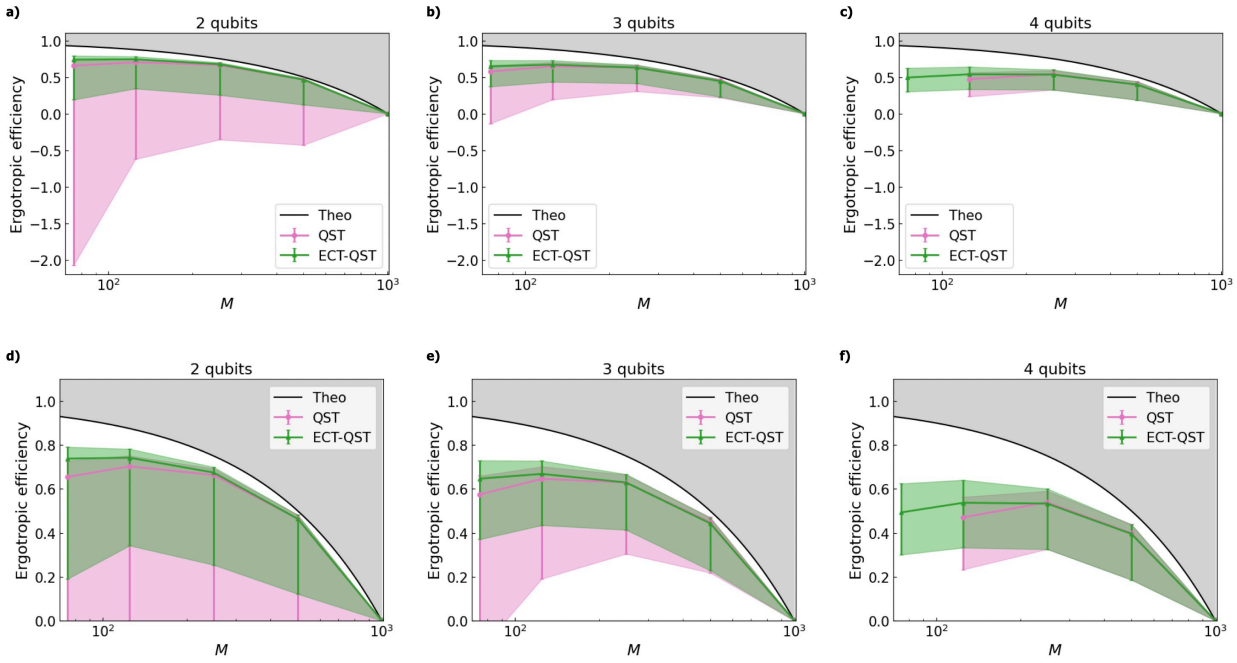


Figure 5.4: **Ergotropic efficiency of QST and ECT-QST.** We show the ergotropic efficiency of QST (pink line) and ECT-QST (green line) for **a), d)** 2, **b), e)** 3, and **c), f)** 4 qubits. For each density matrix, we fixed $N = 10^3$ and used M copies to reconstruct the state and then compute the ergotropic efficiency on the remaining $(N - M)$ according to (5.6). The average was computed over the 2,000 density matrices composing the dataset of each number of qubits. The error bars represent one standard deviation from the average, and were computed asymmetrically, that is, the upper (lower) error bar quantifies the standard deviation of the data that are above (below) the average, respectively. Plots on the same row have the same scale on the vertical axis. Figure to appear in [6].

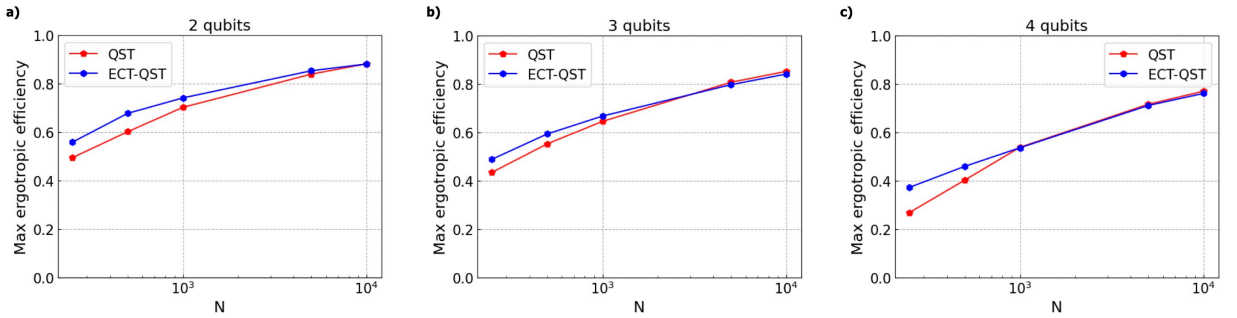


Figure 5.5: **Maximum of ergotropic efficiency of QST and ECT-QST as a function of N.** We show the maximum of ergotropic efficiency $(\max_M \eta_E)$ as a function of the number N of total copies of the state available for QST (red line) and ECT-QST (blue line) for **a)** 2, **b)** 3, and **4)** qubits, respectively. Figure to appear in [6].

The previous results depend on the total amount of resources, that is, different values of N might result in different performances of the two reconstruction protocols. For this reason, we studied the maximum η_E achievable as a function of N. We report the results in Figure 5.5. We note the same behavior of the three regimes with low, intermediate, and high statistics as in the previous results. Although evident for 3 and 4 qubits, in the 2-qubit case the crossing point after which QST performs better than ECT-QST can barely be observed, for it is very close to the rightmost point of the plot. These results further confirm that the ECT-QST reconstruction protocol is more efficient than QST for quantum work extraction in the regime of scarce resources.

As a final note, in Figure 5.6 we report the same quantities shown in Figures 5.3, 5.4, and 5.5 for the following non-degenerate Ising Hamiltonian with an exponentially decaying field:

$$H = -J \sum_{i=1}^{n_q} Z_i Z_{i+1} - h_i \sum_{i=1}^{n_q} Z_i, \tag{5.8}$$

with $h_i = 0.5 h_{i-1}$ and $h_0 = 2$. The results closely resemble those achieved for the degenerate Ising Hamiltonian.

5.4 Final remarks

In this chapter, we revisited the long-standing problem of Maxwell's demon and proposed a quantum version that accounts for realistic constraints, namely partial information and finite resources. This PIFR quantum Maxwell demon provides a framework for exploring the thermodynamic role of information under practical conditions. We introduced two figures of merit to assess the effectiveness of our approach, the reconstruction efficiency and the ergotropic efficiency, which together capture the balance between accurate state characterization and extractable work.

We compared QST and ECT-QST as reconstruction protocols within this framework. The results show that the relationship between state reconstruction and work extraction is far from straightforward. In particular, the maximum extractable work depends sensitively on how resources are distributed between these two tasks, and this optimal allocation is itself state-dependent.

A closer look revealed three distinct statistical regimes. In the high-statistics regime, QST outperforms ECT-QST, as the large number of copies makes full tomography feasible. In the intermediate regime, the two approaches yield comparable results, reflecting a balance between measurement completeness and resource limitations. In contrast, in the low-statistics regime, ECT-QST has a clear advantage. Its strength lies in its ability to identify the most informative measurement settings, thereby channeling scarce resources into the directions that matter most. This selective strategy not only enhances average reconstruction and ergotropic efficiencies but also makes the protocol viable in scenarios where QST would fail altogether, such as with only $M = 75$ copies of a four-qubit state.

Despite this success, the performance gap between ECT-QST and the theoretical optimum widens as statistics decrease. This indicates that there is still room for significant improvement. More advanced reconstruction methods, such as modern deep learning approaches, alternative likelihood functions, or refined strategies to distribute the available copies among measurement settings, could potentially reduce this gap. In particular,

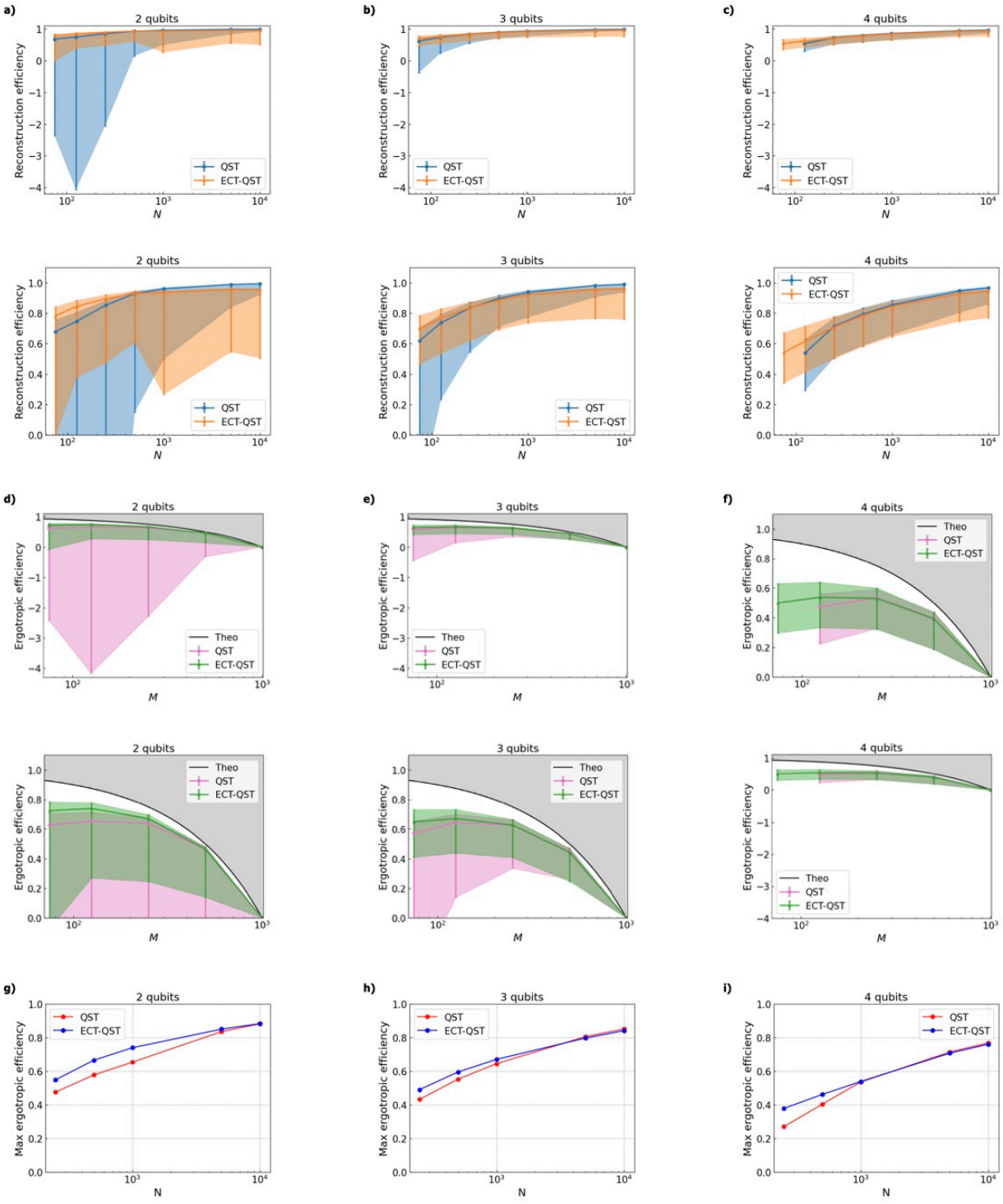


Figure 5.6: **Results of QST and ECT-QST on quantum work extraction for the non-degenerate Hamiltonian** (5.8)**.** We show the same quantities as in Figures 5.3, 5.4, and 5.5 for the non-degenerate Hamiltonian of (5.8). Panels **a) - c)**, **d) - f)**, and **g) - i)** show the results of η_R, η_E, and $\max_M \eta_E$, respectively. Figure to appear in [6].

further development of adaptive or weighted measurement schemes appears especially promising. Future studies will focus on optimizing these strategies and testing their scalability to larger systems. Another avenue for improvement lies in the size of the datasets used to validate the protocols. Our results suggest that 2,000 density matrices may be insufficient for the four-qubit case, hinting that larger datasets could yield more reliable conclusions and uncover subtler trends. This points to the importance of systematic investigations into the size of the dataset as a parameter in its own right.

Our findings underscore the thermodynamic value of selective measurement and illustrate how careful resource allocation can enhance both information gain and energetic performance. Beyond their conceptual significance, these results have direct implications for the design of quantum technologies operating under strict resource constraints. They highlight the thermodynamic advantage of judicious resource management and measurement selection, suggesting concrete strategies to minimize energetic costs while maximizing efficiency. In this way, our work contributes to the broader effort to understand and optimize the energetic footprint of quantum technologies, with relevance for quantum thermodynamics, metrology, and computation.

Conclusions

In this research, we investigated and developed original and efficient QST protocols. Our main goal was to develop QST methods with some desirable features, such as no assumptions on the state to be reconstructed, a reduction of the number of required observables, computational efficiency, and experimental feasibility. The core idea to address these objectives is the introduction of a threshold parameter that provides a balanced compromise between the accuracy of state reconstruction and the number of required observables.

In the first chapter of the manuscript, we introduced the essential elements of quantum mechanics and QST, and reviewed the literature and the state of the art about QST. In the second chapter, we studied the first protocol we developed, called threshold QST or tQST for short, analyzing the results from numerical simulations and two experimental implementations on a superconducting and a photonic platform. We reconstructed different kind of states of up to 7 qubits, from random to maximally entangled. The results showed that tQST can require substantially fewer projective measurements than QST without compromising the fidelity of the reconstruction, as demonstrated by the average fidelity between the tQST reconstructions and 7-qubit target states implemented on IBMQ above 0.87.

Several elements of the proposed protocol could be generalized or replaced to evaluate its efficacy in different contexts. Thus, in the third chapter we studied how to embed a permutation-equivariant deep learning model into the tQST protocol and how to extend it to quantum state characterization tasks, in particular purity estimation. We trained and tested our model on noiseless and noisy data up to 4 qubits, demonstrating the robustness of the proposed model to noise and experimental imperfections. Although our model has yet to surpass MLE in overall performance, it offers a key advantage: the ability to reconstruct the underlying noiseless state from noisy measurement outcomes, while MLE can only recover the noisy state. As a result, MLE performance deteriorates with increasing noise, while our model remains robust. Moreover, inference times after training are comparable between our model and MLE. In the fourth chapter, inspired by compressed sensing, we generalized tQST to systems composed of qudits where we can implement measurement settings. This new protocol, called ECT-QST, requires fewer settings than those of traditional QST, and can reconstruct up to 7 qubits without a significant decrease in the fidelity. We numerically validated ECT-QST on computer-generated data and experimentally tested it on the same two platforms as tQST. All these results allowed us to conclude that we reached the purposes established at the beginning of this investigation, that is, the development of efficient QST protocols that allow one to reconstruct high-dimensional quantum states. This was also possible due to the public Github libraries that we developed and our experimental collaborators could use independently without the need for specific training.

Finally, in the fifth chapter we considered QST and ECT-QST in the framework of quantum thermodynamics, specifically to study the problem of quantum work extraction. We developed a quantum version of the Maxwell demon that deals with finite resources (number of available copies of the state) and partial information about the state, called the PIFR quantum Maxwell demon. This resembles the classical Maxwell demon that observes the state and, accordingly, performs an action on the state. We demonstrated that the maximum extractable work depends on how the available copies of the states are distributed between the state reconstruction and the work extraction tasks. In this context, ECT-QST achieved a higher average efficiency than QST in a regime of low statistics or scarce resources. This happens because ECT-QST can identify a meaningful core set of measurement settings that allows an optimal allocation of the resources among the settings.

We believe that several directions can be explored to improve the proposed protocols. For example, the combined use of tQST or ECT-QST with other existing protocols can lead to further advantages. For instance, shadow tomography might help in understanding, and eventually improving, the sample complexity of tQST and ECT-QST. Finding ways to integrate available information on the state, such as matrix product or tensor network structures, should result in further reduction of the number of required observables. More specialized machine learning and deep learning models, such as attention-based ones, might generalize our models to estimate other properties than the purity and scale our models to many-body quantum states. Furthermore, one could look for other platforms on which to experimentally apply the proposed protocols, such as those based on trapped ions, neutral atoms, or systems where information is encoded in continuous degrees of freedom. In this respect, one might envision to introduce a threshold to reconstruct representation of states such as the Wigner function, and reduce the number of experimental measurements required to carry out this task.

In conclusion, this work has successfully achieved the goals we set out to accomplish. The collaboration with scientists from the European Centre for Theoretical Studies in Nuclear Physics and Related Areas (ECT*, Fondazione Bruno Kessler), Sapienza Università di Roma, and Università di Pavia was invaluable, enriching both the scope and impact of our research. It has been especially rewarding to see our method adopted and applied by others, highlighting its relevance and potential within the broader scientific community.

Appendix A

Supplementary Information on IBMQ experiments

In this appendix we provide further information about our experiments of tQST and ECT-QST on the IBMQ platform.

A.1 IBMQ device specifications and performance metrics

We first report detailed specifications and performance characteristics of the device used in our experiments for ECT-QST, the 7-qubit IBMQ `Lagos` quantum processor. Figure A.1 shows the coupling map of `Lagos` processor, that is, how the seven qubits are coupled with each other. Specifically, qubits 1 and 5 are coupled to three other qubits, qubit 3 to two other qubits, and finally qubits 0, 2, 4, and 6 to only one qubit. In Table A.1 we report the qubit frequency, the coherence times ($T1$ and $T2$), and the measurement fidelity, which represent the probability of correct qubit readout. In Table A.2 we report the single- and two-qubit gate fidelities (Gate F) and the gate operation time, or gate length. We computed the gate fidelity from the reported gate error as: Gate $F = 1 -$ Gate Error.

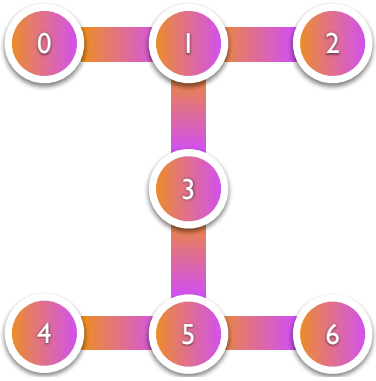


Figure A.1: **Coupling map of the IBMQ Lagos processor.** Each circle represents a different qubit, with a number at its center that identifies the specific qubit. A line between two circles represents the coupling between the two corresponding qubits.

Qubit	Frequency (GHz)	$T1\,(\mu s)$	$T2\,(\mu s)$	Measurement F
0	5.24	125.05	42.56	0.9872
1	5.10	87.70	67.16	0.9911
2	5.19	146.86	87.11	0.9893
3	4.99	140.72	132.11	0.9957
4	5.29	184.47	33.19	0.9927
5	5.18	100.70	94.19	0.9900
6	5.06	123.59	95.05	0.9939

Table A.1: **Qubit properties of the IBMQ Lagos processor.** For each qubit, numbered from 0 to 6, we report the frequency, the coherence times $T1$ and $T2$, and the measurement fidelity (Measurement F).

Gate	Qubits	Gate F	Gate length (ns)
I	0 - 6	0.9995 - 0.9998	35.6
RZ	0 - 6	1.0000	0.0
$\sqrt{X}$	0 - 6	0.9995 - 0.9998	35.6
X	0 - 6	0.9995 - 0.9998	35.6
CX	$0 \rightleftarrows 1$	0.9919	305.8 / 341.3
	$1 \rightleftarrows 2$	0.9944	291.6 / 327.1
	$1 \rightleftarrows 3$	0.9945	298.7 / 334.2
	$3 \rightleftarrows 5$	0.9945	298.7 / 334.2
	$4 \rightleftarrows 5$	0.9857	327.1 / 362.7
	$5 \rightleftarrows 6$	0.9926	256.0 / 291.6

Table A.2: **Gate fidelities and lengths for single- and two-qubit gates.** We report the gate fidelity (Gate F) and the gate length for each single- and two-qubit used in our experiments. The name of the gates are reported according to the `qiskit` convention. The first (second) value of the gate length for the CX gate refers to the rightward (leftward) direction of the corresponding second column.

A.2 Generation of W states

The authors of [403] propose a circuit with logarithmic time complexity to generate n_q-qubit W states. This circuit minimizes depth and is based on a fundamental two-qubit gate block $B(p)$, with $(0 < p < 1)$, shown in Figure A.2. The block consists of a controlled-$G(p)$ rotation, equivalent to a controlled rotation $U(2\arccos\sqrt{p}, 0, 0, 0)$, on the second qubit, followed by an inverted CNOT, that is, a CNOT with the second qubit as control qubit. In practice, we need only the following relations:

$$B(p)\,|00\rangle = |00\rangle\,, \tag{A.1a}$$

$$B(p)\,|10\rangle = \sqrt{p}\,|10\rangle + \sqrt{1-p}\,|01\rangle\,. \tag{A.1b}$$

In the appendix of the same paper [403] the authors present an algorithm to construct the circuit to generate n_q-qubit W states. However, this algorithm does not actually generate W states for certain numbers of qubits. We verified it up to 20 qubits and discovered

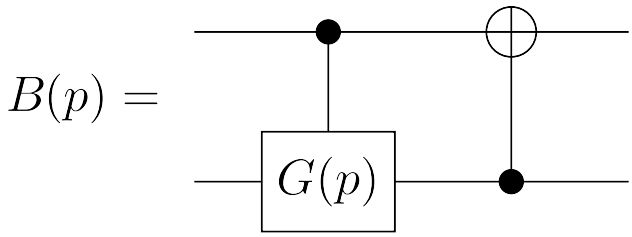


Figure A.2: **Fundamental block to generate W states.** It is composed of a controlled-$G(p)$ rotation on the second qubit, followed by an inverted CNOT.

that the proposed circuit generates an incorrect density matrix for n_q =10, 14, 18, 19 and 20. With reference to the original text, the correct algorithm can be obtained by:

- Removing line 4;
- Replace line 5 with: "Each leaf (n, m) generates an upper child $(f(n/2), n)$ and a lower child $(f(m/2 - n/2), m - n)$, where the function f returns the lower integer part of its argument";
- Removing lines 8 and 9 (as swapping is no longer necessary).

In all of our simulations, we used the correct version of the algorithm. However, while the target W states have ${}^{\otimes n_q}\langle 0| \rho_{\text{target}} |0\rangle^{\otimes n_q} = 0$, the W states generated with this algorithm show a non-vanishing population of the ground state $|0\rangle^{\otimes n_q}$, as shown in Figure A.3. This effect is due to relaxation processes that favor the population of the ground state: the longer the state preparation, the more pronounced the relaxation [403]. This has two effects on our reconstructions. On the one hand, for $n_q \leq 7$ the ground state component $|0\rangle^{\otimes n_q}$ remains below the threshold, and the fidelity between the ECT-QST reconstruction and the target state is higher than the fidelity between IBMQ-QST reconstruction and the target state. On the other hand, the fidelity between the reconstructed and target state decreases as n_q increases as discussed in Section 4.7, due to the increasing relative importance of the ground state.

A.3 Full results for $n_q = 4, \ldots, 7$-qubit random states

We present the reconstruction with ECT-QST of pure random states generated by depth-3 random circuits for systems with n_q = 4, 5, 6, and 7, with different numbers of non-vanishing diagonal elements. Specifically, we selected 8 random circuits for n_q = 4, 5, and 6, while we selected 13 random circuits for n_q = 7. In general, we note that the fidelity between the ECT-QST and IBMQ-QST reconstructions is slightly better than that between the tQST and IBMQ-QST reconstructions, while using significantly fewer measurement settings but requiring more projective measurements, as reported in Tables A.3 - A.6. ECT-QST is more robust against noise when used to reconstruct real states, such as GHZ and W states.

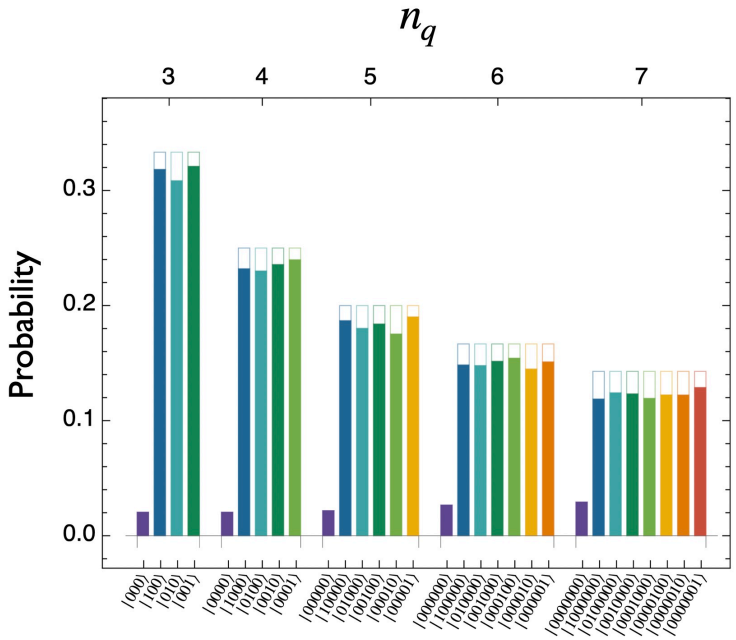


Figure A.3: **Populations of ground state of W states on IBMQ.** We show the probability of the expected non-vanishing diagonal elements and the ground state of W states generated on IBMQ with the correct version of the logarithmic time complexity circuit of [403]. On the horizontal axis, for each number n_q of qubits we report the labels of the ground state and the expected non-vanishing diagonal elements. Relaxation effects drive the non-zero population of the ground state, whose relative importance (with respect to the expected non-vanishing diagonal elements) becomes more and more relevant as n_q increases.

		IBMQ-QST			tQST				ECT-QST			
n_q	$\{\rho_{ii}\} \neq 0$	$\mathcal{F}^{\rho_{\text{IBMQ-QST}}}_{\rho_{\text{target}}}$	$\lvert S\rvert$	$\mathcal{N}_{\text{IBMQ-QST}}$	$\mathcal{F}^{\rho_{\text{tQST}}}_{\rho_{\text{IBMQ-QST}}}$	$\mathcal{F}^{\rho_{\text{tQST}}}_{\rho_{\text{target}}}$	$\lvert S_{\text{tQST}}\rvert$	$\mathcal{N}_t$	$\mathcal{F}^{\rho_{\text{ECT}}}_{\rho_{\text{IBMQ-QST}}}$	$\mathcal{F}^{\rho_{\text{ECT}}}_{\rho_{\text{target}}}$	$\lvert S'_t\rvert$	$\mathcal{N}_{\text{ECT}}$
4	2	0.964	81	1,296	0.966	0.960	3	18	0.978	0.973	3	48
	4	0.942	81	1,296	0.926	0.892	9	28	0.957	0.952	7	112
	6	0.961	81	1,296	0.966	0.963	21	46	0.977	0.975	10	160
	7	0.951	81	1,296	0.969	0.949	37	66	0.971	0.961	14	224
	8	0.956	81	1,296	0.967	0.962	17	32	0.974	0.965	15	240
	12	0.963	81	1,296	0.974	0.963	79	194	0.975	0.972	18	288
	14	0.943	81	1,296	0.953	0.942	81	232	0.960	0.951	18	288
	16	0.952	81	1,296	0.963	0.954	81	246	0.966	0.959	18	288

Table A.3: **Comparison of IBMQ-QST, tQST, and ECT-QST on 4-qubit depth-3 random circuits implemented on an IBMQ processor.** The circuits are characterized by different diagonal fillings, that is, different numbers of non-zero diagonal elements, reported in the second column. $|S| = 3^{n_q}$ is the number of measurement setting required by IBMQ-QST. $\mathcal{N}_{\text{IBMQ-QST}} = |S| \times 2^{n_q} = 6^{n_q}$ is the number of projective measurements performed with IBMQ-QST. $|S_{\text{tQST}}|$ is the number of measurement settings required to measure the projectors suggested by tQST. $\mathcal{N}_t$ is the number of projectors required by tQST. $|S'_t|$ is the number of measurement settings suggested by ECT-QST after pruning. $\mathcal{N}_{\text{ECT}} = |S'_t| \times 2^{n_q}$ is the number of projective measurements performed with ECT-QST.

		IBMQ-QST			tQST				ECT-QST			
n_q	$\{\rho_{ii}\} \neq 0$	$\mathcal{F}^{\rho_{\mathrm{IBMQ-QST}}}_{\rho_{\mathrm{target}}}$	$\lvert S\rvert$	$\mathcal{N}_{\mathrm{IBMQ-QST}}$	$\mathcal{F}^{\rho_{\mathrm{tQST}}}_{\rho_{\mathrm{IBMQ-QST}}}$	$\mathcal{F}^{\rho_{\mathrm{tQST}}}_{\rho_{\mathrm{target}}}$	$\lvert S_{\mathrm{tQST}}\rvert$	$\mathcal{N}_t$	$\mathcal{F}^{\rho_{\mathrm{ECT}}}_{\rho_{\mathrm{IBMQ-QST}}}$	$\mathcal{F}^{\rho_{\mathrm{ECT}}}_{\rho_{\mathrm{target}}}$	$\lvert S'_t\rvert$	$\mathcal{N}_{\mathrm{ECT}}$
5	4	0.919	243	7,776	0.919	0.921	41	84	0.949	0.951	18	576
	6	0.938	243	7,776	0.954	0.937	37	84	0.962	0.952	19	608
	7	0.916	243	7,776	0.927	0.920	65	116	0.947	0.943	25	800
	8	0.946	243	7,776	0.950	0.947	115	212	0.965	0.964	29	928
	12	0.906	243	7,776	0.926	0.903	79	176	0.932	0.937	26	832
	16	0.897	243	7,776	0.923	0.907	93	190	0.931	0.924	26	832
	24	0.926	243	7,776	0.949	0.929	227	544	0.953	0.964	32	1,024
	32	0.919	243	7,776	0.933	0.922	243	1,024	0.942	0.959	26	832

Table A.4: Same quantities as in Table A.3 but for 5-qubit depth-3 random circuits.

		IBMQ-QST			tQST				ECT-QST			
n_q	$\{\rho_{ii}\} \neq 0$	$\mathcal{F}^{\rho_{\mathrm{IBMQ-QST}}}_{\rho_{\mathrm{target}}}$	$\lvert S\rvert$	$\mathcal{N}_{\mathrm{IBMQ-QST}}$	$\mathcal{F}^{\rho_{\mathrm{tQST}}}_{\rho_{\mathrm{IBMQ-QST}}}$	$\mathcal{F}^{\rho_{\mathrm{tQST}}}_{\rho_{\mathrm{target}}}$	$\lvert S_{\mathrm{tQST}}\rvert$	$\mathcal{N}_t$	$\mathcal{F}^{\rho_{\mathrm{ECT}}}_{\rho_{\mathrm{IBMQ-QST}}}$	$\mathcal{F}^{\rho_{\mathrm{ECT}}}_{\rho_{\mathrm{target}}}$	$\lvert S'_t\rvert$	$\mathcal{N}_{\mathrm{ECT}}$
6	8	0.929	729	46,656	0.927	0.926	137	292	0.952	0.966	32	2,048
	9	0.823	729	46,656	0.806	0.813	91	168	0.869	0.898	25	1,600
	12	0.929	729	46,656	0.930	0.933	141	336	0.950	0.943	27	1,728
	16	0.900	729	46,656	0.909	0.900	259	586	0.939	0.949	36	2,304
	24	0.879	729	46,656	0.903	0.877	535	1,654	0.923	0.928	37	2,368
	32	0.843	729	46,656	0.866	0.843	569	1,830	0.882	0.939	41	2,624
	48	0.870	729	46,656	0.883	0.862	565	1,860	0.914	0.937	41	2,624
	64	0.910	729	46,656	0.931	0.909	529	1,306	0.944	0.957	42	2,688

Table A.5: Same quantities as in Table A.3 but for 6-qubit depth-3 random circuits.

		IBMQ-QST			tQST				ECT-QST			
n_q	$\{\rho_{ii}\} \neq 0$	$\mathcal{F}^{\rho_{\text{IBMQ-QST}}}_{\rho_{\text{target}}}$	$\lvert S\rvert$	$\mathcal{N}_{\text{IBMQ-QST}}$	$\mathcal{F}^{\rho_{\text{tQST}}}_{\rho_{\text{IBMQ-QST}}}$	$\mathcal{F}^{\rho_{\text{tQST}}}_{\rho_{\text{target}}}$	$\lvert S_{\text{tQST}}\rvert$	$\mathcal{N}_t$	$\mathcal{F}^{\rho_{\text{ECT}}}_{\rho_{\text{IBMQ-QST}}}$	$\mathcal{F}^{\rho_{\text{ECT}}}_{\rho_{\text{target}}}$	$\lvert S'_t\rvert$	$\mathcal{N}_{\text{ECT}}$
7	16	0.728	2,187	279,936	0.727	0.684	13	144	0.782	0.776	7	896
	18	0.870	2,187	279,936	0.849	0.864	91	234	0.907	0.910	41	5,248
	20	0.818	2,187	279,936	0.776	0.779	65	202	0.845	0.852	28	3,584
	24	0.841	2,187	279,936	0.822	0.832	227	504	0.884	0.903	44	5,632
	27	0.816	2,187	279,936	0.759	0.780	109	264	0.859	0.867	33	4,224
	28	0.865	2,187	279,936	0.854	0.849	99	256	0.892	0.913	31	3,968
	32	0.827	2,187	279,936	0.820	0.833	805	2,414	0.882	0.935	52	6,656
	36	0.863	2,187	279,936	0.848	0.862	511	1,630	0.888	0.924	44	5,632
	42	0.817	2,187	279,936	0.779	0.784	165	350	0.837	0.866	42	5,376
	48	0.851	2,187	279,936	0.844	0.828	325	646	0.898	0.925	50	6,400
	64	0.888	2,187	279,936	0.893	0.886	289	702	0.914	0.915	44	5,632
	72	0.909	2,187	279,936	0.909	0.901	451	1,218	0.921	0.926	59	7,552
	96	0.862	2,187	279,936	0.837	0.837	289	582	0.886	0.895	56	7,168

Table A.6: Same quantities as in Table A.3 but for 7-qubit depth-3 random circuits.

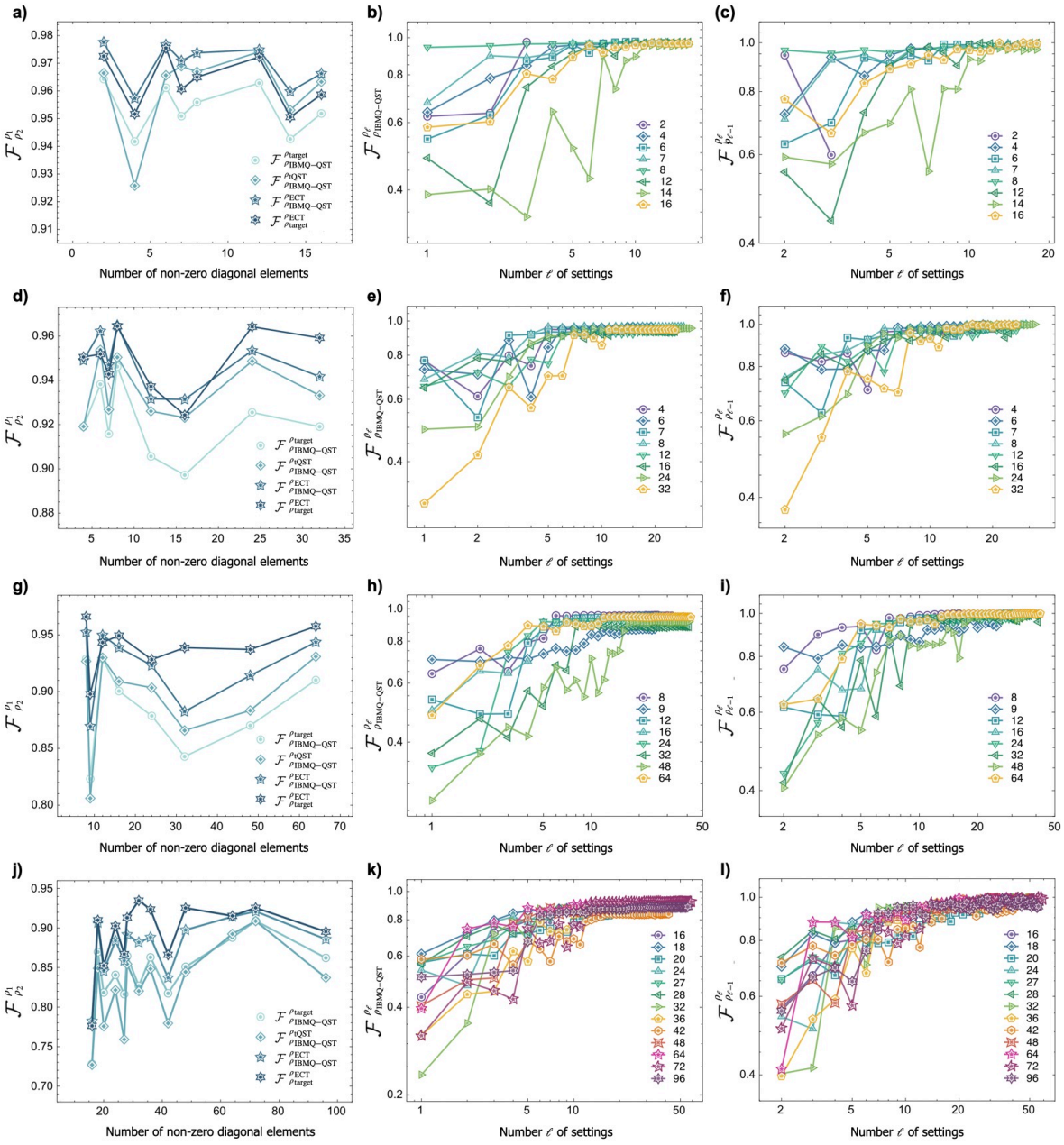


Figure A.4: **Plots of the results of ECT-QST for n_q-qubit depth-3 random circuits on IBMQ, with $n_q = 4, 5, 6, 7$. a)** Fidelities $\mathcal{F}^{\rho_{\rm target}}_{\rho_{\rm IBMQ-QST}}$, $\mathcal{F}^{\rho_{\rm tQST}}_{\rho_{\rm IBMQ-QST}}$, $\mathcal{F}^{\rho_{\rm ECT}}_{\rho_{\rm target}}$, and $\mathcal{F}^{\rho_{\rm ECT}}_{\rho_{\rm IBMQ-QST}}$ for 4-qubit random states, implemented on IBMQ depth-3 random circuits, as a function of the number of non-vanishing diagonal elements. **b), c)**: Fidelities $\mathcal{F}^{\rho_\ell}_{\rho_{\rm IBMQ-QST}}$ and $\mathcal{F}^{\rho_\ell}_{\rho_{\ell-1}}$ for the 4-qubit random states of panel **a)** as the settings identified by ECT-QST are progressively measured by decreasing weight, respectively. Each curve represents a random state with a different number of non-vanishing diagonal elements, whose symbol is reported in the legend. **d) - e) - f), g) - h) - i), j) - k) - l)**: Same as in panels **a) - c)** for $n_q = 5, 6, 7$, respectively.

Appendix B

Supplementary Information on photonic experiments

In this appendix, we provide further information about our experiments of tQST and ECT-QST on the photonic platform.

B.1 Reconfigurable integrated photonic processor

The integrated photonic circuit was fabricated using the femtosecond laser writing (FLW) platform [415]. This technique has several peculiar properties, such as permitting one to write waveguides at speeds of several centimeters per second, enabling the fabrication of complex circuits with throughputs as high as thousands per day. Furthermore, the platform benefits from minimal post-processing and no need for advanced clean room facilities, which contribute to a scalable and cost-effective manufacturing process.

The processor features a waveguide layout based on a rectangular mesh of Mach-Zehnder interferometers (MZIs) [315], which allows the implementation of arbitrary linear transformations on eight input and output modes. Each MZI unit consists of two sequential directional couplers and includes two thermal phase shifters (i.e., resistive microheaters): one for tuning the relative phase between the modes and the other for adjusting the output intensity distribution.

The waveguides were optimized for single-mode operation at wavelengths compatible with single-photon sources based on quantum dots, that is, 927 nm, and inscribed using a multi-pass technique on a Corning EAGLE XG boro-aluminosilicate glass substrate as reported in [318]. The fabrication process is followed by a thermal annealing step to relieve stresses introduced during irradiation and minimize optical losses [416].

The micro-heaters were manufactured using a two-step photolithography process, followed by selective etching of chromium (resistive elements) and copper (metal connections) [417]. Additional micro-structuring of the chip surface, namely laser ablation of deep thermal isolation trenches around the micro-heaters, was implemented to enhance power efficiency and reduce thermal crosstalk [418]. Each thermal phase shifter dissipates, on average, a maximum of 38 mW when inducing a full 2π phase shift. However, during typical operation, that is, when the device is configured with varying phases, the total power consumption across all micro-heaters stays below 1.5 W, ensuring compatibility with thermo-electric cooling. Regarding the reconfiguration speed, the 10-90%

rise/fall time of a single MZI is approximately 50 ms. Before each new measurement, we conservatively introduced a waiting time of several hundred milliseconds when applying new current settings to mitigate the effects of slow thermal crosstalk. This additional delay, however, does not block the typical data acquisition pipeline of a quantum experiment, since the data collection takes place on a much longer time scale to gather sufficient photon statistics.

The device has a footprint of 15 mm $\times$ 85 mm and features input/output fiber arrays that are directly glued to the chip to facilitate optical interfacing. Waveguide fan-in and fan-out structures were designed and fabricated to enable a smooth transition from the 127 µm fiber pitch to the 80 µm inter-waveguide spacing used within the circuit. The curvature radius of the waveguides is 20 mm, selected as the minimum value that does not introduce additional losses (bending losses).

The total fiber-to-fiber insertion loss is less than 3 dB, corresponding to a total transmission of approximately 50%. This value includes both propagation losses within the integrated circuit and coupling losses at the chip facets. Considering that the coupling losses with fiber arrays is approximately 0.5 dB per facet, the insertion losses attributable to the integrated photonic circuit itself are about 2 dB. This corresponds to an average optical loss of 0.25 dB per MZI layer.

Finally, the processor was calibrated using a coherent light source at the design wavelength of 927 nm as described in [316]. The calibration procedure consists of characterizing the relation:

$$\phi = \phi_0 + \alpha I^2, \tag{B.1}$$

with the phase vector ϕ linearly related to the square of the current vector I through the tuning the matrix coefficient α and the static phase vector ϕ_0. This approach allows us to effectively account for thermal crosstalk among the phase shifters. Nonlinear effects are neglected because they are not significant within the operating temperature range of the phase shifters. The processor was operated using multichannel driving electronics capable of setting the electrical currents with a precision of approximately 1 µA, corresponding to a phase resolution better than 1 mrad over the full tuning range from 0 to 2π. The calibration accuracy was assessed by implementing 100 Haar-randomly selected unitary transformations and evaluating the amplitude fidelity using the expression:

$$\mathcal{F}_{\text{ampl}}\left(U_{\text{set}}, U_{\text{exp}}\right) = \frac{1}{M}\text{Tr}\left(|U_{\text{set}}^{\dagger}||U_{\text{exp}}|\right), \tag{B.2}$$

with $M = 8$ is the number of modes, U_{exp} is the experimentally reconstructed matrix (moduli only) and U_{set} is the target unitary. The resulting average fidelity was ~ 0.991.

B.2 Interferometer programming for state generation

In this section, we discuss the interferometer programming procedure to generate different states encoded in the dual rail logic in post-selection. Consider the scenario where n_q input photons are injected into a M-mode linear interferometer, described by a $M \times M$ unitary matrix U that defines the input-output relations for bosonic operators as $b_j = \sum_i U_{ji} a_i$. Let $|S\rangle = |s_1, s_2, \ldots, s_M\rangle$ be the generic input state, or input configuration, in the Fock representation, with s_i the number of photons injected into the input port i. Let

then $|T\rangle = |t_1, t_2, \ldots, t_M\rangle$ be the generic output state, or output configuration, in the Fock representation, with t_j the number of photons detected by detector j. According to the bosonic transformation rule in a linear optical circuit, the transition amplitude from an input configuration $|S\rangle$ to an output configuration $|T\rangle$ reads:

$$\beta_{S\to T} = \frac{\mathrm{per}(U_{S,T})}{\sqrt{\prod_i s_i! \prod_j t_j!}}, \tag{B.3}$$

with $U_{S,T}$ the $n_q \times n_q$ matrix obtained by selecting t_j times the jth row of U and s_i times the ith column of U. Here, $\mathrm{per}(A)$ is the permanent of a $n_q \times n_q$ complex matrix A, defined as:

$$\mathrm{per}(A) = \sum_{\sigma \in S_{n_q}} \prod_{i=1}^{n_q} a_{i,\sigma(i)} \tag{B.4}$$

with S_{n_q} the (super)set of all possible permutations of the set $\{1, \ldots, n_q\}$, that is, the symmetric group of order n_q. For example, consider a 6-mode circuit described by a 6×6 unitary transformation U, $n_q = 2$ input photons and input and output configurations $|S\rangle = |1, 1, 0, 0, 0, 0\rangle$, $|T\rangle = |0, 0, 1, 1, 0, 0\rangle$, respectively. Then, $U_{S,T}$ is given by:

$$U_{S,T} = \begin{pmatrix} U_{3,1} & U_{3,2} \\ U_{4,1} & U_{4,2} \end{pmatrix}. \tag{B.5}$$

If we inject the input state $|S\rangle$ into the interferometer, then the output state will be a superposition of all the output configurations as follows:

$$|\psi\rangle = \sum_T \beta_{S\to T} |T\rangle . \tag{B.6}$$

When using this bosonic platform to encode n_q qubits in the dual rail logic with a $M = 2n_q$-mode interferometer, one needs to post-select those configurations satisfying the correct condition for the dual rail logic, that is, one photon in each of mode pairs (1,2), (3,4), ..., $(2n_q - 1, 2n_q)$. In our case, the input photons are injected in the odd ports of the interferometer, and thus the input state has $s_{2l-1} = 1$ and $s_{2l} = 0$, with $l = 1, \ldots, n_q$. Let DR be the set of possible output configurations T that satisfy the condition specified above for the dual rail logic. Then the post-selected output state reads:

$$|\psi'\rangle = \sum_{T\in \mathrm{DR}} c_T |T\rangle = \sum_{T\in \mathrm{DR}} \frac{\beta_{S\to T}}{\sqrt{p}} |T\rangle = \sum_{T\in \mathrm{DR}} \frac{\mathrm{per}(U_{S,T})}{\left(\sqrt{\prod_i s_i! \prod_j t_j!}\right)\left(\sqrt{\sum_{T\in \mathrm{DR}} |\beta_{S\to T}|^2}\right)} |T\rangle . \tag{B.7}$$

Here, p is the post-selection probability that acts as a normalization condition:

$$p = \sum_{T\in \mathrm{DR}} |\beta_{S\to T}|^2. \tag{B.8}$$

Thus, it is possible to generate different n_q-qubit states in the dual rail logic according to this procedure. This scheme relies on multi-photon interference; thus, the correct generation of a quantum state is sensitive to the presence of partial photon distinguishability

Observable	θ	ϕ
$\hat{\sigma}_x$	$\pi/2$	0
$\hat{\sigma}_y$	$\pi/2$	$\pi/2$
$\hat{\sigma}_z$	π	0

Table B.1: **Implementation of single-qubit projectors on a MZI.** Values of θ and ϕ required to program the unitary matrix in (B.9) to implement the projectors onto the eigenbasis of $\hat{\sigma}_i$, with $i \in \{x, y, z\}$.

(see Section B.3). In our experiment, such a procedure has been used to define the unitary transformation required to generate in post-selection all the resource states tested in our experiment. However, we note that the tQST protocol is completely oblivious to these calculations since all the processes are performed directly on the experimental data and no theoretical calculations are required.

This approach requires implementation of unitary transformations between the modes, which corresponds to the one performed in the first six layers of Reconfigurable Beam Splitters (RBSs) of our device. More specifically, the state preparation layers are used to implement the transformation U by tuning the parameters θ, ϕ of the RBSs. These parameters are realized in the device through a Mach-Zehnder interferometer (see also Figure 2.14), corresponding to the sequence of a phase shift ϕ (top mode), a balanced directional coupler, a second phase shift θ (top mode), and a second balanced directional coupler. This elementary cell is described by a matrix of the form:

$$U_{\mathrm{RBS}} = \begin{pmatrix} e^{i\phi}\sin(\theta/2) & \cos(\theta/2) \\ e^{i\phi}\cos(\theta/2) & -\sin(\theta/2) \end{pmatrix}. \tag{B.9}$$

A sequence of RBSs, applied according to the specific layout of the circuit, leads to a $2n_q \times 2n_q$ matrix $U = U(\{\theta_i\}, \{\phi_j\})$. This unitary matrix provides the values of the state coefficients generated in post-selection on the output modes when used in (B.7). Analogously, a single layer of RBSs acting on the qubit-mode pairs is used to measure the state in the different bases. More specifically, we need to implement in the qubit formalism the projectors corresponding to the Pauli operators $\{\hat{\sigma}_z, \hat{\sigma}_x, \hat{\sigma}_y\}$, such that single-qubit states are projected on the eigenvectors of the Pauli operators: $\{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |R\rangle, |L\rangle\}$. We can implement these operators by programming a single elementary cell U_{RBS} of our interferometer according to the values of the parameters θ and ϕ reported in Table B.1.

We now discuss the generation of random states with equally spaced values of the Gini index. The schemes illustrating how the circuit is programmed are shown in Figure B.1. We generated a set of random states for both $n_q = 2$ and $n_q = 3$ qubits, by randomly selecting the parameters $\{\theta_i\}$ and $\{\phi_j\}$, and calculating the corresponding state according to (B.7). Then, we selected 40 (10) states for $n_q = 2$ ($n_q = 3$) qubits with corresponding Gini index covering the whole range for such a parameter with equal spacing. This allowed us to select a set of states with different levels of sparsity and thus to test the protocol under different conditions.

We used the same encoding strategy to generate maximally entangled states, using the layouts reported in Figure B.2. For the Bell state $|\Psi^+\rangle$ and the GHZ states $|\mathrm{GHZ}_3\rangle$, $|\mathrm{GHZ}_4\rangle$, we identify the correct setting for the RBSs in the state preparation layer following the

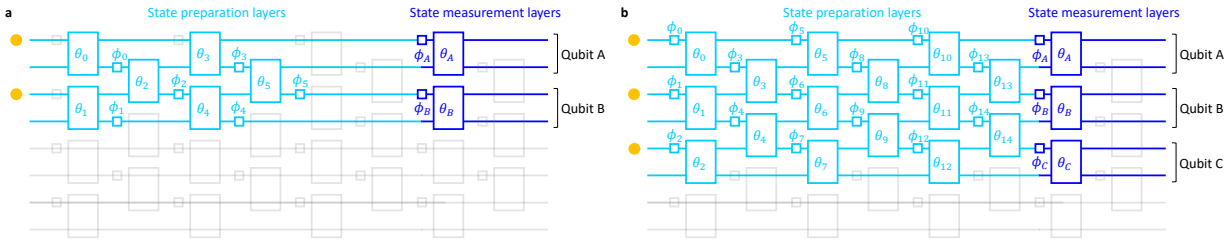


Figure B.1: **Layout of the interferometer programming for the generation of random states with variable Gini index.** Internal scheme to program the device for the generation and measurement of random state of **a** $n_q = 2$ qubits and **b** $n_q = 3$. The cyan boxes represent the RBSs actually programmed to generate the states, while blue boxes correspond to the elements for the state measurement. Grey boxes are RBSs configured to act as identity, and the grey horizontal lines are unused modes. Figure from [4].

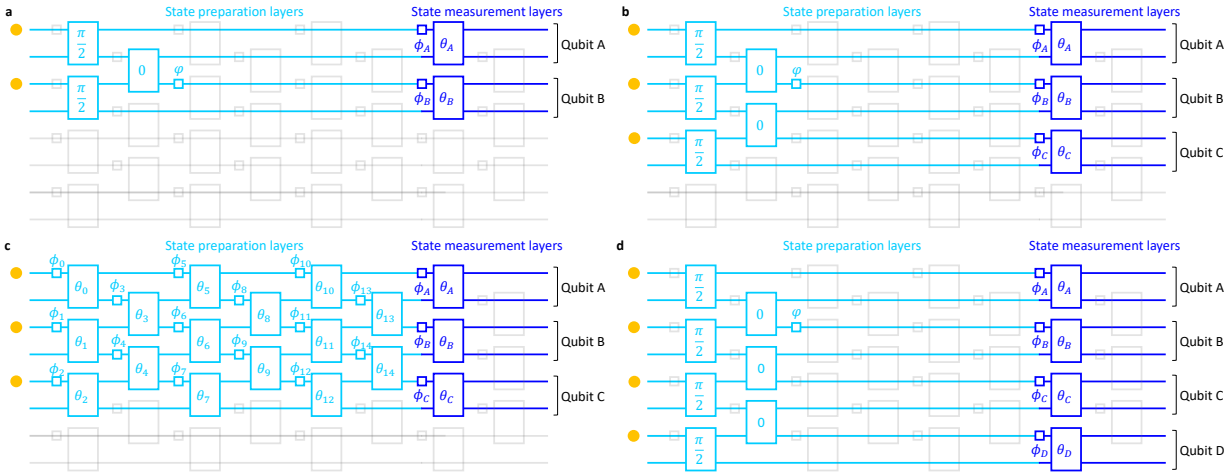


Figure B.2: **Layout of the interferometer programming for the generation of maximally entangled states.** Internal scheme to program the device for the generation and measurement of the following maximally entangled states: **a** $|\Psi^+\rangle$, **b** $|\mathrm{GHZ}_3\rangle$, **c** $|\mathrm{W}_3\rangle$, and **d** $|\mathrm{GHZ}_4\rangle$. The cyan boxes represent the RBSs actually programmed to generate the states, while blue boxes correspond to the elements for the state measurement. Grey boxes are RBSs configured to act as the identity, and the grey horizontal lines are unused modes. For $|\Psi^+\rangle$ and for the GHZ states, the values for the RBSs angles θ are reported in the cyan boxes, while for $|\mathrm{GHZ}_3\rangle$ are reported in Table B.2. Figure from [4].

approach of [318]. Generation of this class of states can be performed by setting all RBSs in the first layer to act as balanced beam splitters, while the RBSs in the second layers are all set to act as swap operations between the modes. This configuration leads to the generation of the target states, upon post-selecting those combinations satisfying the dual rail logic. The relative phase between the two superposition terms can be set by tuning a phase φ acting in one of the modes. Regarding the state $|\mathrm{W}_3\rangle$, its generation requires the full set of 6 generation layers and an additional set of phase shifts which are inserted by modifying phases ϕ_A and ϕ_C at the input of the measurement layers (see Tab. B.2). Finally, in Table B.3, we reported the corresponding expected success probabilities for the generation schemes of the maximally entangled states.

B.3 Modeling experimental imperfections

In Sections 2.10 and 4.9, we compared the reconstructed density matrices with an expectation calculated using a theoretical model that takes into account the main experimental imperfections. We now detail the model mentioned above.

The first class of imperfections in the apparatus arises from two aspects related to the quantum dot. On a first note, it is necessary to take into account multi-photon emission from the source. This corresponds to a small probability of having two photons in the same temporal bin. In our case, the extra photon has to be attributed to unfiltered residual light from the excitation laser, and is thus fully distinguishable with respects to those emitted from the quantum dot. This noise process is modeled by considering that, in each time bin, the state is described by an effective density matrix of the form: $\rho = p_0 |0\rangle\langle 0| + p_1 |1\rangle\langle 1| + p_2 |1, \tilde{1}\rangle\langle 1, \tilde{1}|$, with $\langle 1, \tilde{1}|$ representing the presence of two photons in distinguishable internal states and (p_0, p_1, p_2) the probabilities of having, respectively, 0, 1 or 2 photons in the time bin. In our case, p_0 at the emission stage is small, while p_1 and p_2 can be evaluated from the amount of measured single-photon signal and from the second-order correlation function $g^{(2)}(0) = 2p_2/(p_1 + 2p_2)^2$. The value of $g^{(2)}(0)$ can be retrieved via a Hanbury Brown and Twiss experiment [419]. The second imperfection arising from the source is related to the partial indistinguishability between the emitted photons. This is modeled via the Gram-matrix formalism [420]. In our case, the photons emitted from the quantum dot are characterized by a real-valued Gram-matrix, with elements $S_{ij} = \sqrt{M_{ij}}$, with $M_{ij} = |\langle \psi_i | \psi_j \rangle|^2$ the overlap between photons (i, j) [321]. The overlaps can be estimated from the visibilities V_{ij}^{HOM} obtained from pairwise Hong-Ou-Mandel experiments between each pair of photons [421]. In the presence of multi-photon emission with additional distinguishable noise photons, the overlaps M_{ij} are obtained from the visibilities, in the limit of low $g^{(2)}(0)$, as $M_{ij} = [V_{ij}^{\mathrm{HOM}} + g^{(2)}(0)]/[1 - g^{(2)}(0)]$ [422].

The second class of noise effects arises from the integrated photonic processor. In particular, imperfections in the device may lead to the implementation of a unitary matrix that is different from the one necessary for the generation of the target state. As discussed previously, the RBS elementary cell is implemented via a Mach-Zehnder interferometer comprising two balanced directional couplers. Small imperfections in the fabrication process lead to couplers having reflectivities slightly different than 0.5, which in our case are found in the interval [0.50, 0.58]. This affects the minimum and maximum values that the overall reflectivity of the RBS can reach via its programmability. In the ideal scenario, the overall reflectivity $R(\theta)$ of the RBS is found to be [see Eq. (B.9)] $R(\theta) = \sin^2(\theta/2)$. In con-

trast, if the directional couplers of the elementary cell are characterized by reflectivities r_1 and r_2, the reflectivity of the RBS reads $R(\theta) = r_1 r_2 + t_1 t_2 - 2\sqrt{r_1 r_2 t_1 t_2}\cos\theta$. Hence, the corresponding minimum values $R_{\min}$ and $R_{\max}$ achievable by tuning θ are, respectively:

$$R_{\min} = r_1 r_2 + t_1 t_2 - 2\sqrt{r_1 r_2 t_1 t_2}, \tag{B.10}$$

$$R_{\max} = r_1 r_2 + t_1 t_2 + 2\sqrt{r_1 r_2 t_1 t_2}, \tag{B.11}$$

where $t_i = 1 - r_i,\ i = 1, 2$. If $r_1, r_2 \neq 0.5$, this prevents reaching values close to $R \sim 0$ and $R \sim 1$ in the RBS, thus adding small errors in the state generation process.

We also take into account losses in the apparatus. In our case, the losses are found to be almost balanced between the modes and can thus be included in the model as a single round of loss after the quantum dot source [321, 423]. The only relevant contribution of unbalanced losses can be found in the different detection efficiencies of the employed detector, which have been directly corrected in the experimental data via an appropriate renormalization of the measured counts. Indeed, this is not a noise effect in the state generation process, but arises only at the detection/verification stage. Finally, we did not insert the effect of dark counts in the model, for their effect on n_q-fold coincidences in our experiment is negligible with respect to all other sources of imperfections, and in parallel no correction on the data was performed to take into account accidental coincidences.

Having included all these noise effects, the expectations of the density matrices are obtained by the following procedure. For each state to be tested, we calculated the different output probability distributions corresponding to a circuit programming that implements (i) the state generation unitary and (ii) all possible combinations of unitaries corresponding to the measurement of the Pauli operators at the measurement layers. Combining this set of probability distributions leads to a prediction for the state density matrix.

B.4 Additional data analyses and figures of merit

In this section, we report additional analyses for the experimental data reported in Section 2.10. Specifically, for the sake of completeness, in Figure B.3 we show the imaginary parts of the reconstructed density matrices of the maximally entangled states whose real parts are shown in Figure 2.16. All target states are real, that is, the imaginary part of the corresponding density matrices is zero. The slight deviations of the reconstructed imaginary parts with respect to the target ones must be attributed to experimental imperfections in the unitary transformations for both state generation and measurement stages. Finally, as we did for the random states, we complement the analysis reported in the main text by showing in Figure B.4 the fidelities $\mathcal{F}^{(\alpha)}_{\mathrm{t,m}}$ between the states reconstructed with tQST and the states expected from the model, as a function of the number of projectors $\mathcal{N}$.

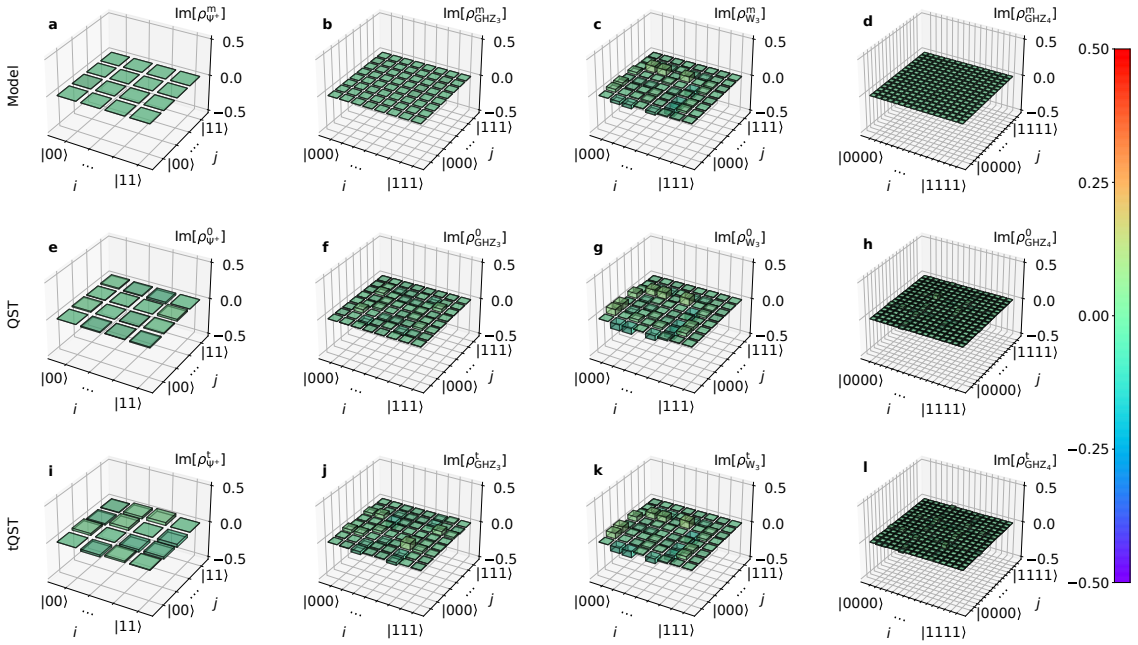


Figure B.3: **Imaginary part of the density matrices of the maximally entangled states.** We show the imaginary part of the density matrices of $|\Psi^+\rangle$ for $n_q = 2$, $|\mathrm{GHZ}_3\rangle$ and $|\mathrm{W}_3\rangle$ for $n_q = 3$ and $|\mathrm{GHZ}_4\rangle$ for $n_q = 4$ (from left to right). **a - d**: expected density matrices $(\rho_\alpha^{\mathrm{m}})$ estimated from the model taking into account experimental imperfections. **e - h**: density matrices (ρ_α^{0}) reconstructed with QST. **i - l**: density matrices $(\rho_\alpha^{\mathrm{t}})$ reconstructed with tQST. The index α labels the states as $\alpha = \Psi^+, \mathrm{GHZ}_3, \mathrm{W}_3, \mathrm{GHZ}_4$. On the right part of the figure, we report the colormap for the density matrix bars, equal for all panels **a-l**. Figure from [4].

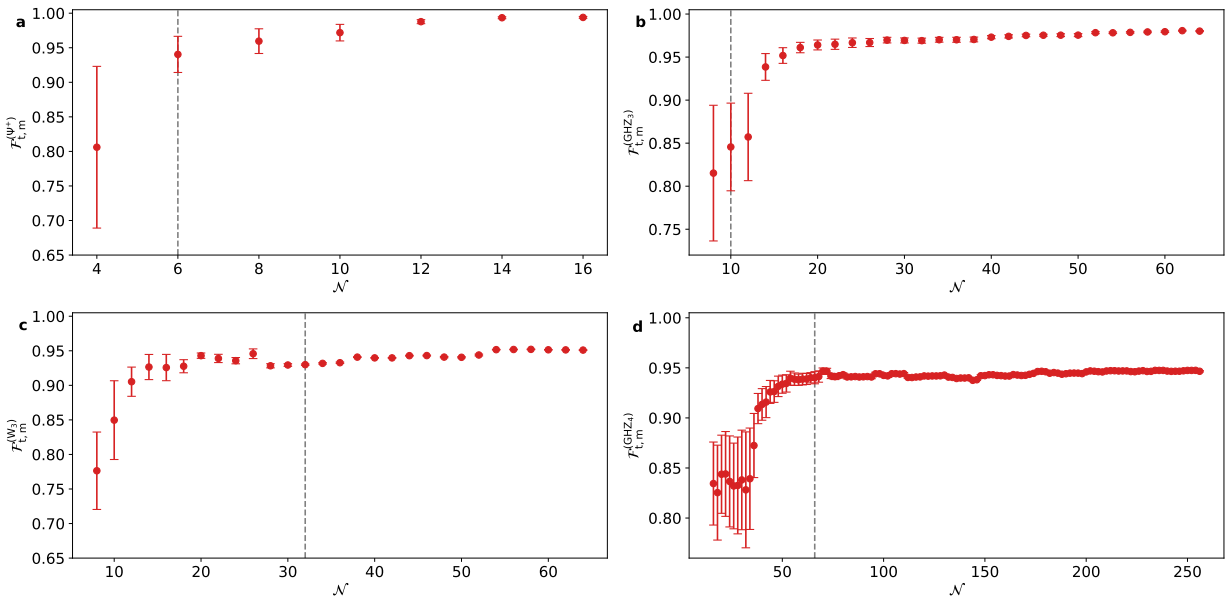


Figure B.4: **Additional figures of merit on the reconstruction of maximally entangled states.** Fidelity $\mathcal{F}^{(\alpha)}_{\mathrm{t,m}}$ with α = **a)** Ψ^+, **b)** GHZ_3, **c)** W_3, and **d)** GHZ_4. In all plots, the vertical dashed line correspond to the number of projectors obtained using the Gini threshold. Error bars on the fidelity estimations are computed by standard bootstrapping techniques, considering the standard deviation of the fidelities computed from a set of independent solutions to the state reconstruction problem. Each solution is obtained by initializing the optimization algorithm with different initial seeds, that is, by inserting a different guess for the starting point of the maximum likelihood algorithm. Figure from [4].

Parameters settings for the generation of a $\|\mathrm{W}_3\rangle$ state, in radians														
θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}	θ_{13}	θ_{14}
3.141	3.141	1.570	3.141	1.230	1.570	3.141	1.570	1.230	3.141	1.570	0	3.141	0	0
ϕ_0	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}	ϕ_{11}	ϕ_{12}	ϕ_{13}	ϕ_{14}
0	0	0	0	3.141	2.094	0	4.712	2.094	2.094	0.523	0	0	0	0

Table B.2: **Settings for the generation of a $|\mathrm{W}_3\rangle$ state.** We report the values of the parameters $\{\theta_i\}$ and $\{\phi_i\}$ to be inserted into the circuit according to the layout of Figure B.2b. The generation of this state requires two additional phase shifts in modes 1 and 5, before the measurement layers. In our experiment, this is performed by shifting the measurement phases ϕ_A and ϕ_C of the quantities $\Delta\phi_A = 2.618$, $\Delta\phi_C = -2.094$.

State	p
$\|\Psi^+\rangle$	0.5
$\|\mathrm{GHZ}_3\rangle$	0.25
$\|\mathrm{W}_3\rangle$	0.1111
$\|\mathrm{GHZ}_4\rangle$	0.125

Table B.3: **Expected success probabilities of generating the maximally entangled states.** We report the expected success probabilities p to generate the maximally entangled states in post-selection.

Appendix C

Supplementary Information on pruning in ECT-QST

C.1 Components of the rank-3 tensors of Section 4.5.1.

In this appendix, we show all the components of the rank-3 tensor $\boldsymbol{A}$ of the examples illustrated in 4.5.1. In Figure C.1 we show all the components of $\boldsymbol{A}$ of the state (4.12), which we report here for convenience:

$$|\Psi\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{3}}|02\rangle + \frac{1}{\sqrt{12}}|11\rangle + \frac{i}{\sqrt{12}}|12\rangle\,, \tag{C.1}$$

while in Figure C.2 we show all the components of $\boldsymbol{A}$ of the state (4.16) reported here below:

$$|\Phi\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{3}}|02\rangle + \frac{1}{\sqrt{12}}|10\rangle + \frac{i}{\sqrt{12}}|12\rangle\,. \tag{C.2}$$

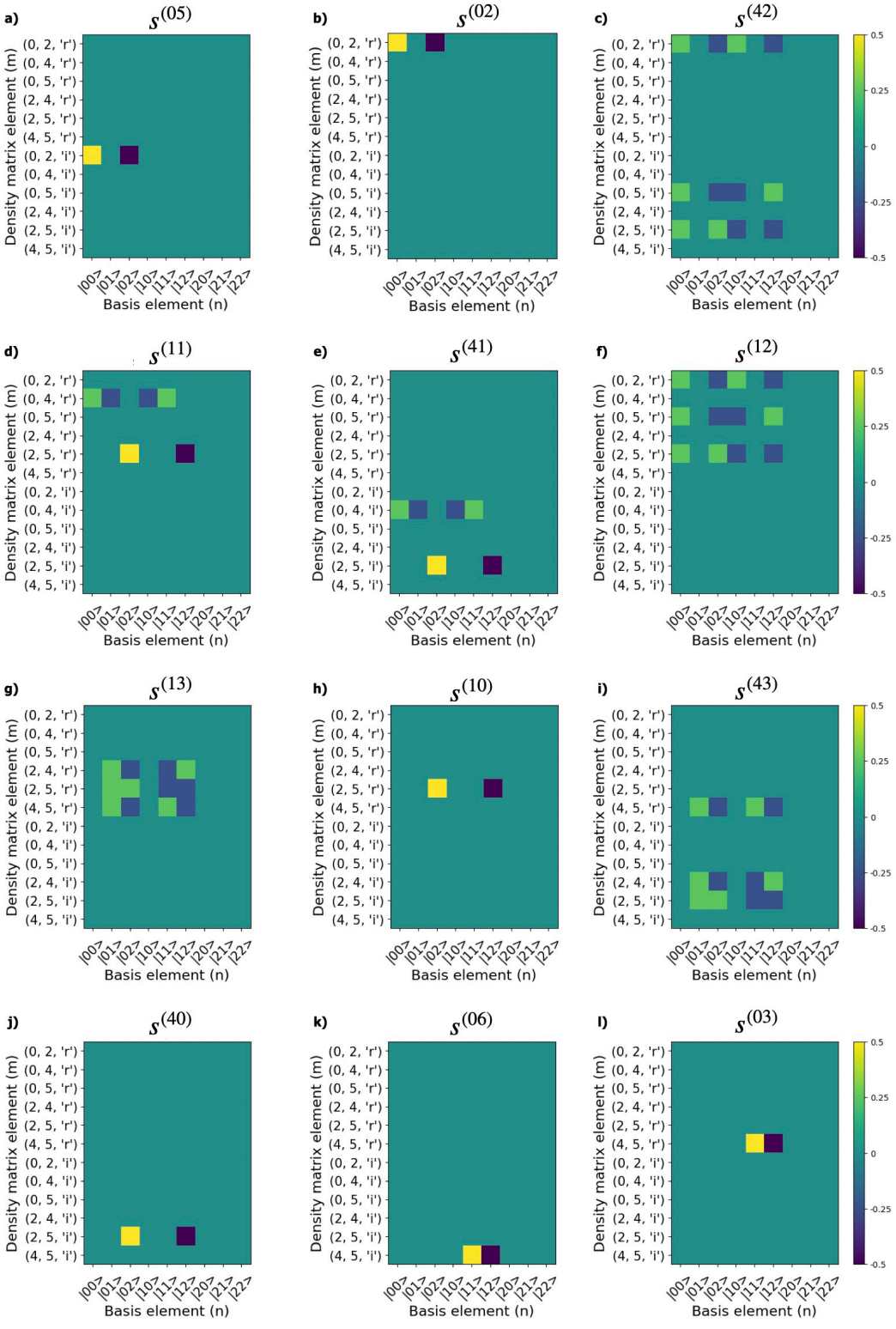


Figure C.1: **Components of the rank-3 tensor A of the state** (4.12)**.** We show the components of the rank-3 tensor $\boldsymbol{A}$ of the state (4.12) corresponding to the settings of (4.13) **a)** $A_{mn}^{(s^{(05)})}$, **b)** $A_{mn}^{(s^{(02)})}$, **c)** $A_{mn}^{(s^{(42)})}$, **d)** $A_{mn}^{(s^{(11)})}$, **e)** $A_{mn}^{(s^{(41)})}$, **f)** $A_{mn}^{(s^{(12)})}$, **g)** $A_{mn}^{(s^{(13)})}$, **h)** $A_{mn}^{(s^{(10)})}$, **i)** $A_{mn}^{(s^{(43)})}$, **j)** $A_{mn}^{(s^{(40)})}$, **k)** $A_{mn}^{(s^{(06)})}$, and **l)** $A_{mn}^{(s^{(03)})}$. The name of the setting is also reported in the title of each panel. We show only the real part for all the elements of the imaginary part are vanishing.

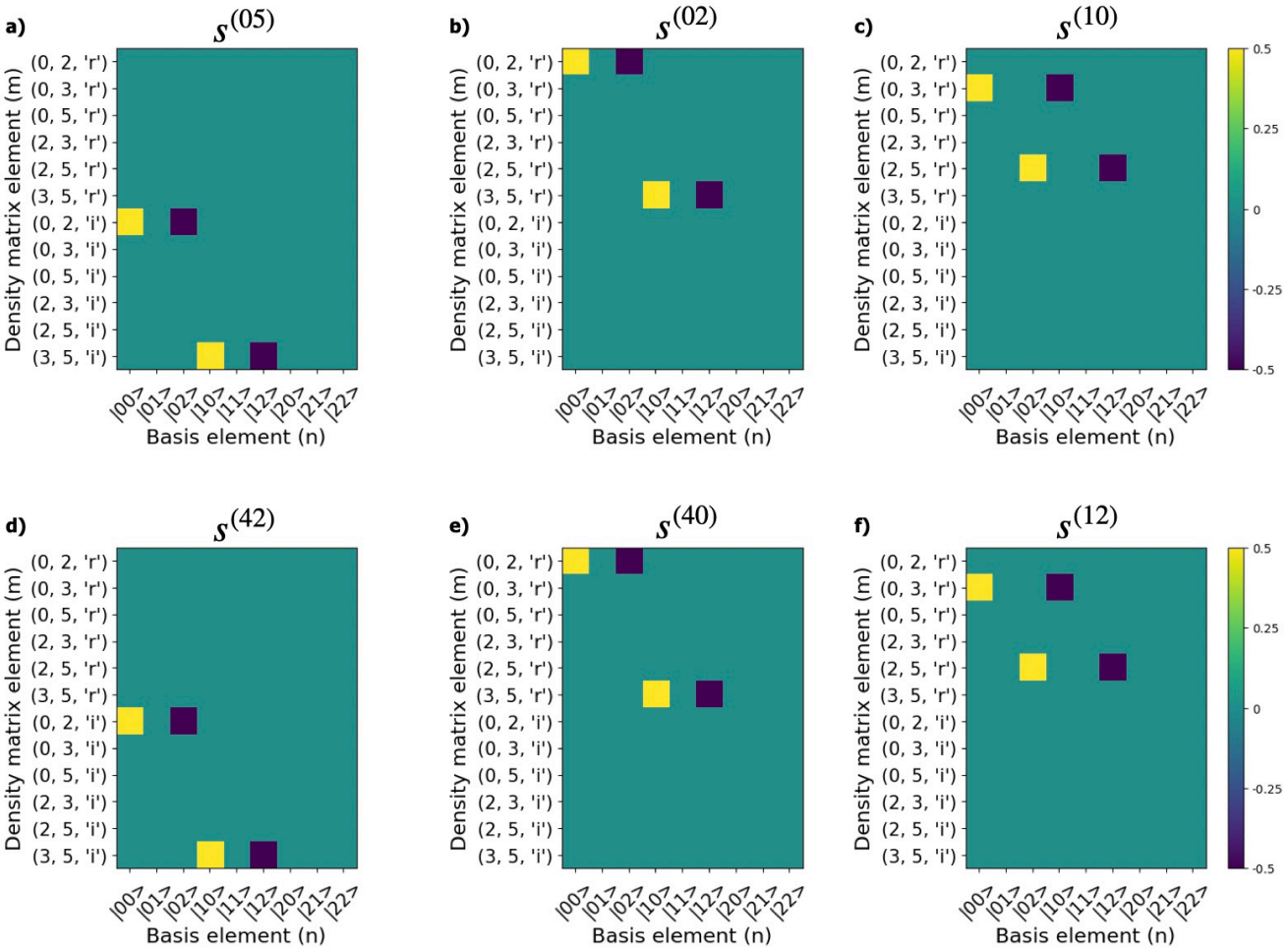


Figure C.2: **Components of the rank-3 tensor A of the state** (4.16)**.** We show the components of the rank-3 tensor $\boldsymbol{A}$ of the state (4.16) corresponding to the settings of (4.17) **a)** $A_{mn}^{(s^{(05)})}$, **b)** $A_{mn}^{(s^{(02)})}$, **c)** $A_{mn}^{(s^{(10)})}$, **d)** $A_{mn}^{(s^{(42)})}$, **e)** $A_{mn}^{(s^{(40)})}$, **f)** $A_{mn}^{(s^{(12)})}$. The name of the setting is also reported in the title of each panel. We show only the real part for all the elements of the imaginary part are vanishing.

Bibliography

[1] Daniele Binosi, Giovanni Garberoglio, Diego Maragnano, Maurizio Dapor, and Marco Liscidini. A tailor-made quantum state tomography approach. *APL Quantum*, 1(3), 2024.

[2] Giovanni Garberoglio, Maurizio Dapor, Diego Maragnano, Marco Liscidini, and Daniele Binosi. Enhanced compressive threshold quantum state tomography for qudit systems. *Physical Review A*, 111(3):032436, 2025.

[3] Diego Maragnano, Claudio Cusano, and Marco Liscidini. A permutation-equivariant deep learning model for quantum state characterization. *APL Machine Learning*, 3(2), 2025.

[4] Eugenio Caruccio, Diego Maragnano, Giovanni Rodari, Davide Picus, Giovanni Garberoglio, Daniele Binosi, Riccardo Albiero, Niki Di Giano, Francesco Ceccarelli, Giacomo Corrielli, et al. Experimental verification of threshold quantum state tomography on a fully-reconfigurable photonic integrated circuit. *npj Quantum Information*, 11(1):173, 2025.

[5] Eugenio Caruccio, Diego Maragnano, Giovanni Rodari, Davide Picus, Giovanni Garberoglio, Daniele Binosi, Riccardo Albiero, Niki Di Giano, Francesco Ceccarelli, Giacomo Corrielli, Nicolò Spagnolo, Roberto Osellame, Maurizio Dapor, Marco Liscidini, and Fabio Sciarrino. Experimental verification of enhanced compressive threshold quantum state tomography on a fully-reconfigurable photonic integrated circuit. *In preparation*.

[6] Giacomo Guarnieri, Diego Maragnano, Matteo Piccolini, and Marco Liscidini. Quantum work extraction from only partial information and finite resources. *In preparation*.

[7] Diego Maragnano, Daniele Binosi, Giovanni Garberoglio, Claudio Cusano, Maurizio Dapor, and Marco Liscidini. Threshold quantum state tomography. In *2023 23rd International Conference on Transparent Optical Networks (ICTON)*, pages 1–4. IEEE, 2023.

[8] Daniele Binosi, Diego Maragnano, Giovanni Garberoglio, Maurizio Dapor, and Marco Liscidini. Threshold quantum state tomography. In *2024 Conference on Lasers and Electro-Optics (CLEO)*, pages 1–2. IEEE, 2024.

[9] Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.

[10] David J Griffiths and Darrell F Schroeter. *Introduction to quantum mechanics*. Cambridge university press, 2018.

[11] Sheldon Axler. *Linear algebra done right*. Springer Nature, 2024.

[12] Erwin Kreyszig. *Introductory functional analysis with applications*. John Wiley & Sons, 1991.

[13] Wolfgang Dür, Guifre Vidal, and J Ignacio Cirac. Three qubits can be entangled in two inequivalent ways. *Physical Review A*, 62(6):062314, 2000.

[14] Si Chen, Li-Chao Peng, Y.-P. Guo, X.-M. Gu, X. Ding, R.-Z. Liu, J.-Y. Zhao, X. You, J. Qin, Y.-F. Wang, Yu-Ming He, Jelmer J. Renema, Yong-Heng Huo, Hui Wang, Chao-Yang Lu, and Jian-Wei Pan. Heralded three-photon entanglement from a single-photon source on a photonic chip. *Phys. Rev. Lett.*, 132:130603, Mar 2024.

[15] Mohammad Ghaderibaneh, Himanshu Gupta, and CR Ramakrishnan. Generation and distribution of ghz states in quantum networks. In *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*, volume 1, pages 1120–1131. IEEE, 2023.

[16] Jorge Miguel-Ramiro, Ferran Riera-Sàbat, and Wolfgang Dür. Quantum repeater for w states. *PRX Quantum*, 4:040323, Nov 2023.

[17] Christian L Degen, Friedemann Reinhard, and Paola Cappellaro. Quantum sensing. *Reviews of modern physics*, 89(3):035002, 2017.

[18] Luca Pezze, Augusto Smerzi, Markus K Oberthaler, Roman Schmied, and Philipp Treutlein. Quantum metrology with nonclassical states of atomic ensembles. *Reviews of Modern Physics*, 90(3):035005, 2018.

[19] Mark Hillery, Vladimír Bužek, and André Berthiaume. Quantum secret sharing. *Physical Review A*, 59(3):1829, 1999.

[20] Thomas Decker, Dominik Janzing, and Thomas Beth. Quantum circuits for single-qubit measurements corresponding to platonic solids. *International Journal of Quantum Information*, 2(03):353–377, 2004.

[21] Max S. Kaznady and Daniel F. V. James. Numerical strategies for quantum tomography: Alternatives to full optimization. *Phys. Rev. A*, 79:022109, Feb 2009.

[22] Daniel F. V. James, Paul G. Kwiat, William J. Munro, and Andrew G. White. Measurement of qubits. *Phys. Rev. A*, 64:052312, Oct 2001.

[23] G Mauro D'Ariano, Matteo GA Paris, and Massimiliano F Sacchi. Quantum tomography. *Advances in imaging and electron physics*, 128(10.1016):S1076–5670, 2003.

[24] Konrad Banaszek, Marcus Cramer, and David Gross. Focus on quantum tomography. *New Journal of Physics*, 15(12):125020, 2013.

[25] Artur Czerwinski. Selected concepts of quantum state tomography. *Optics*, 3(3):268–286, 2022.

[26] Valentin Gebhart, Raffaele Santagati, Antonio Andrea Gentile, Erik M Gauger, David Craig, Natalia Ares, Leonardo Banchi, Florian Marquardt, Luca Pezzè, and Cristian Bonato. Learning quantum systems. *Nature Reviews Physics*, 5(3):141–156, 2023.

[27] W Pauli. Handbüch der physik, edited by h. geiger and k. scheel, 1933.

[28] Edwin C Kemble. The fundamental principles of quantum mechanics: with elementary applications. 1937.

[29] W. Gale, E. Guth, and G. T. Trammell. Determination of the quantum state by measurements. *Phys. Rev.*, 165:1434–1436, Jan 1968.

[30] Stefan Weigert. Pauli problem for a spin of arbitrary length: A simple method to determine its wave function. *Phys. Rev. A*, 45:7688–7696, Jun 1992.

[31] Stefan Weigert. How to determine a quantum state by measurements: The pauli problem for a particle with arbitrary potential. *Physical Review A*, 53(4):2078, 1996.

[32] Arkadiusz Orl, Harry Paul, et al. Phase retrieval in quantum mechanics. *Physical Review A*, 50(2):R921, 1994.

[33] Ugo Fano. Description of states in quantum mechanics by density matrix and operator techniques. *Reviews of modern physics*, 29(1):74, 1957.

[34] Eyvind H Wichmann. Density matrices arising from incomplete measurements. *Journal of Mathematical Physics*, 4(7):884–896, 1963.

[35] William Band and James L Park. The empirical determination of quantum states. *Foundations of Physics*, 1(2):133–144, 1970.

[36] James L Park and William Band. A general theory of empirical state determination in quantum physics: Part i. *Foundations of Physics*, 1(3):211–226, 1971.

[37] William Band and James L Park. A general method of empirical state determination in quantum physics: Part ii. *Foundations of Physics*, 1(4):339–357, 1971.

[38] E Prugovečki. Information-theoretical aspects of quantum measurement. *International Journal of Theoretical Physics*, 16(5):321–331, 1977.

[39] William Band and James L. Park. Quantum state determination: Quorum for a particle in one dimension. *American Journal of Physics*, 47(2):188–191, 02 1979.

[40] Antoine Royer. Measurement of quantum states and the wigner function. *Foundations of physics*, 19:3–32, 1989.

[41] K Vogel and H Risken. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Physical Review A*, 40(5):2847, 1989.

[42] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani. Measurement of the wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum. *Phys. Rev. Lett.*, 70:1244–1247, Mar 1993.

[43] Asher Peres and William K. Wootters. Optimal detection of quantum information. *Phys. Rev. Lett.*, 66:1119–1122, Mar 1991.

[44] KRW Jones. Principles of quantum inference. *Annals of Physics*, 207(1):140–170, 1991.

[45] K. R. W. Jones. Fundamental limits upon the measurement of state vectors. *Phys. Rev. A*, 50:3682–3699, Nov 1994.

[46] Ulf Leonhardt. Quantum-state tomography and discrete wigner function. *Phys. Rev. Lett.*, 74:4101–4105, May 1995.

[47] Serge Massar and Sandu Popescu. Optimal extraction of information from finite quantum ensembles. *Physical review letters*, 74(8):1259, 1995.

[48] Radoslav Derka, Vladimir Buzek, and Artur K Ekert. Universal algorithm for optimal estimation of quantum states from finite ensembles via realizable generalized measurement. *Physical Review Letters*, 80(8):1571, 1998.

[49] J. I. Latorre, P. Pascual, and R. Tarrach. Minimal optimal generalized quantum measurements. *Phys. Rev. Lett.*, 81:1351–1354, Aug 1998.

[50] Richard D. Gill and Serge Massar. State estimation for large ensembles. *Phys. Rev. A*, 61:042312, Mar 2000.

[51] Zhibo Hou, Huangjun Zhu, Guo-Yong Xiang, Chuan-Feng Li, and Guang-Can Guo. Achieving quantum precision limit in adaptive qubit state tomography. *npj Quantum Information*, 2(1):1–5, 2016.

[52] Bo Qi, Zhibo Hou, Yuanlong Wang, Daoyi Dong, Han-Sen Zhong, Li Li, Guo-Yong Xiang, Howard M Wiseman, Chuan-Feng Li, and Guang-Can Guo. Adaptive quantum state tomography via linear regression estimation: Theory and two-qubit experiment. *npj Quantum Information*, 3(1):19, 2017.

[53] Luciana Pereira, Leonardo Zambrano, Jean Cortés-Vega, Sebastián Niklitschek, and Aldo Delgado. Adaptive quantum tomography in high dimensions. *Physical Review A*, 98(1):012339, 2018.

[54] GI Struchalin, EV Kovlakov, SS Straupe, and SP Kulik. Adaptive quantum tomography of high-dimensional bipartite systems. *Physical Review A*, 98(3):032330, 2018.

[55] Leonardo Zambrano, Luciano Pereira, Sebastián Niklitschek, and Aldo Delgado. Estimation of pure quantum states in high dimension at the limit of quantum accuracy through complex optimization and statistical inference. *Scientific Reports*, 10(1):12781, 2020.

[56] John A. Smolin, Jay M. Gambetta, and Graeme Smith. Efficient method for computing the maximum-likelihood quantum state from measurements with additive gaussian noise. *Phys. Rev. Lett.*, 108:070502, Feb 2012.

[57] Z. Hradil. Quantum-state estimation. *Phys. Rev. A*, 55:R1561–R1564, Mar 1997.

[58] K. Banaszek, G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi. Maximum-likelihood estimation of the density matrix. *Phys. Rev. A*, 61:010304, Dec 1999.

[59] D Mogilevtsev, Z Hradil, and J Peřina. Homodyne reconstruction of density matrix in fock-state basis: deterministic versus maximum likelihood approach. *Journal of Modern Optics*, 44(11-12):2261–2269, 1997.

[60] Zdenek Hradil, Johann Summhammer, and Helmut Rauch. Quantum tomography as normalization of incompatible observations. *Physics Letters A*, 261(1-2):20–24, 1999.

[61] Z. Hradil, J. Summhammer, G. Badurek, and H. Rauch. Reconstruction of the spin state. *Phys. Rev. A*, 62:014101, Jun 2000.

[62] J. Řeháček, Z. Hradil, and M. Ježek. Iterative algorithm for reconstruction of entangled states. *Phys. Rev. A*, 63:040303, Mar 2001.

[63] Andrew G. White, Daniel F. V. James, Philippe H. Eberhard, and Paul G. Kwiat. Nonmaximally entangled states: Production, characterization, and utilization. *Phys. Rev. Lett.*, 83:3103–3107, Oct 1999.

[64] G. Molina-Terriza, A. Vaziri, J. Řeháček, Z. Hradil, and A. Zeilinger. Triggered qutrits for quantum communication protocols. *Phys. Rev. Lett.*, 92:167903, Apr 2004.

[65] Alexander I Lvovsky. Iterative maximum-likelihood reconstruction in quantum homodyne tomography. *Journal of Optics B: Quantum and Semiclassical Optics*, 6(6):S556, 2004.

[66] Jaroslav Řeháček, Zden ěk Hradil, E. Knill, and A. I. Lvovsky. Diluted maximum-likelihood algorithm for quantum tomography. *Phys. Rev. A*, 75:042108, Apr 2007.

[67] Rainer Reichle, Dietrich Leibfried, Emanuel Knill, Joseph Britton, R Bradford Blakestad, John D Jost, Christopher Langer, Roee Ozeri, Signe Seidelin, and David J Wineland. Experimental purification of two-atom entanglement. *Nature*, 443(7113):838–841, 2006.

[68] Douglas Soares Gonçalves, Márcia A. Gomes-Ruggiero, and Carlile Lavor. Global convergence of diluted iterations in maximum-likelihood quantum tomography. *Quantum Inf. Comput.*, 14(11-12):966–980, 2014.

[69] Joseph B Altepeter, Evan R Jeffrey, and Paul G Kwiat. Photonic state tomography. *Advances in atomic, molecular, and optical physics*, 52:105–159, 2005.

[70] R. T. Thew, K. Nemoto, A. G. White, and W. J. Munro. Qudit quantum-state tomography. *Phys. Rev. A*, 66:012303, Jul 2002.

[71] Hartmut Häffner, Wolfgang Hänsel, CF Roos, Jan Benhelm, D Chek-al Kar, M Chwalla, T Körber, UD Rapol, M Riebe, PO Schmidt, et al. Scalable multiparticle entanglement of trapped ions. *Nature*, 438(7068):643–646, 2005.

[72] Douglas Soares Gonçalves, Márcia A. Gomes-Ruggiero, Carlile Lavor, Osvaldo Jiménez Farias, and Paulo H. Souto Ribeiro. Local solutions of maximum likelihood estimation in quantum state tomography. *Quantum Inf. Comput.*, 12(9-10):775–790, 2012.

[73] Tillmann Baumgratz, Alexander Nüßeler, Marcus Cramer, and Martin B Plenio. A scalable maximum likelihood method for quantum state tomography. *New Journal of Physics*, 15(12):125004, 2013.

[74] Douglas Soares Gonçalves, Márcia A Gomes-Ruggiero, and Carlile Lavor. A projected gradient method for optimization over density matrices. *Optimization Methods and Software*, 31(2):328–341, 2016.

[75] Jiangwei Shang, Zhengyun Zhang, and Hui Khoon Ng. Superfast maximum-likelihood reconstruction for quantum tomography. *Phys. Rev. A*, 95:062336, Jun 2017.

[76] Chien-Ming Lin, Yu-Ming Hsu, and Yen-Huan Li. Maximum-likelihood quantum state tomography by soft-bayes. *arXiv preprint arXiv:2012.15498*, 2020.

[77] Robin Blume-Kohout. Optimal, reliable estimation of quantum states. *New Journal of Physics*, 12(4):043034, 2010.

[78] Robin Blume-Kohout. Hedged maximum likelihood quantum state estimation. *Phys. Rev. Lett.*, 105:200504, Nov 2010.

[79] Christopher Ferrie and Robin Blume-Kohout. Minimax quantum tomography: Estimators and relative entropy bounds. *Physical Review Letters*, 116(9):090407, 2016.

[80] Travis L. Scholten and Robin Blume-Kohout. Behavior of the maximum likelihood in quantum state tomography. *New Journal of Physics*, 20(2):023050, 2018.

[81] R Derka, V Buzek, G Adam, and PL Knight. From quantum bayesian inference to quantum tomography. *arXiv preprint quant-ph/9701029*, 1997.

[82] Vladimır Buzek, G Drobny, R Derka, G Adam, and H Wiedemann. Quantum state reconstruction from incomplete data. *arXiv preprint quant-ph/9805020*, 1998.

[83] V Bužek, Radoslav Derka, G Adam, and Peter L Knight. Reconstruction of quantum states of spin systems: From quantum bayesian inference to quantum tomography. *Annals of Physics*, 266(2):454–496, 1998.

[84] Ruediger Schack, Todd A Brun, and Carlton M Caves. Quantum bayes rule. *Physical Review A*, 64(1):014305, 2001.

[85] Filippo Neri. Quantum bayesian methods and subsequent measurements. *Physical Review A—Atomic, Molecular, and Optical Physics*, 72(6):062306, 2005.

[86] Fuyuhiko Tanaka and Fumiyasu Komaki. Bayesian predictive density operators for exchangeable quantum-statistical models. *Physical Review A—Atomic, Molecular, and Optical Physics*, 71(5):052323, 2005.

[87] Th Baier, Dénes Petz, KM Hangos, and Attila Magyar. Comparison of some methods of quantum state estimation. In *Quantum probability and infinite dimensional analysis*, pages 64–78. World Scientific, 2007.

[88] Roman Schmied. Quantum state tomography of a single qubit: comparison of methods. *Journal of Modern Optics*, 63(18):1744–1758, 2016.

[89] Hsuan-Hao Lu, Joseph M Lukens, Nicholas A Peters, Brian P Williams, Andrew M Weiner, and Pavel Lougovski. Quantum interference and correlation control of frequency-bin qubits. *Optica*, 5(11):1455–1460, 2018.

[90] Hsuan-Hao Lu, Karthik V Myilswamy, Ryan S Bennink, Suparna Seshadri, Mohammed S Alshaykh, Junqiu Liu, Tobias J Kippenberg, Daniel E Leaird, Andrew M Weiner, and Joseph M Lukens. Bayesian tomography of high-dimensional on-chip biphoton frequency combs with randomized measurements. *Nature Communications*, 13(1):4338, 2022.

[91] Emma M Simmerman, H-H Lu, Andrew M Weiner, and Joseph M Lukens. Efficient compressive and bayesian characterization of biphoton frequency spectra. *Optics letters*, 45(10):2886–2889, 2020.

[92] Joseph M Lukens and Ali Passian. Bayesian inference for plasmonic nanometrology. *Physical Review A*, 104(5):053501, 2021.

[93] Joseph C Chapman, Joseph M Lukens, Bing Qi, Raphael C Pooser, and Nicholas A Peters. Bayesian homodyne and heterodyne tomography. *Optics Express*, 30(9):15184–15200, 2022.

[94] Ferenc Huszár and Neil MT Houlsby. Adaptive bayesian quantum tomography. *Physical Review A—Atomic, Molecular, and Optical Physics*, 85(5):052120, 2012.

[95] Konstantin S Kravtsov, Stanislav S Straupe, Igor V Radchenko, Neil MT Houlsby, Ferenc Huszár, and Sergei P Kulik. Experimental adaptive bayesian tomography. *Physical Review A—Atomic, Molecular, and Optical Physics*, 87(6):062122, 2013.

[96] Joseph M Lukens, Kody JH Law, and Ryan S Bennink. A bayesian analysis of classical shadows. *npj Quantum Information*, 7(1):113, 2021.

[97] Jochen Rau. Evidence procedure for efficient quantum-state tomography. *Phys. Rev. A*, 82:012104, Jul 2010.

[98] Christopher Granade, Joshua Combes, and DG Cory. Practical bayesian tomography. *New Journal of Physics*, 18(3):033024, 2016.

[99] The Tien Mai and Pierre Alquier. Pseudo-bayesian quantum tomography with rank-adaptation. *Journal of Statistical Planning and Inference*, 184:62–76, 2017.

[100] DS Gonçalves, CLN Azevedo, C Lavor, and MA Gomes-Ruggiero. Bayesian inference for quantum state tomography. *Journal of Applied Statistics*, 45(10):1846–1871, 2018.

[101] Sanjaya Lohani, Joseph M Lukens, Daniel E Jones, Thomas A Searles, Ryan T Glasser, and Brian T Kirby. Improving application performance with biased distributions of quantum states. *Physical Review Research*, 3(4):043145, 2021.

[102] Sanjaya Lohani, Joseph M Lukens, Atiyya A Davis, Amirali Khannejad, Sangita Regmi, Daniel E Jones, Ryan T Glasser, Thomas A Searles, and Brian T Kirby. Demonstration of machine-learning-enhanced bayesian quantum state estimation. *New Journal of Physics*, 25(8):083009, 2023.

[103] Jiangwei Shang, Yi-Lin Seah, Hui Khoon Ng, David John Nott, and Berthold-Georg Englert. Monte carlo sampling from the quantum state space. i. *New Journal of Physics*, 17(4):043017, 2015.

[104] Yi-Lin Seah, Jiangwei Shang, Hui Khoon Ng, David John Nott, and Berthold-Georg Englert. Monte carlo sampling from the quantum state space. ii. *New Journal of Physics*, 17(4):043018, 2015.

[105] Joseph M Lukens, Kody JH Law, Ajay Jasra, and Pavel Lougovski. A practical and efficient approach for bayesian quantum state estimation. *New Journal of Physics*, 22(6):063038, 2020.

[106] Hanson H. Nguyen, Kody J. H. Law, and Joseph M. Lukens. Unorthodox parallelization for bayesian quantum state estimation, 2025.

[107] Brian P Williams and Pavel Lougovski. Quantum state estimation when qubits are lost: a no-data-left-behind approach. *New Journal of Physics*, 19(4):043003, 2017.

[108] Jochen Rau. Inferring the gibbs state of a small quantum system. *Physical Review A—Atomic, Molecular, and Optical Physics*, 84(1):012101, 2011.

[109] Christopher Ferrie. Quantum model averaging. *New Journal of Physics*, 16(9):093035, 2014.

[110] Jochen Rau. Appearance of gibbs states in quantum-state tomography. *Physical Review A*, 90(6):062114, 2014.

[111] Richard Kueng and Christopher Ferrie. Near-optimal quantum tomography: estimators and bounds. *New Journal of Physics*, 17(12):123013, 2015.

[112] Christopher Granade, Christopher Ferrie, Ian Hincks, Steven Casagrande, Thomas Alexander, Jonathan Gross, Michal Kononenko, and Yuval Sanders. Qinfer: Statistical inference software for quantum applications. *Quantum*, 1:5, 2017.

[113] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[114] Christian G Graff and Emil Y Sidky. Compressive sensing in medical imaging. *Applied optics*, 54(8):C23–C44, 2015.

[115] Jérôme Bobin, Jean-Luc Starck, and Roland Ottensamer. Compressed sensing in astronomy. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):718–726, 2008.

[116] Jacob N Sanders, Semion K Saikin, Sarah Mostame, Xavier Andrade, Julia R Widom, Andrew H Marcus, and Alán Aspuru-Guzik. Compressed sensing for multidimensional spectroscopy experiments. *The journal of physical chemistry letters*, 3(18):2697–2702, 2012.

[117] Dharmendra Gurve, Denis Delisle-Rodriguez, Teodiano Bastos-Filho, and Sridhar Krishnan. Trends in compressive sensing for eeg signal processing applications. *Sensors*, 20(13):3703, 2020.

[118] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105:150401, Oct 2010.

[119] David Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory*, 57(3):1548–1566, 2011.

[120] David Gross and Vincent Nesme. Note on sampling without replacing from a finite collection of matrices, 2010.

[121] Vladimir Koltchinskii. Von neumann entropy penalization and low-rank matrix estimation. 39(6):2936–2973, 2011.

[122] Yi-Kai Liu. Universal low-rank matrix recovery from pauli measurements. *Advances in Neural Information Processing Systems*, 24, 2011.

[123] Steven T Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New Journal of Physics*, 14(9):095022, 2012.

[124] Yazhen Wang. Asymptotic equivalence of quantum state tomography and noisy matrix completion. *Ann. Statist.*, 41(5):2462–2504, 2013.

[125] Amir Kalev, Robert L Kosut, and Ivan H Deutsch. Quantum tomography protocols with positivity are compressed sensing protocols. *npj Quantum Information*, 1(1):1–6, 2015.

[126] Charles H Baldwin, Ivan H Deutsch, and Amir Kalev. Strictly-complete measurements for bounded-rank quantum-state tomography. *Physical Review A*, 93(5):052105, 2016.

[127] Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. *Applied and Computational Harmonic Analysis*, 42(1):88–116, 2017.

[128] Revanth Badveli, Vinayak Jagadish, R Srikanth, and Francesco Petruccione. Compressed-sensing tomography for qudits in hilbert spaces of non-power-of-two dimensions. *Physical Review A*, 101(6):062328, 2020.

[129] Wei-Tao Liu, Ting Zhang, Ji-Ying Liu, Ping-Xing Chen, and Jian-Min Yuan. Experimental quantum state tomography via compressed sampling. *Physical review letters*, 108(17):170403, 2012.

[130] Francesco Tonolini, Susan Chan, Megan Agnew, Alan Lindsay, and Jonathan Leach. Reconstructing high-dimensional two-photon entangled states via compressive sensing. *Scientific reports*, 4(1):6542, 2014.

[131] Adrian Steffens, CA Riofrío, Will McCutcheon, Ingo Roth, Bryn A Bell, Alex McMillan, MS Tame, JG Rarity, and Jens Eisert. Experimentally exploring compressed sensing quantum tomography. *Quantum Science and Technology*, 2(2):025005, 2017.

[132] Carlos A Riofrio, David Gross, Steven T Flammia, Thomas Monz, Daniel Nigg, Rainer Blatt, and Jens Eisert. Experimental quantum compressed sensing for a seven-qubit system. *Nature communications*, 8(1):15305, 2017.

[133] A Smith, CA Riofrío, BE Anderson, H Sosa-Martinez, IH Deutsch, and PS Jessen. Quantum state tomography by continuous measurement and compressed sensing. *Physical Review A—Atomic, Molecular, and Optical Physics*, 87(3):030102, 2013.

[134] D Ahn, YS Teo, H Jeong, D Koutnỳ, J Řeháček, Z Hradil, G Leuchs, and LL Sánchez-Soto. Adaptive compressive tomography: A numerical study. *Physical Review A*, 100(1):012346, 2019.

[135] Daekun Ahn, Yong Siah Teo, Hyunseok Jeong, Frédéric Bouchard, Felix Hufnagel, Ebrahim Karimi, D Koutnỳ, J Řeháček, Zdenek Hradil, Gerd Leuchs, et al. Adaptive compressive tomography with no a priori information. *Physical Review Letters*, 122(10):100404, 2019.

[136] Jano Gil-Lopez, Yong Siah Teo, Syamsundar De, Benjamin Brecht, Hyunseok Jeong, Christine Silberhorn, and Luis L Sánchez-Soto. Universal compressive tomography in the time-frequency domain. *Optica*, 8(10):1296–1305, 2021.

[137] Franz J Schreiber, Jens Eisert, and Johannes Jakob Meyer. Tomography of parametrized quantum states. *arXiv preprint arXiv:2407.12916*, 2024.

[138] Martin Kliesch, Richard Kueng, Jens Eisert, and David Gross. Improving compressed sensing with the diamond norm. *IEEE transactions on information theory*, 62(12):7445–7463, 2016.

[139] A Shabani, RL Kosut, M Mohseni, H Rabitz, MA Broome, MP Almeida, A Fedrizzi, and AG White. Efficient measurement of quantum dynamics via compressive sensing. *Physical review letters*, 106(10):100401, 2011.

[140] Andrey V Rodionov, Andrzej Veitia, R Barends, J Kelly, Daniel Sank, J Wenner, John M Martinis, Robert L Kosut, and Alexander N Korotkov. Compressed sensing quantum process tomography for superconducting quantum gates. *Physical Review B*, 90(14):144504, 2014.

[141] Akshay Gaikwad, Arvind, and Kavita Dorai. Efficient experimental characterization of quantum processes via compressed sensing on an nmr quantum processor. *Quantum Information Processing*, 21(12):388, 2022.

[142] Dmitrii Dobrynin, Lorenzo Cardarelli, Markus Müller, and Alejandro Bermudez. Compressed-sensing lindbladian quantum tomography with trapped ions. *arXiv preprint arXiv:2403.07462*, 2024.

[143] Anastasios Kyrillidis, Amir Kalev, Dohyung Park, Srinadh Bhojanapalli, Constantine Caramanis, and Sujay Sanghavi. Provable compressed sensing quantum state tomography via non-convex methods. *npj Quantum Information*, 4(1):36, 2018.

[144] Alexandra Carpentier, Jens Eisert, David Gross, and Richard Nickl. Uncertainty quantification for matrix compressed sensing and quantum tomography problems. In *High Dimensional Probability VIII: The Oaxaca Volume*, pages 385–430. Springer, 2019.

[145] Muzhou Ma, Steven T Flammia, John Preskill, and Yu Tong. Learning k-body hamiltonians via compressed sensing. *arXiv preprint arXiv:2410.18928*, 2024.

[146] ID Ivonovic. Geometrical description of quantal state determination. *Journal of Physics A: Mathematical and General*, 14(12):3241, 1981.

[147] William K Wootters and Brian D Fields. Optimal state-determination by mutually unbiased measurements. *Annals of Physics*, 191(2):363–381, 1989.

[148] Julian Schwinger. Unitary operator bases. *Proceedings of the National Academy of Sciences*, 46(4):570–579, 1960.

[149] SN Filippov and VI Man'Ko. Mutually unbiased bases: tomography of spin states and the star-product scheme. *Physica Scripta*, 2011(T143):014010, 2011.

[150] A. B. Klimov, C. Muñoz, A. Fernández, and C. Saavedra. Optimal quantum-state reconstruction for cold trapped ions. *Phys. Rev. A*, 77:060303, Jun 2008.

[151] Niklas Rohling and Guido Burkard. Tomography scheme for two spin-1 2 qubits in a double quantum dot. *Physical Review B—Condensed Matter and Materials Physics*, 88(8):085402, 2013.

[152] RBA Adamson and Aephraim M Steinberg. Improving quantum state estimation with mutually unbiased bases. *Physical review letters*, 105(3):030406, 2010.

[153] Gustavo Lima, Leonardo Neves, R Guzmán, Esteban S Gómez, WAT Nogueira, Aldo Delgado, A Vargas, and Carlos Saavedra. Experimental quantum tomography of photonic qudits via mutually unbiased basis. *Optics Express*, 19(4):3542–3552, 2011.

[154] D. Giovannini, J. Romero, J. Leach, A. Dudley, A. Forbes, and M. J. Padgett. Characterization of high-dimensional entangled systems via mutually unbiased measurements. *Phys. Rev. Lett.*, 110:143601, Apr 2013.

[155] Hao Yuan, Zheng-Wei Zhou, and Guang-Can Guo. Quantum state tomography via mutually unbiased measurements in driven cavity qed systems. *New Journal of Physics*, 18(4):043013, 2016.

[156] Yu Wang and Wei Cui. Classical shadow tomography with mutually unbiased bases. *Phys. Rev. A*, 109:062406, Jun 2024.

[157] U. Seyfarth, L. L. Sánchez-Soto, and G. Leuchs. Practical implementation of mutually unbiased bases using quantum circuits. *Phys. Rev. A*, 91:032102, Mar 2015.

[158] Wang Yu and Wu Dongsheng. An efficient quantum circuit construction method for mutually unbiased bases in n-qubit systems. *arXiv preprint arXiv:2311.11698*, 2023.

[159] Bandyopadhyay, Boykin, Roychowdhury, and Vatan. A new proof for the existence of mutually unbiased bases. *Algorithmica*, 34:512–528, 2002.

[160] Jay Lawrence, Časlav Brukner, and Anton Zeilinger. Mutually unbiased binary observable sets on n qubits. *Phys. Rev. A*, 65:032320, Feb 2002.

[161] Michel Planat, Haret C Rosu, and Serge Perrine. A survey of finite algebraic geometrical structures underlying mutually unbiased quantum measurements. *Foundations of Physics*, 36(11):1662–1680, 2006.

[162] Vincenzo D'Ambrosio, Filippo Cardano, Ebrahim Karimi, Eleonora Nagali, Enrico Santamato, Lorenzo Marrucci, and Fabio Sciarrino. Test of mutually unbiased bases for six-dimensional photonic quantum systems. *Scientific reports*, 3(1):2726, 2013.

[163] AB Klimov, Gunnar Björk, and LL Sánchez-Soto. Optimal quantum tomography of permutationally invariant qubits. *Physical Review A—Atomic, Molecular, and Optical Physics*, 87(1):012109, 2013.

[164] Thomas Durt, Berthold-Georg Englert, Ingemar Bengtsson, and Karol Życzkowski. On mutually unbiased bases. *International journal of quantum information*, 8(04):535–640, 2010.

[165] A. Klappenecker and M. Rotteler. Mutually unbiased bases are complex projective 2-designs. In *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.*, pages 1740–1744, 2005.

[166] Amir Kalev and Gilad Gour. Mutually unbiased measurements in finite dimensions. *New Journal of Physics*, 16(5):053038, 2014.

[167] Andrei-emanuel Dragomir and Radu Ionicioiu. Sic-povm tomography in integrated photonics chips. *Romanian Reports in Physics*, 75(4), 2023.

[168] Gerhard Zauner. Grundzüge einer nichtkommutativen designtheorie. *Ph. D. dissertation, PhD thesis*, 1999.

[169] Joseph M Renes, Robin Blume-Kohout, Andrew J Scott, and Carlton M Caves. Symmetric informationally complete quantum measurements. *Journal of Mathematical Physics*, 45(6):2171–2180, 2004.

[170] Isaac H. Kim. Quantumness, generalized 2-desing and symmetric informationally complete POVM. *Quantum Inf. Comput.*, 7(8):730–737, 2007.

[171] Bernhard Bodmann and John Haas. A short history of frames and quantum designs. *arXiv preprint arXiv:1709.01958*, 2017.

[172] D Marcus Appleby. Symmetric informationally complete–positive operator valued measures and the extended clifford group. *Journal of Mathematical Physics*, 46(5), 2005.

[173] R Salazar, D Goyeneche, A Delgado, and C Saavedra. Constructing symmetric informationally complete positive-operator-valued measures in bloch space. *Physics Letters A*, 376(4):325–329, 2012.

[174] David Marcus Appleby, Christopher A Fuchs, and Huangjun Zhu. Group theoretic, lie algebraic and jordan algebraic formulations of the sic existence problem. *arXiv preprint arXiv:1312.0555*, 2013.

[175] Alexey E Rastegin. Notes on general sic-povms. *Physica Scripta*, 89(8):085101, 2014.

[176] Gilad Gour and Amir Kalev. Construction of all general symmetric informationally complete measurements. *Journal of Physics A: Mathematical and Theoretical*, 47(33):335302, 2014.

[177] Huangjun Zhu. Super-symmetric informationally complete measurements. *Annals of Physics*, 362:311–326, 2015.

[178] Blake C Stacey. *A First Course in the Sporadic SICs*. Springer, 2021.

[179] Andrew James Scott and Markus Grassl. Symmetric informationally complete positive-operator-valued measures: A new computer study. *Journal of Mathematical Physics*, 51(4), 2010.

[180] A. J. Scott. Sics: Extending the list of solutions, 2017.

[181] Tuan-Yow Chien. *Equiangular lines, projective symmetries and nice error frames*. PhD thesis, ResearchSpace@ Auckland, 2015.

[182] Christopher A Fuchs, Michael C Hoang, and Blake C Stacey. The sic question: History and state of play. *Axioms*, 6(3):21, 2017.

[183] Marcus Appleby, Tuan-Yow Chien, Steven Flammia, and Shayne Waldron. Constructing exact symmetric informationally complete measurements from numerical solutions. *Journal of Physics A: Mathematical and Theoretical*, 51(16):165302, 2018.

[184] Mark D De Burgh, Nathan K Langford, Andrew C Doherty, and Alexei Gilchrist. Choice of measurement sets in qubit tomography. *Physical Review A—Atomic, Molecular, and Optical Physics*, 78(5):052122, 2008.

[185] Huangjun Zhu. Quantum state estimation with informationally overcomplete measurements. *Physical Review A*, 90(1):012115, 2014.

[186] N. Bent, H. Qassim, A. A. Tahir, D. Sych, G. Leuchs, L. L. Sánchez-Soto, E. Karimi, and R. W. Boyd. Experimental realization of quantum tomography of photonic qudits via symmetric informationally complete positive operator-valued measures. *Phys. Rev. X*, 5:041006, Oct 2015.

[187] Artur K Ekert, Carolina Moura Alves, Daniel KL Oi, Michał Horodecki, Paweł Horodecki, and Leong Chuan Kwek. Direct estimations of linear and nonlinear functionals of a quantum state. *Physical review letters*, 88(21):217901, 2002.

[188] Todd A Brun. Measuring polynomial functions of states. *arXiv preprint quant-ph/0401067*, 2004.

[189] Steven T Flammia and Yi-Kai Liu. Direct fidelity estimation from few pauli measurements. *Physical review letters*, 106(23):230501, 2011.

[190] Hyunho Cha and Jungwoo Lee. Efficient sampling for pauli measurement-based shadow tomography in direct fidelity estimation. *arXiv preprint arXiv:2501.03512*, 2025.

[191] Koenraad MR Audenaert and Martin B Plenio. When are correlations quantum?—verification and quantification of entanglement by simple measurements. *New Journal of Physics*, 8(11):266, 2006.

[192] Harald Wunderlich and Martin B Plenio. Quantitative verification of entanglement and fidelities from incomplete measurement data. *Journal of Modern Optics*, 56(18-19):2100–2105, 2009.

[193] Huangjun Zhu, Yong Siah Teo, and Berthold-Georg Englert. Minimal tomography with entanglement witnesses. *Physical Review A—Atomic, Molecular, and Optical Physics*, 81(5):052339, 2010.

[194] Da-Jian Zhang and DM Tong. Krylov shadow tomography: Efficient estimation of quantum fisher information. *Physical Review Letters*, 134(11):110802, 2025.

[195] Scott Aaronson. Shadow tomography of quantum states. In *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*, pages 325–338, 2018.

[196] Scott Aaronson. The learnability of quantum states. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463(2088):3089–3114, 2007.

[197] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.

[198] GI Struchalin, Ya A Zagorovskii, EV Kovlakov, SS Straupe, and SP Kulik. Experimental estimation of quantum state properties from classical shadows. *PRX Quantum*, 2(1):010307, 2021.

[199] Ting Zhang, Jinzhao Sun, Xiao-Xu Fang, Xiao-Ming Zhang, Xiao Yuan, and He Lu. Experimental quantum state measurement with classical shadows. *Physical Review Letters*, 127(20):200501, 2021.

[200] Hsin-Yuan Huang, Michael Broughton, Jordan Cotler, Sitan Chen, Jerry Li, Masoud Mohseni, Hartmut Neven, Ryan Babbush, Richard Kueng, John Preskill, et al. Quantum advantage in learning from experiments. *Science*, 376(6598):1182–1186, 2022.

[201] Hong-Ye Hu and Yi-Zhuang You. Hamiltonian-driven shadow tomography of quantum states. *Physical Review Research*, 4(1):013054, 2022.

[202] Hong-Ye Hu, Soonwon Choi, and Yi-Zhuang You. Classical shadow tomography with locally scrambled quantum dynamics. *Physical Review Research*, 5(2):023027, 2023.

[203] Kaifeng Bu, Dax Enshan Koh, Roy J Garcia, and Arthur Jaffe. Classical shadows with pauli-invariant unitary ensembles. *npj Quantum Information*, 10(1):6, 2024.

[204] H Chau Nguyen, Jan Lennart Bönsel, Jonathan Steinberg, and Otfried Gühne. Optimizing shadow tomography with generalized measurements. *Physical Review Letters*, 129(22):220502, 2022.

[205] Atithi Acharya, Siddhartha Saha, and Anirvan M Sengupta. Shadow tomography based on informationally complete positive operator-valued measure. *Physical Review A*, 104(5):052418, 2021.

[206] Ahmed A Akhtar, Hong-Ye Hu, and Yi-Zhuang You. Scalable and flexible classical shadow tomography with tensor networks. *Quantum*, 7:1026, 2023.

[207] You Zhou and Zhenhuan Liu. A hybrid framework for estimating nonlinear functions of quantum states. *npj Quantum Information*, 10(1):62, 2024.

[208] Simon Becker, Nilanjana Datta, Ludovico Lami, and Cambyse Rouzé. Classical shadow tomography for continuous variables quantum systems. *IEEE Transactions on Information Theory*, 2024.

[209] Jonathan Conrad, Jens Eisert, and Steven T Flammia. Chasing shadows with gottesman-kitaev-preskill codes. *arXiv preprint arXiv:2411.00235*, 2024.

[210] Srilekha Gandhari, Victor V Albert, Thomas Gerrits, Jacob M Taylor, and Michael J Gullans. Precision bounds on continuous-variable state tomography using classical shadows. *PRX Quantum*, 5(1):010346, 2024.

[211] Jonathan Kunjummen, Minh C Tran, Daniel Carney, and Jacob M Taylor. Shadow process tomography of quantum channels. *Physical Review A*, 107(4):042403, 2023.

[212] Robbie King, David Gosset, Robin Kothari, and Ryan Babbush. Triply efficient shadow tomography. *PRX Quantum*, 6:010336, Feb 2025.

[213] Xu-Jie Peng, Qing Liu, Lu Liu, Ting Zhang, You Zhou, and He Lu. Experimental shadow tomography beyond single-copy measurements. *Physical Review Applied*, 23(1):014075, 2025.

[214] Max McGinley and Michele Fava. Shadow tomography from emergent state designs in analog quantum simulators. *Physical Review Letters*, 131(16):160601, 2023.

[215] Zhen Qin, Joseph M Lukens, Brian T Kirby, et al. Enhancing quantum state reconstruction with structured classical shadows. *arXiv preprint arXiv:2501.03144*, 2025.

[216] Hsin-Yuan Huang. Learning quantum states from their classical shadows. *Nature Reviews Physics*, 4(2):81–81, 2022.

[217] Andreas Elben, Steven T Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoît Vermersch, and Peter Zoller. The randomized measurement toolbox. *Nature Reviews Physics*, 5(1):9–24, 2023.

[218] Géza Tóth, Witlef Wieczorek, David Gross, Roland Krischek, Christian Schwemmer, and Harald Weinfurter. Permutationally invariant quantum tomography. *Physical review letters*, 105(25):250403, 2010.

[219] Tobias Moroder, Philipp Hyllus, Géza Tóth, Christian Schwemmer, Alexander Niggebaum, Stefanie Gaile, Otfried Gühne, and Harald Weinfurter. Permutationally invariant state reconstruction. *New Journal of Physics*, 14(10):105001, 2012.

[220] Christian Schwemmer, Géza Tóth, Alexander Niggebaum, Tobias Moroder, David Gross, Otfried Gühne, and Harald Weinfurter. Experimental comparison of efficient tomography schemes for a six-qubit state. *Physical review letters*, 113(4):040503, 2014.

[221] Marcus Cramer, Martin B Plenio, Steven T Flammia, Rolando Somma, David Gross, Stephen D Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. Efficient quantum state tomography. *Nature communications*, 1(1):149, 2010.

[222] M. Cramer and M. B. Plenio. Reconstructing quantum states efficiently, 2010.

[223] Ben P Lanyon, Christine Maier, Milan Holzäpfel, Tillmann Baumgratz, Cornelius Hempel, Petar Jurcevic, Ish Dhand, AS Buyskikh, Andrew J Daley, Marcus Cramer, et al. Efficient tomography of a quantum many-body system. *Nature Physics*, 13(12):1158–1162, 2017.

[224] Yuan-Yuan Zhao, Zhibo Hou, Guo-Yong Xiang, Yong-Jian Han, Chuan-Feng Li, and Guang-Can Guo. Experimental demonstration of efficient quantum state tomography of matrix product states. *Optics Express*, 25(8):9010–9018, 2017.

[225] Steven T. Flammia, David Gross, Stephen D. Bartlett, and Rolando Somma. Heralded polynomial-time quantum state tomography, 2010.

[226] Olivier Landon-Cardinal, Yi-Kai Liu, and David Poulin. Efficient direct tomography for matrix product states, 2010.

[227] Tillmann Baumgratz, David Gross, Marcus Cramer, and Martin B Plenio. Scalable reconstruction of density matrices. *Physical review letters*, 111(2):020401, 2013.

[228] Alexander Lidiak, Casey Jameson, Zhen Qin, Gongguo Tang, Michael B Wakin, Zhihui Zhu, and Zhexuan Gong. Quantum state tomography with tensor train cross approximation. *arXiv preprint arXiv:2207.06397*, 2022.

[229] Steven T Flammia, Andrew Silberfarb, and Carlton M Caves. Minimal informationally complete measurements for pure states. *Foundations of Physics*, 35:1985–2006, 2005.

[230] Nan Li, Christopher Ferrie, Jonathan A Gross, Amir Kalev, and Carlton M Caves. Fisher-symmetric informationally complete measurements for pure states. *Physical review letters*, 116(18):180402, 2016.

[231] Artur Czerwinski. Quantum tomography of pure states with projective measurements distorted by experimental noise. *arXiv preprint arXiv:2012.13402*, 2020.

[232] Warley Alves and Leonardo Neves. Ptychographic estimation of pure multiqubit states in a quantum device. *APL Quantum*, 1(4), 2024.

[233] D Goyeneche, G Cañas, S Etcheverry, ES Gómez, GB Xavier, G Lima, and A Delgado. Five measurement bases determine pure quantum states on any dimension. *Physical review letters*, 115(9):090401, 2015.

[234] Claudio Carmeli, Teiko Heinosaari, Michael Kech, Jussi Schultz, and Alessandro Toigo. Stable pure state quantum tomography from five orthonormal bases. *Europhysics Letters*, 115(3):30001, 2016.

[235] L Zambrano, L Pereira, and A Delgado. Improved estimation accuracy of the 5-bases-based tomographic method. *Physical Review A*, 100(2):022340, 2019.

[236] Dylan H Mahler, Lee A Rozema, Ardavan Darabi, Christopher Ferrie, Robin Blume-Kohout, and Aephraim M Steinberg. Adaptive quantum state tomography improves accuracy quadratically. *Physical review letters*, 111(18):183601, 2013.

[237] Christopher Granade, Christopher Ferrie, and Steven T Flammia. Practical adaptive quantum tomography. *New Journal of Physics*, 19(11):113017, 2017.

[238] Hyeok Hwang, JaeKyung Choi, and Eunseong Kim. Adaptive quantum tomography in an indistinct measurement system with superconducting circuits. *Physical Review Applied*, 20(6):064007, 2023.

[239] Hannah Lange, Matjaž Kebrič, Maximilian Buser, Ulrich Schollwöck, Fabian Grusdt, and Annabelle Bohrdt. Adaptive quantum state tomography with active learning. *Quantum*, 7:1129, 2023.

[240] Christopher Ferrie. Self-guided quantum tomography. *Physical review letters*, 113(19):190404, 2014.

[241] Robert J Chapman, Christopher Ferrie, and Alberto Peruzzo. Experimental demonstration of self-guided quantum tomography. *Physical review letters*, 117(4):040402, 2016.

[242] Zhibo Hou, Jun-Feng Tang, Christopher Ferrie, Guo-Yong Xiang, Chuan-Feng Li, and Guang-Can Guo. Experimental realization of self-guided quantum process tomography. *Physical Review A*, 101(2):022317, 2020.

[243] Markus Rambach, Mahdi Qaryan, Michael Kewming, Christopher Ferrie, Andrew G White, and Jacquiline Romero. Robust and efficient high-dimensional quantum state tomography. *Physical Review Letters*, 126(10):100402, 2021.

[244] Syed Tihaam Ahmad, Ahmad Farooq, and Hyundong Shin. Self-guided quantum state tomography for limited resources. *Scientific Reports*, 12(1):5092, 2022.

[245] Ahmad Farooq, Muhammad Asad Ullah, Junaid ur Rehman, Kyesan Lee, and Hyundong Shin. Self-guided quantum state learning for mixed states. *Quantum Information Processing*, 21(7):243, 2022.

[246] Laura Serino, Markus Rambach, Benjamin Brecht, Jacquiline Romero, and Christine Silberhorn. Self-guided tomography of time-frequency qudits. *Quantum Science and Technology*, 2024.

[247] A. I. Lvovsky and M. G. Raymer. Continuous-variable optical quantum-state tomography. *Rev. Mod. Phys.*, 81:299–332, Mar 2009.

[248] Francesco Anna Mele, Antonio Anna Mele, Lennart Bittel, Jens Eisert, Vittorio Giovannetti, Ludovico Lami, Lorenzo Leone, and Salvatore FE Oliviero. Learning quantum states of continuous variable systems. *arXiv preprint arXiv:2405.01431*, 2024.

[249] Antonio Anna Mele and Yaroslav Herasymenko. Efficient learning of quantum states prepared with few fermionic non-gaussian gates. *arXiv preprint arXiv:2402.18665*, 2024.

[250] Akito Kawasaki, Ryuhoh Ide, Hector Brunel, Takumi Suzuki, Rajveer Nehra, Katsuki Nakashima, Takahiro Kashiwazaki, Asuka Inoue, Takeshi Umeki, Fumihiro China, et al. High-rate generation and state tomography of non-gaussian quantum states for ultra-fast clock frequency quantum processors. *arXiv preprint arXiv:2402.17408*, 2024.

[251] Roger Alfredo Kögler, Gabriel Couto Rickli, Renato Ribeiro Domeneguetti, Xingchen Ji, Alexander L Gaeta, Michal Lipson, Marcelo Martinelli, and Paulo Nussenzveig. Quantum state tomography in a third-order integrated optical parametric oscillator. *Optics Letters*, 49(11):3150–3153, 2024.

[252] Luis Roa, Carla Hermann-Avigliano, Roberto Salazar, AB Klimov, B Burgos, and A Delgado. Petal-shape probability areas: complete quantum state discrimination. *arXiv preprint arXiv:0808.0725*, 2008.

[253] C Paiva-Sánchez, E Burgos-Inostroza, O Jiménez, and A Delgado. Quantum tomography via equidistant states. *Physical Review A—Atomic, Molecular, and Optical Physics*, 82(3):032115, 2010.

[254] D Martínez, MA Solís-Prosser, G Cañas, O Jiménez, A Delgado, and G Lima. Experimental quantum tomography assisted by multiply symmetric states in higher dimensions. *Physical Review A*, 99(1):012336, 2019.

[255] Tao Xin, Dawei Lu, Joel Klassen, Nengkun Yu, Zhengfeng Ji, Jianxin Chen, Xian Ma, Guilu Long, Bei Zeng, and Raymond Laflamme. Quantum state tomography via reduced density matrices. *Physical review letters*, 118(2):020401, 2017.

[256] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[257] Edwin T Jaynes. Information theory and statistical mechanics. ii. *Physical review*, 108(2):171, 1957.

[258] V Bužek, G Adam, and G Drobnỳ. Reconstruction of wigner functions on different observation levels. *annals of physics*, 245(1):37–97, 1996.

[259] Rishabh Gupta, Rongxin Xia, Raphael D Levine, and Sabre Kais. Maximal entropy approach for quantum state tomography. *PRX Quantum*, 2(1):010318, 2021.

[260] Jordan Cotler and Frank Wilczek. Quantum overlapping tomography. *Physical review letters*, 124(10):100401, 2020.

[261] Zhengning Yang, Shihao Ru, Lianzhen Cao, Nikolay Zheludev, and Weibo Gao. Experimental demonstration of quantum overlapping tomography. *Physical Review Letters*, 130(5):050804, 2023.

[262] Marco Liscidini and JE Sipe. Stimulated emission tomography. *Physical review letters*, 111(19):193602, 2013.

[263] B Fang, Marco Liscidini, JE Sipe, and VO Lorenz. Multidimensional characterization of an entangled photon-pair source via stimulated emission tomography. *Optics express*, 24(9):10013–10019, 2016.

[264] Yang Xu, Saumya Choudhary, and Robert W Boyd. Stimulated emission tomography for efficient characterization of spatial entanglement. *Physical Review Research*, 6(4):L042047, 2024.

[265] Adrian Steffens, Mathis Friesdorf, T Langen, B Rauer, T Schweigler, Robert Hübener, J Schmiedmayer, CA Riofrío, and Jens Eisert. Towards experimental quantum-field tomography with ultracold atoms. *Nature Communications*, 6(1):7663, 2015.

[266] James G Titchener, Markus Gräfe, René Heilmann, Alexander S Solntsev, Alexander Szameit, and Andrey A Sukhorukov. Scalable on-chip quantum state tomography. *npj Quantum Information*, 4(1):19, 2018.

[267] GM D'Ariano and P Lo Presti. Quantum tomography for measuring experimentally the matrix elements of an arbitrary quantum operation. *Physical review letters*, 86(19):4195, 2001.

[268] Joseph B Altepeter, David Branning, Evan Jeffrey, TC Wei, Paul G Kwiat, Robert T Thew, Jeremy L O'Brien, Michael A Nielsen, and Andrew G White. Ancilla-assisted quantum process tomography. *Physical Review Letters*, 90(19):193601, 2003.

[269] Yaakov S Weinstein, Timothy F Havel, Joseph Emerson, Nicolas Boulant, Marcos Saraceno, Seth Lloyd, and David G Cory. Quantum process tomography of the quantum fourier transform. *The Journal of chemical physics*, 121(13):6117–6133, 2004.

[270] Jeremy L O'Brien, Geoff J Pryde, Alexei Gilchrist, Daniel FV James, Nathan K Langford, Timothy C Ralph, and Andrew G White. Quantum process tomography of a controlled-not gate. *Physical review letters*, 93(8):080502, 2004.

[271] M Mičuda, M Miková, I Straka, M Sedlák, M Dušek, M Ježek, and J Fiurášek. Tomographic characterization of a linear optical quantum toffoli gate. *Physical Review A*, 92(3):032312, 2015.

[272] SH Myrskog, JK Fox, MW Mitchell, and AM Steinberg. Quantum process tomography on vibrational states of atoms in an optical lattice. *Physical Review A—Atomic, Molecular, and Optical Physics*, 72(1):013615, 2005.

[273] Giacomo Torlai, Christopher J Wood, Atithi Acharya, Giuseppe Carleo, Juan Carrasquilla, and Leandro Aolita. Quantum process tomography with unsupervised learning and tensor networks. *Nature Communications*, 14(1):2858, 2023.

[274] Arefur Rahman, Noah I Wasserbeck, Zachary Goisman, Rhea P Fernandes, Brian T Kirby, Muneer Alshowkan, Chris Kurtz, and Joseph M Lukens. Deployed quantum link characterization via bayesian ancilla-assisted process tomography. *Optics Letters*, 50(8):2478–2481, 2025.

[275] Jeff S Lundeen, Alvaro Feito, Hendrik Coldenstrodt-Ronge, Kenny L Pregnell, Ch Silberhorn, Timothy C Ralph, Jens Eisert, Martin B Plenio, and Ian A Walmsley. Tomography of quantum detectors. *Nature Physics*, 5(1):27–30, 2009.

[276] Lijian Zhang, Hendrik B Coldenstrodt-Ronge, Animesh Datta, Graciana Puentes, Jeff S Lundeen, Xian-Min Jin, Brian J Smith, Martin B Plenio, and Ian A Walmsley. Mapping coherence in measurement via full quantum tomography of a hybrid optical detector. *Nature Photonics*, 6(6):364–368, 2012.

[277] Lijian Zhang, Animesh Datta, Hendrik B Coldenstrodt-Ronge, Xian-Min Jin, Jens Eisert, Martin B Plenio, and Ian A Walmsley. Recursive quantum detector tomography. *New Journal of Physics*, 14(11):115005, 2012.

[278] Chandra M Natarajan, Lijian Zhang, Hendrik Coldenstrodt-Ronge, Gaia Donati, Sander N Dorenbos, Val Zwiller, Ian A Walmsley, and Robert H Hadfield. Quantum detector tomography of a time-multiplexed superconducting nanowire single-photon detector at telecom wavelengths. *Optics express*, 21(1):893–902, 2013.

[279] Samuele Grandi, Alessandro Zavatta, Marco Bellini, and Matteo GA Paris. Experimental quantum tomography of a homodyne detector. *New Journal of Physics*, 19(5):053015, 2017.

[280] Yanzhu Chen, Maziar Farahzad, Shinjae Yoo, and Tzu-Chieh Wei. Detector tomography on ibm quantum computers and mitigation of an imperfect measurement. *Physical Review A*, 100(5):052315, 2019.

[281] Mamoru Endo, Tatsuki Sonoyama, Mikihisa Matsuyama, Fumiya Okamoto, Shigehito Miki, Masahiro Yabuno, Fumihiro China, Hirotaka Terai, and Akira Furusawa. Quantum detector tomography of a superconducting nanostrip photon-number-resolving detector. *Optics Express*, 29(8):11728–11738, 2021.

[282] Nathan Wiebe, Christopher Granade, Christopher Ferrie, and David G Cory. Hamiltonian learning and certification using quantum resources. *Physical review letters*, 112(19):190501, 2014.

[283] Nathan Wiebe, Christopher Granade, Christopher Ferrie, and David Cory. Quantum hamiltonian learning using imperfect quantum resources. *Physical Review A*, 89(4):042314, 2014.

[284] Kihwan Kim, M-S Chang, Simcha Korenblit, Rajibul Islam, Emily E Edwards, James K Freericks, G-D Lin, L-M Duan, and Christopher Monroe. Quantum simulation of frustrated ising spins with trapped ions. *Nature*, 465(7298):590–593, 2010.

[285] Yves Salathé, Mintu Mondal, Markus Oppliger, Johannes Heinsoo, Philipp Kurpiers, Anton Potočnik, Antonio Mezzacapo, Urtzi Las Heras, Lucas Lamata, Enrique Solano, et al. Digital quantum simulation of spin models with circuit quantum electrodynamics. *Physical Review X*, 5(2):021027, 2015.

[286] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O'brien. A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5(1):4213, 2014.

[287] Dominik Hangleiter, Ingo Roth, Jonáš Fuksa, Jens Eisert, and Pedram Roushan. Robustly learning the hamiltonian dynamics of a superconducting quantum processor. *Nature Communications*, 15(1):9595, 2024.

[288] Eyal Bairey, Itai Arad, and Netanel H. Lindner. Learning a local hamiltonian from local measurements. *Phys. Rev. Lett.*, 122:020504, Jan 2019.

[289] Xiao-Liang Qi and Daniel Ranard. Determining a local hamiltonian from a single eigenstate. *Quantum*, 3:159, 2019.

[290] Eyal Bairey, Chu Guo, Dario Poletti, Netanel H Lindner, and Itai Arad. Learning the dynamics of open quantum systems from their steady states. *New Journal of Physics*, 22(3):032001, 2020.

[291] Jeongwan Haah, Robin Kothari, and Ewin Tang. Optimal learning of quantum hamiltonians from high-temperature gibbs states. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 135–146. IEEE, 2022.

[292] Tim J. Evans, Robin Harper, and Steven T. Flammia. Scalable bayesian hamiltonian learning, 2019.

[293] Bo Qi, Zhibo Hou, Li Li, Daoyi Dong, Guoyong Xiang, and Guangcan Guo. Quantum state tomography via linear regression estimation. *Scientific reports*, 3(1):3496, 2013.

[294] Zhibo Hou, Han-Sen Zhong, Ye Tian, Daoyi Dong, Bo Qi, Li Li, Yuanlong Wang, Franco Nori, Guo-Yong Xiang, Chuan-Feng Li, et al. Full reconstruction of a 14-qubit state within four hours. *New Journal of Physics*, 18(8):083036, 2016.

[295] Pierre Alquier, Cristina Butucea, Mohamed Hebiri, Katia Meziani, and Tomoyuki Morimae. Rank-penalized estimation of a quantum system. *Physical Review A—Atomic, Molecular, and Optical Physics*, 88(3):032113, 2013.

[296] Cristina Butucea, Mădălin Guţă, and Theodore Kypraios. Spectral thresholding quantum tomography for low rank states. *New Journal of Physics*, 17(11):113050, 2015.

[297] Madalin Guţă, Jonas Kahn, Richard Kueng, and Joel A Tropp. Fast state tomography with optimal error bounds. *Journal of Physics A: Mathematical and Theoretical*, 53(20):204001, 2020.

[298] Júlia Barberà-Rodríguez, Leonardo Zambrano, Antonio Acín, and Donato Farina. Boosting projective methods for quantum process and detector tomography. *Physical Review Research*, 7(1):013208, 2025.

[299] Ming-Chien Hsu, En-Jui Kuo, Wei-Hsuan Yu, Jian-Feng Cai, and Min-Hsiu Hsieh. Quantum state tomography via nonconvex riemannian gradient descent. *Physical Review Letters*, 132(24):240804, 2024.

[300] Manuel Sebastian Torres Hernandez. Quantum state tomography with gradient descent. 2024.

[301] Akshay Gaikwad, Manuel Sebastian Torres, Shahnawaz Ahmed, and Anton Frisk Kockum. Gradient-descent methods for fast quantum state tomography. *arXiv preprint arXiv:2503.04526*, 2025.

[302] A Utreras-Alarcón, M Rivera-Tapia, S Niklitschek, and A Delgado. Stochastic optimization on complex variables and pure-state quantum tomography. *Scientific Reports*, 9(1):16143, 2019.

[303] Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 913–925, 2016.

[304] Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 899–912, 2016.

[305] Steven T Flammia and Ryan O'Donnell. Quantum chi-squared tomography and mutual information testing. *Quantum*, 8:1381, 2024.

[306] Anurag Anshu and Srinivasan Arunachalam. A survey on the complexity of learning quantum states. *Nature Reviews Physics*, 6(1):59–69, 2024.

[307] Daniele Binosi, Giovanni Garberoglio, Diego Maragnano, Maurizio Dapor, and Marco Liscidini. thresholdqst: A python library to implement threshold quantum state tomography for qubits. https://github.com/liscidini-group/thresholdqst, 2024.

[308] Niall Hurley and Scott Rickard. Comparing measures of sparsity. *IEEE Transactions on Information Theory*, 55(10):4723–4741, 2009.

[309] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[310] Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, 16(5):1190–1208, 1995.

[311] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641, 2017.

[312] Patrick J Coles, M Cerezo, and Lukasz Cincio. Strong bound between trace distance and hilbert-schmidt distance for low-rank states. *Physical Review A*, 100(2):022103, 2019.

[313] Qiskit contributors. Qiskit: An open-source framework for quantum computing, 2023.

[314] Naoki Kanazawa, Daniel J Egger, Yael Ben-Haim, Helena Zhang, William E Shanks, Gadi Aleksandrowicz, and Christopher J Wood. Qiskit experiments: A python package to characterize and calibrate quantum computers. *Journal of Open Source Software*, 8(84):5329, 2023.

[315] William R Clements, Peter C Humphreys, Benjamin J Metcalf, W Steven Kolthammer, and Ian A Walmsley. Optimal design for universal multiport interferometers. *Optica*, 3(12):1460–1465, 2016.

[316] Ciro Pentangelo, Niki Di Giano, Simone Piacentini, Riccardo Arpe, Francesco Ceccarelli, Andrea Crespi, and Roberto Osellame. High-fidelity and polarization-insensitive universal photonic processors fabricated by femtosecond laser writing. *Nanophotonics*, 13(12):2259–2270, 2024.

[317] Pieter Kok, William J Munro, Kae Nemoto, Timothy C Ralph, Jonathan P Dowling, and Gerard J Milburn. Linear optical quantum computing with photonic qubits. *Reviews of Modern Physics*, 79(1):135–174, 2007.

[318] Mathias Pont, Giacomo Corrielli, Andreas Fyrillas, Iris Agresti, Gonzalo Carvacho, Nicolas Maring, Pierre-Emmanuel Emeriau, Francesco Ceccarelli, Ricardo Albiero, Paulo HD Ferreira, N Somaschi, J Senellart, I Sagnes, M Morassi, A Lemaître, P Senellart, F Sciarrino, M Liscidini, N Belabas, and R Osellame. High-fidelity four-photon ghz states on-chip. *npj Quantum Information*, 10:50, 2024.

[319] Mathias Pont, Riccardo Albiero, Sarah E Thomas, Nicolò Spagnolo, Francesco Ceccarelli, Giacomo Corrielli, Alexandre Brieussel, Niccolo Somaschi, Hêlio Huet, Abdelmounaim Harouri, Aristide Lemaître, Isabelle Sagnes, Nadia Belabas, Fabio Sciarrino, Roberto Osellame, Pascale Senellart, and Andrea Crespi. Quantifying n-photon indistinguishability with a cyclic integrated interferometer. *Physical Review X*, 12(3):031033, 2022.

[320] Giovanni Rodari, Carlos Fernandes, Eugenio Caruccio, Alessia Suprano, Francesco Hoch, Taira Giordani, Gonzalo Carvacho, Riccardo Albiero, Niki Di Giano, Giacomo Corrielli, Francesco Ceccarelli, Roberto Osellame, Daniel J. Brod, Leonardo Novo, N Spagnolo, Ernesto F. Galväo, and Fabio Sciarrino. Experimental observation of counter-intuitive features of photonic bunching, 2024.

[321] Giovanni Rodari, Leonardo Novo, Riccardo Albiero, Alessia Suprano, Carlos T. Tavares, Eugenio Caruccio, Francesco Hoch, Taira Giordani, Gonzalo Carvacho, Marco Gardina, Niki Di Giano, Serena Di Giorgio, Giacomo Corrielli, Francesco Ceccarelli, Roberto Osellame, N Spagnolo, Ernesto F. Galväo, and Fabio Sciarrino. Semi-device independent characterization of multiphoton indistinguishability, 2024.

[322] Ian Goodfellow. Deep learning, 2016.

[323] Christopher M Bishop and Hugh Bishop. *Deep learning: Foundations and concepts*. Springer Nature, 2023.

[324] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

[325] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, 1985.

[326] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

[327] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[328] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 2002.

[329] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[330] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[331] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

[332] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[333] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training. 2018.

[334] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186, 2019.

[335] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[336] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

[337] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[338] Geoffrey E Hinton, Terrence J Sejnowski, and David H Ackley. Boltzmann machines: Constraint satisfaction networks that learn. 1984.

[339] David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski. A learning algorithm for boltzmann machines. *Cognitive science*, 9(1):147–169, 1985.

[340] Geoffrey E Hinton, Terrence J Sejnowski, et al. Learning and relearning in boltzmann machines. *Parallel distributed processing: Explorations in the microstructure of cognition*, 1(282-317):2, 1986.

[341] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

[342] Zi Cai and Jinguo Liu. Approximating quantum many-body wave functions using artificial neural networks. *Phys. Rev. B*, 97:035116, Jan 2018.

[343] Jun Gao, Lu-Feng Qiao, Zhi-Qiang Jiao, Yue-Chi Ma, Cheng-Qiu Hu, Ruo-Jing Ren, Ai-Lin Yang, Hao Tang, Man-Hong Yung, and Xian-Min Jin. Experimental machine learning of quantum states. *Physical review letters*, 120(24):240501, 2018.

[344] Tao Xin, Sirui Lu, Ningping Cao, Galit Anikeeva, Dawei Lu, Jun Li, Guilu Long, and Bei Zeng. Local-measurement-based quantum state tomography via neural networks. *npj Quantum Information*, 5(1):109, 2019.

[345] Adriano Macarone Palmieri, Egor Kovlakov, Federico Bianchi, Dmitry Yudin, Stanislav Straupe, Jacob D. Biamonte, and Sergei Kulik. Experimental neural network enhanced quantum tomography. *npj Quantum Information*, 6(1), February 2020.

[346] Dominik Koutný, Libor Motka, Zden ěk Hradil, Jaroslav Řeháček, and Luis L. Sánchez-Soto. Neural-network quantum state tomography. *Phys. Rev. A*, 106:012409, Jul 2022.

[347] Nhan T Luu, Thang C Truong, and Duong T Luu. Universal quantum tomography with deep neural networks. *arXiv preprint arXiv:2407.01734*, 2024.

[348] Sanjaya Lohani, Brian T Kirby, Michael Brodsky, Onur Danaci, and Ryan T Glasser. Machine learning assisted quantum state estimation. *Machine Learning: Science and Technology*, 1(3):035007, 2020.

[349] Tobias Schmale, Moritz Reh, and Martin Gärttner. Efficient quantum state tomography with convolutional neural networks. *npj Quantum Information*, 8(1):115, 2022.

[350] Yihui Quek, Stanislav Fort, and Hui Khoon Ng. Adaptive quantum state tomography with neural networks. *npj Quantum Information*, 7(1):105, 2021.

[351] Andrea Rocchetto, Edward Grant, Sergii Strelchuk, Giuseppe Carleo, and Simone Severini. Learning hard quantum distributions with variational autoencoders. *npj Quantum Information*, 4(1):28, 2018.

[352] Shahnawaz Ahmed, Carlos Sánchez Muñoz, Franco Nori, and Anton Frisk Kockum. Quantum state tomography with conditional generative adversarial networks. *Physical review letters*, 127(14):140502, 2021.

[353] Yan Zhu, Ya-Dong Wu, Ge Bai, Dong-Sheng Wang, Yuexuan Wang, and Giulio Chiribella. Flexible learning of quantum states with generative query neural networks. *Nature communications*, 13(1):6222, 2022.

[354] Xuegang Li, Wenjie Jiang, Ziyue Hua, Weiting Wang, Xiaoxuan Pan, Weizhou Cai, Zhide Lu, Jiaxiu Han, Rebing Wu, Chang-Ling Zou, et al. Experimental demonstration of reconstructing quantum states with generative models. *arXiv preprint arXiv:2407.15102*, 2024.

[355] Peter Cha, Paul Ginsparg, Felix Wu, Juan Carrasquilla, Peter L McMahon, and Eun-Ah Kim. Attention-based quantum tomography. *Machine Learning: Science and Technology*, 3(1):01LT01, 2021.

[356] Lu Zhong, Chu Guo, and Xiaoting Wang. Quantum state tomography inspired by language modeling. *arXiv preprint arXiv:2212.04940*, 2022.

[357] Hailan Ma, Zhenhong Sun, Daoyi Dong, Chunlin Chen, and Herschel Rabitz. Tomography of quantum states from structured measurements via quantum-aware transformer. *arXiv preprint arXiv:2305.05433*, 2023.

[358] Adriano Macarone Palmieri, Guillem Müller-Rigat, Anubhav Kumar Srivastava, Maciej Lewenstein, Grzegorz Rajchel-Mieldzioć, and Marcin Płodzień. Enhancing quantum state tomography via resource-efficient attention-based neural networks. *Physical Review Research*, 6(3):033248, 2024.

[359] Abhishek Abhishek. Benchmarking attention-based quantum state tomography.

[360] Ya-Dong Wu, Yan Zhu, Yuexuan Wang, and Giulio Chiribella. Learning quantum properties from short-range correlations using multi-task networks. *Nature Communications*, 15(1):8796, 2024.

[361] Alistair WR Smith, Johnnie Gray, and MS Kim. Efficient quantum state sample tomography with basis-dependent neural networks. *PRX Quantum*, 2(2):020348, 2021.

[362] Victor Wei, WA Coish, Pooya Ronagh, and Christine A Muschik. Neural-shadow quantum state tomography. *Physical Review Research*, 6(2):023250, 2024.

[363] Francisco Albarrán-Arriagada, Juan C Retamal, Enrique Solano, and Lucas Lamata. Measurement-based adaptation protocol with quantum reinforcement learning. *Physical Review A*, 98(4):042315, 2018.

[364] Alexey A Melnikov, Hendrik Poulsen Nautrup, Mario Krenn, Vedran Dunjko, Markus Tiersch, Anton Zeilinger, and Hans J Briegel. Active learning machine learns to create new quantum experiments. *Proceedings of the National Academy of Sciences*, 115(6):1221–1226, 2018.

[365] Shang Yu, Francisco Albarrán-Arriagada, Juan Carlos Retamal, Yi-Tao Wang, Wei Liu, Zhi-Jin Ke, Yu Meng, Zhi-Peng Li, Jian-Shun Tang, Enrique Solano, et al. Reconstruction of a photonic qubit state with reinforcement learning. *Advanced Quantum Technologies*, 2(7-8):1800074, 2019.

[366] Yan Zhu, Tailong Xiao, Guihua Zeng, Giulio Chiribella, and Ya-Dong Wu. Controlling unknown quantum states via data-driven state representations. *arXiv preprint arXiv:2406.05711*, 2024.

[367] Syed Tihaam Ahmad, Ahmad Farooq, and Hyundong Shin. Genetic algorithm based quantum state tomography. pages 468–469, 2022.

[368] Zheng An, Jiahui Wu, Muchun Yang, D. L. Zhou, and Bei Zeng. Unified quantum state tomography and hamiltonian learning: A language-translation-like approach for quantum systems. *Phys. Rev. Appl.*, 21:014037, Jan 2024.

[369] Xiaoqian Zhang, Maolin Luo, Zhaodi Wen, Qin Feng, Shengshi Pang, Weiqi Luo, and Xiaoqi Zhou. Direct fidelity estimation of quantum states using machine learning. *Physical Review Letters*, 127(13):130503, 2021.

[370] G Koolstra, N Stevenson, S Barzili, L Burns, K Siva, S Greenfield, W Livingston, A Hashim, RK Naik, JM Kreikebaum, et al. Monitoring fast superconducting qubit dynamics using a neural network. *Physical Review X*, 12(3):031017, 2022.

[371] Valeria Cimini, Marco Barbieri, Nicolas Treps, Mattia Walschaers, and Valentina Parigi. Neural networks for detecting multimode wigner negativity. *Physical Review Letters*, 125(16):160504, 2020.

[372] Shahnawaz Ahmed, Carlos Sánchez Muñoz, Franco Nori, and Anton Frisk Kockum. Classification and reconstruction of optical quantum states with deep neural networks. *Physical Review Research*, 3(3):033278, 2021.

[373] Ying Zuo, Chenfeng Cao, Ningping Cao, Xuanying Lai, Bei Zeng, and Shengwang Du. Optical neural network quantum state tomography. *Advanced Photonics*, 4(2):026004–026004, 2022.

[374] Hsun-Chung Wu, Hsien-Yi Hsieh, Zhi-Kai Xu, Hua Li Chen, Zi-Hao Shi, Po-Han Wang, Popo Yang, Ole Steuernagel, Chien-Ming Wu, and Ray-Kuang Lee. Machine learning enhanced quantum state tomography on fpga. *arXiv preprint arXiv:2501.04327*, 2025.

[375] Mario Krenn, Jonas Landgraf, Thomas Foesel, and Florian Marquardt. Artificial intelligence and machine learning for quantum technologies. *Phys. Rev. A*, 107:010101, Jan 2023.

[376] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.

[377] Ivan Glasser, Nicola Pancotti, Moritz August, Ivan D. Rodriguez, and J. Ignacio Cirac. Neural-network quantum states, string-bond states, and chiral topological states. *Phys. Rev. X*, 8:011006, Jan 2018.

[378] Giuseppe Carleo, Yusuke Nomura, and Masatoshi Imada. Constructing exact representations of quantum many-body systems with deep neural networks. *Nature communications*, 9(1):5322, 2018.

[379] Giacomo Torlai and Roger G Melko. Latent space purification via neural density operators. *Physical review letters*, 120(24):240503, 2018.

[380] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. Neural-network quantum state tomography. *Nature physics*, 14(5):447–450, 2018.

[381] Juan Carrasquilla, Giacomo Torlai, Roger G Melko, and Leandro Aolita. Reconstructing quantum states with generative models. *Nature Machine Intelligence*, 1(3):155–161, 2019.

[382] Marcel Neugebauer, Laurin Fischer, Alexander Jäger, Stefanie Czischek, Selim Jochim, Matthias Weidemüller, and Martin Gärttner. Neural-network quantum state tomography in a two-qubit experiment. *Phys. Rev. A*, 102:042604, Oct 2020.

[383] Giacomo Torlai, Brian Timar, Evert PL Van Nieuwenburg, Harry Levine, Ahmed Omran, Alexander Keesling, Hannes Bernien, Markus Greiner, Vladan Vuletić, Mikhail D Lukin, et al. Integrating neural networks with a quantum simulator for state reconstruction. *Physical review letters*, 123(23):230504, 2019.

[384] Roger G Melko, Giuseppe Carleo, Juan Carrasquilla, and J Ignacio Cirac. Restricted boltzmann machines in quantum physics. *Nature Physics*, 15(9):887–892, 2019.

[385] Giacomo Torlai and Roger G Melko. Machine-learning quantum states in the nisq era. *Annual Review of Condensed Matter Physics*, 11(1):325–344, 2020.

[386] Juan Carrasquilla. Machine learning for quantum matter. *Advances in Physics: X*, 5(1):1797528, 2020.

[387] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4):045002, 2019.

[388] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[389] Siamak Ravanbakhsh, Jeff Schneider, and Barnabas Poczos. Equivariance through parameter-sharing. In *International Conference on Machine Learning*, pages 2892–2901. PMLR, 2017.

[390] Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *International Conference on Machine Learning*, pages 2747–2755. PMLR, 2018.

[391] Dmitry Yarotsky. Universal approximations of invariant maps by neural networks. *Constructive Approximation*, pages 1–68, 2021.

[392] Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and equivariant graph networks. In *International Conference on Learning Representations*, pages 1–14, 2018.

[393] Erik Henning Thiede, Truong Son Hy, and Risi Kondor. The general theory of permutation equivariant neural networks and higher order graph variational encoders. *arXiv preprint arXiv:2004.03990*, 2020.

[394] Richard HR Hahnloser, Rahul Sarpeshkar, Misha A Mahowald, Rodney J Douglas, and H Sebastian Seung. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *nature*, 405(6789):947–951, 2000.

[395] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. JMLR Workshop and Conference Proceedings, 2011.

[396] Joseph B Altepeter, Daniel FV James, and Paul G Kwiat. 4 qubit quantum state tomography. *Quantum state estimation*, pages 113–145, 2004.

[397] Paul R Cohen and Adele E Howe. How evaluation guides ai research: The message still counts more than the medium. *AI magazine*, 9(4):35–35, 1988.

[398] Jocelyn Sietsma and Robert JF Dow. Creating artificial neural networks that generalize. *Neural networks*, 4(1):67–79, 1991.

[399] Chris M Bishop. Training with noise is equivalent to tikhonov regularization. *Neural computation*, 7(1):108–116, 1995.

[400] Guozhong An. The effects of adding noise during backpropagation training on a generalization performance. *Neural computation*, 8(3):643–674, 1996.

[401] Marco Benedetti and Enrico Ventura. Training neural networks with structured noise improves classification and generalization. *Journal of Physics A: Mathematical and Theoretical*, 57(41):415001, 2024.

[402] https://github.com/gioGarbe/ECT-QST.

[403] Diogo Cruz, Romain Fournier, Fabien Gremion, Alix Jeannerot, Kenichi Komagata, Tara Tosic, Jarla Thiesbrummel, Chun Lam Chan, Nicolas Macris, Marc-André Dupertuis, and Clément Javerzac-Galy. Efficient quantum algorithms for ghz and w states, and implementation on the IBM quantum computer. *Advanced Quantum Technologies*, 2(5-6), apr 2019.

[404] H. Bombin and M. A. Martin-Delgado. Topological quantum distillation. *Phys. Rev. Lett.*, 97:180501, 2006.

[405] D. Nigg, M. Müller, E. A. Martinez, P. Schindler, M. Hennrich, T. Monz, M. A. Martin-Delgado, and R. Blatt. Quantum computations on a topologically encoded qubit. *Science*, 345(6194):1253742, 2014.

[406] James Clerk Maxwell. *Theory of Heat*. Longmans, Green, and Co., 1871. See Chapter 12, where the demon thought experiment is introduced.

[407] Marian von Smoluchowski. Experimentell nachweisbare, der üblichen thermodynamik widersprechende molekularphänomene. *Physikalische Zeitschrift*, 13:1069–1080, 1912.

[408] Leo Szilard. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Zeitschrift für Physik*, 53:840–856, 1929.

[409] Rolf Landauer. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5(3):183–191, 1961.

[410] Charles H Bennett. Demons, engines and the second law. *Scientific American*, 257(5):108–117, 1987.

[411] Charles H Bennett. Notes on landauer's principle, reversible computation, and maxwell's demon. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 34(3):501–510, 2003.

[412] Rolf Landauer. Information is physical. *Physics Today*, 44(5):23–29, 1991.

[413] A. de Oliveira Junior, Jonatan Bohr Brask, and Rafael Chaves. A friendly guide to exorcising maxwell's demon. *PRX Quantum*, 6:030201, Aug 2025.

[414] A.E.Allahverdyan, R.Balian, and Th.M.Nieuwenhuizen. Maximal work extraction from finite quantum systems. *Europhysics Letters*, 67(4):565–571, 2004.

[415] Giacomo Corrielli, Andrea Crespi, and Roberto Osellame. Femtosecond laser micromachining for integrated quantum photonics. *Nanophotonics*, 10(15):3789–3812, 2021.

[416] Giacomo Corrielli, Simone Atzeni, Simone Piacentini, Ioannis Pitsios, Andrea Crespi, and Roberto Osellame. Symmetric polarization-insensitive directional couplers fabricated by femtosecond laser writing. *Optics Express*, 26(12):15101–15109, 2018.

[417] Riccardo Albiero, Ciro Pentangelo, Marco Gardina, Simone Atzeni, Francesco Ceccarelli, and Roberto Osellame. Toward higher integration density in femtosecond-laser-written programmable photonic circuits. *Micromachines*, 13(7):1145, 2022.

[418] Francesco Ceccarelli, Simone Atzeni, Ciro Pentangelo, Francesco Pellegatta, Andrea Crespi, and Roberto Osellame. Low power reconfigurability and reduced crosstalk in integrated photonic circuits fabricated by femtosecond laser micromachining. *Laser & Photonics Reviews*, 14(10):2000024, 2020.

[419] R Hanbury Brown and R Q Twiss. A new type of interferometer for use in radio astronomy. *Phylosophical Magazine*, 45:663–683, 1954.

[420] Malte C Tichy. Sampling of partially distinguishable bosons and the relation to the multidimensional permanent. *Phys. Rev. A*, 91(2):022316, 2015.

[421] C. K. Hong, Z. Y. Ou, and L. Mandel. Measurement of subpicosecond time intervals between two photons by interference. *Phys. Rev. Lett.*, 59:2044–2046, Nov 1987.

[422] H. Ollivier, S. E. Thomas, S. C. Wein, I. Maillette de Buy Wenniger, N. Coste, J. C. Loredo, N. Somaschi, A. Harouri, A. Lemaitre, I. Sagnes, L. Lanco, C. Simon, C. Anton, O. Krebs, and P. Senellart. Hong-Ou-Mandel interference with imperfect single photon sources. *Phys. Rev. Lett.*, 126:063602, Feb 2021.

[423] Michal Oszmaniec and Daniel J Brod. Classical simulation of photonic linear optics with lost particles. *New J. Phys.*, 20:092002, 2018.