Ingo Roth

SEMI-DEVICE-DEPENDENT QUANTUM SYSTEM IDENTIFICATION

Dissertation zur Erlangung des Grades eines
Doktors der Naturwissenschaften vorgelegt am
Fachbereich Physik der Freien Universität Berlin

Berlin, 2022

Erstgutachter:	Prof. Dr. Jens Eisert
Zweitgutachter:	Prof. Dr. Jean-Pierre Seifert
Tag der Disputation:	8. September 2023

Ingo Roth

SEMI-DEVICE-DEPENDENT QUANTUM SYSTEM IDENTIFICATION

Contents

Publications

Core material of this thesis has already been published in the following articles and preprints:

[1] I. Roth, R. Kueng, S. Kimmel, Y.-K. Liu, D. Gross, J. Eisert, and M. Kliesch. “Recovering Quantum Gates from Few Average Gate Fidelities”. *Phys. Rev. Lett.* **121** (2018). arXiv: 1803.00572 [quant-ph].

[2] I. Roth, J. Wilkens, D. Hangleiter, and J. Eisert. “Semi-device-dependent blind quantum tomography”. 2020. arXiv: 2006.03069 [quant-ph].

[3] D. Hangleiter, I. Roth, J. Eisert, and P. Roushan. “Precise Hamiltonian identification of a superconducting quantum processor”. 2021. arXiv: 2108.08319 [quant-ph].

[4] M. Kliesch and I. Roth. “Theory of Quantum System Certification”. *PRX Quantum* **2** (2021), 010201. arXiv: 2010.05925 [quant-ph].

We further present results published in the following articles and preprints:

[5] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi. “Quantum certification and benchmarking”. *Nat. Rev. Phys.* **2** (2020), 382–390. arXiv: 1910.06343 [quant-ph].

[6] J. Haferkamp, F. Montealegre-Mora, M. Heinrich, J. Eisert, D. Gross, and I. Roth. “Quantum homeopathy works: Efficient unitary designs with a system-size independent number of non-Clifford gates”. 2020. arXiv: 2002.09524 [quant-ph].

[7] H. Wilming, M. Goihl, I. Roth, and J. Eisert. “Entanglement-Ergodic Quantum Systems Equilibrate Exponentially Well”. *Phys. Rev. Lett.* **123** (2019). arXiv: 1802.02052 [quant-ph].

[8] J. Haferkamp, C. Bertoni, I. Roth, and J. Eisert. “Emergent statistical mechanics from properties of disordered random matrix product states”. *PRX Quantum* **2** (2021), 040308. arXiv: 2103.02634 [quant-ph].

[9] J. Helsen, I. Roth, E. Onorati, A. H. Werner, and J. Eisert. “A general framework for randomized benchmarking”. accepted in PRX Quantum. 2020. arXiv: 2010.07974 [quant-ph].

[10] D. Hangleiter, I. Roth, D. Nagaj, and J. Eisert. “Easing the Monte Carlo sign problem”. *Sci. Adv.* **6** (2020), eabb8341. arXiv: 1906.02309 [quant-ph].

In addition, the following publications and preprints were authored that are not presented in the thesis:

[11] A. Steffens, C. A. Riofrío, W. McCutcheon, I. Roth, B. A. Bell, A. McMillan, M. S. Tame, J. G. Rarity, and J. Eisert. “Experimentally exploring compressed sensing quantum tomography”. *Quantum Sci. and Technol.* **2** (2017), 025005. arXiv: 1611.01189 [quant-ph].

[12] I. Roth, M. Kliesch, G. Wunder, and J. Eisert. “Reliable recovery of hierarchically sparse signals”. In: *Proceedings of the third "international Traveling Workshop on Interactions between Sparse models and Technology" (iTWIST'16).* 2016, 36–38. arXiv: 1609.04167 [cs.NA].

[13] I. Roth, A. Flinth, R. Kueng, J. Eisert, and G. Wunder. “Hierarchical restricted isometry property for Kronecker product measurements”. In: *56th Annual Allerton Conference on Communication, Control, and Computing (Allerton).* 2018, 632–638. arXiv: 1801.10433 [cs.IT].

[14] I. Roth, M. Kliesch, A. Flinth, G. Wunder, and J. Eisert. “Reliable Recovery of Hierarchically Sparse Signals for Gaussian and Kronecker Product Measurements”. *IEEE Trans. Signal Process.* **68** (2020), 4002–4016. arXiv: 1612.07806 [cs.IT].

[15] G. Wunder, I. Roth, R. Fritschek, and J. Eisert. “HiHTP: A custom-tailored hierarchical sparse detector for massive MTC”. In: *2017 51st Asilomar Conference on Signals, Systems, and Computers.* 2017, 1929–1934.

[16] G. Wunder, I. Roth, R. Fritschek, B. Groß, and J. Eisert. “Secure Massive IoT Using Hierarchical Fast Blind Deconvolution”. In: *2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW).* 2018, 119–124. arXiv: 1801.09628 [cs.IT].

[17] G. Wunder, I. Roth, A. Flinth, M. Barzegar, S. Haghighatshoar, G. Caire, and G. Kutyniok. “Hierarchical Sparse Channel Estimation for Massive MIMO”. In: *WSA 2018; 22nd International ITG Workshop on Smart Antennas.* 2018, 1–8. arXiv: 1803.10994 [cs.IT].

[18] G. Wunder, S. Stefanatos, A. Flinth, I. Roth, and G. Caire. "Low-Overhead Hierarchically-Sparse Channel Estimation for Multiuser Wideband Massive MIMO". *IEEE Transactions on Wireless Communications* **18** (2019), 2186–2199. arXiv: 1806.00815 [cs.IT].

[19] G. Wunder, I. Roth, R. Fritschek, and J. Eisert. "Performance of Hierarchical Sparse Detectors for Massive MTC". 2018. arXiv: 1806.02754 [cs.IT].

[20] B. Groß, A. Flinth, I. Roth, J. Eisert, and G. Wunder. "Hierarchical sparse recovery from hierarchically structured measurements". In: *IEEE Stat. Sig. Proc. Workshop (SSP)*. 2021, 531–535. arXiv: 2105.03169 [cs.IT].

[21] J. Eisert, A. Flinth, B. Groß, I. Roth, and G. Wunder. "Hierarchical compressed sensing". 2021. arXiv: 2104.02721 [cs.IT].

[22] A. Goeßmann, M. Götte, I. Roth, R. Sweke, G. Kutyniok, and J. Eisert. "Tensor network approaches for learning non-linear dynamical laws". NeurIPS2020-Tensorworkshop, Dec. 2020. arXiv: 2002.12388 [math.NA].

[23] R. Brieger, I. Roth, and M. Kliesch. "Compressive gate set tomography". 2021. arXiv: 2112.05176 [quant-ph].

The code for the algorithms and numerical simulations presented in Chapters 3, 6 and 7 is available as:

[24] J. Wilkens, D. Hangleiter, and I. Roth. Gitlab repository at https://gitlab.com/wilkensJ/blind-quantum-tomography. 2020.

[25] I. Roth, R. Kueng, S. Kimmel, Y.-K. Liu, D. Gross, J. Eisert, and M. Kliesch. *Quantum process tomography with average gate fidelities*. GitHub repository https://github.com/MartKl/Quantum_process_tomography_with_average_gate_fidelities. 2017.

[26] I. Roth and D. Hangleiter. *Signease – A python package to ease the quantum Monte Carlo sign problem*. Gitlab repository at https://gitlab.com/ingo.roth/signease. 2019.

1 Introduction

We are witnessing rapid progress in the experimental abilities to manipulate physical systems in their inner quantum properties such as state superposition and entanglement.[1] Most importantly, we begin to have precise control over complex quantum systems on scales that are out of reach of simulations on the existing even most-powerful classical computing devices [Aru+19]. Harnessing their computational power promises the development of digital quantum computers that solve important problems much faster than any classical computer [NC10; Fey86]. Other envisioned applications of quantum technologies include the study of complex phases of matter in analogue simulations [CZ12] and cryptographically secure communication [Ací+18]. Quantum technology promises highly useful devices with diverse domains of application ranging from fundamental research to commercial businesses.

With the advent of these novel technologies comes the necessity for characterizing their functioning. Precise characterization of quantum devices is an ubiquitous task in their development. But also in the long term, it is indispensable for the diagnostic of more advanced devices during run-time. Characterization yields information about the actual inner-working of a quantum device such as its imperfections. The resulting understanding is crucial in order to interpret the output of a quantum device and eventually to improve the device's performance.

The characterization of quantum devices is a particularly daunting task in the interesting regime of high complexity: For a complex quantum device full characterization of a current state or the prediction of its time-evolution quickly exhausts the available classical computing power, in both time and memory. But these two tasks are crucial routines for many straight-forward approaches to characterization. For example, a direct approach to certifying the correct functioning of a quantum device is to compare its output with the prediction obtained

[1] We are witnessing … in the last decades.] The paragraphs are based on a text published in Ref. [4] as an introduction to quantum certification protocols.

by a classical simulation of the same task. Obstacles originating in the complexity of quantum devices are already encountered in the characterization and calibration of the measurement devices that are used for the characterization tasks. Ironically, it is the same complexity that makes quantum technology powerful that hinders their characterization. This challenging prospective has motivated extensive effort in developing characterization tools for quantum devices in the last decades [5].

In this thesis, we study approaches to the identification of models of the abstract physical layer of a quantum device. The abstract physical layer is modelled by quantum states, processes, gate sets and generators of time evolutions. The focus shall be on approaches that relax the assumptions on the quantum measurement device that is used in the data acquisition. Typically, inaccuracy in the implementation and characterization of the measurement device reduce the precision of the result of a characterization protocol. We study methods for the identification of quantum states, processes, gate sets and Hamiltonians that are to some degree robust against imperfections in the measurement device. We refer to such protocols as *semi-device-dependent*. The emphasis of this work is on a rigorous study of the resource requirements of the protocol and in particular the scaling of the resources. An omni-present theme in this work is the efficient exploitation of *structural assumption* on the quantum system that allow for a reduction in resource requirements. We aim at giving rigorous mathematical guarantees for the protocols that delineate their realm of applicability in practice. The analytical work is accompanied by evaluations of the performance of the protocols in numerical simulations. In the last part of the thesis we demonstrate one of the methods on experimental data from a superconducting qubit device. Occasionally, the development of technical methods will briefly bring us to further applications and questions outside the field of quantum characterization, such as the requirements for the equilibration of quantum systems, and the nature of the sign problem arising in the classical Monte-Carlo simulations of quantum systems.

In the centre of the thesis are three different identification tasks:

1. The identification of a low-rank quantum *state* from measurements taken with a partially uncalibrated measurement device—*Blind quantum state tomography.*

2. The identification of a unitary quantum process from data that is robust against imperfections in the state preparation and measurement device—*Compressive randomized benchmarking tomography*

3. The high-precision and robust identification of non-interacting Hamiltonians in an analogue quantum simulation—*Hamiltonian identification for analogue simulations*.

These tasks of quantum system identification can be seen as special instances of tasks that aim at gaining information about a quantum system. In the following section we outline a unifying, abstract framework to more broadly classify and discuss quantum characterization protocols. Taking this abstract view point allows us to develop a convenient language to present our identification schemes and compare them to other approaches in the literature.

1.1 Semi-device-dependent quantum characterization—an overview

Quantum characterization is the task of extracting information from a quantum system. Such information can for example be a detailed description of the system's state, specific quantities such as an entanglement measure between two subsystems, or a certificate ensuring the compliance of a quantum device with its specification.

Some quantum characterization tasks are among the most fundamental challenges in the physical sciences itself. For example, the question to infer a Hamiltonian that governs a system's time evolution from measured data is essential to uncovering the laws of nature. Results in quantum characterization can potentially provide valuable tools to advance our understanding of nature as well as identifying fundamental obstacles to gaining knowledge about nature itself, e.g. [CEW12]. The need for diagnostic tools in the quantum technologies in addition provides a very concrete motivation to enhance our abilities of characterizing quantum systems. Concrete technological challenges are also the motivation of the protocols in this work.

In the context of quantum technologies, one is interested in characterizing *quantum devices*. A full-fledged quantum device is described using multiple layers of abstraction from the physical layer over, e.g. abstract physical and logical gate layers, to an application layer, see Figure 1.1. On the 'lowest' physical layer of quantum devices is a physical system that can be prepared and manipulated coherently in its quantum states. The idea of digital quantum computers [NC10] is to control two states of a quantum system, giving rise to the two-dimensional state space of a quantum bit (qubit), and construct arrays of such two-level systems. Gate-based quantum computing further implements a set of coherent operations to manipulate the states of multiple qubit systems simultaneously, allowing for the creation of entanglement in between. The most prominent implementation for qubits and gates today are trapped atomic ions [BW08; MK13; Hom+09] and super-conducting circuits [NPT99; DS13; Che+14]. Other candidates include nitrogen vacancy centres [HA08] and integrated photonics [Lan+09]. In this work we already substantially abstract from the underlying physical system of a quantum device. We describe a quantum device in terms of quantum states and processes that already live in and operate on the abstract two-level Hilbert spaces introducing an *abstract physical layer*. Thereby we model the imperfect functioning of the device in its abstract description independent of the underlying physical system. At the same time this makes the characterization protocols and theoretical results widely applicable to a variety of different quantum technology platforms provided that they already allow for the required degree of abstraction. In the following we will often refer to the abstract physical layer simply as the physical layer.

On top of the abstract physical layer comes the manipulation of the quantum system in terms of quantum gates, a set of unitary operation that constitute the basic computational operations of the quantum computing device. In the circuit model a computation of a quantum computer is described by a sequence of different quantum gates (a circuit) operating on the registers of qubits. In the plain vanilla model, the qubit registers are initialized in a certain coherent quantum states and after the execution of the circuit a measurement is performed. It is widely believed that the envisioned quantum algorithms require accurately implemented large circuits of gates and can only be run on a quantum computer that, on top of the physical qubit and gate layers, implement a logical layer. Here multiple physical qubits encode the information of a single logical qubit such that the system can be corrected against errors that appear when manipulating the physical qubits [CTV17]. On top of such a stack of layers, one can build an application layer that actually implements the specific tasks of the device, such as running a quantum algorithm [NC10]. Note that the layer stack sketched here is adequate to abstractly model current quantum devices for our purposes. Other task might require a considerably more refined description, especially as the technological development progresses.

When a device already comes with multiple layers of abstraction, one can also characterize the device on the higher levels. A simple quantum device—say, a device that prepares a photon pair in a certain quantum state on demand—might be fully described by its physical layer that coincides with its application layer. For

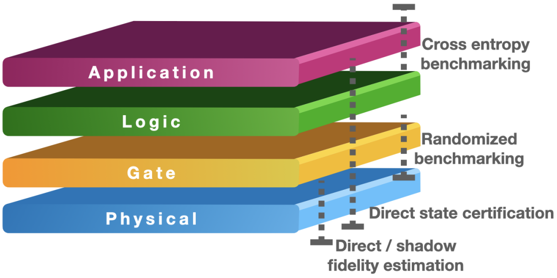

Application
Logic
Gate
Physical
Cross entropy benchmarking
Randomized benchmarking
Direct state certification
Direct / shadow fidelity estimation

Figure 1.1: A complex quantum device comprises multiple abstraction layers. Different protocols aim at certifying the functioning of the device on different layers. NISQ devices are not expected to feature a powerful logical gate layer. Instead applications are directly tailored to the physical gate layer.

simple devices, thus, the terms *quantum system characterization* and *quantum device characterization* are often used interchangeably.

Current and near-term quantum computing devices are still expected to be fairly noisy and of intermediate size, so-called noisy and intermediate scale quantum (NISQ) devices [Pre18]. NISQ devices only allow for few abstractions above the physical layer in terms of (noisy) gate sets. This becomes apparent when looking at one of the major milestones, that was recently achieved [Aru+19], in the development of quantum computers, so-called *quantum (computational) supremacy* [Pre13]. In the past decade theoretical evidence has been collected that it is intractable to produce samples from the measurement output distribution of certain ensembles of quantum states on a classical computer [AA11; BJS10; TD04; BH13; Han+18; Bou+18; Haf+20a; Dal+20; MT19; MT20]. The idea of quantum supremacy demonstrations is to prepare quantum states of such ensembles and perform the measurements on a quantum device, thus, outperforming current classical computers in this specific task. Proposals exist in many flavours for different platforms [Boi+18; BJS10; BMZ16; Mor17; BMS17; BFK18; GWD17; Ber+18; FU16; FH16; MB17]. The task is obviously custom-tailored to the way quantum computing devices work and not necessarily understood to have another useful application beyond the demonstration of quantum supremacy. In our abstract model, the application layer of such a NISQ device performing a quantum supremacy demonstration is described by its output measurement distribution and directly docks to the physical gate layer.

Numerous fields within the quantum sciences have tackled the problem of characterization on different levels of abstraction and from a variety of perspectives. Over the last decades, a large landscape of different protocols has been developed. These protocols operate under distinct assumptions and resource requirements that are well-motivated by the different perspectives. For example, certifying the correct function of a small-scale quantum device used in basic research allows one to invest sizeable effort, might aim at plenty of discriminative information, and can rely on a precise model of the physics of the device. A very different example is the certification of a server, correctly performing a quantum computation, by a remote client with standard desktop hardware. Such a protocol should be light-weight on the client-side and not rely on a detailed model of the server. Nonetheless, as we outline below identifying the information gain, the underlying model assumptions and resource requirements allows one to get a panoramic view of the landscape of quantum characterization protocols in an abstract framework.[2]

Characterization can refer to multiple different tasks which in the literature is referred to by different names and sometimes contradictory names.[3] For this reason, it is instructive to begin with briefly collecting the different tasks of quantum characterization, we have encountered in the literature. This glossary also makes the narrower focus of this thesis on identification tasks more precise. We do not claim that the list is exhaustive.

The tasks of quantum device characterization

Identification ($\supset$ tomography). Given a set, a hypothesis class, of elements each potentially describing the device, identification is the task to determine the *unique* element describing the device from data. For example, a device producing a physical system in a certain quantum state, a state-preparation device, is described by the quantum state it produces. *Quantum state tomography* [Hra97; Jam+01] refers to the identification task of, given measurement data on multiple output states of the device, to determine the quantum state it produces. The hypothesis class can be further

[2] We described this framework together with colleagues in the overview article, Ref. [5]. The version presented here is refined and adapted to the context of the thesis and profited greatly from repeated discussions with Dominik Hangleiter and Nathan Walk. The basic organizing principles of the framework are widely used in the presentation of many works and to a large degree common knowledge in the field. See also Ref. [Fla17].

[3] Wallman, Emerson & Hincks [WFH18] coined the term quantum characterization, verification and validation (QCVV) for a collection of protocols in the context of digital quantum computation. We perceive the emphasis in QCVV on *verification* and *validation* among other potentially relevant tasks such as estimation, identification and benchmarking that all qualify as quantum characterization as rather unmotivated. For this reason we refer to the broad field as quantum (device) characterization.

restricted, e.g. assuming that the state is pure [Gro+10]. The task of quantum state tomography, the related quantum process tomography [CN97] and Hamiltonian identification take the key roles in this thesis. This allows us to be very brief at this point.

Learning. A related task to identification is learning. Given a hypothesis class, learning is the task to determine *an* element within the class that reproduces the measurement outcomes including those that the device has not yet seen. In contrast to identification, learning is not required to have a unique result. Ref. [CT21] gives an introduction into neural-networks-based methods for quantum characterization and provides references to recent seminal examples.

Estimation. Instead of asking for an entire description of a quantum state or process, one might be interested in a single quantity, e.g. the expectation value of an observable or the distance to a reference state in a certain measure. We refer to a protocol that infers a specific quantity of the device as an *estimation protocol.* Of course there exist many quantities one can be interested in and accordingly a plethora of estimation protocols. In Ref. [4] we review a couple of estimation protocols for different distance measures on quantum states and processes. Fidelity measures with respect to pure reference states and unitary processes in particular are linear functions on the state and process giving rise to particularly simple protocols. In this context, there are two flavours of protocols. One common set of techniques uses *importance sampling* to classically select the measurements that reveal most information about the quantity, e.g. [FL11]. The second class, nowadays sometimes referred to as *shadow estimation* [HKP20; 4; 9], fixes an informationally complete and even over-complete frame for the measurement. Thereby, in the measured samples, information of the quantum state is revealed with high-probability and a large class of quantities can be estimated from the samples.

Tomography protocols will typically build upon an estimation protocol that produces the classical input to the reconstruction protocol. Such input can be expectation values of observables or so-called average gate fidelities, see Chapter 5.

Certification ($\supset$ verification, property testing) is the task of either 'accepting' or 'rejecting' the hypothesis that the device is functioning correctly according to a given specification. As there are many ways to write a specification for a quantum device on the different levels of abstraction, there are multiple types of certification tasks arising. We give a panoramic tour through the zoo of existing certification protocols for quantum devices in the review Ref. [5]. Abstractly, robust notions of certification typically make use of a measure of quality that quantifies the deviation from the ideal implementation. The specification is then formulated as a threshold ϵ for this measure of quality. If the deviation according to the measure of quality is above the threshold the ϵ-certification protocol should 'reject' with high probability and 'accept' an ideal implementation, see e.g. Ref. [4, Section B] or Ref. [Han20] for a formal definition. On the physical layer the device is abstractly modelled by quantum states and processes. The ideal functioning of the device might require it to arrive in a specific quantum state after a sequence of operations or to apply a specific quantum operation on demand. Theoretically attractive measures of qualities for the states and processes are the trace distance and the diamond norm that have the operational interpretation of worst-case measures. A selection of prominent examples of certification protocols for the physical layer of NISQ devices is presented in Ref. [4].

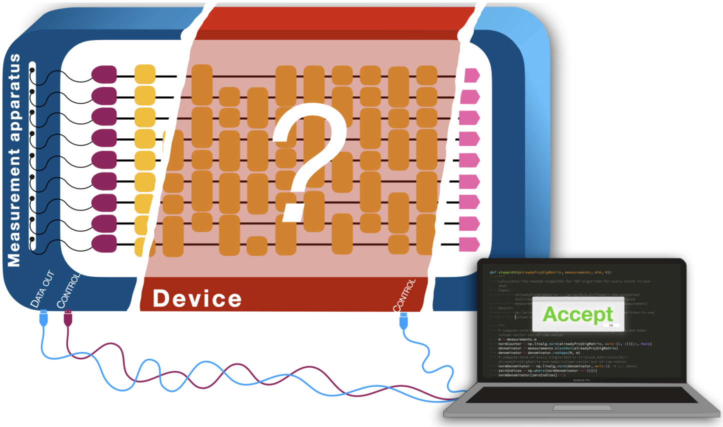
Another way to formulate a specification in this context is to require the device to arrive at a state within a certain subset of all quantum states, i.e. that the state have a certain property. Protocols for testing for such properties are reviewed in Ref. [Md16].

Borrowing from the slang of software development, certification on the application level is also referred to as *verification*. Ref. [GKK19] provides a review of existing approaches for verifying quantum computations by a client on remote devices that are close to being able to accurately perform a universal set of operations.

Benchmarking is the task of comparing the performance of two devices. Benchmarking especially provides pragmatic impetus towards measures of quality that are not directly interpretable on the physical layer. Instead, for the benchmarking of quantum devices it suffices to implicitly define a reproducible performance measure directly in terms of a protocol that estimates it. The only requirement is that the measure is expected to be correlated with the performance in practically relevant tasks.

Validation is the task of ensuring that the specification that was certified is suitable for the intended application of the device.

The protocols for certain tasks often can also be used for other tasks that require less information. For example, a tomographic estimate of a quantum state



```
def stepwidth(alreadyProjbigMatrix, measurements, dim, N):
    """
    calculates the needed stepwidth for SDT algorithm for every block in one
    shot
    Input:
            alreadyProjbigMatrix    [array(N,d,d)[float]] the projected
            multivec                                       lated
            measurem                                       easurement]
    Return:
            mu [arra                                       written in one
            column v

    """
    # compute norm o                                       x and make
    column vector out of row vector
    m = measurements.m
    normCounter = np.linalg.norm(alreadyProjbigMatrix, axis=(2, 1))[::, None]
    denominator = measurements.blockDot(alreadyProjbigMatrix)
    denominator = denominator.reshape(N, m)
    # compute norm of every single matrix in block_matrix(curlA)*
    alreadyProjbigMatrix and make column vector out of row vector
    normDenominator = np.linalg.norm(denominator, axis=1)  # [::,None]
    zeroIndices = np.where(normDenominator == 0)[0]
    normDenominator[zeroIndices] = 1
```

Figure 1.2: The theoretical description of protocols makes use of the distinction into device, measurement apparatus and classical processor.

(identification) can be used to estimate a distance measure to a target state (estimation) that in turn provides a certificate for the related specification (certification). We have ordered the tasks in the list above roughly according to their information gain.

Anatomy of quantum characterization protocols

Theoretically[4], it is convenient to describe a protocol as involving three distinct objects, Fig. 1.2: the *device* that is under scrutiny, the *measurement apparatus* that the protocol employs, and the *classical processor*. The classical processor is a classical computing device that might take care of potentially required pre- and post-processing tasks for the device control and the processing of the output data to arrive at a certificate or even communicates with the device and measurement apparatus in multiple rounds of an interactive protocol. The distinction between the device and the measurement apparatus requires a more detailed discussion and is the concern of the remainder of this section.

There exists a class of protocols where all quantum parts of the device are regarded as a single device that is not subject to any assumptions. In particular,

[4] Theoretically, . . . the classical post-processing time complexity] The following two sections are based on Section I.A of Ref. [4]. Beyond the already published material, they contain a considerably extended discussion of the different regimes of device-dependence and of the figures of merits beyond the context of certification.

they do not involve a somehow characterized separate quantum measurement apparatus. Such protocols are referred to as *device-independent* . The violation of the Bell inequality can certify the presence of non-local correlations ensuring the cryptographic security of quantum key distribution [Aci+07; Pir+09; MPA11]. More general, the presence of sufficiently high entanglement and other properties can be exploited to arrive at device-independent certification protocols for specific quantum states and processes, so-called *self-testing protocols* [SB20]. In the context of quantum communication protocols the effort of fully device-independent protocols can be significantly reduced by introducing mild assumptions on the device such as bounds on the system dimension [Gal+10; PB11; LVB11; Li+11; Li+12]. Such protocols are called semi-device-*in*dependent.

Device-independent and semi-device-independent protocols are limited in the tasks they can solve and information they provide. In particular, such certification protocols can only accept very precisely functioning devices out of the reach of today's quantum computing devices [SB20].

Less paranoid in their assumptions, approaches for the majority of characterization tasks introduce a measurement apparatus to the model. The device and measurement apparatus are not necessarily physically distinct devices. For example, complex measurements on a digital quantum computer are implemented by running a circuit of gates and subsequently performing a set of native measurements. Similarly, the preparation of a desired input state might require running a short circuit on an initial state that can be prepared more directly, e.g. the systems ground state. For example, if we want to characterize the noise process associated to a specific gate, all the operations implementing the measurement and the state preparation might be regarded as being performed by the measurement apparatus. The implementation of the specific gate we want to characterize is *the device* in this example.

The choice of the splitting into the measurement apparatus and the device can be ambiguous and can yield different formulations of equivalent sets of assumptions. In practice, one is often more confident about the correct functioning of a part of the setup that 'performs the measurements' in comparison to the parts that have to be characterized. This motivates to regard the more trusted part of the setup as the measurement apparatus, distinct from the device under scrutiny. At the same time, the split allows one to be fairly conservative in the modelling and assumptions describing the device and aim at a high information gain.

The theoretical formulation of many protocols assume a perfectly working measurement apparatus. In analogy, we refer to such protocols as fully (measurement) *device-dependent*. In practice, it can become an obstacle to balance the required precision of the measurement apparatus with the resource requirements of protocols. For example, a linear shadow estimation protocol [HKP20] for pure-state fidelities of a quantum state employs measurements in different bases [4, Section II.J]. In order to function with a minimal number of state copies, these bases must be related by global unitary rotations of the quantum system. Such measurements can be implemented with a quantum circuits of polynomially many local gates in the number of qubits. For this reason, on actual quantum computing hardware the noise associated with the implementation of such measurements is in many cases considerably higher than the imprecision of the state preparation that one wants to characterize. Resorting to measurements that are simpler and, thus, more precisely implementable unavoidably increases the number of state copies required to arrive at the same precision of the estimate. In consequence, the actually reachable precision of such experiments is limited. Even more severely, in quantum system identification imprecisions in the measurements of device-dependent protocols can significantly deteriorate the precision of the estimates. For example, an error in the measurement, such as an unknown rotation, can be explained by assuming the quantum system was in a rotated state already before the measurement. In this way the error in the measurement can directly enter into the error of a device-dependent quantum state identification protocol. This can reduce the precision of certain protocols dramatically, even to the point where they become infeasible in practice.

This brings us to the class of protocols that are in the focus of this work. One can overcome the limitations in precision and resource demands by relaxing the assumption of a perfect measurement apparatus in the protocol. We call protocols that exhibit some degree of robustness against imprecisions in the measurement apparatus without decreasing the accuracy of the estimates *semi-device dependent*. For NISQ devices and in the further improvement of the quantum technologies scalable semi-device dependent protocols are arguably the most important approaches to characterization as they take the imprecisions of the available hardware into account. An example of semi-device dependent estimation protocols are so-called randomized benchmarking[5] protocols [EAŻ05a; Lév+07; Dan+09; 9] that have become a de-facto standard in assessing the quality of gate implementation of digital quantum computers [Erh+19; McK+19; Aru+19].

[5] Under certain assumptions, randomized benchmarking protocols estimate the so-called average fidelity of a gate set, see discussion in Chapter 5. Thus, randomized benchmarking protocols can and are often regarded as *estimation* protocols even though the name *randomized benchmarking* [Kni+08] stems from the motivation to use the estimated quantity to compare the quality of different devices, without further need for a physical interpretation beyond being the reproducible output of the protocol.

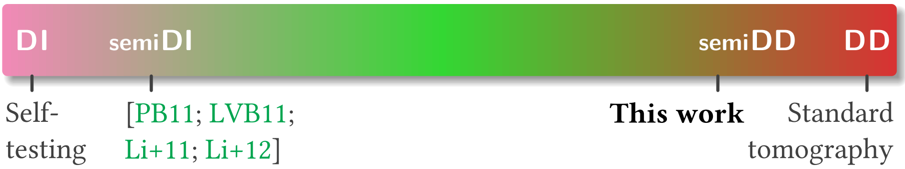


Figure 1.3: Illustration of the spectrum between fully device-independent (DI) and fully device-dependent (DD) quantum system characterization methods such as self-testing and standard tomography, respectively. Semi-device-independent (semiDI) methods relax the stringent requirements of full device-independence. Coming from the opposite end, semi-device-dependent (semiDD) schemes relax assumptions on the measurement apparatus, such as calibration requirements. (The graphic is adopted from Ref. [2])

All together, we arrive at a classification of characterization protocols in a spectrum, illustrated in Fig. 1.3. Device-independent protocols, not involving a measurement apparatus and making assumptions about the device, constitute one end of the spectrum and can be relaxed by introducing assumptions on the device yielding semi-device independent protocols. The characterization tasks for NISQ devices that can be addressed in this regime are limited. Jumping ahead of making more and more assumptions, one introduces a fully trusted, ideally working part of the device, the measurement apparatus. This opposite end of the spectrum, the device-dependent regime, is the natural theoretical idealization of an experimentalist who has already built considerable trust in the working of parts of his setup before proceeding. This idealization, however, severely limits the realm of applicability of these methods in practice, constituting the need for semi-device dependent protocols.

In our discussion above we have already seen that the 'spectrum of assumptions' from device-independence to device-dependence is not the only axis along which we can order and compare different protocols. We now describe the other figures of merit that can be used to assess the realm of applicability of different protocols.

Figures of merit

The landscape of protocols can be roughly organized according to three 'axes': The first axis comprises the set of *assumptions* that are imposed on the device and measurement apparatus, that we introduced above, to guarantee the functioning of the protocol.

A second axis summarizes the *complexity* or amount of the resources that the protocol consumes. A protocol will require a certain number of different measurement settings that the measurement apparatus has to implement, its *measurement complexity*. Each implementation of each of the measurement settings involves operations of certain complexities, the *quantum measurement complexity*. The quantum measurement complexity collects several notions of quantifying the resources that are required to perform a measurement. A concrete example might be the depth of a circuit of local gates required to reduce the measurement to a native measurement of the device. What a meaningful measure is crucially depends on the experimental platform. On fault-tolerant quantum computation hardware one might eventually only regard non-transversal gates as non-free resources for the measurement. In order to arrive at statistical estimates in a protocol, one requires a total number of repetitions of device invocations referred to as the *sample complexity*.

The practical indication of the different complexity measures again depend on the device under consideration. While in a photonics experiment that can achieve high repetition rates the sampling complexity of a protocol is often not a limiting factor. Instead, reducing the number of different measurement settings, e.g., the number of different configurations of optical elements on the table, can be highly desirable [11]. In contrast, a computing device using trapped ions typically features much lower repetition rate. Here, changing the measurement setting for every invocation might not be problematic and the protocol's sampling complexity becomes crucial. For NISQ devices, however, the practical limitations on the quantum measurement complexity are an even more important consideration.

Furthermore, in the face of the desired complexity of the quantum devices, a particularly critical figure of merit for a protocol is its *classical processing complexity*, i.e. the demands in space and time of the classical processing tasks. The exponential scaling of the configuration space of the quantum systems in terms of its constituents gives rise to classical post-processing tasks that quickly exhaust the available classical computing power. In many tasks of quantum system identification a scaling in the overall Hilbert space dimension is unavoidable. In these instances the exact scaling, i.e. the degree of the polynomial, plays a crucial role. To put things into perspective, a simple protocol to reconstruct a quantum state on a d-dimensional quantum system that makes use of a semi-definite program to take the positivity constraint into account, implemented with an off-the-shelf solver, uses $O(d^6)$ storage for its objective variable. For a 5-qubit system such an algorithm already allocates more than 4GB (in single floating precision 4bytes). Larger systems are already out of the reach of current desktop hardware and 8 or 10-qubit reconstruction would allocate over 1PB and over 4.5EB of storage, respectively. In contrast, a more efficient non-convex optimization to recover a pure quantum state requires only $O(d)$ storage bringing the storage cost of a 30-qubit system down to about 4GB. For quantum channels that already scale with d^4 in their degrees of freedom the classical processing power becomes even faster a scarce resource. Already a square-root improvement is therefore of practical importance, significantly extending a protocol's applicability.

For our present scope, the mentioned notions of complexity are the most important ones and will be in the focus of our discussion. Note however that the list is by far not complete. For example, interactive protocols might be compared in terms of challenging demands in the timing of the device's control or the complexity of communication.

The third and final axis is the *information gain* of the protocol. Intuitively, a quantum system identification protocol that outputs a quantum state as an estimate for a prepared arbitrary quantum state extracts more information about the device than an estimation protocol for the distance of the prepared state to a certain fixed target state. The precise information gain depends on the measures of quality. Different measures of quality have different discriminatory power among the hypothesis class that models the device compatible with the protocol's set of assumptions. Concomitant with less information gain of a task, it is conceivable that one can design a protocol with significantly less complexity compared to one gaining more information. Theoretically quantifying and analysing the information gain in performing a characterization task allows one to derive lower bounds on the complexity of any protocol for this task. As we saw above with the example of the storage complexity of classical post-processing, designing protocols with optimal complexity can be crucial for its application in a practical setting. Unfortunately, often protocols achieving the optimal scaling in one complexity measure, say, the sampling complexity, might impose theoretical limits on the achievable complexity in another measure, e.g. the classical post-processing time complexity.

1.2 Overview of the results

After broadening our scope to quantum characterization tasks in general and en passant defining the title of this work within its context, we now come to our focus: *semi-device dependent quantum system identification.* Here, we revisit the three tasks of identifying low-rank states, unitary quantum processes and non-interacting Hamiltonians with resource-efficient (complexity) and semi-device-dependent (assumptions) protocols. In the remainder of the introduction, we give the specific motivation for each of these tasks, briefly review the approaches that exist in the literature and condensely summarize our own results. This shall serve as a first overview of this thesis that omits the technical details going into the derivations of the results. Correspondingly, the only formulas that appear here are statement of the scaling behaviour of the resource requirements of the protocol, that constitute main results of this work.

1.2.1 Blind quantum state tomography

One of the most basic diagnostic tasks in the development of quantum technologies is the identification of a quantum states from experimentally measured data, commonly referred to as *quantum state tomography*.[6] Indeed, at the heart of every quantum computation is the preparation of a quantum state. Quantum state tomography can therefore provide valuable information for improving quantum devices beyond a mere certification or benchmarking of their correct functioning. It has been a de-facto standard in research on many precisely controllable quantum systems [Ste+06; Lob+08; Pry+03; Bar+12; Sch+13; McC+16].

However, in any such endeavour one encounters the following fundamental challenge: In order to arrive at an accurate state estimate, most tomography schemes rely on measurement devices that are calibrated to a very high precision. At the same time, a precise and detailed characterization of a measurement device requires an accurate state preparation. But improving the accuracy of the state preparation using tomographic information was our goal to begin with. We are trapped in a vicious cycle. This vicious cycle, depicted in Figure 1.4, constitutes a fundamental obstacle to the improvement of quantum devices.[7]

A make-or-break question is therefore: Is there any hope to break this cycle at all? In other words, can one perform quantum state tomography *blindly*, that is, without full knowledge of the measurement apparatus to begin with? Or even better, can one simultaneously infer a quantum state and learn certain unknown calibration parameters of the measurement apparatus in a *self-calibrating tomography scheme* [Bra+12]? A simple parameter count indicates that this is typically impossible by just measuring a set of mutually orthogonal observables: While an arbitrary quantum state in a d-dimensional Hilbert space is characterized by

[6] One of the most ... estimates in practice.] is based on the introduction of Ref. [2]. Individual passages are adopted verbatim.

[7] We are grateful to Susane Calegari for kindly providing drawings used the figure.

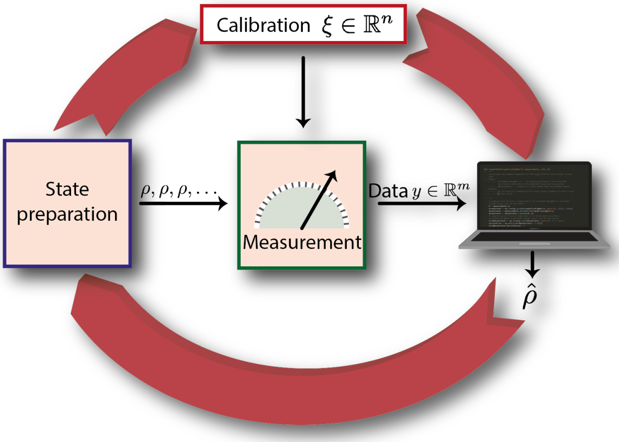

Calibration $\xi \in \mathbb{R}^n$
State preparation
$\rho, \rho, \rho, \ldots$
Measurement
Data $y \in \mathbb{R}^m$
$\hat{\rho}$

Figure 1.4: In the quest to engineer high fidelity quantum technologies one encounters a vicious cycle: Extracting *actionable advice* to correct for error in the state preparation requires accurate quantum state estimation. The accuracy of a state estimate crucially relies on the precise calibration of the measurement device. But the calibration can ultimately only be tested and improved if high fidelity quantum states are provided.

$d^2 - 1$ many real parameters, at the same time, the number of linearly independent measurements in this space implies that we can learn at most d^2 independent parameters. This leaves room for a single additional (calibration) parameter only. This prohibits even slightly relaxing the requirement of a complete and accurate characterization of the measurement device towards a partially uncalibrated device. Tomography of an arbitrary quantum state is therefore typically intrinsically measurement *device-dependent* in this sense.

A couple of *semi-device-dependent* approaches to the problem in specific settings exist in the literature. In Ref. [Mog10] it has been argued that single photon detectors can be simultaneously calibrated during state tomography under the assumptions that the state is squeezed extending the mindset of Ref. [MŘH09]. Ref. [MŘH12] provides a more extensive discussion of potential classes of states and the recent Ref. [Sim+19] derives error bars. All these schemes combine specific measurement models with a restricted hypothesis class of states in the identification.

Another approach to more flexibly assume partial information about the state preparation device and the measurement apparatus is the *Gram matrix completion* proposed in Refs. [Sta12; Sta14]. Here, one assumes that certain entries of the measurement and the state in a basis as well as some expectation values are already precisely known a priori. A correlation matrix that simultaneously encode the measurement, the state and the expectation values is then completed from a subset of known indices. We are however not aware of a concrete experimental situation where this type of partial knowledge naturally arises. So-called data-pattern tomography [ŘMH10] avoids the calibration of the measurement device by assuming that the device can accurately prepare some well-controlled reference states such as coherent states [Mot+14]. For the reconstruction the data is compared to previously determined signatures of the reference states.

In spite of these promising approaches, the question whether self-calibrating tomography is possible in more generic settings and without severe restrictions of the hypothesis set remains elusive. Ref. [Bra+12] has reported the experimental demonstration of simultaneously reconstructing a quantum state together with certain unknown unitary rotations associated to the measurement device via maximum likelihood estimation in a linear optics setting.

Here we develop a framework in which we can prove that blind tomography of quantum states for generic measurements models is possible by making a natural assumption about the quantum state, namely that it is of low rank. To this end, we formulate the *blind tomography problem* as the recovery task of a highly structured signal. We use a general model of the measurement apparatus which applies to a variety of relevant experiments: we model the measurements as depending linearly on the unknown parameters of the possible calibration errors. In many situations the daunting uncertainty about the device calibration is small and can be approximated as a linear deviation from an empirically known calibration baseline. We illustrate our measurement model and the underlying assumption with a concrete model inspired by implementation of quantum computing in trapped ions, in Section 3.1.

Our formulation of the blind tomography problem takes the form of a linear inverse problem where the solution is assumed to have a certain low-rank structure. Linear inverse problems under structure assumptions are studied in the mathematical discipline of *compressed sensing* [FR13]. In compressed sensing one studies the reconstruction of sparse vectors and low-rank matrices or tensors from linear data in the regime where without the structure assumptions the problem is under-determined. In general, optimally exploiting the structure in the inverse problems can be shown to be NP-hard. Nonetheless, for generic measurements drawn from suitable random ensembles one can prove that different algorithmic approaches such as convex solvers or projective gradient descent algorithms converge to the correct solution of the problem with high probability, such theorems are referred to as *recovery guarantees*.

For the tomography of low-rank quantum states there exist device-dependent compressed sensing schemes. Compressed sensing algorithms, e.g. nuclear-norm minimization, where introduces in order to reduce the measurement complexity while still ensuring an efficient classical post-processing in the Hilbert space dimension [Gro+10; Gro11; Liu11; KKD15; Kue15; Kab+16]. These schemes come with theoretical guarantees and have also been successfully employed in experiments [11; Rio+17; Sha+11]. The practical applicability of compressed sensing tomography schemes rests on their robustness and stability against various imperfections of the experimental setup. Small deviations from the compressive model assumption and additive errors to the measurement outcomes, e.g. induced by finite statistics, are reflected in a proportional and only slightly enhanced estimation error. Still, the state tomography schemes rely on measurement apparata that are calibrated to very high precision. Blind tomography aims at relaxing the device dependence. In distinction, in the previous schemes low-rank assumptions on the quantum state were considered to reduce the complexity of a tomography scheme, giving rise to an important *quantitative improvement*. Here, we aim at exploiting the low-rank structure to make blind tomography possible in the first place—a *qualitative improvement* over the known schemes. We find that the signal structure of the blind tomography problem is not directly admissible to compressed sensing algorithms. More formally we show as a first result that already the projection onto the class of solutions of the blind-tomography problem is NP-hard and, thus, a projective gradient algorithm is not efficient. To circumvent this bottleneck we consider a relaxation of the blind tomography problem to the problem of de-mixing a sum of low-rank matrices from linear measurements. An efficient algorithmic solution to the de-mixing problem also solves the blind tomography problem. We state an efficient projective gradient algorithm for the de-mixing problem and analyse by building on and extending the framework of hierarchical compressed sensing.

The price to pay in the relaxation of the problem is a worse measurement complexity: While one might hope to recover a rank-r state in Hilbert space dimension d and n calibration parameters with order of $dr + n$ measurement settings, solving the corresponding de-mixing problem requires ndr measurement settings. This price is acceptable if the calibration involves only a comparatively small number of calibration parameter compared to the Hilbert space dimension. To remedy this situation, our scheme allows one to exploit yet another structure, significantly extending the realm of applicability and efficiency of the scheme, namely the *sparsity* of the calibration. Physically, this structural property amounts to the assumption that only a small number s out of the many possible calibration errors has occurred in the specific experiment.

We derive theoretical guarantees using insights from hierarchical compressed sensing, a special instance of model-based compressed sensing [Bar+10]. A compressed sensing framework for the reconstruction of hierarchically structured signals was developed in a series of works [14; 13; 12; 15; 16; 17; 18; 19; 20]. A summary of this line of work can be found in the book chapter [21]. This work is independently motivated by different applications in classical machine-type communications. But the problems arising in machine-type communications feature similar underlying mathematical structures as the quantum tomography and results readily carry over.

For the sparse de-mixing problem, we show that the proposed algorithm successfully converges to the correct solution under the assumption that the measurement acts close to isometrically on the set of structured solutions. Furthermore, for a generic measurement ensemble we show that the required isometry property is satisfied with high-probability for a number of measurements that scales as $s\log(n/s)+drs$. The logarithmic dependency on the number of potential calibration parameters n makes this scheme scalable in n and allows one to work with flexible models of the systematic measurement errors and calibration corrections. We complement our analytical results with numerical studies that demonstrate the feasibility of the scheme in relevant parameter regimes. Furthermore, we numerically investigate a more pragmatic algorithmic approach using constrained alternating minimization to exploit the full structure of the blind tomography problem. We find that the alternating minimization is capable of solving the blind tomography problem for a realistic measurement and calibration model. Analytical recovery guarantees for this algorithm are, however, not easily attainable with the presented proof strategies.

Our analytical work initiated from the conceptual question of whether we can provably break the vicious tomography cycle in generic situations. We give an affirmative answer to this question by developing a semi-device-dependent tomography scheme with a provable recovery guarantee for low-rank quantum states. To this end, we consider unstructured and formally generic measurement models that allow us to apply the compressed sensing techniques for optimal measurement complexities. Our numerical results collect significant evidence that the newly developed approach provides a readily usable and flexible tool that is expected to be capable of increasing the precision of tomographic estimates in practice.

1.2.2 Compressive quantum process tomography

After studying the tomography of the static property of a quantum device, its state, we now turn our attention to identifying the description of devices that manipulate the quantum states. This task is referred to as quantum process tomography (QPT) [CN97]. The characterization of quantum processes is a versatile tool to diagnose errors and noise of a quantum device that manipulates quantum states.

In standard quantum process tomography protocols, the experiment inputs a set of input states and performs a set of measurements on the results of the process acting on each of the input states. If the input states do not entangle the system with ancillary systems that are also subjected to measurement, one calls such protocols *prepare-and-measure* schemes. Such schemes are generally considered simpler to realize. The output data are, thus, again statistical estimation of the output of a linear map acting on the quantum process. Using vector-space isomorphisms, the Choi-Jamiołkowski isomorphism, one can generalize results on quantum state tomography to process tomography. The restriction to prepare-and-measure schemes readily translates to a unit-rank condition of the linear measurement operators that has to be taken into account in the analysis [Fla+12].

In terms of our abstract model introduced above in Section 1.1 p. 19, the state-preparation and the measurement comprise the measurement apparatus and the part of the experiment performing the quantum process is the device. In the context of quantum process tomography imperfections in the measurement apparatus are commonly referred to as SPAM (state preparation and measurement) errors. For the characterization of gates of a simple digital quantum processor it has been observed that SPAM errors make quantum process tomography insufficiently inaccurate in practice [EAŻ05b; MGE11]. Roughly speaking, the operations involving multiple gates that are required for the state preparation and measurement are more complicated and, thus, more prone to error than the individual gate under scrutiny. Thus, the most important flavour of semi-device-dependence for the characterization of quantum processes is *SPAM (error) robustness*.

This fundamental obstruction for making a well-motivated cut between the measurement apparatus and the device in characterizing quantum gates, has motivated the introduction of self-consistent tomography methods for gate sets [Mer+13; Blu+13; Gre15; Blu+17; COB20]. In *gate set tomography (GST)* one extracts a full description of the entire gate set, the state preparation and the measurement from the observed output statistics of gate sequences. Concomitant with the information gain—full tomographic information of an entire gate set, the input state and measurement—gate set tomography is resource-intense in the number of sequences, overall samples and the classical post-processing. This remains true for schemes that incorporate structure assumptions such as compressive GST [23] or extensive prior knowledge such as linear [Gu+21] and Bayesian approaches [Eva+22]. To date, it remains an open problem to equip gate set tomography with rigorous complexity and error bounds.

We here take another route to the SPAM-robust characterization of the noise processes associated to gate sets: It has been realized that considering random sequences of gates one can, under certain assumptions, efficiently extract the average gate fidelity of a gate-independent noise-process of a gate set [EAŻ05a; Lév+07; Kni+08; Dan+09; MGE11]. By combining the output of multiple such *randomized benchmarking (RB) protocols*, one can extract tomographic information about the noise process [Kim+14]. Such RB tomography schemes naturally inherit the SPAM-robustness of the RB protocol used in the data acquisition.

The classical pre-/post-processing of RB protocols is efficient in the number of qubits if the gate-set features an underlying tractable group structure. The arguably most prominent example in quantum computing of such a group structure is exhibited by the Clifford group. The Clifford group is a finite group with elements characterized by a number of bits that scale quadratically in the number of qubits. This structure allows one to efficiently simulate circuits of Clifford unitaries[8] on a classical computer by means of the Gottesman-Knill theorem [NC10]. Their practical relevance for quantum computing is further rooted in their comparatively simple realization in fault-tolerant architectures, that can perform error-correction of qubits on a logic layer [Got09; CTV17]. In resource theories of quantum computing Clifford operations are therefore often considered a free resource and circuit complexities quantify only the number of non-Clifford gates [Vei+14; HC17].

It has been an open problem[9] to combine the robustness of RB tomography protocols with the efficiency of compressed sensing protocols that exploit the low-rank structure of coherent noise processes. Obtaining actionable advice regarding coherent errors is of particular practical interest as such errors can often be

[8] Efficient simulation further requires the input state to be of low stabilizer rank and the measurement to be Pauli-basis measurements.

[9] It has been an open problem …formulation of our results.] is based on the introduction of Ref. [1]. Individual passages are adopted verbatim.

corrected by experimental control. Correcting coherent error is additionally motivated by fault-tolerant quantum computation: Refs. [Kue+16; Wal15] indicate that it is coherent errors that lead to an enormous mismatch between average errors, which are estimated by randomized benchmarking and worst-case errors reflected by fault-tolerance thresholds.

In devising compressive versions of RB protocols, it turns out that the proper design of the measurements is crucial [KL17]. In contrast to the computationally tractable group structure of RB protocols, compressed sensing methods typically require measurements with less structure in this context, in that their fourth-order moments are close to those of the uniform Haar measure. Thus, the key technical question is whether the seemingly conflicting requirements of sufficient randomness and desired structure in the measurements can be combined.

We here show that the answer is indeed yes. We demonstrate that Clifford-group-based measurements are also sufficiently unstructured that they can be used for compressed sensing. Thus, we develop methods for such quantum process tomography that are resource efficient, robust with respect to SPAM, and use measurements that are already routinely acquired in many experiments. In more detail, we provide procedures for the reconstruction from so-called average gate fidelities (AGFs), which are the quantities that are measured in randomized benchmarking under suitable assumptions. It was established that the unital part of general quantum channels can be reconstructed from AGFs relative to a maximal linearly independent subset of Clifford-group operations [Kim+14]. We generalize this result by noting that the Clifford group can be replaced by an arbitrary unitary 2-design and also explicitly provide an analytic form of the reconstruction—basically reproving a result by Ref. [Sco08].

Our main result of this part is a practical reconstruction procedure for quantum channels that are close to being unitary. Let d be the Hilbert space dimension, so that a unitary quantum channel can be described by roughly d^2 scalar parameters. The protocol is rigorously guaranteed to succeed using essentially order of d^2 AGFs with respect to randomly drawn Clifford gates, and we also prove it to be stable against errors in the AGF estimates. In this way, we generalize a previous recovery guarantee [KL17] from AGFs with 4-designs to ones with the more relevant Clifford gates. Conversely, we also prove that the sample complexity of our reconstruction procedure is optimal in a simplified measurement setting.

Following further along the question of what information can be obtained from multiple AGFs, we also find a new interpretation of the *unitarity* [Wal+15]—a figure of merit that captures the coherence of noise. We show that this quantity can be estimated directly from AGFs rather than simulating purity measurements as proposed in Ref. [Wal+15]. This potentially provides an alternative route to unitary estimation compared to the RB protocols [DHW19].

Our main technical contributions are results for the second and fourth moments of AGF measurements with random Clifford gates. For the second moment, we provide an explicit formula improving over the previous lower bound [KL17]. In the case of trace-preserving and unital maps, our analysis gives rise to a tight frame condition. In order to prove a bound on the fourth moment, we derive—as a more universal new technical tool—a general integration formula for the fourth-order diagonal tensor representation of the Clifford group. The proof, presented in Chapter 4 builds on recent results on the representation theory of the multiqubit Clifford group [Zhu+16; HWW18; GNW21]. Our result is the Clifford analogue to Collins's integration formula for the unitary group [Col03; CS06] for fourth orders.

Chapter 4 also discusses further results and applications of random ensembles arising from different measures on the unitary group. We briefly summarize results on the general construction of Clifford and unitary designs with higher moments from random circuits, that were derived with collaborators in Ref. [6]: We find that a random circuit with order of n^2k^9 one and two-local Clifford gates acting only on adjacent qubits approximate a unitary k-design in relative error. This establishes that one can directly work with random Clifford circuits instead of compiling random multi-qubits Clifford unitaries in many applications. We furthermore give a bound for the number of non-Clifford gates required in a random circuit in order to construct unitary k-designs with $k \geq 3$. The overall scaling of our construction improves on the previous construction of Ref. [BHH16b]. Deviating a bit from the focus on quantum device characterization, we briefly present further applications of locally random ensembles of quantum states in the study of the equilibration of quantum systems and summarize the results of Refs. [7; 8] in Section 4.4. We formulate a supposedly weak condition for the energy eigenstates of a Hamiltonian that we have shown to guarantee the equilibration of the system even when initialized in product states. We further present ensembles of states that generically fulfil the condition.

The derivation of our main results on compressive RB tomography are presented in Chapter 6. In the preceding Chapter 5 we provide a detailed introduction and review of RB techniques. In this context we briefly explain our result on establishing the general data form of RB experiments under broad circumstance [9]. We further derive sample complexity statements for extracting average gate fidelities in idealized settings.

1.2.3 Hamiltonian identification

The third part of the thesis studies the precise identification of non-interacting Hamiltonians in an analogue quantum simulation experiment. Generally put, finding the most parsimonious dynamical equations that govern the time evolution of a physical system is arguably one of *the* central tasks in physics. Another more fundamental problem is maybe solely presented in the task of identifying the relevant physical degrees of freedoms to begin with.[10] The multitude of successful approaches over centuries to this key problem can be described as well-motivated heuristics guided by trained expert intuition and domain knowledge, and ultimately challenging luck—the common scientific method that is brilliantly summarized in Feyerabend's famous sentiment *"Anything goes."* [Fey75]. By making informed guesses about idealized models that feature simple rules of interactions we hope to capture the essence of the physics and describe laboratory as good as we can. Fanned by the advances in computational power, the automation of essential parts of physics discovery in data-driven top-down approaches has been envisioned [CM87; SL09; Man+17; 22; Ite+20]. Data-driven Hamiltonian inference is a central task in this endeavour and is therefore not only interesting from a technological but also from a foundational perspective. It touches on the fundamental question: Is it possible to directly infer laws of nature from data? Answers to this question are elusive for complex natural physical systems which can neither be accessed nor probed with sufficient accuracy and control.

Analogue quantum simulators present a very different, novel playground for these questions. An analogue quantum simulator is a physical system which is designed to accurately realize an idealized Hamiltonian model under precisely controlled conditions [CZ12]. Examples include the Bose-Hubbard Hamiltonian which describes a cold-atom setup [Jak+98; Gre+02], and quantum Ising or Heisenberg models which can be implemented in ion-trap setups [BR12; Zha+17], arrays of Rydberg atoms in optical tweezers [Bar+16; Ber+17], and in superconducting qubit architectures [Rou+17]. Compared to a universal, fault-tolerant quantum computer that can perform a digital simulation, analogue quantum simulation is possible in considerably simpler experiments. With such experiments, so the hope, we could shed light on fundamental questions of physics that have remained elusive to the standard methods of inference, such as the

[10] Generally put, finding the most . . . today's devices.] As of writing the text is not yet published elsewhere. A variation of this text was authored together with Dominik Hangleiter and is going to be published as the introduction of an extended manuscript on the presented work.

mechanism behind high-temperature superconductivity [Köh+05] and the persistence of many-body localization in higher dimensions [Cho+16].

Already when building an analogue simulator we must understand the simulator system extremely well to be able to reduce any unwanted interactions and sources of error. Such understanding is typically built by characterizing individual components of the system and reducing the noise sources one after another and has yielded unprecedented understanding and control of the physics of quantum simulation platforms in recent years [CZ12; BDN12; BR12; Ací+18]. In the previous parts, we have already encountered a rich toolkit for the diagnostic and characterization of individual qubits and gates of universal digital quantum computers such as the robust randomized benchmarking of entire gate sets or efficient compressive tomographic methods. Then, trust in the performance of large circuits can be inferred from composing well-characterized constituents. Such techniques from the digital regime can to some extent be used for characterizing components of analogue simulators [SLP11; Hol+15; Der+20]. But when using a physical system for an analogue computation much more stringent requirements on the accuracy apply. Now, it is necessary to capture and understand the machine's dynamics that is relevant for the computation as a whole—in other words, on its application layer.

The experimenters' thorough understanding of the interactions and sources of errors and control over the complex systems that motivates analogue simulations in the first place also allows for a controlled reductionistic approach to data-driven Hamiltonian inference. For example, one can attempt to isolate individual physical excitations and solely observe their free dynamics. This is a very different starting point compared to natural physical systems that do not allow for reducing complexity to this extent. From a phenomenological view point of a condensed matter physicist, the study of such simple systems might even be regarded as 'boring'. From a technological view point, thoroughly exploring Hamiltonian inference problems in this newly accessible regime is, however, of utmost importance and a necessary first step for their precise certification and characterization.

Several methods have been proposed for estimating Hamiltonian parameter from data[11]: The key approaches that we will be also our starting point, is the tracking of the dynamics of single excitations [BMN09; BMN11; BM09; DPK09; WM10; BY12] and using tools from Fourier analysis for processing of time trace data

[11] Several methods …superconducting qubit platforms.] is based on a section of Ref. [3] with verbatim adoptions.

[SKO04; Col+05; Col+06; CDH06; SOD08; SO09; OS12]. For general Hamiltonian models, one can then exploit algebraic structure of the Hamiltonian terms to simplify the data processing problem [ZS14; SC17]. We will work out a particular simple example of this type. Other approaches to the problem include learning a Hamiltonian from a single eigenstate or its steady state [GG18; QR19; CC18; BAL19; Bai+20] using quantum-quenches [LZH20; Cze21], or with general-purpose machine-learning methods [Gra+12; WDD15; Val+19; BSH21; Kra+19; Che+21a].

Hamiltonian parameter estimation has been demonstrated for fixed instances of two- and three-qubit Hamiltonians in nuclear magnetic resonance (NMR) experiments [Lap+12; HLL17; Che+21b; Zha+21]. However, these work do not yet live up to the demands of a robust and scalable recovery of Hamiltonians arising in superconducting qubits, trapped ions or cold atoms in optical lattices. In contrast to NMR, such platforms involves beyond incoherent noise sizeable systematic errors in the state preparation and the measurement. Very recently, learning of two-qubit dissipative Lindblad dynamics, i.e. a characterization of certain errors, has been demonstrated [Flu+20; Sam+21; OKC21] on superconducting qubit platforms.

In this work, we follow the outlined bottom-up imperative for studying Hamiltonian inference problems from analogue simulators starting with a 'simple' instance. We explore the Hamiltonian identification problem of non-interacting dynamical systems in a setting motivated by the diagnostic requirements of analogue simulations using gmon transmon qubit processors. We find that taking the structure of the problem seriously, gives rise to a rich signal processing problem already for this simple dynamics. We develop a measurement protocol and signal processing pipeline that allows to fully exploit the problem's inherent structure. To this end, we make use of super-resolution and de-noising algorithms for tone finding [RPK86; Fan16; CF13; CF14] and manifold optimization over the orthogonal group [EAS98; AEK09; AMS09]. In turn, the structural constraint allow us to devise a method that has the necessary robustness against unavoidable imperfections in the state preparation and measurements of today's devices.

The approach of this part of the thesis is distinct from the previous two parts. In the first two parts, we developed methods and equipped them with strong theoretical guarantees. We only demonstrated their practicality in numerical simulations. Here, we devise a method that is custom-tailored to a concrete experimental setup and demonstrate its working in an actual experiment. In turn, we here leave an in-depth theoretical analysis of the method's performance to future research.

The Hamiltonian dynamics is implemented on a Google Sycamore chip by our collaborators at Google AI. The chip consists of coupled superconducting qubits arranged in a two-dimensional array. Since our Hamiltonian dynamics is limited to a single excitation, the qubits can truthfully represent bosons and fermions equally. From measured data, we recover all Hamiltonian parameters for five and six coupled qubits using our method and achieve sub-MHz precision. Averaging the deviation of the recovered Hamiltonian parameter and the target Hamiltonian corresponding to a specific qubit on the chip over multiple Hamiltonian instances, we can associate a performance benchmark to the chip component. Using such benchmarks, we construct a spatial map for the implementation error for a grid of 27 qubits. Beyond thereby providing a first framework for quantifying the implementation accuracy of analogue dynamic simulators, the method also establishes diagnostic toolkit for understanding, calibrating and improving the device. For example, for the specific experiment we find that the effects of ramping phases where the devices parameters rapidly but not instantaneously are driven to their values implementing the Hamiltonian, are the dominant sources of errors.

In our bottom-up approach to Hamiltonian identification we consider a setting that is easily accessible to classical simulations. We conclude the chapter with a complementary discussion on the obstacles arising in the classical simulations of quantum systems using Monte-Carlo techniques [Hir+82; Tro+03; Pol12; Tro+10]. Here we again deviate from the narrower focus on quantum device characterization and briefly present the framework of Ref. [10] for the systematic *easing* of the so-called Monte-Carlo sign problem. Very concisely stated, the Monte-Carlo sign problem refers to the observation that positive entries in the basis expansion of a Hamiltonian ultimately lead to an unfavourable increase in the sampling complexity of Monte-Carlo estimators for, e.g. the expectation value of observables in the corresponding Gibbs state. We introduce a computationally tractable measure of *non-stoquasticity* for a Hamiltonian in a basis and discuss how it captures the complexity of Monte-Carlo simulations. This measure can be optimized over tractable basis transformations using the same manifold optimization techniques that we employed in our Hamiltonian identification algorithm. At the same time we establish the complexity-theoretic limitations of a systematic approach to easing the sign problem.

1.3 Structure of the thesis

We conclude the introduction with an overview over the structure of the thesis. In the following chapter we introduce the mathematical objects and notations that are featured in the results of the thesis. We give some perspective on the metric structures and their interpretation on the spaces of quantum states and processes. A particular emphasis is on reviewing the theory of Haar random unitary matrices as they and their de-randomizations in terms of unitary designs will be important in the second part of the thesis. In Chapter 3 we study the first semi-device-dependent task: *blind quantum state tomography*. Chapters 4, 5 and 6 together constitute the second part of the thesis. To set the stage for the second task of *compressive randomized benchmarking tomography*, the subject of Chapter 6, we prepend two separate chapters: Chapter 4 takes a look at the properties of relevant random ensembles within the unitary group. We here derive the integration formula for the Clifford group that we need in Chapter 6, discuss the scaling of local random circuits that reproduce the properties of random Clifford and general unitaries, and present further applications of locally random ensembles of quantum states and unitaries in the context of equilibration. Chapter 5 subsequently gives an introduction and review of randomized benchmarking techniques. Here, we also briefly explain our framework guaranteeing the general data form of RB protocols and derive sampling complexities that we use to establish the optimality of compressive randomized benchmarking tomography. The third task of *Hamiltonian identification for analogue simulators* is presented in Chapter 7. At the end of the chapter we provide a brief perspective on easing the Monte-Carlo sign problem. We close each of the chapters that present our results on the three semi-device-dependent identification tasks with concluding remarks on the respective topic.

2 Preliminaries

Many results of the thesis are mathematical theorems that build on rigorous mathematical definitions of the underlying concepts. To set the stage, we briefly collect some basic mathematical concepts that appear in the field of quantum characterization and the quantum information sciences more generally. These preliminaries will function as a reference for the notation used later on and fundamental results that we invoke throughout the thesis. The presentation is based on material that we authored for the pedagogical tutorial Ref. [4]. For this reason, we are occasionally slightly more generous in the selection of material and depth of explanations than necessarily required for the presentation of the later results.

2.1 Mathematical objects of quantum mechanics

In order to discuss quantum states we set up some mathematical notation.[1][2] We focus on finite-dimensional quantum mechanics in accordance with our emphasis on digital quantum computing. Hence, we assume all vector spaces to be finite-dimensional. The space of *linear operators* from a vector space V to a vector space W is denoted by $\mathrm{L}(V, W)$, and we set $\mathrm{L}(V) := \mathrm{L}(V, V)$. A *Hilbert space* is a vector space with an inner product $\langle \cdot, \cdot \rangle$. Let $\mathcal{H}$ and $\mathcal{K}$ be complex Hilbert spaces throughout the tutorial. We denote the *adjoint* of an operator $X \in \mathrm{L}(\mathcal{H}, \mathcal{K})$ by $X^\dagger$, i.e. $\langle k, Xh \rangle = \langle X^\dagger k, h \rangle$ for all $h \in \mathcal{H}$ and $k \in \mathcal{K}$.

As customary in physics, we will use the bra-ket-notation (Dirac notation): We denote vectors by ket-vectors $|\psi\rangle \in \mathcal{H}$ and linear functionals on $\mathcal{H}$ by bra-vectors $\langle\psi|$, which are elements of the dual space $\mathcal{H}^*$. Furthermore, we understand ket-vectors and bra-vectors with the same label as being related by the

[1] The definitions, notation and results we present in the preliminaries are standard in the field of quantum information and can—if not mentioned otherwise—be found, e.g., in Refs. [NC10; Wat11; Hal13; Bha97].

[2] Most parts of the preliminaries are taken from the tutorial Ref. [4] that was authored together with Martin Kliesch. Some parts of the tutorial were based on lecture notes by Martin Kliesch available as Ref. [Kli19].

canonical isomorphism induced by the inner product. In bra-ket notation we frequently drop tensor-product operators to shorten the notation, e.g. $|\psi\rangle\,|\phi\rangle := |\psi\rangle \otimes |\phi\rangle \in \mathcal{K} \otimes \mathcal{H}$ or $|\psi\rangle\langle\psi| := |\psi\rangle \otimes \langle\psi| \in \mathcal{K} \otimes \mathcal{H}' \cong \mathrm{L}(\mathcal{K}, \mathcal{H})$ for $|\psi\rangle \in \mathcal{K}$ and $|\phi\rangle \in \mathcal{H}$.

To describe the state of a quantum system we require the notion of *density operators*. The real subspace of *self-adjoint* operators, $X = X^\dagger$, is denoted by $\mathrm{Herm}(\mathcal{H}) \subset \mathrm{L}(\mathcal{H})$ and the convex cone of *positive semidefinite* operators by $\mathrm{Pos}(\mathcal{H}) := \{X \in \mathrm{Herm}(\mathcal{H}) \mid \langle\psi|\, X \,|\psi\rangle \geq 0\}$. The *trace* of an operator $X \in \mathrm{L}(\mathcal{H})$ is $\mathrm{Tr}[X] := \sum_i \langle i|\, X \,|i\rangle$, where $\{\,|i\rangle\} \subset \mathcal{H}$ is an arbitrary orthonormal basis of $\mathcal{H}$. The vector space $\mathrm{L}(\mathcal{H})$ is itself a Hilbert space endowed with the Hilbert-Schmidt (trace) inner-product

$$\langle X, Y\rangle := \mathrm{Tr}[X^\dagger Y]\,. \tag{2.1}$$

The set of *density operators* is defined as $\mathcal{D}(\mathcal{H}) := \{\rho \in \mathrm{Pos}(\mathcal{H}) : \ \mathrm{Tr}[\rho] = 1\}$. We sometimes simply decorate the set of density operators with the dimension $d = \dim \mathcal{H}$ of the underlying Hilbert space and write $\mathcal{D}^{(d)}$.

Outcomes of a quantum measurement are modelled by random variables. Abstractly, a *random variable* is defined as a measurable function from a probability space to a measurable space $\mathcal{X}$. Here, we will exclusively be concerned with two types of random variable: (i) Those that take values in a finite, discrete set $\mathcal{X} \cong [n] := \{1, \dots, n\}$ (understood as the measurable space with its power set as the σ-algebra) and (ii)) those that take values in the reals $\mathcal{X} = \mathbb{R}$ (with the standard Borel σ-algebra generated by the open sets). In practice, the underlying probability space is often left implicit and one describes a random variable X taking values in $\mathcal{X}$ directly by its probability distribution $\mathbb{P}$ that assigns a probability to an element of the σ-algebra of $\mathcal{X}$. For example, for a random variable X taken values in $\mathbb{R}$ and $I \subset \mathbb{R}$ an interval, we write $\mathbb{P}[X \in I]$ for the probability of X assuming a value in I. Abstractly speaking, $\mathbb{P}$ is the push-forward of the measure of the probability space to $\mathcal{X}$ induced by the random variable X. Thus, $\mathbb{P}$ is sufficient to describe X. The underlying probability space is, however, important to define correlations between multiple random variables which are understood to be defined on the same probability space.

The probability distribution of a discrete random variable X taking values in a finite set $\mathcal{X} \cong [n]$ is characterized by its *probability mass function* $p_X : [n] \to [0, 1]$, $k \mapsto p_X(k) := \mathbb{P}[X = k] := \mathbb{P}(X \in \{k\})$. A real random variable X is characterized by its *(cumulative) probability distribution* $P_X : \mathbb{R} \to [0, 1]$, $x \mapsto P_X(x) := \mathbb{P}[X < x] := \mathbb{P}[X \in (\infty, x)]$ or in case it is absolutely continuous by its *probability density* $p_X : \mathbb{R} \to [0, 1]$, $x \mapsto p_X(x) := \frac{d}{dt}\big|_x P(t)$. Note that if a discrete random variable takes values in a discrete subset of $\mathbb{R}$ we can also assign a non-continuous (cumulative) probability distribution.

The most general way to define a linear map from density operators $\mathcal{D}(\mathcal{H})$ to random variables is by means of a *positive operator valued measure (POVM)*. A POVM is a map from (the σ-algebra) of $\mathcal{X}$ to $\operatorname{Pos}(\mathcal{H})$. For a discrete random variable X taking values in $[n]$ a POVM is uniquely defined by a set of *effects* $\{E_i \in \operatorname{Pos}(\mathcal{H})\}_{i=1}^n$ with

$$\sum_{i=1}^{n} E_i = \mathbb{1}_{\mathcal{H}}\,, \tag{2.2}$$

where $\mathbb{1}_{\mathcal{H}} \in \mathrm{L}(\mathcal{H})$ denotes the identity operator. Strictly speaking the POVM is the map on the power set of $[n]$ that extends $k \mapsto E_k$ additively. It is convenient and common to refer to the set of effects as the POVM. A POVM M (with effects) $\{E_i \in \operatorname{Pos}(\mathcal{H})\}_{i=1}^n$ induces a map from $\mathcal{D}(\mathcal{H})$ to random variables. To this end, we associate to ρ the random variable M_ρ with probability mass function $p_{\mathsf{M}_\rho}(k) \coloneqq \langle \rho, E_k \rangle$.

These are the ingredients to formalize the *static* postulates of quantum theory.

Postulates (quantum states and measurements):

- A quantum system is associated with a (separable) complex Hilbert space $\mathcal{H}$.
- The state of a quantum system, its *quantum state*, is described by a density operator $\rho \in \mathcal{D}(\mathcal{H})$
- A *measurement* with potential outcomes in a finite, discrete set $O \cong [n]$ is described by a POVM M with effects $\{E_i\}_{i\in[n]}$.
- If a quantum system is in the state $\rho \in \mathcal{D}(\mathcal{H})$ and the measurement M is performed the observed outcome is a realization of the random variable M_ρ associated to ρ by M.

The set $\mathcal{D}(\mathcal{H})$ is convex. Its extremal points are rank-one operators. A quantum state $\rho \in \mathcal{D}(\mathcal{H})$ of unit rank is called a *pure* state. In particular, there exist a state vector $|\psi\rangle \in \mathcal{H}$ such that $\rho = |\psi\rangle\langle\psi|$. The state vector associated to a pure quantum state is only unique up to a phase factor. A general quantum state is therefore a convex combination of the form $\sum_i p_i \, |\psi_i\rangle\langle\psi_i|$, where p is a *probability vector*, i.e., an entry-wise non-negative vector $p \in \mathbb{R}^d$, $p \geq 0$ that is normalized, i.e., $\sum_i p_i = 1$. A quantum state that is not pure is called *mixed*.

The Hilbert space associated a composite quantum system consisting of two quantum systems with Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, is the tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$. The operators $\mathrm{L}(\mathcal{H}_1)$ can be embedded into $\mathrm{L}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ by $A \mapsto A \otimes \mathbb{1}$. Dually to that, for any state $\rho \in \mathcal{D}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ we can introduce its *reduction* to system 1 by demanding that

$$\mathrm{Tr}[\rho\,(A \otimes \mathbb{1})] = \mathrm{Tr}[\rho_1 A]\,. \tag{2.3}$$

The reduced state captures all information of ρ that can be obtained from measuring system 1 alone. The *partial trace* $\mathrm{Tr}_2 : \mathrm{L}(\mathcal{H}_1 \otimes \mathcal{H}_2) \to \mathrm{L}(\mathcal{H}_1)$ that linearly extends the action $X \otimes Y \mapsto X\,\mathrm{Tr}[Y]$ maps the state ρ to its reduced state ρ_1. By $\mathbb{F} \in \mathrm{L}(\mathcal{H} \otimes \mathcal{H})$ we denote the *flip operator* (or *swap operator*) that is defined by linearly extending

$$\mathbb{F}\,|\psi\rangle\,|\phi\rangle := |\phi\rangle\,|\psi\rangle\ . \tag{2.4}$$

In a basis $\{\,|i\rangle\}_{i=1}^{\dim(\mathcal{H})}$ of $\mathcal{H}$, we can express $|\psi\rangle \in \mathcal{H} \otimes \mathcal{H}$ by a coefficient matrix $A \in \mathbb{C}^{\dim \mathcal{H} \times \dim \mathcal{H}}$ as $|\psi\rangle = \sum_{i,j} A_{ij}\,|i\rangle\,|j\rangle$. The coefficient matrix of $\mathbb{F}\,|\psi\rangle$ is given by the matrix transpose $A^\intercal$ of A with entries $(A^\intercal)_{i,j} = A_{j,i}$.

2.2 Distance measures for quantum states

In this section we introduce some ‘natural’ measures on quantum states. One main application of these measures within this work is to specify the deviation of the estimate of a quantum state identification protocol from the actual state that generated the data.

To this end, recall that any *normal* operator $X \in \mathrm{L}(\mathcal{H})$, i.e., any operator that commutes with its adjoint, $[X, X^\dagger] := XX^\dagger - X^\dagger X = 0$, can be written in spectral composition $X = \sum_i x_i P_i$, where $x_i \in \mathbb{C}$ are its eigenvalues and $P_j = P_j^2 \in \mathrm{Pos}(\mathcal{H})$ the corresponding spectral projectors. There are several useful norms of an operator $X \in \mathrm{L}(\mathcal{H}, \mathcal{K})$. For any operator $X \in \mathrm{L}(\mathcal{H}, \mathcal{K})$ between two Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, the operator $X^\dagger X$ is positive semidefinite, i.e., in $\mathrm{Pos}(\mathcal{H})$. In consequence, it has a positive semidefinite square root $|X| := \sqrt{X^\dagger X} \in \mathrm{Pos}(\mathcal{H})$.

The *spectral norm* (a.k.a. operator norm) $\|X\|_\infty \in \mathbb{R}_+$ of X is defined to be the largest eigenvalue of $|X|$. The *trace norm* is $\|X\|_1 := \mathrm{Tr}[|X|]$ and the *Frobenius norm* $\|X\|_\mathrm{F} := \sqrt{\mathrm{Tr}[|X|^2]} = \sqrt{\mathrm{Tr}[X^\dagger X]}$. These norms can be defined in a variety of equivalent ways: The spectral norm coincides with the norm induced by the ℓ_2-norm on $\mathcal{H}$ via $\|X\|_\infty = \sup_{\|v\|_{\ell_2} \leq 1} \|Xv\|_{\ell_2}$, a manifestation of the

Rayleigh principle. The Frobenius norm is induced by the *Hilbert-Schmidt inner product* (2.1). It can also be expressed in terms of the matrix representation of X as $\|X\|_{\mathrm{F}} = \sum_{i,j} |X_{ij}|^2$. Finally, all three norms are instance of the *Schatten p-norms* that are directly defined as ℓ_p-norms on the singular value spectrum. The singular value spectrum $\sigma(X)$ of X is defined as the eigenvalue spectrum of $|X|$ and the ℓ_p-norms are given by $\|x\|_{\ell_p} := (\sum_i |x_i|^p)^{1/p}$. This gives rise to the unitarily invariant Schatten p-norm $\|X\|_p := \|\sigma(X)\|_{\ell_p}$ and $\|\cdot\|_\infty$, $\|\cdot\|_1$, and $\|\cdot\|_{\mathrm{F}}$ are the Schatten p-norms with $p = \infty, 1, 2$, respectively.

The Euclidean inner product is bounded by ℓ_p-norms by the Hölder inequality that states that for all $x, y \in \mathbb{C}^d$ and pairs $p, q \in \{1, 2, \dots, \infty\}$ with $p^{-1} + q^{-1} = 1$ (understanding $1/\infty = 0$) it holds that

$$|\langle x, y\rangle| \leq \|x\|_{\ell_p} \|x\|_{\ell_q}. \tag{2.5}$$

The Hölder inequality generalizes the Cauchy-Schwarz inequality where $p = q = 2$. The Schatten p-norms inherit a *matrix Hölder inequality* from the Hölder inequality: Let $X, Y \in \mathrm{L}(\mathcal{H}, \mathcal{K})$ and p, q as before, then

$$|\langle X, Y\rangle| \leq \left\|X^\dagger Y\right\|_1 \leq \|X\|_p \|Y\|_q. \tag{2.6}$$

The Hölder inequality directly follows from the von Neumann inequality $\operatorname{Tr}|AB| \leq \langle \sigma(A), \sigma(B)\rangle$ where the singular value spectra $\sigma(A)$ and $\sigma(B)$ are each in descending order [Bha97]. Furthermore, the Schatten p-norms inherit the ordering of the ℓ_p-norms, $\|X\|_\infty \leq \ldots \leq \|X\|_2 \leq \ldots \leq \|X\|_1$ for all X. Norm bounds in reversed order will in general introduce dimensional factors. For low-rank matrices these bounds can be tightened.

Lemma 1 (Reversed norm bounds). *For all $X \in \mathrm{L}(\mathcal{H}, \mathcal{K})$ it holds that*

$$\|X\|_1 \leq \sqrt{\operatorname{rank}(X)}\, \|X\|_{\mathrm{F}} \leq \operatorname{rank}(X)\, \|X\|_\infty. \tag{2.7}$$

Proof. Let $X \in \mathrm{L}(\mathcal{H}, \mathcal{K})$ and $r = \operatorname{rank}(X)$. We can always write $X = XP_r$ with P_r a rank-r projector onto the orthogonal complement of the kernel of X. Now by the matrix Hölder inequality (2.6) $\|X\|_1 = \|XP_r\|_1 \leq \|P_r\|_{\mathrm{F}} \|X\|_{\mathrm{F}} = \sqrt{r}\, \|X\|_{\mathrm{F}}$. For the second inequality, bound again using the matrix Hölder inequality $|\operatorname{Tr}(X^\dagger X)| \leq \left\|X^\dagger X\right\|_1 \leq \|P_r\|_1 \left\|X^\dagger X\right\|_\infty = r\, \|X\|_\infty^2$. Taking the square root we conclude that $\|X\|_{\mathrm{F}} \leq \sqrt{r}\, \|X\|_\infty$ from which the second inequality follows. □

A natural metric on quantum states is the *trace distance* $\mathrm{dist}_{\mathrm{Tr}} : \mathcal{D}(\mathcal{H}) \times \mathcal{D}(\mathcal{H}) \to \mathbb{R}_+$,

$$\mathrm{dist}_{\mathrm{Tr}}(\tilde{\rho}, \rho) = \frac{1}{2} \left\| \rho - \tilde{\rho} \right\|_1 . \tag{2.8}$$

We have already seen that compared to the other Schatten p-norms the trace norm is the largest, i.e., the most 'pessimistic' distance measure. Furthermore, the trace norm has an operational interpretation in terms of the distinguishability of quantum states by dichotomic measurements.

Proposition 2 (Operational interpretation of the trace distance)**.** *Let $\rho, \sigma \in \mathcal{D}(\mathcal{H})$. It holds that*

$$\mathrm{dist}_{\mathrm{Tr}}(\rho, \sigma) = \sup_{0 \leq P \leq \mathbb{1}} \mathrm{Tr}[P(\rho - \sigma)] . \tag{2.9}$$

Furthermore, the supremum is attained for the orthogonal projector P^+ onto the positive part of $\rho - \sigma$.

We refer to Ref. [4] for a proof of the proposition.

Given two quantum states the optimal dichotomic POVM measurement $\{P, \mathbb{1} - P\}$ to distinguish the two states is the POVM that maximizes the probability of measuring the outcome associated to P in one state and minimizes the same probability for the other state. Of course exchanging the role of P and $\mathbb{1} - P$ works equivalently. We can think of the achievable differences in probabilities as a measure for the distinguishability of ρ and σ. Proposition 2 shows that the trace distance of two states coincides with the maximal distinguishability by any dichotomic POVM measurements. By measuring multiple iid. copies one can amplify the distinguishability of a single shot measurement.

Let us introduce another important distance measure on quantum states. The *(squared*[3]*) fidelity* of two quantum state $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ is defined as

$$\mathrm{F}(\rho, \sigma) := \left\| \sqrt{\rho}\sqrt{\sigma} \right\|_1^2 . \tag{2.10}$$

One can rewrite

$$\left\| \sqrt{\rho}\sqrt{\sigma} \right\|_1 = \mathrm{Tr}\left[\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}} \right] . \tag{2.11}$$

[3] Some authors define the fidelity as $\left\| \sqrt{\rho}\sqrt{\sigma} \right\|_1$ without the square. For this reason, one might want to refer to the expression of (2.10) explicitly as the squared fidelity to avoid confusion. For brevity, we however call F simply the fidelity hereinafter.

While not any more directly evident from (2.11), the fidelity is symmetric as is apparent from (2.10).

The fidelity is more precisely not a measure of ‘distance’ for two quantum states but of ‘closeness’. In particular, $\mathrm{F}(\rho,\rho)=1$, which can be seen to be the maximal values of $\mathrm{F}(\rho,\sigma)$ for all $\rho,\sigma\in\mathcal{D}(\mathcal{H})$. Hence, $0\leq\mathrm{F}(\rho,\sigma)\leq 1$ on $\mathcal{D}(\mathcal{H})$. Often it is convenient to work with the *infidelity* $1-\mathrm{F}(\rho,\sigma)$ as the complementary measure of ‘distance’.

If at least one of the states ρ or σ is pure, say $\rho=|\psi\rangle\langle\psi|$, then

$$\mathrm{F}(\rho,\sigma)=\langle\psi|\,\sigma\,|\psi\rangle=\mathrm{Tr}[\rho\sigma]=\langle\rho,\sigma\rangle\,, \tag{2.12}$$

which follows from (2.11). Furthermore, for both states being pure we have $\mathrm{F}(|\psi\rangle\langle\psi|\,,|\phi\rangle\langle\phi|)=|\langle\psi|\phi\rangle|^2$ for all $|\psi\rangle\langle\psi|\,,|\phi\rangle\langle\phi|\in\mathcal{D}(\mathcal{H})$. Thus, for pure states the fidelity is the overlap of the states and can be related to the angle between the state vectors. We will in fact mostly encounter the case where at least one of the states is pure and mostly work with (2.12) instead of (2.10).

The fidelity is related to the trace distance as follows.

Proposition 3 (Fuchs-van-de-Graaf inequalities [Fv99, Theorem 1])**.** *For any states $\rho,\sigma\in\mathcal{D}(\mathcal{H})$*

$$1-\sqrt{\mathrm{F}(\rho,\sigma)}\leq\frac{1}{2}\left\|\rho-\sigma\right\|_1\leq\sqrt{1-\mathrm{F}(\rho,\sigma)}\,. \tag{2.13}$$

Since the Fuchs-van-de-Graaf inequalities are not explicitly dependent on the Hilbert-space dimension one can regard the trace-distance and fidelity as equivalent measures of quality in many applications. Note however that the square-root on the right-hand side can still make a painstaking difference in practice. Aiming at a trace-norm distance of 10^{-3} can in the worst case require an infidelity of 10^{-6}. This can be a crucial difference when it comes to the feasibility of certification. Importantly, the square-root scaling is unavoidable for pure states.

Lemma 4 (Fuchs-van-de-Graaf inequality for pure states)**.** *The upper bound of the Fuchs-van-de-Graaf inequality for pure states $|\psi\rangle\langle\psi|\,,|\phi\rangle\langle\phi|\in\mathcal{D}(\mathcal{H})$ is tight, i.e.*

$$\left\|\,|\psi\rangle\langle\psi|-|\phi\rangle\langle\phi|\,\right\|_p=2^{1/p}\sqrt{1-|\langle\psi|\phi\rangle|^2}\,. \tag{2.14}$$

Proof. Denote $X := |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|$. We have $\mathrm{Tr}[X] = 0$ and $\mathrm{rank}(X) \in \{0, 2\}$. Hence, X has two eigenvalues $\lambda > 0$ and $-\lambda < 0$. This implies that $\lambda^2 = \|X\|_{\mathrm{F}}^2 /2 = 1-1\,|\langle\psi|\phi\rangle|^2$, as directly follows by writing $\|X\|_{\mathrm{F}}^2$ as a Hilbert-Schmidt inner product. From the eigenvalues one can calculate $\|X\|_p$ as Schatten p-norm, which yields the result. □

In Lemma 4 we showed that the upper bound of (2.13) is tight for pure states. Conversely, one might hope for more mixed states to arrive at an improved scaling closer to the lower-bound of (2.13). We will review such a bound in the analogous discussion of distance measures of quantum channels, Theorem 22 in Section 2.4.

2.3 Random states and unitaries

Random ensembles of quantum states and unitary matrices find ubiquitous applications in quantum information processing [CN16] and, in particular, in quantum characterization protocols. Roughly speaking, random unitary operations together with a fixed quantum measurement allow one to quickly gain information about the entire state space. One main technical ingredients to the compressive randomized protocol devised in Chapter 6 is a novel integration formula for the Clifford group. To set the stage, we review in this section the mathematical preliminaries that are required to understand and prove our result.

Arguably the simplest probability distribution on the unitary group $\mathrm{U}(d)$ is given by the *Haar measure* $\mu_{\mathrm{U}(d)}$. In general, for a compact Lie-group the Haar measure is the unique left and right invariant probability measure. It generalizes the notion of a uniform measure. In applications one is often interested in random variables that are polynomials in matrix elements of a Haar-random unitary U and its complex-conjugate $U^\dagger$. In this case, also all moments of the random variable are the expected value of such polynomials. In this section we will introduce the mathematical theory required to explicitly calculate such moments. To this end, we observe that any polynomial $p_t(U, U^\dagger)$ of degree k can be written as the contraction with two matrices $A, B \in \mathbb{C}^{dk\times dk}$,

$$p_k(U, U^\dagger) = \mathrm{Tr}[BU^{\otimes k}A(U^\dagger)^{\otimes k}]\,. \tag{2.15}$$

This motivates to define the *kth moment operator* of a probability measure μ on

$\mathrm{U}(d)$ as $\mathcal{M}_\mu^{(k)} : \mathbb{C}^{dk\times dk} \to \mathbb{C}^{dk\times dk}$,

$$\begin{aligned}\mathcal{M}_\mu^{(k)}(A) &= \mathbb{E}_{U\sim\mu}\big[U^{\otimes k} A (U^\dagger)^{\otimes k}\big] \\ &= \int_{\mathrm{U}(d)} U^{\otimes k} A (U^\dagger)^{\otimes k} \mathrm{d}\mu(U).\end{aligned} \tag{2.16}$$

If we have an expression for the kth moment operator for the Haar measure $\mu_{\mathrm{U}(d)}$, we can calculate the expectation value of arbitrary polynomials $p_k(U, U^\dagger)$ over $U \sim \mu_{\mathrm{U}(d)}$ by a linear contraction (2.15).

The crucial property that characterizes the kth moment operator of $\mu_{\mathrm{U}(d)}$ is the following: Consider a fixed unitary $U \in \mathrm{U}(d)$. Then a short calculation exploiting the unitary invariance of the Haar measure reveals that

$$U^{\otimes k} \mathcal{M}_{\mu_{\mathrm{U}(d)}}^{(k)}(A) = \mathcal{M}_{\mu_{\mathrm{U}(d)}}^{(k)}(A) U^{\otimes k}. \tag{2.17}$$

We find that $\mathcal{M}_{\mu_{\mathrm{U}(d)}}^{(k)}(A)$ commutes with every unitary U raised to the kth tensor power.

For a set of endomorphisms $\mathcal{A} \subset \mathrm{L}(W)$ on a vector space W one calls the set

$$\mathrm{comm}(\mathcal{A}) = \{B \in \mathrm{L}(W) \mid BA = AB \quad \forall A \in \mathcal{A}\} \tag{2.18}$$

of all endomorphisms that commute with all elements of $\mathcal{A}$ the *commutant of* $\mathcal{A}$. The following lemma establishes that not only does $\mathcal{M}_{\mu_{\mathrm{U}(d)}}^{(k)}(A)$ commute with every unitary of the form $U^{\otimes k}$, but it is in fact the orthogonal projector onto the commutant of $\mathcal{A} = \{U^{\otimes k} | U \in \mathrm{U}(d)\}$, where orthogonality is understood with respect to the Hilbert-Schmidt inner product (2.1). As will become motivated shortly, we refer to

$$\Delta_d^k : \mathrm{U}(d) \to \mathrm{U}(d^k)\,, \qquad U \mapsto U^{\otimes k} \tag{2.19}$$

as the *diagonal representation* of $\mathrm{U}(d)$.

Lemma 5 (kth moment operator). *The kth moment operator $\mathcal{M}_{\mu_{\mathrm{U}(d)}}^{(k)}$ is the orthogonal projector onto $\mathrm{comm}(\Delta_d^k(\mathrm{U}(d)))$, the commutant of the k-order diagonal representation of $\mathrm{U}(d)$.*

Proof. With (2.17) we have established that the range of $\mathcal{M}_{\mu_{\mathrm{U}(d)}}^{(k)}$ is in $\mathrm{comm}(\Delta_d^k(\mathrm{U}(d)))$. The converse also holds since for $A \in \mathrm{comm}(\Delta_d^k(\mathrm{U}(d)))$ we calculate that $\mathcal{M}_{\mu_{\mathrm{U}(d)}}^{(k)}(A) = A\mathcal{M}_{\mu_{\mathrm{U}(d)}}^{(k)}(\mathbb{1}) = A$. Thus, it remains to check the orthogonality $(\mathcal{M}_{\mu_{\mathrm{U}(d)}}^{(k)})^\dagger = \mathcal{M}_{\mu_{\mathrm{U}(d)}}^{(k)}$. The orthogonality requirement follows in very few lines of calculation using linearity and cyclicity of the trace. □

The commutant of the diagonal representation of the unitary group can be characterized using a powerful result from representation theory: Schur-Weyl duality. We start by reviewing some basic definitions and results from representation theory.

2.3.1 Representation theory and twirling

An introduction into representation theory can be found, e.g. in the books [Sim96] and [GW00]. Let G be a group and $\mathrm{U}(\mathcal{H})$ the unitary group of a Hilbert space $\mathcal{H}$. A *(unitary) linear representation* $R : G \to \mathrm{U}(\mathcal{H})$ is a group homomorphism, i.e. a linear map fulfilling $f(gg') = f(g)f(g')$ for all $g, g' \in G$. One can generalize the notion of eigenspace and diagonalization of vector space homomorphisms to representations. A subspace $V \subset \mathcal{H}$ is *invariant under* R if $R(g)V \subseteq V$ for all $g \in G$. A representation R is *irreducible* if its only invariant subspaces are the trivial subspaces, $\mathcal{H}$ and $\{0\}$. One can decompose a representation into its irreducible components giving rise to a block-diagonal form (see e.g. [Sim96, Theorem II.2.3]).

Proposition 6 (Decomposition into irreps). *Let $R : G \to \mathrm{U}(\mathcal{H})$ be a unitary representation of a group G on a finite-dimensional Hilbert space $\mathcal{H}$. Then there exist irreducible representations $(R_i, \mathcal{H}_i)$ of G such that*

$$\mathcal{H} = \bigoplus_i \mathcal{H}_i \quad \text{and} \quad R(g) = \bigoplus_i R_i(g)\,. \tag{2.20}$$

For two irreducible representations R, R' appearing in the decomposition (2.20) there might exist a unitary transformation such that $R(g) = SR'(g)S^\dagger$ for g. The number m of such *unitarily equivalent* representations is called the *multiplicity* of an irreducible representation in the decomposition. If all irreducible representations are mutually inequivalent, we say that the decomposition is *multiplicity free.*

The most fundamental theorem of representation theory is Schur's lemma.

Theorem 7 (Schur's lemma). *Let $R : G \to \mathrm{U}(\mathcal{H})$ and $\tilde{R} : G \to \mathrm{U}(\tilde{\mathcal{H}})$ be two irreps of G on finite dimensional Hilbert spaces $\mathcal{H}$ and $\tilde{\mathcal{H}}$. If $A \in \mathrm{L}(\mathcal{H}, \tilde{\mathcal{H}})$ satisfies*

$$AR(g) = \tilde{R}(g)A \qquad \forall g \in G \tag{2.21}$$

then either $A = 0$ or R_1 and R_2 are unitarily equivalent up to a constant factor.

A proof can be found, e.g. in Ref. [4].

A more general perspective on the moment operator is the twirling over a group with the diagonal representation[4]. More generally, for a unitary representation $R : G \to \mathrm{U}(V)$ of a subgroup $G \subset \mathrm{U}(d)$ carried by a vector space V, we define $T_R : \mathrm{L}(V) \to \mathrm{L}(V)$ ("twirling") as

$$T_R(A) = \int_G R(g) A R(g)^\dagger d\mu_G(g), \tag{2.22}$$

where μ_G is the invariant measure induced by the Haar measure on $\mathrm{U}(d)$. We have $\mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}} = T_{\Delta^k_d}$.

The argument of the proof of Lemma 5 now applies more generally and yields the analogous statement for arbitrary representation of groups equipped with a Haar measure, e.g. the uniform measure on a finite subgroup. We will more specifically rely on the following well-known properties of T_R that are straightforward to verify:

Lemma 8 (Properties of T_R). *Let R be a unitary representation of a subgroup $G \subseteq \mathrm{U}(d)$. Then, for all $A \in \mathrm{L}(V)$ and $B \in \mathrm{Comm}(R(G))$, the map T_R (defined in Eq. (2.22)) fulfils*

$$\mathrm{Tr}(T_R(A)) = \mathrm{Tr}(A), \tag{2.23}$$

$$T_R(AB) = T_R(A)B, \tag{2.24}$$

$$T_R(A) \in \mathrm{Comm}(R(G)). \tag{2.25}$$

2.3.2 Schur-Weyl duality and the commutant of the diagonal action

To calculate the moments of random variables depending on Haar-random unitaries, we are interested in understanding the commutant of the diagonal representation of the unitary group. Formally, we define the diagonal representation of $\mathrm{U}(d)$ on $(\mathbb{C}^d)^{\otimes k}$ as

$$\Delta^k_d : \mathrm{U}(d) \to \mathrm{U}\Big((\mathbb{C}^d)^{\otimes k}\Big) \tag{2.26}$$

by linearly extending the action

$$\Delta^k_d(U)(|\psi_1\rangle \otimes \cdots |\psi_k\rangle) := (U\,|\psi\rangle_1) \otimes \cdots (U\,|\psi_k\rangle)\,. \tag{2.27}$$

[4] This paragraph is taken from Section A of the supplemental material of Ref. [1].

The representation Δ_d^k has a duality relation with another well-known representation on $\mathbb{C}^k$: the representation π_k of the *symmetric group* $\mathfrak{S}_k$ permuting the k tensor components:

$$\begin{aligned}&\pi_k : \mathfrak{S}_k \to \mathrm{U}\Big((\mathbb{C}^d)^{\otimes k}\Big)\,,\\ &\pi_k(\sigma)(\left|\psi_1\right\rangle \otimes \dots \otimes \left|\psi_k\right\rangle) := \left|\psi_{\sigma^{-1}(1)}\right\rangle \otimes \dots \otimes \left|\psi_{\sigma^{-1}(k)}\right\rangle\,.\end{aligned} \tag{2.28}$$

We note that $\pi_k(\sigma)$ and $\Delta_d(U)$ commute for any $\sigma \in \mathfrak{S}_k$ and $U \in \mathrm{U}(d)$.

Let us consider the following two irreducible representations of the symmetric group which appear in the decomposition, Proposition 6, of π_k for any k. We call $\left|\Psi\right\rangle \in (\mathbb{C}^d)^{\otimes k}$ *symmetric* if $\pi_k(\sigma)\left|\Psi\right\rangle = \left|\Psi\right\rangle$ for all $\sigma \in \mathfrak{S}_k$ and *antisymmetric* if $\pi_k(\sigma)\left|\Psi\right\rangle = \operatorname{sign}(\sigma)\left|\Psi\right\rangle$ for all $\sigma \in \mathfrak{S}_k$. The *symmetric subspace* $\mathcal{H}_{\mathrm{sym}^k}$ and *antisymmetric subspace* $\mathcal{H}_{\wedge^k}$ of $(\mathbb{C}^d)^{\otimes k}$ are the subspaces consisting of all symmetric and all antisymmetric vectors, respectively.

Lemma 9 (Symmetric subspace)**.** *The orthogonal projectors onto the symmetric and antisymmetric subspace are*

$$P_{\mathrm{sym}^k} = \frac{1}{k!}\sum_{\sigma\in\mathfrak{S}_k} \pi_k(\sigma) \quad \textit{and} \quad P_{\wedge^k} = \frac{1}{k!}\sum_{\sigma\in\mathfrak{S}_k} \operatorname{sign}(\sigma)\pi_k(\sigma)\,, \tag{2.29}$$

respectively. The dimension of the symmetric subspace $P_{\mathrm{sym}^k}(\mathbb{C}^d)^{\otimes k}$ *is*

$$\operatorname{Tr}[P_{\mathrm{sym}^k}] = \binom{k+d-1}{d-1}\,. \tag{2.30}$$

Proof. The first statement follows, e.g., for P_{sym^k} by realizing that any symmetric vector in the range of P_{sym^k} and that this operator in indeed a projector, i.e., that P_{sym^k} is self-adjoint and $P_{\mathrm{sym}^k}P_{\mathrm{sym}^k} = P_{\mathrm{sym}^k}$.

The second statement is a combinatorial one. The trace of the symmetric projector is the number of ways to distribute k indistinguishable particles (bosons) into d boxes (modes), i.e., the dimension of the corresponding subspace of the bosonic subspace, which is known to be given by the binomial coefficient. □

In physics, we ubiquitously use the fact that any matrix can be decomposed into the direct sum of a symmetric and antisymmetric matrix. In other word, under the action of π_2 we have the irreducible decomposition

$$(\mathbb{C}^d)^{\otimes 2} = \mathcal{H}_{\mathrm{sym}^2} \oplus \mathcal{H}_{\wedge^2}\,. \tag{2.31}$$

This decomposition has a generalization to π_k with $k > 2$, the *Schur-Weyl* decomposition. The Schur-Weyl decomposition relies on a duality relation between the commuting representations Δ_d^k and π_k. The representations Δ_d^k and π_k span each other's commutant as algebras.

Theorem 10 (Schur-Weyl duality [GW00, Theorem 4.2.10])**.** *For the two commuting representations* (2.28) *and* (2.27) *it holds that*

$$\operatorname{comm}(\Delta_d^k(U(d))) = \operatorname{span}\{\pi_k(\mathfrak{S}_k)\} \tag{2.32}$$

and

$$\operatorname{comm}(\pi_k(\mathfrak{S}_k)) = \operatorname{span}\{\Delta_d^k(U(d))\}\,. \tag{2.33}$$

By Schur's lemma such a duality relation implies that the multiplicity spaces of the irreducible representation of one representation are irreducible representations of the dual representation and vice versa. In other words, $\mathbb{C}^d$ decomposes into multiplicity-free representations of the combined action $\mathrm{U}(d) \times \mathfrak{S}_k$. In order to state this composition, we write $\lambda = (\lambda_1, \lambda_2 \ldots, \lambda_{l(\lambda)}) \vdash k$ for a partition of k into $l(\lambda)$ non-increasing integers with $\lambda_1 \geq 1$ and fulfilling

$$k = \sum_{i=1}^{l(\lambda)} \lambda_i\,. \tag{2.34}$$

Such partitions of integers label the irreducible representations of the symmetric group and the diagonal representation. As a consequence of Schur-Weyl duality one can prove.

Theorem 11 (Schur-Weyl decomposition [GW00, Theorem 9.1.2])**.** *The action of* $\mathrm{U}(d) \times \mathfrak{S}_k$ *on* $(\mathbb{C}^d)^{\otimes k}$ *given by the commuting representations* (2.28) *and* (2.27) *is multiplicity-free and* $(\mathbb{C}^d)^{\otimes k}$ *decomposes into irreducible components as*

$$(\mathbb{C}^d)^{\otimes k} \cong \bigoplus_{\lambda \vdash k, l(\lambda) \leq d} W_\lambda \otimes S_\lambda, \tag{2.35}$$

where $\mathrm{U}(d)$ *acts non-trivially only on* W_λ*, the* Weyl modules*, and* $\mathfrak{S}_k$ *acts non-trivially only on* S_λ*, the* Specht modules*. For any* $k \geq 2$*, both* $\mathcal{H}_{\mathrm{sym}^k}$ *and* $\mathcal{H}_{\wedge^k}$ *occur as components in the direct sum* (2.35).

Schur-Weyl duality, Theorem 10, and the resulting decomposition, Theorem 11, give a simple characterization of the commutant of the diagonal action. The relation (2.32) allows one to derive an expression for the k-moment operator $\mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}$

as the orthogonal projector onto the span of the symmetric group. But one has to be careful since $\{\pi_k^d(\sigma)\}_{\sigma\in\mathfrak{S}_k}$ is not an orthonormal basis. Note that it only becomes an orthogonal set asymptotically for large k which can be exploited in some applications, e.g. [BHH16b].

Before deriving a general expression for $\mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}$, we take a look at prominent special cases that often arise in quantum information, namely, $k = 2$ and the projection restricted to symmetric endomorphisms as its input. We begin with the second moment, $k = 2$.

Proposition 12 (Second moment operator). *For an operator $A \in \mathrm{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$, $d \geq 2$, it holds that*

$$\mathcal{M}^{(2)}_{\mu_{\mathrm{U}(d)}}(A) = c_{\mathrm{sym}^2} P_{\mathrm{sym}^2} + c_{\wedge^2} P_{\wedge^2} \tag{2.36}$$

with $c_{\mathrm{sym}^2} = \frac{2}{d(d+1)} \operatorname{Tr}[A P_{\mathrm{sym}^2}]$ *and* $c_{\wedge^2} = \frac{2}{d(d-1)} \operatorname{Tr}[A P_{\wedge^2}]$.

Proof. From Lemma 5 and Theorem 10 we know that $\mathcal{M}^{(2)}_{\mu_{\mathrm{U}(d)}}(A)$ is a linear combination of the identity $\mathbb{1}$ and the swap operator $\mathbb{F}$ from (2.4). For $\mathfrak{S}_2$ the expansion of the projectors onto the symmetric and antisymmetric subspace, (2.29), can be inverted yielding $\mathrm{Id} = P_{\mathrm{sym}^2} + P_{\wedge^2}$ and $\mathbb{F} = P_{\mathrm{sym}^2} - P_{\wedge^2}$. This establishes the form of (2.36). Since P_{sym^2} and $P_{\wedge^2}$ are mutually orthogonal projectors and $\mathcal{M}^{(2)}_{\mu_{\mathrm{U}(d)}}$ is an orthogonal projector the coefficients are given by $c_{\mathrm{sym}^2} = \operatorname{Tr}[A P_{\mathrm{sym}^2}] / \operatorname{Tr}[P_{\mathrm{sym}^2}] = \frac{2}{d(d+1)} \operatorname{Tr}[A P_{\mathrm{sym}^2}]$ and $c_{\wedge^2}$ analogously. □

Second, we allow for arbitrary k but restrict the input of $\mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}$ to endomorphisms that are itself symmetric, i.e., of product form. In this case we also find an orthogonal decomposition as given by the following lemma.

Lemma 13 (Moment operator on symmetric operators). *For any operator $A \in \mathrm{L}(\mathbb{C}^d)$ it holds that*

$$\mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}(A^{\otimes k}) = \sum_{\lambda\vdash k, l(\lambda)\leq d} c_\lambda P_\lambda \,, \tag{2.37}$$

with P_λ the orthogonal projector onto $W_\lambda \otimes S_\lambda$ and $c_\lambda = \operatorname{Tr}(P_\lambda A^{\otimes k}) / \operatorname{Tr}(P_\lambda)$. *Furthermore, if the operator A is of* unit rank, *then*

$$\mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}(A^{\otimes k}) = c P_{\mathrm{sym}^k} \;, \tag{2.38}$$

with $c = \operatorname{Tr}(P_{\mathrm{sym}^k} A^{\otimes k}) / \operatorname{Tr}(P_{\mathrm{sym}^k})$.

Proof. We fix some $A \in \mathrm{L}(\mathbb{C}^d)$ and denote $E := \mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}(A^{\otimes k})$. By the definition of the moment operator (2.16), $E = \int_{\mathrm{U}(d)} (UAU^\dagger)^{\otimes k} \mathrm{d}\mu_{\mathrm{U}(d)}(U)$ and it becomes apparent that E commutes with $\pi^d_k(\sigma)$ for any $\sigma \in \mathfrak{S}_k$. In other words, $E \in \operatorname{comm} \Delta^k_d(\mathrm{U}(d)) \cap \operatorname{comm} \pi^d_k(\mathfrak{S}_k)$ by Lemma 5. By Schur's lemma (Theorem 7) and the Schur-Weyl decomposition (2.35), we thus conclude that E acts proportionally to the identity on every Weyl module W_λ and Specht module S_λ. Denoting the orthogonal projector onto $W_\lambda \otimes S_\lambda$ as P_λ, the operator E permits the decomposition $E = \sum_{\lambda \vdash k, l(\lambda) \leq d} c_\lambda P_\lambda$ with $c_\lambda \in \mathbb{C}$. Since the projectors are onto mutually orthogonal the coefficients are given by $c_\lambda = \operatorname{Tr}(A^{\otimes k} P_\lambda) / \operatorname{Tr}(P_\lambda)$. This establishes the lemma's first assertion for E. Finally, for unit rank A, i.e. $A = |\psi\rangle\langle\phi|$ with $|\psi\rangle, |\phi\rangle \in \mathbb{C}^d$, we observe that $P_{\mathrm{sym}^k} A^{\otimes k} P_{\mathrm{sym}^k} = P_{\mathrm{sym}^k} |\psi\rangle^{\otimes n} \langle\phi|^{\otimes n} P_{\mathrm{sym}^k} = A^{\otimes k}$. Hence, $c_\lambda = 0$ for all λ that do not correspond to the symmetric subspace. This leaves us with the lemma's second expression for E and unit rank A. □

In many applications in quantum information the statements of Proposition 12 and Lemma 13 suffice. In particular, Lemma 13 is sufficient to derive statements for prominent ensembles random states in the next section. Going beyond this, a general expression in terms of so-called Weingarten functions [Wei78] is derived by Collins and Sniady [CS06]. We here summarize the derivation that serves as the blue-print for deriving the analogous result for the fourth-moment operator for the Clifford group in Chapter 4.[5]

We denote the dimension of the Weyl modules W_λ by D_λ and the dimensions of the Specht modules S_λ by d_λ. Let P_λ be the orthogonal projections onto $W_\lambda \otimes S_\lambda$ and the character of the irreducible representation π^λ_k carried by S_λ be $\chi^\lambda(\pi) := \operatorname{Tr}(\pi^\lambda_k(\pi))$. The orthogonal projectors can be written as

$$P_\lambda = \frac{d_\lambda}{k!} \sum_{\sigma \in \mathfrak{S}_k} \chi^\lambda(\sigma) \pi_k(\sigma), \tag{2.39}$$

see, e.g. Ref. [Wig59, Eq. (12.10)]. In terms of these projectors $\mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}$ can be calculated using the following theorem.

[5] Going beyond this, . . . of the theorem.] The material of the remainder of the section has been published as Section A.1 of the supplemental material of Ref. [1]. It is altered to use a consistent notation.

Theorem 14 (Integration over the unitary group $\mathrm{U}(d)$)**.** *Let* $A \in \mathrm{L}((\mathbb{C}^d)^{\otimes k}$. *Then,*

$$\mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}(A) = \frac{1}{k!} \sum_{\tau \in \mathfrak{S}_k} \mathrm{Tr}(A\pi_k(\tau))\, \pi_k(\tau^{-1}) \sum_{\lambda \vdash k,\; l(\lambda) \leq d} \frac{d_\lambda}{D_\lambda} P_\lambda. \tag{2.40}$$

This formula differs slightly from the original statement presented in Ref. [CS06]. The more common formulation presented there follows from evaluating the expression of Theorem 14 using a standard tensor basis of $\mathrm{L}((\mathbb{C}^d)^{\otimes k})$. However, here we have opted for a presentation of Theorem 14 that is easier to generalize beyond the full unitary group.[6]

In the remainder of this section, we present a proof of Theorem 14 following the strategy of Ref. [CS06]. Let $V = (\mathbb{C}^d)^{\otimes k}$. In general, the direct sum of endomorphisms acting on the irreducible representations of a group is isomorphic to the group ring which consists of formal (complex) linear combinations of the group elements [FH91, Propositon 3.29]. We denote the group ring of $\mathfrak{S}_k$ by $\mathbb{C}[S_k]$. To derive an explicit expression of the coefficient of the expansion of $\mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}(A)$ in $\mathbb{C}[S_k]$, we introduce the map $\Phi : \mathrm{L}(V) \to \mathrm{L}(V)$

$$\Phi(A) = \sum_{\sigma \in \mathfrak{S}_k} \mathrm{Tr}(A\pi_k(\sigma^{-1}))\pi_k(\sigma). \tag{2.41}$$

We will make use of the following properties of the map Φ.

Lemma 15 (Properties of Φ)**.** *For all* $A \in \mathrm{L}(V)$ *and* $B \in \mathrm{Comm}(\Delta^k_d)$

$$\Phi(A) = \Phi(\mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}(A)), \tag{2.42}$$

$$\Phi(B) = B\Phi(\mathrm{Id}), \tag{2.43}$$

$$\Phi(\mathrm{Id})^{-1} = \frac{1}{k!} \sum_{\lambda \vdash k, l(\lambda) \leq d} \frac{d_\lambda}{D_\lambda} P_\lambda. \tag{2.44}$$

Proof. 1. Since $\pi_k(\sigma^{-1})$ is in $\mathrm{Comm}(\Delta^k_d)$ for all $\sigma \in \mathfrak{S}_k$, we can apply Lemma 8 to get

$$\begin{aligned} \mathrm{Tr}(\mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}(A)\pi_k(\sigma^{-1})) &= \mathrm{Tr}(\mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}(A\pi_k(\sigma^{-1}))) \\ &= \mathrm{Tr}(A\pi_k(\sigma^{-1}))\,, \end{aligned} \tag{2.45}$$

which establishes the first statement.

[6] This way of stating the result of Ref. [CS06] was brought to our attention by study notes of K. Audenaert.

2. Since the commutant is isomorphic to the group ring, it suffices to proof the statement for all $B = \pi_k(\tau)$ with $\tau \in \mathfrak{S}_k$. In this case, using the cyclicity of the trace for the first equality, we find

$$\begin{aligned}\Phi(\pi_k(\tau)) &= \sum_{\sigma\in\mathfrak{S}_k} \operatorname{Tr}(\pi_k(\sigma^{-1})\pi_k(\tau))\pi_k(\sigma)\\ &= \sum_{\sigma\in\mathfrak{S}_k} \operatorname{Tr}(\pi_k(\tau\sigma^{-1}))\pi_k(\sigma)\\ &= \sum_{\sigma\in\mathfrak{S}_k} \operatorname{Tr}(\pi_k(\sigma^{-1}))\pi_k(\sigma\tau)\\ &= \pi_k(\tau)\sum_{\sigma\in\mathfrak{S}_k} \operatorname{Tr}(\pi_k(\sigma^{-1}))\pi_k(\sigma).\end{aligned} \tag{2.46}$$

Here we have used that $\pi_k(\tau\sigma) = \pi_k(\sigma)\pi_k(\tau)$ for all $\tau, \sigma \in \mathfrak{S}_k$.

3. Using Theorem 11, we can rewrite $\Phi(\mathrm{Id})$ as

$$\begin{aligned}\Phi(\mathrm{Id}) &= \sum_{\sigma\in\mathfrak{S}_k} \operatorname{Tr}(\pi_k(\sigma^{-1}))\pi_k(\sigma)\\ &= \sum_{\sigma\in\mathfrak{S}_k}\sum_{\lambda\vdash k, l(\lambda)\leq d} D_\lambda \operatorname{Tr}(\pi_\lambda(\sigma^{-1}))\pi_k(\sigma)\\ &= \sum_{\lambda\vdash k, l(\lambda)\leq d} D_\lambda \sum_{\sigma\in\mathfrak{S}_k} \chi^\lambda(\sigma)\pi_k(\sigma).\end{aligned} \tag{2.47}$$

The explicit expression (2.39) for the projectors identifies $\Phi(\mathrm{Id})$ as

$$\Phi(\mathrm{Id}) = k! \sum_{\lambda\vdash k, l(\lambda)\leq d} \frac{D_\lambda}{d_\lambda} P_\lambda. \tag{2.48}$$

Since the $\{P_\lambda\}$ are a complete set of orthogonal projectors, the inverse of $\Phi(\mathrm{Id})$ is given by

$$\Phi(\mathrm{Id})^{-1} = \frac{1}{k!} \sum_{\lambda\vdash k, l(\lambda)\leq d} \frac{d_\lambda}{D_\lambda} P_\lambda. \tag{2.49}$$

□

We are now in position to give a concise proof of Theorem 14:

Proof of Theorem 14. From Eqs. (2.42) and (2.43) we conclude $\Phi(A) = \Phi(T_{\Delta_d^k}(A)) = T_{\Delta_d^k}(A)\Phi(\mathrm{Id})$ and, thus, $T_{\Delta_d^k}(A) = \Phi(A)\Phi(\mathrm{Id})^{-1}$. Inserting the expression (2.44) for $\Phi(\mathrm{Id})^{-1}$ and the definition (2.41) of Φ yields the expression of the theorem. □

2.3.3 Uniformly random state vectors

One can also define a uniform distribution on pure quantum states in multiple equivalent ways. First, one can draw randomly from the complex sphere $\mathbb{S}(\mathbb{C}^d)$, i.e. the set of normalized vectors in $\mathbb{C}^d$. Indeed, there is a unique uniform probability measure $\mu_{\mathbb{S}(\mathbb{C}^d)}$ on $\mathbb{S}(\mathbb{C}^d)$ that is invariant under the canonical action of $\mathrm{U}(d)$ on $\mathbb{C}^d$. By definition we see that a column $|\psi\rangle = U|0\rangle$ of a Haar-randomly drawn unitary $U \sim \mu_{\mathrm{U}(d)}$ is distributed according to $\mu_{\mathbb{S}(\mathbb{C}^d)}$. Finally, we can switch to density matrices by factoring out a global phase. In more detail, the *complex projective space* $\mathbb{C}\mathbf{P}^{d-1} := \mathbb{S}(\mathbb{C}^d)/\,\mathrm{U}(1)$ is the set of state vectors modulo a phase in $\mathrm{U}(1)$, which can be identified with the set of pure density matrices $\mathbb{C}\mathbf{P}^{d-1} \subset \mathcal{D}(\mathbb{C}^d)$. It also has a uniform unitarily invariant probability distribution: a uniformly random pure state $|\psi\rangle\langle\psi|$ can be obtained by drawing $|\psi\rangle \sim \mu_{\mathbb{S}(\mathbb{C}^d)}$.

We can calculate the moments of polynomials that depend on states drawn uniformly from $\mu_{\mathbb{S}(\mathbb{C}^d)}$ using the moment operator $\mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}$. To this end, note that any polynomial $p_k(|\psi\rangle, \langle\psi|)$ of degree k in the component of each $|\psi\rangle$ and $\langle\psi|$ can be written as a contraction of $|\psi\rangle\langle\psi|^{\otimes k}$ with some operator in $\mathrm{L}(\mathbb{C}^{d^k})$. For this reason the following lemma summarizes everything we need.

Lemma 16 (Moment operator of random states). *Let $K^{(k)}_{\mu_{\mathbb{S}(\mathbb{C}^d)}}$ be the moment-operator for $|\psi\rangle \sim \mu_{\mathbb{S}(\mathbb{C}^d)}$ explicitly defined by*

$$K^{(k)}_{\mu_{\mathbb{S}(\mathbb{C}^d)}} := \int_{\mathbb{S}(\mathbb{C}^d)} (|\psi\rangle\langle\psi|)^{\otimes k} \mathrm{d}\mu_{\mathbb{S}(\mathbb{C}^d)}(\psi)\,. \tag{2.50}$$

It holds that

$$K^{(k)}_{\mu_{\mathbb{S}(\mathbb{C}^d)}} = \frac{k!(d-1)!}{(k+d-1)!} P_{\mathrm{sym}^k}\,, \tag{2.51}$$

where P_{sym^k} is the projector (2.29) onto the symmetric subspace.

Proof. As $\mu_{\mathbb{S}(\mathbb{C}^d)}$ is $\mathrm{U}(d)$-invariant, we find $K^{(k)}_{\mu_{\mathbb{S}(\mathbb{C}^d)}} = \mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}((|\psi\rangle\langle\psi|)^{\otimes k})$. Lemma 13 thus implies that $K^{(k)}_{\mu_{\mathbb{S}(\mathbb{C}^d)}} = cP_{\mathrm{sym}^k}$ with

$$c = \frac{\mathrm{Tr}(P_{\mathrm{sym}^k}(|\psi\rangle\langle\psi|)^{\otimes k})}{\mathrm{Tr}(P_{\mathrm{sym}^k})}.$$

Since P_{sym^k} acts trivially on $|\psi\rangle$, and it is normalized, the enumerator evaluates to one. The denominator is the dimension of P_{sym^k} given by (2.30). □

2.3.4 Unitary, spherical and complex-projective k-designs

With our excursion to representation theory we have derived expressions to calculate the moments of random variables on uniformly random states and unitaries. The very same results can be also used for certain other interesting probability distributions. To this end, note that if we only want to control the first t moments of a random variable that is a polynomial of degree ℓ in a random state or unitary, then our calculation will only involve the moment operators $\mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}$ for $k \leq t\ell$. In many applications it is sufficient to control the expectation value and the variance of low-degree polynomials. In these cases, any probability distribution that reproduces the first couple of moments of the uniform distributions can be used without changing the mathematical expressions. This idea is formalized by the definition of k-designs [Dan+09; GAE07].

Definition 1 (Unitary k-design)**.** A distribution μ on the unitary group $\mathrm{U}(d)$ is a *unitary k-design* if its kth moment operator (2.16) equals the corresponding moment operator of the Haar measure,

$$\mathcal{M}^{(k)}_{\mu} = \mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}. \tag{2.52}$$

A subset $\{U_1, \dots, U_{n_{\mathrm{G}}}\} \subset \mathrm{U}(d)$ is called a *unitary k-design* if its uniform distribution is one.

Note that by definition, any unitary k-design is also a unitary $(k-1)$-design for $k \geq 2$. A famous example of a unitary design in the context of quantum computing is the Clifford group.

The Clifford group The (qubit) Pauli matrices are

$$X := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y := \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad Z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{2.53}$$

and we refer to tensor products $W_1 \otimes \dots \otimes W_n$ with $W_i \in \{X, Y, Z, \mathrm{Id}\}$ for all $i \in [n]$ as n-qubit *Pauli strings*. The group generated by all n-qubit Pauli strings and $\mathrm{i}\,\mathrm{Id}$ is the *Pauli group* $\mathcal{P}_n \subset \mathrm{U}(2^n)$. The n-qubit *Clifford group* $\mathrm{Cl}_n = \mathrm{Cl}(2^n) =\subset \mathrm{U}(2^n)$ is the stabilizer of the Pauli group $\mathcal{P}_n$,

$$\mathrm{Cl}_n := \{U \in \mathrm{U}(2^n; \mathbb{Q}) : \; U\mathcal{P}_n U^\dagger \subset \mathcal{P}_n\}\,, \tag{2.54}$$

where it is common to restrict to unitary matrices with complex rational entries, here denoted by $\mathrm{U}(d;\mathbb{Q}) := \mathrm{U}(d) \cap (\mathbb{Q}^{d\times d} + \mathrm{i}\mathbb{Q}^{d\times d})$, so that Cl_n becomes a finite group. The n-qubit Clifford group is generated by the single qubit Hadamard gate H and the phase gate S given by (see, e.g. [NC10, Theorem 10.6])

$$\mathrm{H} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \text{and} \quad \mathrm{S} = \begin{pmatrix} 1 & \\ & \mathrm{i} \end{pmatrix} \tag{2.55}$$

together with the two-qubit controlled NOT (CNOT) gate

$$\mathrm{CNOT} = |0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes \sigma_x. \tag{2.56}$$

acting locally on any qubit. Together with the $\mathrm{T} = \sqrt{\mathrm{S}}$ gate the Clifford group is a universal gate set (see, e.g. [NC10, Section 4.5.3]).

The Clifford group constitutes a unitary 3-design but 'fails gracefully' to be a unitary 4-design [Web16; Zhu17; Zhu+16]. Being a subgroup of the unitary group the commutant of the diagonal action of the Clifford group for $k > 3$ is, thus, a strictly larger space than the span of the permutation group. Deriving an expression for the 4-order moment operator is a result of Chapter 4.

Note that analogously to unitary designs, we can define spherical k-designs and state k-designs [AE07; RS07]. For a distribution μ on the complex sphere $\mathbb{S}(\mathbb{C}^d)$ we define the *kth moment-operator* as

$$K_\mu^{(k)} := \int_{\mathbb{S}(\mathbb{C}^d)} (|\psi\rangle\langle\psi|)^{\otimes k} \mathrm{d}\mu(\psi)\,. \tag{2.57}$$

Definition 2 (Complex spherical/projective/state k-design)**.** A distribution μ on $\mathbb{S}(\mathbb{C}^d)$ is a *spherical k-design* if

$$K_\mu^{(k)} = K_{\mu_{U(d)}}^{(k)}. \tag{2.58}$$

Furthermore, a subset $\mathbb{S}(\mathbb{C}^d)$ is called a *spherical k-design* if its uniform distribution is one. The corresponding distribution of pure density matrices $|\psi\rangle\langle\psi|$ is called a *complex projective k-design*.

Analogously to the relation of the uniform measure on $\mathrm{U}(d)$ and $\mathbb{S}(\mathbb{C}^d)$, a rather obvious but important example of a spherical k-design is an orbit of a unitary k-design. If μ is a unitary k-design for $\mathrm{U}(d)$ and $|\psi\rangle \in \mathbb{C}^d$ then the induced distribution $\tilde{\mu}$ given by $U\ |\psi\rangle$ with $U \sim \mu$, is a complex spherical k-design. One can use this relation to see that the Clifford group being a unitary 3-design implies the analogous statement for stabilizer states, see also Ref. [KG15] for a direct proof. Other examples for spherical designs that play essential roles in quantum system characterization are *mutually unbiased bases (MUBs)* [Ivo81; WF89; KR05] and *symmetric, informationally complete (SIC) POVMs* [Zau99; Ren+04].

2.4 Quantum processes and measures of quality

In Section 2.2 we modelled the static properties of a quantum device by its quantum state. The main function of quantum devices roots of course in their ability to accurately manipulate their quantum state and, thus, the quantum information encoded therein. The manipulation of quantum states can be generally described by means of *quantum processes.* A quantum process is a linear map from density operators to density operators preserving defining properties. Formally, we require the following definitions: Let $\mathcal{H}, \mathcal{K}$ be finite-dimensional Hilbert spaces. We denote by $\mathbb{L}(\mathcal{H}, \mathcal{K}) := \mathrm{L}(\mathrm{L}(\mathcal{H}), \mathrm{L}(\mathcal{K}))$ the vector space of linear maps from linear operators on $\mathcal{H}$ to the ones on $\mathcal{K}$. We will frequently refer to elements in $\mathbb{L}(\mathcal{H}, \mathcal{K})$ simply as *maps*. We abbreviate $\mathbb{L}(\mathcal{H}) := \mathbb{L}(\mathcal{H}, \mathcal{H})$ and analogous definitions. An important subsets of $\mathbb{L}(\mathcal{H}, \mathcal{K})$ are the one that preserve properties of quantum states: We say $\mathcal{X} \in \mathbb{L}(\mathcal{H}, \mathcal{K})$ is *Hermicity-preserving* if $\mathcal{X}(\mathrm{Herm}(\mathcal{H})) \subset \mathrm{Herm}(\mathcal{K})$, *positive* if $\mathcal{X}(\mathrm{Pos}(\mathcal{H})) \subset \mathrm{Pos}(\mathcal{K})$, *completely positive (CP)* if $\mathcal{X} \otimes \mathbb{1}_{L(\mathcal{H}')}$ is positive for all Hilbert spaces $\mathcal{H}'$ and $\mathbb{1}_{L(\mathcal{H}')}$ their identity map, *trace-preserving* if $\mathrm{Tr}[\mathcal{X}(X)] = \mathrm{Tr}[X]$ for all $X \in \mathrm{L}(\mathcal{H})$.

If we want to restrict a map such that the image of every quantum state is a quantum state even when composing it with a larger system the map must be *completely positive and trace-preserving (CPT).* Such CPT maps are called *quantum channels*, and we write $\mathrm{CPT}(\mathcal{H}, \mathcal{K}) \subset \mathrm{CP}(\mathcal{H}, \mathcal{K})$ for their set. Note that the set of completely positive maps $\mathrm{CP}(\mathcal{H}, \mathcal{K})$ is a convex cone and $\mathrm{CPT}(\mathcal{H}, \mathcal{K})$ a convex set in $\mathbb{L}(\mathcal{H}, \mathcal{K})$.

For our approach to process tomography in Chapter 6 of particular importance is the notion of *unital maps*. Unital maps have the identity as their fix-point, $\mathcal{X}(\mathbb{1}_{\mathcal{H}}) = \mathbb{1}_{\mathcal{K}}$. Note that $\mathcal{X}$ is unital if and only if $\mathcal{X}^\dagger$, its adjoint w.r.t. the Hilbert-Schmidt inner product, is trace-preserving. When working with unital maps we always implicitely assume that the map is hermicity-preserving and has domain restricted to the hermitian matrices. To emphasize this we introduce the short-hand notation $H_d \coloneqq \mathrm{Herm}(\mathcal{H})$ for the set of Hermitian operators on a d-dimensional Hilbert space $\mathcal{H}$ and denote the set of unital and trace-preserving maps by $\mathrm{L}_{\mathrm{u,tp}}(H_d)$, its linear-hull by $\mathrm{L}_{\overline{\mathrm{u,tp}}}(H_d)$.

A proper-functioning quantum device is typically expected to coherently manipulate its quantum state. In particular, we are interested in implementing rotations in the pure state space. Such operations are described by a unitary $U \in \mathrm{U}(\mathcal{H})$ acting adjointly on $\mathrm{L}(\mathcal{H})$ as

$$\mathcal{U}(X) \coloneqq UXU^\dagger \,. \tag{2.59}$$

We will generally use the corresponding calligraphic letters to denote the adjoint representation of a unitary. It is easy to see that $\mathcal{U}$ is a quantum channel. Unitary channels are invertible, and the inverses are again unitary channels.

A matrix $H \in \mathrm{Herm}(\mathcal{H})$ generates a one-parameter of family of unitaries via $U : t \mapsto \exp(-iHt)$. We recall the postulate of unitary dynamics: The time-evolution of a quantum system's state is described by acting with the unitary channel $\mathcal{U}(t)$ defined through its governing Hamiltonian H.

Going beyond the unitary operations and time-evolution, quantum channels model the imperfect, noisy functioning of a quantum device. An important quantum channel and frequent model for noise processes appearing in quantum technologies is the depolarizing channel. The *(quantum) depolarizing channel* $\mathcal{D}_p : \mathrm{L}(\mathbb{C}^d) \to \mathrm{L}(\mathbb{C}^d)$ with parameter $p \in [0,1]$ is the linear map defined by

$$\mathcal{D}_p(X) \coloneqq pX + (1-p)\,\mathrm{Tr}[X]\frac{\mathbb{1}}{d} \,. \tag{2.60}$$

2.4.1 The Choi-Jamiołkowski isomorphism

Since quantum process are linear maps of vectors spaces of linear maps on a Hilbert space, there exist multiple canonical isomorphisms to other vector spaces. A particular essential one is the Choi-Jamiołkowski isomorphism [Jam72; Cho75] as it identifies CP maps with positive semidefinite operators. It is constructed by a concatenation of the standard canonical vector space isomorphisms: Let

$V^* := \mathrm{L}(V, \mathbb{C})$ denote the dual space of V. We have the standard identification $\mathrm{L}(V) = V \otimes V^*$. If V is equipped with an inner product $\langle \cdot | \cdot \rangle$, we can furthermore identify $V \overset{\text{hc}}{\cong} V^*$, i.e. in Dirac notation the identification of $|v\rangle$ with $\langle v|$. This allows us to identify $\mathbb{L}(\mathcal{H}, \mathcal{K}) = \mathrm{L}(\mathcal{K}) \otimes \mathrm{L}(\mathcal{H})^* = \mathcal{K} \otimes \mathcal{K}^* \otimes \mathcal{H}^* \otimes \mathcal{H} \cong \mathcal{K} \otimes \mathcal{H}^* \otimes \mathcal{K}^* \otimes \mathcal{H} = \mathrm{L}(\mathcal{K} \otimes \mathcal{H}^*) \overset{\text{hc}}{\cong} \mathrm{L}(\mathcal{K} \otimes \mathcal{H})$. This identification defines the *Choi-Jamiołkowski isomorphism* $\mathfrak{C} : \mathbb{L}(\mathcal{H}, \mathcal{K}) \to \mathrm{L}(\mathcal{K} \otimes \mathcal{H})$. An explicit expression for the Choi-Jamiołkowski isomorphism with respect to a basis is in terms of the *Choi matrix* of $\mathcal{X} \in \mathbb{L}(\mathcal{H}, \mathcal{K})$

$$\mathfrak{C}(\mathcal{X}) = \mathcal{X} \otimes \mathrm{id}(|\mathbb{1}\rangle\langle\mathbb{1}|)\,, \tag{2.61}$$

where $|\mathbb{1}\rangle = \sum_i |i\rangle\,|i\rangle$ is the unnormalized maximally entangled state with $\{\,|i\rangle\}$ an orthonormal basis of $\mathcal{H}$. Direct calculation reveals that

$$\mathrm{Tr}[B\mathcal{X}(A)] = \mathrm{Tr}[(B \otimes A^\intercal)\,\mathfrak{C}(\mathcal{X})] \tag{2.62}$$

for all $\mathcal{X} \in \mathbb{L}(\mathcal{H}, \mathcal{K})$, $A \in \mathrm{L}(\mathcal{H})$ and $B \in \mathrm{L}(\mathcal{K})$.

The importance of the Choi-Jamiołkowski isomorphism in quantum information stems from the fact that the properties of CPT maps are well-captured by properties of their Choi matrix as summarized by the following theorem.

Theorem 17 (CPT conditions [Wat18, Chapter 2.2]). *For any map $\mathcal{X} \in \mathbb{L}(\mathcal{H}, \mathcal{K})$ the following equivalences hold:*

(i) $\mathcal{X}$ is trace-preserving if and only if $\mathrm{Tr}_{\mathcal{K}}[\mathfrak{C}(\mathcal{X})] = \mathbb{1}$.

(ii) $\mathcal{X}$ is Hermicity-preserving if and only if $\mathfrak{C}(\mathcal{X})$ is Hermitian.

(iii) $\mathcal{X}$ is completely positive if and only if $\mathfrak{C}(\mathcal{X})$ is positive semidefinite.

Another important consequence of the complete positivity of a map is the existence of so-called Kraus operators [Wat18, Chapter 2.2]. This gives another item that could be added to Theorem 17: A map $\mathcal{X}$ is a CP map if and only if there are (Kraus) operators $K_1, \dots, K_r \in \mathrm{L}(\mathcal{H}, \mathcal{K})$, where $r = \mathrm{rank}(\mathfrak{C}(\mathcal{X}))$ so that

$$\mathcal{X}(A) = \sum_{i=1}^{r} K_i A K_i^\dagger \tag{2.63}$$

for all $A \in \mathrm{L}(\mathcal{H})$. The map $\mathcal{X}$ is additionally trace-preserving if and only if $\sum_{i=1}^{r} K_i^\dagger K_i = \mathbb{1}$.

Using a different normalization

$$\mathfrak{J}(\mathcal{X}) := \frac{1}{\dim(\mathcal{H})}\,\mathfrak{C}(\mathcal{X}) \tag{2.64}$$

we can associate to a channel $\mathcal{X}$ it's so called *Choi state* $\mathfrak{J}(\mathcal{X}) \in \mathcal{D}(\mathcal{K} \otimes \mathcal{H})$. This motivates the term *channel-state duality* for the Choi-Jamiołkowski isomorphism. However, strictly speaking the Choi-Jamiołkowski is an isomorphism of the convex cones $\mathrm{CP}(\mathcal{H}, \mathcal{K})$ and $\mathrm{Pos}(\mathcal{H} \otimes \mathcal{H})$. In order, to have a (trace-preserving) channel as their preimage a bipartite state has to be maximally entangled over the partition. The identity (2.61) implies that one can in principle prepare the Choi state of a quantum channel by acting with it on a quantum system that is maximally entangled with an ancillary system.

2.4.2 Inner products of maps and fidelity measures

There are a couple of more or less equivalent inner products on $\mathbb{L}(\mathcal{H}, \mathcal{K})$. We start with the canonical inner product, the Hilbert-Schmidt,

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \mathrm{Tr}[\mathcal{X}^\dagger \mathcal{Y}] \tag{2.65}$$

for any $\mathcal{X}, \mathcal{Y} \in \mathbb{L}(\mathcal{H}, \mathcal{K})$. The trace can be defined with an orthonormal basis $\{E_0, E_1, \dots, E_{d^2-1}\}$ of $\mathrm{L}(\mathcal{H})$, lifting the inner product of $\mathrm{L}(\mathcal{H})$:

$$\mathrm{Tr}[\mathcal{X}] = \sum_{i=0}^{d^2-1} \langle E_i, \mathcal{X}(E_i) \rangle = \sum_{i=0}^{d^2-1} \mathrm{Tr}[E_i^\dagger \mathcal{X}(E_i)]\,. \tag{2.66}$$

The Hilbert-Schmidt inner product on $\mathbb{L}(\mathcal{H}, \mathcal{K})$ coincides with the inner product of the corresponding Choi matrices, i.e., for any $\mathcal{X}, \mathcal{Y} \in \mathbb{L}(\mathcal{H}, \mathcal{K})$

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \langle \mathfrak{C}(\mathcal{X}), \mathfrak{C}(\mathcal{Y}) \rangle. \tag{2.67}$$

We now consider the case where $\mathcal{Y}$ is a quantum channel and $\mathcal{X}$ a unitary quantum channel. Then, as we have seen above, $\mathfrak{J}(\mathcal{Y})$ and $\mathfrak{J}(\mathcal{X})$ are quantum states (density matrices). Moreover, $\mathfrak{J}(\mathcal{X})$ is a pure state. In this case, the above Hilbert-Schmidt inner product with the proper normalization is the fidelity measure induced by the state fidelity (2.12) via the Choi-Jamiołkowski isomorphism (2.61),

$$\mathrm{F_e}(\mathcal{X}, \mathcal{Y}) := \mathrm{F}(\mathfrak{J}(\mathcal{X}), \mathfrak{J}(\mathcal{Y})) = \frac{1}{\dim(\mathcal{H})^2} \langle \mathcal{X}, \mathcal{Y} \rangle\,; \tag{2.68}$$

it is referred to as the *entanglement (gate) fidelity*.

In the context of digital quantum computing, another very prominent fidelity measure for quantum processes is the so-called *average gate fidelity*. The *average gate fidelity* (AGF) between maps $\mathcal{X}, \mathcal{Y} \in \mathbb{L}(\mathcal{H}, \mathcal{K})$ is defined as

$$F_{\mathrm{avg}}(\mathcal{X}, \mathcal{Y}) := \int_{\mathbb{S}(\mathbb{C}^d)} \langle \mathcal{X}(|\psi\rangle\langle\psi|), \mathcal{Y}(|\psi\rangle\langle\psi|)\rangle \, \mathrm{d}\mu_{\mathbb{S}(\mathbb{C}^d)}(\psi) \, . \tag{2.69}$$

The measure here is the uniform Haar-invariant probability measure on state vectors of Section 2.3.3. The inner product is the Hilbert-Schmidt inner product of $\mathrm{L}(\mathcal{K})$ not $\mathbb{L}(\mathcal{H}, \mathcal{K})$. From the definition we see that the average gate fidelity $F_{\mathrm{avg}}(\mathcal{X}, \mathcal{Y})$ is a measure of closeness of $\mathcal{X}$ and $\mathcal{Y}$ that compares the action of $\mathcal{X}$ and $\mathcal{Y}$ on pure input states on average. Intuitively, if $\mathcal{X}$ and $\mathcal{Y}$ only deviate in their action on a low-dimensional subspace of $\mathcal{H}$ they can still have an average gate fidelity close to one.

For any $\mathcal{X}, \mathcal{Y} \in \mathbb{L}(\mathcal{H}, \mathcal{K})$

$$F_{\mathrm{avg}}(\mathcal{X}, \mathcal{Y}) = F_{\mathrm{avg}}(\mathrm{id}, \mathcal{X}^\dagger \circ \mathcal{Y}) \, . \tag{2.70}$$

We therefore also write $F_{\mathrm{avg}}(\mathcal{X}) := F_{\mathrm{avg}}(\mathrm{id}, \mathcal{X})$ for $\mathcal{X} \in \mathbb{L}(\mathcal{H})$. The average gate fidelity is intricately related to the Hilbert-Schmidt inner product on $\mathbb{L}(\mathcal{H}, \mathcal{K})$ [HHH99; Nie02] (see also [Kue+16]).

Proposition 18 (Inner product and F_{avg})**.** *For $\mathcal{X}, \mathcal{Y} \in \mathbb{L}(\mathcal{H}, \mathcal{K})$ with $d = \dim \mathcal{H}$ it holds that*

$$\langle \mathcal{X}, \mathcal{Y} \rangle = d(d+1) \, F_{\mathrm{avg}}(\mathcal{X}, \mathcal{Y}) - \langle \mathcal{X}(\mathbb{1}), \mathcal{Y}(\mathbb{1}) \rangle \, . \tag{2.71}$$

This proposition implies that the average gate fidelity is an inner product, i.e., a conjugate symmetric non-degenerate form that is linear in its second argument. It also connects the average gate fidelity to the Frobenius norm induce by the Hilbert-Schmidt inner product. Importantly, this indicates the Frobenius norm on quantum processes is in fact an average case error measure over the input states. For Hermicity-preserving $\mathcal{X}$ and $\mathcal{Y}$ the average gate fidelity is real. Thus, on Hermicity-preserving maps it is *symmetric*, $F_{\mathrm{avg}}(\mathcal{X}, \mathcal{Y}) = F_{\mathrm{avg}}(\mathcal{Y}, \mathcal{X})$.

Associate to the average gate fidelity is the *average error rate* or *average infidelity*,

$$\mathrm{r}(\mathcal{X}, \mathcal{Y}) := 1 - F_{\mathrm{avg}}(\mathcal{X}, \mathcal{Y}) \tag{2.72}$$

that is also real-valued for Hermicity-preserving maps and $\mathrm{r}(\mathcal{X}) := 1 - F_{\mathrm{avg}}(\mathcal{X})$. For unital, completely positive $\mathcal{X}$, the average infidelity can be regarded as a distance to other quantum channels in the following sense:

Lemma 19 (Infidelity as distance measure)**.** *Let $\mathcal{X} \in \mathrm{CP}(\mathcal{H},\mathcal{K})$ be unital. For all $\mathcal{Y} \in \mathrm{CPT}(\mathcal{H},\mathcal{K})$ it holds that $\mathrm{r}(\mathcal{X},\mathcal{Y}) \geq 0$ and, $\mathrm{r}(\mathcal{X},\mathcal{Y}) = 0$ if and only if $\mathcal{X} = \mathcal{Y}$.*

Proof. Using Proposition 18, we have $F_{\mathrm{avg}}(\mathcal{Y}) = \frac{1}{d(d+1)}\langle \mathrm{id}, \mathcal{Y}\rangle + \frac{1}{d+1}$. The overlap of two CP maps can be bounded via the Cauchy-Schwarz inequality as $\langle \mathrm{id}, \mathcal{Y}\rangle \leq \|\mathrm{id}\|_F \|\mathcal{Y}\|_F$ with equality if and only if $\mathcal{Y} = \mathrm{id}$. For $\mathcal{Y} \in \mathrm{CPT}(\mathcal{H})$ it holds that $\|\mathcal{Y}\|_F \leq d^2$ and $\|\mathrm{id}\|_F^2 = d^2$. This can be seen, e.g., from (2.66) choosing a unit-rank basis and applying the Hölder inequality (2.6). Therefore, $\langle \mathrm{id}, \mathcal{Y}\rangle \leq d^2$. We conclude that $F_{\mathrm{avg}}(\mathcal{Y}) \leq 1$ again with equality if and only if $\mathcal{Y} = \mathrm{id}$ which implies the assertion. □

We still have to prove of Proposition 18.

Proof of Proposition 18. By the virtue of (2.70) which also holds for the inner products appearing in (2.71) it suffices to prove the statement for $\mathcal{X} = \mathrm{id}$. Using (2.62) and denoting the transposition map as $T : \mathrm{L}(\mathcal{H}) \to \mathrm{L}(\mathcal{H})$, $A \mapsto A^\intercal$, we can rewrite the average gate fidelity as

$$\begin{aligned} F_{\mathrm{avg}}(\mathrm{id}, \mathcal{Y}) &= \int_{\mathbb{S}(\mathbb{C}^d)} \langle |\psi\rangle\langle\psi|, \mathcal{Y}(|\psi\rangle\langle\psi|)\rangle \,\mathrm{d}\mu_{\mathbb{S}(\mathbb{C}^d)}(\psi) \\ &= \int_{\mathbb{S}^{d-1}} \mathrm{Tr}\left[|\psi\rangle\langle\psi|\, \mathcal{Y}(|\psi\rangle\langle\psi|)\right] \mathrm{d}\mu_{\mathbb{S}(\mathbb{C}^d)}(\psi) \\ &= \int_{\mathbb{S}(\mathbb{C}^d)} \mathrm{Tr}\left[\mathrm{Id} \otimes T\left(|\psi\rangle\langle\psi|^{\otimes 2}\right) \mathfrak{C}(\mathcal{Y})\right] \mathrm{d}\mu_{\mathbb{S}(\mathbb{C}^d)}(\psi). \end{aligned} \tag{2.73}$$

Due to linearity, we can recast this expression with the moment-operator $K^{(k)}_{\mu_{\mathbb{S}(\mathbb{C}^d)}}$ of random states and use the expression we derived in Lemma 16. Then,

$$\begin{aligned} F_{\mathrm{avg}}(\mathrm{id}, \mathcal{Y}) &= \mathrm{Tr}\left[\mathrm{Id} \otimes T\left(K^{(2)}_{\mu_{\mathbb{S}(\mathbb{C}^d)}}\right) \mathfrak{C}(\mathcal{Y})\right] \\ &= \frac{2}{d(d+1)} \mathrm{Tr}\left[\mathrm{Id} \otimes T(P_{\mathrm{sym}^2})\, \mathfrak{C}(\mathcal{Y})\right] \\ &= \frac{1}{d(d+1)} \left(\mathrm{Tr}\left[\mathbb{1}\, \mathfrak{C}(\mathcal{Y})\right] + \mathrm{Tr}\left[|\mathbb{1}\rangle\langle\mathbb{1}|\, \mathfrak{C}(\mathcal{Y})\right]\right), \end{aligned} \tag{2.74}$$

where the last step follows from $P_{\mathrm{sym}^2} = \frac{1}{2}(\mathbb{1} + \mathbb{F})$ with the swap operator $\mathbb{F}$ from (2.4) and $\mathrm{Id} \otimes T(\mathbb{F}) = |\mathbb{1}\rangle\langle\mathbb{1}|$. Using (2.62) this time the other way around, we see that the first summand of (2.74) is $\mathrm{Tr}[\mathbb{1}\, \mathfrak{C}(\mathcal{Y})] = \mathrm{Tr}[\mathbb{1} \otimes \mathbb{1}\, \mathfrak{C}(\mathcal{Y})] =$

$\mathrm{Tr}[\mathcal{Y}(\mathbb{1})] = \langle \mathrm{Id}(\mathbb{1}), \mathcal{Y}(\mathbb{1})\rangle$. From (2.61) it directly follows that $\mathfrak{C}(\mathrm{Id}) = |\mathbb{1}\rangle\langle\mathbb{1}|$. Hence, the second term of (2.74) is equal to $\mathrm{Tr}\left[|\mathbb{1}\rangle\langle\mathbb{1}|\,\mathfrak{C}(\mathcal{Y})\right] = \mathrm{Tr}\left[\mathfrak{C}(\mathrm{Id})\,\mathfrak{C}(\mathcal{Y})\right] = \langle \mathfrak{C}(\mathrm{Id}), \mathfrak{C}(\mathcal{Y})\rangle = \langle \mathrm{Id}, \mathcal{Y}\rangle$. Plugging these two expressions into (2.74) and solving for $\langle \mathrm{Id}, \mathcal{Y}\rangle$ yields the assertion of the proposition. □

If $\mathcal{X}^\dagger\mathcal{Y}$ trace-preserving, (2.71) simplifies to

$$\langle \mathcal{X}, \mathcal{Y}\rangle = d(d+1)\, F_{\mathrm{avg}}(\mathcal{X}, \mathcal{Y}) - d\,, \tag{2.75}$$

or, equivalently,

$$F_{\mathrm{avg}}(\mathcal{X}, \mathcal{Y}) = \frac{\langle \mathcal{X}, \mathcal{Y}\rangle + d}{d(d+1)}\,. \tag{2.76}$$

We conclude that for trace-preserving and unital quantum channels the average gate fidelity and the Hilbert-Schmidt inner product are affinely related with a proportionality constant in $\mathrm{O}(d^{-2})$. This is the same scaling as appearing for the entanglement fidelity in (2.68). More precisely, we find the affine relation between the two fidelities

$$F_{\mathrm{avg}}(\mathcal{X}, \mathcal{Y}) = \frac{d\, \mathrm{F_e}(\mathcal{X}, \mathcal{Y}) + 1}{d+1}\,, \tag{2.77}$$

still assuming $\mathcal{X}^\dagger\mathcal{Y}$ being trace-preserving and one of $\mathcal{X}$ and $\mathcal{Y}$ being a unitary channel. For two unitary channels $\mathcal{U}, \mathcal{V} \in \mathrm{CPT}(\mathcal{H})$ with $U, V \in \mathrm{U}(d)$ we can further simplify (2.76) to

$$F_{\mathrm{avg}}(\mathcal{V}, \mathcal{U}) = \frac{|\,\mathrm{Tr}[V^\dagger U]\,|^2 - d}{d(d+1)}\,. \tag{2.78}$$

Lastly, beside the entanglement fidelity, the Hilbert-Schmidt inner-product, the average gate fidelity, there is another affinely related measure of quality that is particularly convenient to work with in the analysis of randomized benchmarking: the effective depolarizing parameter. Here, we will define the effective depolarizing parameter only for trace-preserving maps via its linear relation to the fidelity. If $\mathcal{X}$ is not trace-preserving one can more generally define it by explicitly first projecting on unital maps. Let $\mathcal{X} \in \mathbb{L}(\mathcal{H}, \mathcal{K})$ be trace-preserving, its *effective depolarizing parameter* is

$$p(\mathcal{X}) := \frac{d\, F_{\mathrm{avg}}(\mathcal{X}) - 1}{d-1}\,. \tag{2.79}$$

To justify its name let us have a look at the depolarizing channel $\mathcal{D}_p$, which was defined in (2.60) as the convex combination of $\mathcal{D}_1 = \mathrm{Id}$ and $\mathcal{D}_0$. The average-gate fidelity of these extremal channel can be quickly calculated to be $F_{\mathrm{avg}}(\mathrm{Id}) = 1$ and $F_{\mathrm{avg}}(\mathcal{D}_0) = \frac{1}{d}$. Thus, $F_{\mathrm{avg}}(\mathcal{D}_p) = p + \frac{1-p}{d}$. Plugging this into the definition of the effective depolarizing parameter (2.79) yields

$$p(\mathcal{D}_p) = p. \tag{2.80}$$

Another affinely related measure that is often used in this context is the $\chi_{0,0}$-entry of the so-called χ-process matrix, see e.g. Ref. [CWE19] for further details.

2.4.3 The diamond norm

The distance measures on quantum channels we encountered so far can be regarded as average error measures. A more pessimistic, worst-case error measure is induced by the trace-norm on operators, the so-called *diamond norm*. It measures the operational distinguishability of quantum channels. Hence, it plays an important role in the characterization of quantum processes. Indeed, also error-correction thresholds require worst-case guarantees without additional assumption on the error model, see e.g. the discussion Refs. [SWS16; Kue+16]. At the same time, characterization protocols that directly come with guarantees in diamond norm are very resource intense and typically practically infeasible. For this reason, the connection of the diamond norm to the already introduced average error measures will be in the focus of this section.

The diamond norm is defined as the *completely boundedness (CB)-completion* of the $(1 \to 1)$-norm that is induced on $\mathbb{L}(\mathcal{H}, \mathcal{K})$ by the trace-norm on $\mathrm{L}(\mathcal{H})$ and $\mathrm{L}(\mathcal{K})$: Let $\mathcal{X} \in \mathbb{L}(\mathcal{H}, \mathcal{K})$,

$$\|\mathcal{X}\|_{1\to 1} := \sup_{\|A\|_1 \leq 1} \|\mathcal{X}(A)\|_1 \ . \tag{2.81}$$

When comparing quantum channels, one can also input a state to a quantum channel that shares entanglement with another quantum system. Such entanglement might be beneficial for distinguishing the channels. This motivates the definition of the *diamond norm* as the CB-completion of the $(1 \to 1)$-norm for $\mathcal{X} \in \mathbb{L}(\mathcal{H})$:

$$\|\mathcal{X}\|_\diamond := \|\mathcal{X} \otimes \mathrm{id}_{\mathcal{H}}\|_{1\to 1} \ . \tag{2.82}$$

It is easy to convince oneself that $\|\mathcal{X}\|_\diamond = \sup_{\mathcal{H}'} \|\mathcal{X} \otimes \mathrm{id}_{\mathcal{H}'}\|_{1\to 1}$, see e.g. [Wat18, Chapter 3.3]. For $\mathcal{X}$ Hermicity-preserving the supremum (2.81) is attained for pure quantum states [Wat18, Theorem 3.51]. By means of Proposition 2, the diamond norm, thus, inherits an operational interpretation from the interpretation of the trace-norm. While the trace-distance quantifies the distinguishability of quantum states, the diamond norm difference between two quantum channels is the worst-case distinguishability over all possible input states taking into account all possibilities of entanglement with an arbitrary ancillary system. Further note that channels are normalized to 1 in $(1 \to 1)$-norm and diamond norm.

Guarantees for many characterization protocols are more directly formulated in norm distances of their Choi states. The following proposition illustrates that this typically yields looser bounds compared to the diamond norm.

Proposition 20 (Diamond norm and trace norm). *For any map $\mathcal{X} \in \mathbb{L}(\mathcal{H}, \mathcal{K})$*

$$\|\mathfrak{J}(\mathcal{X})\|_1 \leq \|\mathcal{X}\|_\diamond \leq \dim(\mathcal{H})\, \|\mathfrak{J}(\mathcal{X})\|_1 \,. \tag{2.83}$$

For a Hermicity-preserving map $\mathcal{X} \in \mathbb{L}(\mathcal{H}, \mathcal{K})$ the upper bound can be tightened to [Nec+18, Corollary 2]

$$\|\mathcal{X}\|_\diamond \leq \dim(\mathcal{H})\, \|\mathrm{Tr}_2[|\,\mathfrak{J}(\mathcal{X})|]\|_\infty \,. \tag{2.84}$$

Proof of Proposition 20. Tracking vector space isomorphism, we have for $\mathcal{X} \in \mathbb{L}(\mathcal{H}, \mathcal{K})$, $A, B \in \mathrm{L}(\mathcal{H})$ and $|A\rangle = A \otimes \mathbb{1}\, |\mathbb{1}\rangle$, $\langle B| = \langle \mathbb{1}\,|\, B \otimes \mathbb{1}$

$$\mathcal{X} \otimes \mathrm{id}_{\mathcal{H}}(|A\rangle\langle B|) = \mathbb{1} \otimes A\, \mathfrak{C}(\mathcal{X}) \mathbb{1} \otimes B \,. \tag{2.85}$$

The supremum of the convex trace-norm ball is attained at its extremal point that can be written as $|A\rangle\langle B|$ with $\|A\|_{\mathrm{F}} = \|B\|_{\mathrm{F}} = 1$. Thus,

$$\begin{aligned} \|\mathcal{X}\|_\diamond &= \sup_{\|A\|_{\mathrm{F}} = \|B\|_{\mathrm{F}} = 1} \{\mathbb{1} \otimes A\, \mathfrak{C}(\mathcal{X}) \mathbb{1} \otimes B\} \\ &\leq \sup_{\|A\|_{\mathrm{F}} = \|B\|_{\mathrm{F}} = 1} \|A\|_\infty \|\mathfrak{C}\, X\|_1 \|B\|_\infty \leq \|\mathfrak{C}\, X\|_1 \,, \end{aligned} \tag{2.86}$$

where we have used Hölder's inequality (2.6), sub-multiplicativity and norm ordering of the Schatten-p-norm. With $\mathfrak{C}(\mathcal{X}) = \dim(\mathcal{H})\, \mathfrak{J}(\mathcal{X})$, the upper bound follows. The lower bound is seen by choosing $A = B = \mathbb{1}/\sqrt{\dim(\mathcal{H})}$. □

It is not difficult to see that the bounds in Proposition 20 are tight, i.e., that there are $\mathcal{X}, \mathcal{Y} \in \mathbb{L}(\mathcal{H}, \mathcal{K})$ so that $\|\mathfrak{J}(\mathcal{X})\|_1 = \|\mathcal{X}\|_\diamond$ and $\|\mathcal{X}\|_\diamond = \dim(\mathcal{H})\, \|\mathfrak{J}(\mathcal{X})\|_1$.

Other distance measures that are frequently featured in characterization protocols for quantum processes are the Hilbert-Schmidt overlap, average gate fidelity or an equivalent quantity. In terms of the infidelity $r(\mathcal{X}) = 1 - F_{\text{avg}}(\mathcal{X})$, the diamond norm and the average gate fidelity are in general related by the following inequalities.

Proposition 21 (Infidelity and diamond norm [WF14, Proposition 9]). *For any $\mathcal{X} \in \mathrm{CPT}(\mathbb{C}^d)$ it holds that*

$$\frac{d+1}{d} r(\mathcal{X}) \leq \frac{1}{2} \|\mathrm{id} - \mathcal{X}\|_\diamond \leq \sqrt{d(d+1) r(\mathcal{X})}\,. \tag{2.87}$$

Proof. The proof combines Proposition 20 with the Fuchs-van-de-Graaf inequality (2.13). Latter yields

$$1 - \mathrm{F}(\mathfrak{J}(\mathrm{Id}), \mathfrak{J}(\mathcal{X})) \leq \frac{1}{2} \| \mathfrak{J}(\mathrm{Id}) - \mathfrak{J}(\mathcal{X}) \|_1 \leq \sqrt{1 - \mathrm{F}(\mathfrak{J}(\mathrm{Id}), \mathfrak{J}(\mathcal{X}))}\,, \tag{2.88}$$

where we already dropped a square-root on the lower bound.

Since $\mathfrak{J}(\mathrm{Id}) = \frac{1}{d} \left|\mathbb{1}\right\rangle\!\left\langle\mathbb{1}\right|$ is of unit rank and Hermitian, it holds that $\mathrm{F}(\mathfrak{J}(\mathrm{Id}), \mathfrak{J}(\mathcal{X})) = \langle \mathfrak{J}(\mathrm{Id}), \mathfrak{J}(\mathcal{X}) \rangle = \mathrm{F}_\mathrm{e}(\mathrm{Id}, \mathcal{X})$. We can cast this in terms of the average gate fidelity via (2.77),

$$\mathrm{F}(\mathfrak{J}(\mathrm{Id}), \mathfrak{J}(\mathcal{X})) = \frac{d+1}{d} F_{\text{avg}}(\mathcal{X}) - \frac{1}{d}\,. \tag{2.89}$$

Plugging (2.89) into (2.88) gives

$$\frac{d+1}{d}(1 - F_{\text{avg}}(\mathcal{X})) \leq \frac{1}{2} \| \mathfrak{J}(\mathrm{Id}) - \mathfrak{J}(\mathcal{X}) \|_1 \leq \sqrt{\frac{d+1}{d}} \sqrt{1 - F_{\text{avg}}(\mathcal{X})}. \tag{2.90}$$

Finally, from Proposition 20 the proposition's assertion follows. □

Proposition 21 leaves us with an unsatisfactory state of affairs in two regards: First, the upper bound of the diamond norm introduces a dimensional factor $\mathrm{O}(d)$. In the context of quantum computing, this leaves us with a potentially large factor scaling exponentially $\mathrm{O}(2^n)$ with the number of qubits n. Second, the upper bound scales with the square-root of the infidelity. For unitary quantum channels one can in fact tighten the lower-bound to $\sqrt{r(\mathcal{X})}$ [Kue+16]. The lower-bound for unitary quantum channels indicates that the square-root scaling is unavoidable in general. Practically, this means that to certify in diamond norm requires a certificate in infidelity that is order of magnitudes smaller. Especially, for small system sizes this can be a key obstacle for the certification of the worst-case performance of quantum processes.

Fortunately, if a quantum process is highly incoherent, i.e. far away from being unitary, one can derive a linear scaling of the diamond-norm distance in the infidelity. The incoherence can be controlled by the so-called unitarity introduced by Wallman *et al.* [Wal+15]. For $\mathcal{X} \in \mathbb{L}(\mathcal{H})$ the *unitarity* is defined as

$$u(\mathcal{X}) = \frac{d}{d-1} F_{\text{avg}}(\mathcal{X}', \mathcal{X}'), \tag{2.91}$$

where $d = \dim \mathcal{H}$ and $\mathcal{X}' \in \mathbb{L}(\mathcal{H})$ is defined as $\mathcal{X}'(A) = \mathcal{X}(A) - \operatorname{Tr}[\mathcal{X}(A)]\mathbb{1}/\sqrt{d}$. One can straightforwardly check that $u(\mathcal{U}) = 1$ for every unitary channel $\mathcal{U}$. On the other hand, in Refs. [Wal+15; Kue+16] a lower-bound on u in terms of the infidelity r was derived for trace-decreasing maps. For $\mathcal{X} \in \mathbb{L}(\mathcal{H})$ and $\operatorname{Tr}(\mathcal{X}(\mathbb{1})) \leq \operatorname{Tr}(\mathbb{1})$ it holds that

$$u(\mathcal{X}) \geq u_{\text{min}} = \left(1 - \frac{d}{d-1} r(\mathcal{X})\right)^2 . \tag{2.92}$$

Kueng *et al.* [Kue+16] established that quantum channels saturating this lower bound exhibit a linear scaling of the diamond norm distance in terms of the infidelity.

Theorem 22 (Worst-case bound for incoherent channels [Kue+16, Theorem 3])**.** *Let $\mathcal{X} \in \operatorname{CPT}(\mathcal{H})$ be unital, $d = \dim(\mathcal{H})$. $\|\operatorname{Id} - \mathcal{X}\|_\diamond \in \mathrm{O}(r(\mathcal{X}))$ if $u = u_{min} + \mathrm{O}(r^2(\mathcal{X}))$.*

We leave it with this qualitative statement and refer to Ref. [Kue+16, Proposition 3] for a quantitative statement. See also Ref. [Wal15].

3 Blind quantum state tomography

Blind quantum state tomography is the task of uniquely identifying a quantum state from the data acquired with a measurement apparatus that is itself not fully characterized.[1] In the introduction we motivated blind quantum state tomography as a way to break the vicious cycle, illustrated in Figure 1.4, that fundamentally limits the precision of quantum state tomography by the achievable accuracy of the measurement device's calibration. Here we show that by exploiting the natural low-rank structure of quantum states we can provably break this vicious cycle for generic measurement models with a close-to-optimal number of measurement settings and efficient classical post-processing.

Provable blind tomography via sparse de-mixing. Let us be slightly more formal in order to first give an overview over the technical methods and contributions of this chapter. In mathematical terms, the blind tomography task that we solve is to infer a vector ξ of n calibration parameters and a rank-r quantum state ρ from data of the form

$$y = \mathcal{B}_\xi(\rho) = \mathcal{A}(\xi \otimes \rho) \tag{3.1}$$

where $\mathcal{B} : \xi, \rho \mapsto \mathcal{B}_\xi(\rho)$ is a bi-linear map describing the measurement model. The measured data y might for example be estimates for the expectation values of observables or probabilities of POVM elements. For the time being, we ignore the error of the estimates induced by finite statistics. It is convenient to regard the data as associated to a structured linear estimation problem: we can equivalently model the measurement map as a linear map $\mathcal{A}$ acting on $\xi \otimes \rho$.

[1] The material of this chapter is published as the preprint [2]. We adjusted the presentation for notational consistency and added references to other parts of the monograph. The research was conducted in close collaboration with Jadwiga Wilkens, Dominik Hangleiter and Jens Eisert. The majority of the numerical implementation was coded by Jadwiga Wilkens under the supervision of the author of the thesis. Furthermore, we thank David Gross, Steven T. Flammia, Christian Krumnow, Robin Harper, Yi-Kai Liu, and Carlos A. Riofrio for inspiring discussions and helpful comments. We are grateful to Alireza Seif and Nobert Linke for valuable comments on realistic error models and Peter Jung for making us aware of Ref. [SW19].

Such structured linear inverse problems are studied in the mathematical discipline of model-based *compressed sensing* [Bar+10; FR13], where efficient algorithms with analytical performance guarantees have been developed. A work horse of compressed sensing that most rapidly solve the relevant inverse problems are so-called *iterative hard thresholding (IHT)* algorithms [BD08].

As a first result of this work, we establish that the key step of an IHT algorithm that solves the blind tomography problem is NP-hard. To overcome this obstacle, we propose an IHT algorithm that solves a slightly relaxed version of the blind tomography problem: the task of de-mixing a sum of n different low-rank quantum states ρ_i, i.e., data of the form

$$y = \mathcal{A}\left(\sum_{i=1}^{n} \xi_i e_i \otimes \rho_i\right), \tag{3.2}$$

where $\{e_i\}_{i=1}^{n}$ denotes the standard orthonormal basis. An efficient IHT algorithm for the de-mixing problem of low-rank matrices was developed and analysed in Ref. [SW19]. This algorithm can be readily adapted to our problem.

But relaxing the blind tomography problem to the de-mixing problem artificially introduces an overhead in the number of unknown degrees of freedom of the problem scaling as $2drn$, and in particular linearly with the number of calibration parameters in the model. This leads to an unfavourable situation in a two-fold manner: First, determining many calibration parameters also requires many measurement settings as the cost per calibration parameter scales with the dimension d of the quantum system. Second, a necessary condition for a well-posed blind de-mixing problem of rank-r with a maximal number of d^2 linearly independent measurements of the form (3.2) is that there are more linearly independent measurements than real parameters, i.e., $2rnd \leq d^2$. This means that the simultaneous determination of a certain number of calibration parameters n can in principle only work for sufficiently large system dimension d in many situations. This causes severe constraints in the achievable self-calibration for small system sizes.

We argue that an additional well-motivated structural assumption can render the blind tomography much more broadly applicable. Our argument is based on the observation that the problem of determining an accurate estimate of the quantum state in the blind setting involves solving two distinct sub-problems: first, one needs to determine which ones of many potential error models of the measurement contribute. Second, one needs to estimate the calibration parameters of these models. Generically, there are many potential models that parametrize, for instance, the deviation of every imperfect implementation of a fixed measurement setting from its ideal implementation.

In this case, the first problem becomes combinatorially costly since many distinct measurement settings need to be simultaneously calibrated. In contrast, in our approach, it is straightforward to solve both tasks simultaneously and even avoid a combinatorial overhead using the built-in relaxations of compressed sensing. We want to allow for many potential errors with associated calibration parameters but of which only small number s actually contribute. This can be cast as the assumption that the calibration vector ξ is s-sparse, i.e., it has only s non-vanishing entries. Of course, we do not assume that we know the support of the vector ξ. This falls naturally into the framework of structured signal recovery. To summarize: We observe data generated by linear measurements acting on $\xi \otimes \rho$ where ξ is an s-sparse vector and ρ is a rank r quantum state.

We are now faced with the recovery problem of de-mixing a *sparse sum* of different *low-rank quantum states*. We show that the projection onto this structure can be efficiently calculated using hierarchical thresholding [14] and therefore circumvents our NP-hardness result. We derive the corresponding iterative hard-thresholding algorithm and prove that it successfully recovers the states ρ_i and the sparse vector ξ provided that the measurement map $\mathcal{A}$ acts isometrically on sparse sums of low-rank states. We further show that generic measurement ensembles with m different measurement settings exhibit this restricted isometry property provided that m scales at least as $srd + s \log n$. Thus, we find that our algorithm solves the blind tomography problem with an overhead in the required number of measurements that scales linearly in s as compared to the number of degrees of freedom in the problem given by $rd + s$. In particular, the number of potential calibration models n enters only logarithmically in the measurement complexity of the scheme. This renders the scheme highly scalable in n providing flexibility in the modelling of systematic measurement errors or calibration corrections. Furthermore, it leaves sufficiently many linearly independent parameters in order to infer a couple of calibration parameters already for comparably small system sizes. We demonstrate the performance of the algorithm for the physically relevant case of measuring Pauli operators that are locally mixed with the unknown calibration parameters.

Practical blind tomography. Going beyond working out the theoretical guarantees, we numerically demonstrate the functioning of the scheme and the mindset behind it. Specifically, we show that the iterative hard-thresholding algorithm solves the blind tomography problem from much fewer samples than competing methods from generic (Gaussian) measurements as well as sub-sampled random Pauli measurements. We then take the theoretical model to the practical test bed and turn to a realistic model of measurement errors given by a coherent over-rotation along some axis. Those measurements have significantly more structure. We observe that the measurement structure together with the sparsity constraints causes the SDT algorithm to frequently get stuck at objective variables with an incorrect support. For this reason, we also study the performance of a more pragmatically minded optimization strategy, namely, constrained alternating minimization that does not require the relaxation to the de-mixing problem. We numerically demonstrate that the blind tomography problem in a realistic setting can be solved using this adapted algorithmic approach. Thereby, we show that exploiting the low-rank structures of quantum states allows one to perform tomography blindly in realistic calibration and measurement models. Therefore, we expect our approach to be directly applicable in a variety of experimental settings that are practically relevant in the quantum technologies.

The remainder of this chapter is organized as follows. In the subsequent Section 3.1, we give a detailed description of a concrete experimental setup that motivates our mathematical formulation of the blind tomography problem. In Section 3.2, we provide the formal definitions of the blind tomography problem and introduce the notation used in the subsequent parts. The details of the sparse demixing algorithm and its variant based on alternating optimization are derived in Section 3.3. On the way, we establish the NP-hardness of the projection associated to the original blind tomography problem. The theorems guaranteeing the performance of the sparse demixing algorithm are explained in Section 3.4. Finally, numerical simulations of the algorithms' performance and its application to practical use cases are shown in Section 3.5 before we conclude with an outlook in Section 3.6.

3.1 Tomography with imperfect Pauli correlation measurements

So far, our description of the measurement scheme has been fairly abstract. In the following, we describe a concrete scenario in which our formalism applies. Consider an ion trap experiment preparing a multi-qubit quantum state ρ. We perform Pauli correlation measurements, i.e., we estimate m expectation values of l-qubit Pauli strings of the form

$$y_0^{(k)} = \mathcal{A}_0(\rho)^{(k)} = \operatorname{Tr}\left[\rho\left(W_1^{(k)} \otimes W_2^{(k)} \otimes \cdots \otimes W_l^{(k)}\right)\right], \tag{3.3}$$

where $W_j^{(k)} \in \{X, Y, Z, \mathrm{Id}\}$ is a Pauli matrix, (2.53), acting on the jth qubit and $k \in [m] := \{1, 2, \dots, m\}$. We refer to $\mathcal{A}_0 : \mathbb{C}^{d \times d} \to \mathbb{R}^m$ as the measurement map or sampling operator.[2].

In many experimental setups, it is natural to implement measurements of a certain Pauli observable—in the case of ion traps Pauli Z—while the other Pauli observables require more effort. A measurement of any other Pauli observable—in the case of ion traps Pauli X and Pauli Y—can then be implemented by applying a suitable sequence of unitary gates prior to the measurement. For example, using addressed laser pulses of different duration one can implement rotations around different axes and thus implement the Hadamard gate H as well as the phase gate S as defined in (2.55). In this way, one can realize measurements in the $X = HZH$ and $Y = SHZHS^\dagger$ basis.

But each application of an additional gate may come with a coherent error *in addition* to the native error associated with the measurement itself. In this way, we end up with different systematic errors for different Pauli observables parametrized by the angles θ, φ of a coherent error given by $\mathrm{e}^{\mathrm{i}\theta X}\mathrm{e}^{\mathrm{i}\varphi Z}$. This gives rise to some probability of actually measuring the expectation value of another local Pauli matrix than the targeted one. For example, consider a coherent error given by a (small) rotation around the Z-axis as given by $\mathrm{e}^{\mathrm{i}\varphi Z}$. The faulty implementation of the Hadamard gate is then given by $\tilde{H} = \mathrm{e}^{\mathrm{i}\varphi Z}H$. Of course, the native Z-measurement is untouched by this coherent error, since no unitary rotation precedes this measurement. However, instead of Y one now actually measures $\tilde{Y} = S\tilde{H}Z\tilde{H}^\dagger S^\dagger = \cos(2\varphi)Y + \sin(2\varphi)X$. At the same time X remains undisturbed.

More generally, we can introduce calibration parameters $\xi_{W \to \tilde{W}}$ measuring the strength of the error that replaces a certain target Pauli matrix W by $\tilde{W}$. For instance, in the above example those parameters are given by $\xi_{Y \to Y} = \cos(2\varphi)$, $\xi_{Y \to X} = \sin(2\varphi)$ and $\xi_{Z \to Z} = \xi_{X \to X} = 1$. For simplicity, we assume that these calibration parameters are identical for different qubit registers. Assuming that errors are not too large, the calibration parameters for the target measurement fulfil $\xi_{W \to W} \approx 1$ or all $W \in \{X, Y, Z\}$. This leaves us with six independent calibration parameters corresponding to the cross-contributions. To construct

[2] Note that one might actually implement projective measurements in the multi-qubit Pauli basis as done, e.g., in Refs. [Rio+17; 11] While such projective measurements contain more information than the Pauli correlation measurement, we restrict ourselves to Pauli expectation values both for the sake of simplicity and to remain in a setting for which theoretical guarantees can be proven [Gro+10; Liu11].

the measurement map $\mathcal{A}$, we start from the definition of the target measurement $\mathcal{A}_0$ in (3.3). From $\mathcal{A}_0$ we can derive calibration measurement components $\mathcal{A}_{W\to\tilde{W}}$ appearing with the coefficient $\xi_{W\to\tilde{W}}$ by replacing all appearances of the Pauli matrix W in the definition of $\mathcal{A}_0$ with $\tilde{W}$. If W appears in a multi-qubit Pauli observable several times the resulting observable is the sum of all Pauli observables generated by replacing only one of the W by $\tilde{W}$, assuming that the coherent errors are small so that the higher-order terms can be neglected. For example, a faulty realization of the observable $ZYZZY$ is now given by $\xi_{Y\to Y}ZYZZY + \xi_{Y\to X}(ZXZZY + ZYZZX)$.

Altogether, to linear order in the calibration parameters $\xi_{W\to\tilde{W}}$ with $W \neq \tilde{W}$ we end up constructing a description of the effective faulty measurement by

$$y = \xi_0\mathcal{A}_0(\rho) + \sum_{W\neq\tilde{W}\in\{X,Y,Z\}} \xi_{W\to\tilde{W}}\mathcal{A}_{W\to\tilde{W}}(\rho), \tag{3.4}$$

which can be written as linear map $\mathcal{A}$ action on $\xi\otimes\rho$ with the vector of calibration parameters $\xi = [\xi_0, \xi_{X\to Y}, \xi_{X\to Z}, \ldots, \xi_{Z\to Y}]^T$. By assumption, we set the parameter $\xi_0 = 1$.

In this measurement model the sparsity assumption is justified if unitary errors in a certain coordinate plane are dominant compared to others thus singling out certain types of calibration measurement components. Importantly, we do not assume that we know which corrections are dominant (i.e., the support of ξ) *a priori*. The measurement model also exemplifies a setting in which one is ultimately limited to measuring a maximal set of d^2 observables. Thus, blind tomography becomes only possible exploiting structure assumptions if one does not allow for different ways of implementing the same measurement that yield different calibration corrections without introducing too many new calibration parameters.

3.2 Formal problem definition

Motivated by this example, we set out to provide a formal definition of the blind tomography problem and the related sparse-de-mixing problem. The notation and terminology introduced in this section allow us to formulate a general signal-processing approach using which the blind tomography can be provably solved. Both, the blind tomography and the sparse de-mixing, problems are linear inverse problems that feature a combination of different compressive structures. These are smaller sets of linear vector spaces, and it will be convenient to introduce some notation to refer to these sets. The prototypical example is the *set of s-sparse real vectors*

$$\Sigma_s^n := \{\xi \in \mathbb{R}^n \mid |\operatorname{supp} \xi| \leq s\} \subset \mathbb{R}^n, \tag{3.5}$$

which is defined by the support $\operatorname{supp} \xi$ of a vector ξ, i.e., the index set of the non-vanishing entries of ξ, having cardinality smaller or equal than s. The set of s-sparse vectors is not a vector space itself but the union of $\binom{n}{s}$ s-dimensional subspaces.

In the realm of quantum mechanics, the non-commutative analogue of sparse vectors, namely low-rank matrices, is important. We denote the *set of complex rank r matrices* by

$$\mathbb{C}_r^{d\times d} := \{x \in \mathbb{C}^{d\times d} \mid \operatorname{rank} x \leq r\}. \tag{3.6}$$

Since we are dealing with quantum states we will restrict our attention to the set $\mathcal{D}^{(d)} \subset \mathbb{C}^{d\times d}$ of trace-normalized, positive semidefinite matrices, i.e., $\rho \geq 0$ and $\operatorname{Tr} \rho = 1$ for all $\rho \in \mathcal{D}^{(d)}$. Our results can be straightforwardly generalized to general matrices without these constraints. We denote the set of rank-r quantum states as $\mathcal{D}_r^{(d)} = \mathcal{D}^{(d)} \cap \mathbb{C}_r^{d\times d}$. In particular, $\mathcal{D}_1^{(d)}$ is the set of pure quantum states. In order to solve the blind tomography problem we need to simultaneously recover an s-sparse real vector ξ and a rank-r quantum state ρ. It is convenient to regard both ξ and ρ as a combined signal $X = \xi \otimes \rho$ and model the measurement including its dependence on the calibration parameter as a linear map $\mathcal{A}$ acting on X. Considering such linear maps instead of bi-linear maps is sometimes referred to as 'lifting' in the compressed sensing literature [ARR14]. For a physicist, 'lifting' is also the natural isomorphism at the heart of the density matrix formulation of quantum mechanics. The signal X is highly structured as it is a tensor product of a sparse vector and a low-rank quantum state. We denote the set of all potential signals as

$$\Omega_{s,r}^{n,d} := \{\xi \otimes x \mid \xi \in \Sigma_s^n,\ x \in \mathcal{D}_r^{(d)}\} \subset \mathbb{C}^{nd\times d}. \tag{3.7}$$

One can regard a signal $X \in \Omega_{s,r}^{n,d}$ as an $nd \times d$ matrix consisting of n blocks of size $d \times d$ stacked on top of each other as depicted in Figure 3.1, where each $d \times d$ block is proportional to the same quantum state ρ and only s of the blocks are non-vanishing.

We are now equipped to concisely state the problem we would like to study.

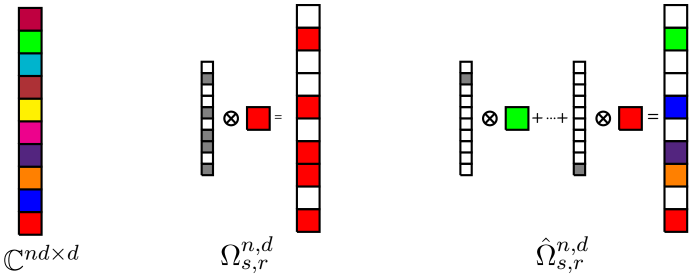


Figure 3.1: The signal sets of the blind tomography and sparse de-mixing problem can be regarded as subsets of $\mathbb{C}^{nd\times d}$, i.e., matrices consisting of n blocks of $d \times d$. For a blind tomography signal in $\Omega^{n,d}_{s,r}$, only s out of the n blocks are non-zero and are proportional to the same rank r matrix. In contrast, a signal of the sparse de-mixing problem in $\hat{\Omega}^{n,d}_{s,r}$ comprises s non-vanishing blocks with potentially different rank r matrices.

Problem 23 (Blind tomography)**.** *Let $\mathcal{A} : \mathbb{C}^{nd\times d} \to \mathbb{R}^m$ be a linear map. Given data $y = \mathcal{A}(X) \in \mathbb{R}^m$ and the linear map $\mathcal{A}$, recover X under the assumption that*

$$X \in \Omega^{n,d}_{s,r}$$

Our approach to algorithmically solving the blind tomography problem makes use of a proxy problem: we relax it to signals that are a bit less restrictively structured

$$\hat{\Omega}^{n,d}_{s,r} := \left\{ \sum_{i=1}^{n} \xi_i e_i \otimes x_i \;\middle|\; \xi \in \Sigma^n_s,\ x_i \in \mathcal{D}^{(d)}_r \ \forall i \in [n] \right\}. \tag{3.8}$$

Both sets $\hat{\Omega}^{n,d}_{s,r}$ and $\Omega^{n,d}_{s,r}$ are subsets of $\mathbb{C}^{nd\times d}$. The difference between them as illustrated in Figure 3.1 is the following: While for $X \in \Omega^{n,d}_{s,r}$ all $d \times d$ blocks are proportional to the same quantum state x, we allow the $d \times d$ blocks of $\hat{X} \in \Omega^{n,d}_{s,r}$ to be proportional to *different quantum states* x_i. Again only s out of the n blocks of $\hat{X}$ are non-vanishing. Analogously to Problem 23, we define the linear inverse problem associated with $\hat{\Omega}$.

Problem 24 (Sparse de-mixing)**.** *Let $\mathcal{A} : \mathbb{C}^{nd\times d} \to \mathbb{R}^m$ be a linear map. Given data $y = \mathcal{A}(X) \in \mathbb{R}^m$ and the linear map $\mathcal{A}$, recover X under the assumption that*

$$X \in \hat{\Omega}^{n,d}_{s,r}.$$

The observed data of the sparse de-mixing problem can be equivalently described as

$$y = \sum_{k=1}^{n} \xi_k \mathcal{A}_k(x_k), \tag{3.9}$$

where we have split up X into trace-normalized $d \times d$ blocks x_k and their norm ξ_k according to the definition of $\hat{\Omega}_{s,r}^{n,d}$. Correspondingly, we can decompose the linear map $\mathcal{A}$ into the set of linear maps $\{\mathcal{A}_k\}_{k=1}^{n}$ where each $\mathcal{A}_k$ acts only on the kth $d \times d$ block of X. From this reformulation it becomes clear that the problem amounts to reconstructing a set of low-rank signals $\{x_k\}_k$ from observing its sparse mixture under linear maps, hence the name sparse de-mixing.

For both the blind-tomography and the sparse-de-mixing problem, we alternatively write each of the n linear maps $\mathcal{A}_k$ in terms of m observables

$$\{A_k^{(i)} \in \mathbb{C}^{d \times d} \mid (A_k^{(i)})^\dagger = A_k^{(i)}\}_{i=1}^{m} \tag{3.10}$$

via

$$\mathcal{A}_k(x_k)^{(i)} = \langle A_k^{(i)}, x_k \rangle \tag{3.11}$$

with the Hilbert-Schmidt inner product $\langle \cdot, \cdot \rangle$.

Note that as long as we consider Hermitian matrices for the measurement $A_k^{(i)}$ and signals x_i, we end up with a real data vector $y \in \mathbb{R}^m$. For applications other than quantum tomography it is straightforward to adopt our proofs and results to real signals or complex-valued measurement maps. Furthermore, for the sake of simplicity we have formulated both recovery problems without noise. More generally, the data can be assumed to be of the form $y = \mathcal{A}(X) + \epsilon$ where ϵ denotes additive, e.g. statistical, noise.

3.3 Algorithm

We now turn to the technical derivation of our algorithm for the blind quantum tomography and the sparse de-mixing problem. Our algorithm builds on primitives developed in the field of compressed sensing. In particular, we generalize the hard thresholding algorithm to accommodate the structural assumptions of both problems. As a first step, we establish the hardness of direct thresholding approaches to the blind tomography problem before stating a tractable algorithm for the sparse de-mixing problem.

Let us be more precise: the blind quantum tomography problem requires different assumptions on two levels. First, we want the signal to be a tensor product $\xi \otimes \rho$, i.e., of rank one. Second, both tensor factors are assumed to be structured. Concretely, we assume ξ to be s-sparse and ρ to be of rank r. We are therefore faced with low-rank structures on two separate levels: first, the block-structured signal as given by the tensor product of calibration vector and quantum state has unit rank. Second, by assumption the target quantum states, i.e., the individual blocks of the signal, have low rank.

It has been observed in the compressed sensing literature that multi-level structures with structured tensor components can be notoriously hard to reconstruct. One prototypical example of this is combined sparsity and low-rankness in the sense that the signal is the tensor product of two sparse vectors, i.e., $X = \xi \otimes \tau$ with $\xi, \tau \in \Sigma_s^n$. This problem is already very similar to the blind tomography problem where one of the sparse vectors is replaced by a low-rank matrix, the quantum state.

The obstacle arising from such structures can be understood from a different perspective present in the compressed sensing literature that is related to different algorithmic approaches. The perhaps most prominent approach in compressed sensing is the convex relaxation of structure-promoting regularizers yielding efficient convex optimization programs. Minimizing the ℓ_1 norm or the Schatten-1 norm is known to solve linear inversion problems involving sparse or low-rank vectors efficiently and sampling optimal, respectively. However, simply combining both regularizers in a convex fashion does not yield a sampling-optimal reconstruction of problems that feature both structures any more [Oym+15].

3.3.1 Hard-thresholding: Ease and hardness of the projection

Another algorithmic approach used in compressed sensing are so-called *hard thresholding algorithms* such as CoSAMP, IHT or HTP [NT08; BD08; Fou11]; see also the textbook [FR13] for an introduction. These are typically iterative procedures that minimize the deviation from the linear constraints in some way or other, e.g. by gradient descent, and in each iteration project onto the structure of the signal. For many compressed sensing problems this is possible because even though recovery problems, such as

$$\underset{\xi}{\operatorname{minimize}} \, \|\mathcal{A}(\xi) - y\|_{\ell_2}^2 \quad \text{subject to } \xi \in \Sigma_s^n, \tag{3.12}$$

are NP-hard [Mag17], the related projection

$$P_{\Sigma_s^n}(\tau) := \operatorname*{arg\,min}_{\xi \in \Sigma_s^n} \|\xi - \tau\|_{\ell_2} \tag{3.13}$$

can be computed efficiently. For the given example of projecting onto s-sparse vectors, this solution is given by the hard-thresholding operation defined as follows: Let $\Sigma_{\max}$ be the set of indices of the s absolutely largest entries of τ. Then,

$$[P_{\Sigma_s^n}(\tau)]_i = \begin{cases} \tau_i & \text{for } i \in \Sigma_{\max} \\ 0 & \text{otherwise.} \end{cases} \tag{3.14}$$

In words, one keeps the largest entries of τ and replaces the other entries by zero. Analogously, the projection of Hermitian matrices onto low-rank matrices can be efficiently calculated by calculating the eigenvalue decomposition and applying $P_{\Sigma_r^d}$ to the eigenvalue vector. Let $X \in \mathbb{C}^{d\times d}$ be a Hermitian matrix with eigenvalue decomposition $X = U \operatorname{diag}(\lambda) U^\dagger$. We define the projection onto positive semi-definite low-rank matrices as

$$P_{\mathcal{D}_r^{(d)}}(X) = U \operatorname{diag}[P_{\Sigma_r^d}(\lambda|_{\geq 0})] U^\dagger, \tag{3.15}$$

where $\lambda|_{\geq 0}$ denotes the restriction of λ to its non-negative entries.

In hard-thresholding algorithms, the problems associated with simultaneously exploiting sparse and low-rank structures are manifest in the computational hardness of computing the respective projections. For the case of unit rank matrix with sparse singular vectors, calculating the projection is the so-called *sparse PCA* problem, i.e., given a matrix $A \in \mathbb{R}^{n\times n}$

$$\operatorname*{minimize}_{\xi,\tau \in \Sigma_s^n} \; \|A - \xi \otimes \tau\|_{\mathrm{F}} \, . \tag{3.16}$$

Indeed, exactly solving this problem in the worst case is NP-hard by a trivial reduction to the CLIQUE problem [Mag17]. But it turns out that the hardness is much worse: one can even make average-case hardness statements based on conjectures regarding the hardness of the *planted clique problem* [BR13a; BR13b; BB19]. Moreover, the sparsePCA problem remains just as hard even when one merely asks for an approximation up to a *constant relative error* [CPR16; Mag17].

As the first technical result of this work, we show that also the projection onto $\Omega_{s,r}^{n,d}$ is an NP-hard problem by reducing it to the sparse PCA problem.

Theorem 25 (Hardness of constrained minimization). *There exists no polynomial time algorithm that calculates*

$$\text{minimize} \quad \|A - X\|_{\mathrm{F}} \quad \text{subject to } X \in \Omega^{n,d}_{s,r}, \tag{3.17}$$

for all $A \in \mathbb{C}^{nd\times d}$ unless $\mathsf{P} = \mathsf{NP}$. This still holds for $s = n$.

We note that our result for exactly computing straightforwardly generalizes to the case of approximating the target function up to constant relative error using results on the approximatability of sparsePCA [CPR16].

Theorem 25 provides a strong indication that a straightforward adaptation of compressed sensing techniques is not feasible. In this work, our way out of this is to sacrifice sampling optimality of the algorithm for a lower runtime and being able to prove analytical performance guarantees. Alternating minimization approaches that make the factorization explicit is also a viable way forward. We provide a detailed description of such an algorithm in Section 3.3.4. But proving global recovery guarantees for non-convex algorithms typically becomes much more involved and often rely on an initialization step with the same computational complexity as the original problem.

In the remainder of the section we prove Theorem 25. As a starting point we state the sparsePCA problem.

Problem 26 (SparsePCA). Input: *Symmetric matrix $A \in \mathbb{R}^{n\times n}$, sparsity s, positive real number $a > 0$.* Question: *Does there exist an s-sparse unit vector $v \in \mathbb{R}^n$ with $v^T A v \geq a$?*

It has been folklore for quite some time that the sparse PCA problem is NP-hard. A formal proof can be found in Ref. [Mag17], where the CLIQUE problem is encoded into instances of sparsePCA. From the hardness of sparsePCA it follows that there does not exist a polynomial time algorithm for the projection onto the set of symmetric, unit rank matrices with sparse eigenvectors. Formally, we have:

Proposition 27 (Hardness of projection onto the set of symmetric, unit rank matrices with sparse eigenvectors). *Given a matrix $A \in \mathbb{R}^{d\times n}$ and $s, \sigma \in \mathbb{N}$, there exist no polynomial time algorithm that calculates*

$$\begin{gathered} \text{minimize} \quad \left\| A - vw^T \right\|_{\mathrm{F}}, \\ \text{subject to } v \in \Sigma^d_\sigma,\, w \in \Sigma^n_s. \end{gathered} \tag{3.18}$$

This still holds for $\sigma = d$.

Proof. It turns out to be sufficient to only consider the case where $\sigma = d$, i.e., only one of the factors is required to be sparse. It is straightforward to see that solving the problem with both vectors being sparse allows one to solve the projection with only one sparse vector: Define

$$A = \begin{pmatrix} \mathbf{0}_{d-\sigma,n} \\ A' \end{pmatrix} \tag{3.19}$$

with $\mathbf{0}_{a,b}$ being an $a \times b$ matrix filled with zeros. It then holds that

$$\min_{v \in \Sigma_\sigma^d, w \in \Sigma_s^n} \left\| A - vw^T \right\|_{\mathrm{F}} = \min_{v' \in \mathbb{C}^\sigma, w \in \Sigma_s^n} \left\| A' - v'w^T \right\|_{\mathrm{F}} . \tag{3.20}$$

We now embed the sparsePCA problem. To do so we first make the normalization of the vectors v, w in the optimization problem explicit to it to a maximization problem over normalized vectors:

$$\begin{aligned} &\min_{v \in \mathbb{R}^\sigma, w \in \Sigma_s^n} \left\| A - vw^T \right\|_{\mathrm{F}}^2 \\ = &\min_{\lambda \in \mathbb{R}, v \in \mathbb{R}^\sigma \cap B_{\ell_2}^\sigma, w \in \Sigma_s^n \cap B_{\ell_2}^n} \left\| A - \lambda vw^T \right\|_{\mathrm{F}}^2 \end{aligned} \tag{3.21}$$

with $B_{\ell_2}^n = \{v \in \mathbb{R}^n \mid \|v\|_{\ell_2} \leq 1\}$ the ℓ_2-norm ball. Solving the optimization problem over λ yields

$$\begin{aligned} &\min_{\lambda \in \mathbb{R}} \left\| A - \lambda vw^T \right\|_{\mathrm{F}}^2 &(3.22)\\ &\quad = \min_{\lambda \in \mathbb{R}} \left\{ \|A\|_{\mathrm{F}}^2 + \lambda^2 \|v\|_{\ell_2}^2 \|w\|_{\ell_2}^2 - 2\lambda \langle w, Av \rangle \right\} &(3.23)\\ &\quad = \|A\|_{\mathrm{F}}^2 - \min_{v \in \mathbb{R}^\sigma \cap B_{\ell_2}^\sigma, w \in \Sigma_s^n \cap B_{\ell_2}^n} \langle w, Av \rangle^2 . &(3.24) \end{aligned}$$

Since A is fixed we conclude that the optimization problem (3.21) is equivalent to

$$\begin{aligned} &\text{maximize } |\langle w, Av \rangle| \\ &\text{subject to } v \in \mathbb{R}^\sigma \cap B_{\ell_2}^\sigma,\ w \in \Sigma_s^n \cap B_{\ell_2}^n . \end{aligned} \tag{3.25}$$

Furthermore, using the Cauchy-Schwarz inequality we find that

$$\max_{v \in \mathbb{R}^\sigma \cap B_{\ell_2}^\sigma, w \in \Sigma_s^n \cap B_{\ell_2}^n} |\langle v, Aw \rangle| = \max_{w \in \Sigma_s^n \cap B_{\ell_2}^n} \|Aw\|_{\ell_2} . \tag{3.26}$$

Now consider an instance of the sparsePCA problem with a symmetric input matrix $B \in \mathbb{R}^{n \times n}$, sparsity s and $a > 0$. W.l.o.g. we can assume that B is a positive matrix since solving the sparsePCA problem for the $B - \min\{0, \lambda_{\min}(B)\} \operatorname{Id}$ shifted by the smallest eigenvalue $\lambda_{\min}(B)$ of B and a shifted correspondingly, allows one to solve the sparsePCA problem for B. For a positive matrix B we find a factorization $B = A^T A$. Hence, deciding whether the maximum over all $w \in \Sigma_s^n$ of $w^T B w$ is larger than a is solved by calculating the maximum of $\|Aw\|_{\ell_2}^2 = w^T B w$. This completes the reduction. □

We are now prepared to prove the related result for the projection onto $\Omega_{s,r}^{n,d}$.

Proof of Theorem 25. Suppose the existed an efficient algorithm that determines the objective value of the projection (3.17). To encode the sparsePCA problem, we choose an instance of A as follows: Let $A' \in \mathbb{R}^{n\times d}$ be a matrix and let A'_i denote the ith row of A. Let e_i be the basis vectors $(e_i)_j = \delta_{i,j}$, with $\delta_{i,j}$ the Kronecker symbol. We choose $A = \sum_{i=1}^n e_i \otimes \operatorname{diag}(A'_i)$, where $\operatorname{diag}(A'_i)$ denotes the diagonal matrix with the ith row of A' on its diagonal. Furthermore, we define $a' = (A'_1, \dots, A'_n) \in \mathbb{R}^{nd}$ to be the vector arising by concatenating all rows of A. By definition an $X \in \Omega_{s,r}^{n,d}$ can be decomposed as $X = \xi \otimes \rho$ with $\xi \in \Sigma_s^n$ and $\rho \in \mathcal{D}_r^{(d)}$. Let $\rho = U \operatorname{diag}(\lambda) U^\dagger$ the eigenvalue decomposition of ρ with a suitable unitary $U \in \mathrm{U}(n)$ and λ the vector of its eigenvalues. Then, we can rewrite

$$
\begin{aligned}
\|A - \xi\otimes\rho\|_{\mathrm{F}}^2 &= \sum_{i=1}^n \left\|\operatorname{diag}(A'_i) - \xi_i \rho)\right\|_{\mathrm{F}}^2 \\
&= \left\|\operatorname{diag}(a') - (\operatorname{Id}_n \otimes U) \operatorname{diag}(\xi\otimes\lambda)(\operatorname{Id}_n \otimes U^\dagger)\right\|_{\mathrm{F}}^2 \\
&= \left\|\operatorname{diag}(a')(\operatorname{Id}_n \otimes U) - (\operatorname{Id}_n \otimes U) \operatorname{diag}(\xi\otimes\lambda)\right\|_{\mathrm{F}}^2 \\
&= \sum_{i,j=1}^{nd} |A'_i - (\xi\otimes\lambda)_j|^2 |(\operatorname{Id}_n \otimes U)_{i,j}|^2,
\end{aligned}
$$

where we have used the unitary invariance of the Frobenius norm in the third step. We can introduce the doubly stochastic matrix W with entries $W_{k,l} = |U_{k,l}|^2$ and relax the optimization to

$$
\begin{aligned}
&\min_{\xi\in\Sigma_s^n, \rho\in\mathcal{D}_r^{(d)}} \|A - \xi\otimes\rho\|_{\mathrm{F}}^2 \qquad (3.27) \\
&\quad\leq \min_{W\in\mathrm{DS}^n,\, \xi\in\Sigma_n^s,\, \lambda\in\Sigma_r^d} \sum_{i,j=1}^{nd} |A'_i - (\xi\otimes\lambda)_j|^2 (\operatorname{Id}_n \otimes W)_{i,j}\,,
\end{aligned}
$$

where W is optimized over all doubly stochastic matrices $\mathrm{DS}^d \subset \mathbb{C}^{d\times d}$. For $\sigma \in \mathfrak{S}_d$, a permutation of the symbols in $[d]$, we denote the corresponding permutation matrix by $\Pi_\sigma : \mathbb{C}^d \to \mathbb{C}^d$, $\xi \mapsto \Pi_\sigma \xi$ with $(\Pi_\sigma \xi)_i = \xi_{\sigma(i)}$. By Birkhoff's theorem, see e.g., Ref. [Bha97, Theorem II.2.3], the set of extremal points of the convex set of doubly stochastic matrices DS^d are the permutation matrices $\Pi_{\mathfrak{S}_d} = \{\Pi_\sigma \mid \sigma \in \mathfrak{S}_d\}$.

Since the optimum is, hence, attained for a permutation matrix $W = \Pi_\sigma$ and $U_{i,j} = (\Pi_\sigma)^{1/2}_{i,j} = (\Pi_\sigma)_{i,j}$ is a unitary matrix, the inequality (3.27) is saturated. Therefore, we conclude that

$$\begin{aligned}
&\min_{\xi\in\Sigma^n_s,\rho\in H^y_r} \left\|A' - \xi \otimes \rho\right\|^2_{\mathrm{F}} \\
&\quad= \min_{\xi\in\Sigma^n_s,\lambda\in\Sigma^d_r,\sigma\in S_d} \|a' - \xi \otimes \Pi_\sigma \lambda\|^2_{\ell_2} \\
&\quad= \min_{\xi\in\Sigma^n_s,\lambda\in\Sigma^d_r} \|a' - \xi \otimes \lambda\|^2_{\ell_2} \\
&\quad= \min_{\xi\in\Sigma^n_s,\lambda\in\Sigma^d_r} \left\|A' - \xi\lambda^T\right\|^2_{\mathrm{F}} .
\end{aligned}$$

Thus, an algorithm calculating the projection onto $\Omega^{n,d}_{s,r}$ for the matrix A chosen here solves the sparsePCA problem for A'. We conclude that there exists no polynomial time algorithm for the problem. □

3.3.2 Relaxing the blind tomography problem: sparse de-mixing

Bi-sparse and low-rank structure can be relaxed to a simple hierarchical sparsity constraint [16; Fou+19]. A vector $\xi \in \mathbb{C}^{Nn}$ consisting of N blocks of size n is called (s,σ)-hierarchically sparse if it has at most s blocks with non-vanishing entries, that themselves are σ-sparse [Spr+10; FHT10; Spr+11; Sim+13]. In the Refs. [12; 13; 14; 21] we develop a hard-thresholding algorithm for this structure together with a framework for proving theoretical recovery guarantees. It has been applied in different contexts [15; 19; 17; 18] including *sparse blind deconvolution* [16] which features the combined low-rank, sparse structure.

Here, we make use of this relaxation approach to solve the blind quantum tomography problem as formalized in Problem 23. At the heart of our approach is the insight that the projection onto $\hat{\Omega}^{n,d}_{s,r}$ can be efficiently computed hierarchically since the n $d \times d$ blocks may be different. This allows one to combine the projection onto Σ^n_s and the projection onto $\mathcal{D}^{(d)}_r$: First, the low-rank projection $P_{\mathcal{D}^{(d)}_r}$ is applied to each of the $d \times d$ blocks of the input matrix X. Subsequently,

Algorithm 1 Projection onto $\hat{\Omega}_{s,r}^{n,d}$

```
Input: $X \in \mathbb{C}^{nd \times d}$.
$Y = 0$
for $k \in [n]$ do
    $n_k = \left\| P_{\mathcal{D}_r^{(d)}}(x_k) \right\|_{\mathrm{F}}$
end for
$W = \operatorname{supp} P_{\Sigma_s^n}(n)$
$Y_W = X_W$.
Output: $Y$ is projection of $X$ onto $\hat{\Omega}_{s,r}^{n,d}$.
```

the sparse projection operator is applied by setting the $n - s$ smallest blocks in Frobenius norm to zero. The resulting algorithm is summarized as Algorithm 1. The computational cost of the projection onto $\hat{\Omega}_{s,r}^{n,d}$ is dominated by the eigenvalue decomposition required to compute the low-rank approximation $\mathcal{D}_r^{(d)}$ of each block. Computing the full eigenvalue decomposition of the $d \times d$ blocks requires computation time of $O(d^3)$ using, e.g., Householder reflections [GL89]. Since we are only interested in the dominant $r \ll d$ eigenvalues, the effort can be reduced to $O(rdw)$ using the Lanczos algorithm, where w is the average number of non-zero elements in a row of a block [GL89]. Using randomized techniques one might be able to further reduce the computational costs [HMT11]. The calculation of the Frobenius norms contributes $O(nd^2)$ flops. The largest blocks can be selected using the quick-select algorithm [Hoa61] in $O(n)$. Note that the low-rank projections and Frobenius norms of all blocks can also be performed in parallel without any modification to the algorithm.

Equipped with an efficient projection for $\hat{\Omega}_{s,r}^{n,d}$, we can construct a structured iterative gradient descent algorithm. This is a variant of the IHT algorithm, that was originally developed for sparse vectors [BD08]. The IHT algorithm is a projective gradient descent algorithm that iteratively alternates gradient steps to optimize the ℓ_2-norm deviation between the data and a projection onto the constraint set. The resulting recovery algorithm for the sparse de-mixing problem is stated as Algorithm 2, the sparse de-mixing thresholding (SDT) algorithm.

The SDT algorithm is closely related to the IHT algorithm for de-mixing low-rank matrices that was developed in Ref. [SW19]. We will refer to this algorithm as the *DT algorithm*. The main difference between our SDT and the DT algorithm of Ref. [SW19] is that the latter does not make the additional spar-

Algorithm 2 SDT algorithm

```
Input: Data $y$, measurement $\mathcal{A}$, sparsity $s$ and rank $r$ of signal
Initialize $X^0 = 0$.
repeat
    Calculate step-widths $\mu^l$
    $X^{l+1} = P_{\hat{\Omega}^{n,d}_{s,r}}\left(X^l + \operatorname{diag}(\mu^l) P_{\mathcal{T}_{X^l}}\left(\mathcal{A}^\dagger\left(y - \mathcal{A}(X^l)\right)\right)\right)$
until stopping criterion is met at $l = l^*$
Output: Recovered signal $X^{l^*}$
```

sity assumptions on the signal. For this reason, the SDT algorithm differs in the projection $P_{\hat{\Omega}^{n,d}_{s,r}}$ that additionally applies the projection $P_{\Sigma^n_s}$ selecting the s dominant blocks. In fact, in the special case of considering non-sparse signals in $\hat{\Omega}^{n,d}_{n,r}$ the SDT algorithm coincides with the DT algorithm.

3.3.3 Details of the SDT algorithm

To be fully self-contained, let us now go through the individual steps of the SDT algorithm and specify the relevant details. Every iteration of the algorithm starts with the computation of $G^l = \mathcal{A}^\dagger(y - \mathcal{A}(X^l))$, the gradient for the ℓ_2-norm deviation $f(X) = \frac{1}{2}\|y - \mathcal{A}(X)\|^2_{\ell_2}$ evaluated at X^l. The algorithm subsequently employs a modification from Ref. [Wei+16] in calculating the steepest gradient inspired by geometrical optimization techniques which leads to a faster convergence [AMS09; Van13]: The set of rank r matrices is an embedded differential manifold in the linear vector space of all matrices. Thus, a direction on this embedded manifold is characterized by a tangent vector on the manifold. While this geometry straight-forwardly generalizes to the set of $nd \times d$ matrices with rank r blocks, due to sparsity constraint $\hat{\Omega}^{n,d}_{s,r}$ fails to be a differential manifold. Nonetheless, we can make use of tangent vectors as 'natural' search directions in our optimization problem for the non-vanishing blocks of X^l that are conforming with a fixed-rank constraint.

The tangent space of rank r matrices at point x is given by the set of matrices that share the same column or row space x [AMS09]. Correspondingly, the tangent space projection of a non-vanishing block of X can be defined as follows: Let $x_k = U_k \Lambda_k U_k^\dagger$ be the eigenvalue decomposition of the kth block of X with Λ_k the diagonal matrix with eigenvalues in decreasing order. Further, let $U_k^{(r)}$

denote the restriction of U_k to its first r columns corresponding to the range of x_k. Then, the tangent space projection acting on g_k the kth block of G is given by

$$P_{\mathcal{T}_X}(G)_k = g_k - (\mathrm{Id} - P_U) g_k (\mathrm{Id} - P_U), \tag{3.28}$$

with $(P_U)_k = U_k^{(r)} (U_k^{(r)})^\dagger$. The entire tangent-space projection $P_{\mathcal{T}_X}(G)$ is defined by acting trivially on the blocks of G corresponding to vanishing blocks of X and as the projection (3.28) otherwise.

As we prove below in generic situations the SDT algorithm converges for a constant step-width set to $\mu_l = 1$ and even without using the tangent space projection. Empirically, a faster convergence is achieved with the tangent space projection and using the following prescription for the step-width calculation: From the projected gradient $G_P^l = P_{\mathcal{T}_X}(G^l)$ in the lth iteration we then calculate the step width for each block individually as

$$\mu_k^l = \frac{\|(G_P^l)_k\|_\mathrm{F}^2}{\|\mathcal{A}((G_P^l)_k)\|_{\ell_2}^2} \tag{3.29}$$

and multiply each block by the corresponding μ_k^l.

In order to have a compact notation, we introduce the diagonal matrix $\mathrm{diag}(\mu^l) = \mathrm{diag}(\mu_l^1, \dots \mu_1^l, \mu_2^l, \dots, \mu_2^l, \dots, \mu_n^l)$ where each step width is repeated d times. The new state of the algorithm, X^{l+1}, is given by the projection of the result of a gradient step with step width μ^l onto the set $\hat{\Omega}_{s,r}^{n,d}$.

Finally, we have to specify a stopping criterion at which the loop of the algorithm is exited. We terminate the algorithms if the objective function is below a specified threshold, i.e.,

$$\frac{\|y - \mathcal{A}(X^l)\|_{\ell_2}}{\|y\|_{\ell_2}} \leq \gamma_\mathrm{break} \tag{3.30}$$

or a maximal number of iteration is reached. If the data vector y has additive noise, γ_break has to be chosen to be larger than the expected norm of the noise. To be less relying on expectations on the noise levels, one can alternatively make use of criteria on the gradient and step width or test for oscillating patterns in the identified support.

Algorithm 3 ALS-BT algorithm

```
Input: Data $y$, measurement $\mathcal{A}$, sparsity $s$ and rank $r$ of signal
1: Initialize $\rho^0$.
2: repeat
3:     $\xi^l = \arg\min_{\xi \in \Sigma_s^n} f_{\mathrm{ALS}}(\xi, \rho^{l-1})$
4:     $\rho^l = \arg\min_{\rho \in \mathbb{C}_r^{d \times d}} f_{\mathrm{ALS}}(\xi^l, \rho)$
5: until stopping criterion is met at $l = l^*$
Output: Recovered signal $\rho^{l^*}$, $\xi^{l^*}$
```

3.3.4 Blind tomography via alternating optimization

A more direct algorithmic approach to the blind tomography problem is to use a constrained *alternating least square (ALS)* optimization. In ALS optimization, one performs a constrained optimization of the objective function

$$f_{\mathrm{ALS}}(\xi, \rho) = \frac{1}{2} \|y - \mathcal{A}(\xi, \rho)\|_{\ell_2}^2, \tag{3.31}$$

with respect to one of the two variables while regarding the respective other variable as constant in an alternating fashion, see Algorithm 3.

We perform the optimization over Σ_s^n, Algorithm 3 Step 3, using the standard IHT algorithm for sparse vector recovery. Note that calculating the linear measurement map for ξ given a fixed ρ simply involves evaluating all calibration measurement blocks individually, i.e. calculating $\mathcal{A}_i(\rho)$ for all $i \in [N]$. Analogously, the low-rank optimization over $\mathbb{C}_r^{d \times d}$, Algorithm 3 Step 4, can be performed with iterative hard-thresholding on the manifold of low-rank matrices. A detailed description of a suitable algorithmic implementation is given by Algorithm 2 in the special case of a single matrix block, i.e. $N, s = 1$. Computing the corresponding linear map acting on ρ for fixed ξ amounts to summing up the individual measurement blocks weighted by their corresponding calibration coefficient.

The ALS optimization requires an initialization with a suitable ρ^0 in order to evaluate the first objective function for optimizing ξ. One method that we found viable is to randomly draw a rank-r state using Haar-random eigenvectors. Note that, in general, constrained ALS optimization can be highly sensitive to the chosen initialization. For this reason, depending on the measurement map and calibration model, alternative initialization strategies might become necessary. As break-off criteria we can again use a bound on the objective function as in (3.30) and an allowed maximal number of iterations.

3.4 Recovery guarantees

We now prove that for certain simple measurement ensembles, the SDT algorithm converges to the optimal solution before we numerically demonstrate its performance in the following section. More precisely, following the outline of model-based compressed sensing [Bar+10; BD09], the SDT algorithm can be equipped with recovery guarantees based on a *restricted isometry property (RIP)* of the measurement ensemble that is custom-tailored to the structure at hand. Intuitively, it seems clear that a measurement map should at least in principle allow for solving the associated linear inverse problem uniquely if it acts as an isometry on signals from the constraint set. So-called RIP constants formalize this intuition:

Definition 3 ($\hat{\Omega}_{s,r}$-RIP)**.** Given a linear map $\mathcal{A} : \mathbb{C}^{nd^2} \to \mathbb{C}^m$, we denote by $\delta_{s,r}$ the smallest $\delta \geq 0$ such that

$$(1 - \delta) \|x\|_{\mathrm{F}}^2 \leq \|\mathcal{A}(x)\|_{\ell_2}^2 \leq (1 + \delta) \|x\|_{\mathrm{F}}^2 \tag{3.32}$$

for all $x \in \hat{\Omega}_{s,r}$.

The constant $\delta_{s,r}$ measures how much the action of $\mathcal{A}$ when restricted to elements of $\hat{\Omega}_{s,r}$ deviates from that of an isometry. Correspondingly, if $\delta_{s,r}$ is sufficiently small we expect this to be sufficient to ensure that the restricted action of $\mathcal{A}$ becomes invertible. In fact, if a measurement map has a sufficiently small RIP constant one can prove the convergence of projective gradient descent algorithms to the correct solution of the structured linear inverse problem.

Theorem 28 (Recovery guarantee)**.** *Let $\mathcal{A} : \mathbb{C}^{nd \times d} \to \mathbb{C}^m$ be a linear map and suppose that the following RIP condition for $\mathcal{A}$ holds*

$$\delta_{3s,3r} < \frac{1}{2}. \tag{3.33}$$

Then, for $X \in \hat{\Omega}_{s,r}$, the sequence (X^l) defined by the SDT algorithm (Algorithm 2) with $\mu^l = 1$ and $P_{\mathcal{T}_{X^l}} = \mathrm{Id}$ with $y = \mathcal{A}(X)$ satisfies, for any $l \geq 0$,

$$\left\|X^l - X\right\|_{\mathrm{F}} \leq \gamma^l \left\|X^0 - X\right\|_{\mathrm{F}}, \tag{3.34}$$

where $\gamma = 2\delta_{3s,3r} < 1$

We establish that the SDT algorithm converges to the correct solution in Frobenius of the sparse de-mixing problem at a rate that is upper bounded by the RIP constant $\delta_{3s,3r}$ of the measurement map. For the sake of simplicity, we analyse the SDT algorithm omitting the tangent space projection and also assuming a constant step widths $\mu^l = 1$. In numerically simulations we observe that making use of the tangent space projection and a more sophisticated heuristic for the step width yields faster convergence and better recovery performance. Making the stronger assumption that the RIP constant is smaller than $\frac{1}{3}$, one can also show that with high-probability the non-trivial choice of the step-width is close to 1 [SW19]. The right-hand side of the RIP condition (3.33) is not expected to be optimal. Typically, one can at least improve the bound to $\frac{1}{\sqrt{3}}$ with a slightly more refined argument. Since we are interested in the parametric scaling here, we choose to present a simpler argument at the cost of slightly worse constants. Furthermore, the statement of Theorem 28 does not account for statistical noise or potential mild violation of the signal constraints. Following standard techniques, we expect that a more complicated noise- and model-robust version of Theorem 28 can be derived, see e.g. [14]. For the current analysis, we are content with the significantly simpler version.

The derivation of recovery guarantees for the IHT algorithm follows largely the same blueprint developed in the original IHT proposal for sparse vectors [BD08], see also Ref. [FR13] for a detailed description of the proof. Here, we are in addition in the comfortable position that Ref. [SW19] already fleshed out the details of the recovery proof for an IHT algorithm for de-mixing low-rank matrices. Choosing $\mu^l = 1$, we give a slightly simpler proof that carefully adapts the one given in Ref. [SW19] to account for the additional sparsity constraint and uses a slightly more concise notation.

To state the proof of Theorem 28, we introduce a bit more notation. Consider $X \in \Omega_{s,r}^{n,d}$. By definition, it can be written as $X = \sum_{i=1}^{n} \xi_i e_i \otimes x_i$ with $\xi \in \Sigma_s^n$ and $x_i \in \mathcal{D}_r^{(d)}$ for all i. Let Q_i be the projector onto the range of x_i. Furthermore, we set $Q_i = 0$ for all i not in the support of ξ. Slightly overloading our notation, we define the projection of every 'block' onto the range of the corresponding 'block' of X as $\mathcal{P}_{\hat{\Omega}(X)}(Y) := P_{\hat{\Omega}(X)} Y P_{\hat{\Omega}(X)}$ with

$$P_{\hat{\Omega}(X)} := \operatorname{diag}(Q_1, \dots, Q_n). \tag{3.35}$$

Note that the projection simultaneously projects onto the "block-wise support" of X.

It is common and useful to rewrite the RIP inequalities such as in Definition 3 as an equivalent spectral condition of restrictions of $\mathcal{A}^\dagger \mathcal{A}$.

Proposition 29. *Let $X \in \hat{\Omega}^{n,d}_{s,r}$ and $\mathcal{A} : \mathbb{C}^{nd \times d} \to \mathbb{R}^m$ a linear map. Then the following two statements are equivalent:*

(i) $\left\| \mathcal{P}_{\hat{\Omega}(X)} \circ (\operatorname{Id} - \mathcal{A}^\dagger \circ \mathcal{A}) \circ \mathcal{P}_{\hat{\Omega}(X)} \right\|_\infty \leq \delta.$

(ii) For all $Y \in \operatorname{range} \mathcal{P}_{\hat{\Omega}(X)}$ it holds that

$$(1 - \delta) \|Y\|_{\mathrm{F}}^2 \leq \|\mathcal{A}(Y)\|_{\mathrm{F}}^2 \leq (1 + \delta) \|Y\|_{\mathrm{F}}^2 . \tag{3.36}$$

Proof. The inequality

$$\begin{aligned} \delta &\geq \left\| \mathcal{P}_{\hat{\Omega}(X)} \circ (\operatorname{Id} - \mathcal{A}^\dagger \circ \mathcal{A}) \circ \mathcal{P}_{\hat{\Omega}(X)} \right\|_\infty && (3.37) \\ &= \max_{Y \in \operatorname{range} \mathcal{P}_{\hat{\Omega}(X)}} \frac{|\langle Y, (\operatorname{Id} - \mathcal{A}^\dagger \circ \mathcal{A}) Y \rangle|}{\|Y\|_{\mathrm{F}}^2} && (3.38) \end{aligned}$$

holds if and only if for all $Y \in \operatorname{range} \mathcal{P}_{\hat{\Omega}(X)}$

$$\delta \|Y\|_{\mathrm{F}}^2 \geq |\, \|Y\|_{\mathrm{F}}^2 - \|\mathcal{A}(Y)\|_{\mathrm{F}}^2 \,|. \tag{3.39}$$

The last bound is equivalent to (3.36). □

Proof of Theorem 28. Let $X \in \hat{\Omega}^{n,d}_{s,r}$ be the matrix to be recovered. Let X^l denote the lth iterate of the vector of matrices in the SDT algorithm (Algorithm 2). Since the algorithm always involves a projection step onto $\hat{\Omega}^{n,d}_{s,r}$ the lth iterate X^l is in $\hat{\Omega}^{n,d}_{s,r}$. Furthermore, we observe that $X + X^l + X^{l+1} \in \hat{\Omega}^{n,d}_{3s,3r}$. For convenience, we denote the projection onto the ("block-wise") joint range and support of X, X^l and X^{l+1} simply by $\mathcal{P}^l := \mathcal{P}_{\hat{\Omega}(X + X^l + X^{l+1})}$ and its orthogonal complement by $\mathcal{P}^l_\perp$. It is crucial for the proof to bound norm deviations restricted to the range of $\mathcal{P}^l$ as this eventually allows us to apply a RIP bound.

We want to show the convergence of the iterates of the algorithm X^l to the correct solution X. In other words, we want to derive a bound of the form

$$\left\|X^{l+1} - X\right\|_{\mathrm{F}} \leq \gamma \left\|X^l - X\right\|_{\mathrm{F}} \tag{3.40}$$

with constant $\gamma < 1$. Note that by the theorem's assumption we set the step width to $\mu^l = 1$ in the following and omit the tangent space projection $P_{\mathcal{T}_{X^l}}$.

We first derive the following consequence of the thresholding operation: Let $G^l := \mathcal{A}^\dagger(y - \mathcal{A}(X^l)) = \mathcal{A}^\dagger \circ \mathcal{A}(X - X^l)$. By the definition of X^{l+1} as the best approximation to $X^l + G^l$ in $\hat{\Omega}^{n,d}_{s,r}$ it holds that

$$\left\|X^{l+1} - \left[X^l + G^l\right]\right\|_{\mathrm{F}} \leq \left\|X - \left[X^l + G^l\right]\right\|_{\mathrm{F}} . \tag{3.41}$$

Since the parts of both sides of the inequality that are not in the kernel of $\mathcal{P}^l_\perp$ coincides, we get the same inequality also for the with $\mathcal{P}^l$ inserted

$$\left\|X^{l+1} - \left[X^l + \mathcal{P}^l(G^l)\right]\right\|_{\mathrm{F}} \leq \left\|X - \left[X^l + \mathcal{P}^l(G^l)\right]\right\|_{\mathrm{F}} . \tag{3.42}$$

With the help of this inequality, we can bound

$$\begin{aligned} \left\|X^{l+1} - X\right\|_{\mathrm{F}} &\leq \left\|X^{l+1} - \left[X^l + \mathcal{P}^l(G^l)\right]\right\|_{\mathrm{F}} \\ &\quad + \left\|X - \left[X^l + \mathcal{P}^l(G^l)\right]\right\|_{\mathrm{F}} \\ &\leq 2\left\|X - \left[X^l + \mathcal{P}^l(G^l)\right]\right\|_{\mathrm{F}} \\ &= 2\left\|\mathcal{M}_1(X^l - X)\right\|_{\mathrm{F}} \\ &\leq 2\left\|\mathcal{M}_1\right\|_\infty \left\|X^l - X\right\|_{\mathrm{F}} , \end{aligned} \tag{3.43}$$

where in the last step we used the definition of G^l, the fact that $\mathcal{P}^l$ acts trivially on $X^l - X$ and defined $\mathcal{M}_1 := \mathcal{P}^l \circ (\mathrm{Id} - \mathcal{A}^\dagger \circ \mathcal{A}) \circ \mathcal{P}^l$. To arrive at the theorem's assertion, we now bound the spectral norm of $\mathcal{M}_1$ using the RIP property of $\mathcal{A}$ and Proposition 29:

$$\left\|\mathcal{M}_1\right\|_\infty = \left\|\mathcal{P}^l \circ (\mathrm{Id} - \mathcal{A}^\dagger \circ \mathcal{A}) \circ \mathcal{P}^l\right\|_\infty \leq \delta_{3s,3r} \tag{3.44}$$

since the range of $\mathcal{P}^l$ is in $\Omega^{n,d}_{3s,3r}$. Using (3.44) in (3.43) completes the proof. □

The pressing next question is, of course, which measurement ensembles actually exhibit the required RIP. Interestingly, it is notoriously hard to give deterministic constructions of measurement maps that are sample optimal and feature the RIP. In fact, already for the RIP for s-sparse vectors there are no sample optimal deterministic measurement maps known to date [FR13]. To further complicate the state of affairs, it is also known to be NP-hard to check whether a given measurement map exhibits the s-sparse RIP with RIP constant small than a given δ [Ban+13].

For this reason, the field of compressed sensing uses probabilistic constructions to arrive at provably sampling optimal measurement maps. Using a random ensemble of measurement maps of sampling optimal dimension one establishes that with high probability a randomly drawn instance will exhibit the RIP property. In other words, one proves that the originally hard linear inverse problem typically becomes easy for a certain measurement ensemble. Arguably the simplest measurement ensemble consists of observables given by i.i.d. chosen random Gaussian matrices. In our setting a fully Gaussian measurement map can be constructed from a set of $\{A_i \in \mathbb{R}^{nd\times d}\}_{i=1}^m$ of m Gaussian matrices with entries draws i.i.d. from the normal distribution $\mathcal{N}(0,1)$ and defining $y^{(l)} = \mathrm{Tr}(A_i X)$.

As a toy model for quantum tomography it is more natural to consider observables drawn from a random ensemble of Hermitian matrices such as the Gaussian unitary ensemble (GUE). Operationally, we define the GUE by drawing a matrix X with complex Gaussian entries, $X_{k,l} \sim \mathcal{N}(0,1) + \mathrm{i}\mathcal{N}(0,1)$, and subsequently projecting X onto Hermitian matrices using $P_{\square\!\square} : X \mapsto \frac{1}{2}(X + X^\dagger)$. For measurement maps from GUE we prove the following statement:

Theorem 30 ($\hat{\Omega}^{n,d}_{s,r}$-RIP for random Hermitian matrices.). *Let $\{A^{(k)}_i\}_{i=1,k=1}^{n,m}$ be a set of Hermitian matrices drawn i.i.d. from the* GUE. *Let $\mathcal{A}$ be the measurement operator defined by $\{A^{(k)}_i\}_{i=1,k=1}^{n,m}$ via Eqs.* (3.9) *and* (3.11)*. Then $\frac{1}{\sqrt{m}}\mathcal{A}$ satisfies the $\hat{\Omega}^{n,d}_{s,r}$-RIP with parameter $\delta_{s,r}$ with probability at least $1-\tau$ provided that*

$$m \geq \frac{C}{\delta_{s,r}^2}\left[s \ln \frac{\mathrm{e}n}{s} + (2d+1)rs \ln \frac{c}{\delta} + \ln \frac{2}{\tau} \right] \tag{3.45}$$

for sufficiently large numerical constants $C, c > 0$.

Before proving the theorem, we first discuss the implications on the asymptotic scaling of the measurement complexity of our approach to the blind tomography problem and the sparse de-mixing problem based on the results for random

Hermitian measurement maps. First, the derived measurement complexity (3.45) is in accordance with the degrees of freedom of signal $X \in \hat{\Omega}^{n,d}_{s,r}$. The second term of $O(drs)$ corresponds to the number of degrees of freedom specifying the s rank-r matrices of dimension d. The first term of $O(s \ln n)$ is the minimal sampling complexity in s for learning the s non-trivial entries and their support [FR13]. Second, in analogy, we expect the optimal number of measurements for the blind tomography problem, i.e., reconstructing signals in $\Omega^{n,d}_{s,r}$ instead of $\hat{\Omega}^{n,d}_{s,r}$, to scale as $O(s \ln n + dr)$. Hence, having a provably efficient algorithms capable of solving the blind tomography as well as the sparse de-mixing problem comes at the cost of an increase in the sampling complexity by an additional factor of s in the second term of the sampling complexity. Most importantly, invoking the sparsity assumption on the calibration vector ξ allows us to get away without a linear increase n of the number of calibration parameters. Thus, the overhead in measurement complexity of our approach to the blind tomography problem is relatively mild.

In fact, the measurement complexity derived for Gaussian measurements can often be used as a guideline for the sampling complexity of other measurement ensembles that are also sufficiently unstructured. However, the proof techniques for model-based compressed sensing that exploit the combination of different structures are not easily translatable to other measurement ensembles. An exception are measurement ensembles that feature a structure that is sufficiently aligned with the signal structure for hierarchically sparse signals, see e.g. Refs. [13; 21]. We leave the study of more involved measurement ensembles to future work.

It remains to prove Theorem 30. Establishing RIP conditions for Gaussian matrices for a set of structured signals typically proceeds in two steps: One first derives a strong concentration result for a single signal in the set using standard concentration of measure. Second, one takes the union bound over the signal set with the help of an ϵ-covering net construction to arrive at the uniform statement of RIP. We can readily adapt this strategy also to GUE.

For the first step, we derive a Gaussian-type concentration result, modifying a standard line of arguments for our example, see, e.g., Ref. [SW19]. The result is summarized as the following lemma:

Lemma 31 (Gaussian-type concentration)**.** *Let $X \in \hat{\Omega}^{n,d}_{s,r}$. Let $\{A_i^{(k)}\}_{i=1,k=1}^{n,m}$ be a set of Hermitian matrices drawn i.i.d. from the* GUE *and $\mathcal{A}$ be the measurement* *operator defined by* $\{A_i^{(k)}\}_{i=1,k=1}^{n,m}$ *via Eqs.* (3.9) *and* (3.11)*. Then, for* $0 < \delta < 1$

$$(1-\delta)\,\|X\|_{\mathrm{F}}^2 \le \frac{1}{m}\,\|\mathcal{A}(X)\|_{\ell_2}^2 \le (1+\delta)\,\|X\|_{\mathrm{F}}^2 \tag{3.46}$$

with probability of at least $1 - 2\mathrm{e}^{-m\delta^2/C_\delta}$ *and constant* $C_\delta \ge 40$.

Our proof essentially follows the argument of Ref. [SW19] for Gaussian measurements and then exploits that the Hermitian blocks of the signal $X \in \hat{\Omega}_{s,r}^{n,d}$ only overlap with the Hermitian part of the Gaussian measurement matrix.

Proof. Let $X \in \hat{\Omega}_{s,r}^{n,d}$ and denote its n $d \times d$ blocks by x_i. Consider a set $\{B_i^{(k)} \in \mathbb{C}^{d\times d}\}_{k,i=1}^{m,n}$ of $m \cdot n$ $d \times d$ matrices with entries independently drawn from the complex-valued normal distribution. Let $A_i^{(k)} := P_{\Box\!\Box} B_i^{(k)}$ be corresponding matrices drawn from the GUE and $\mathcal{A}$ the corresponding measurement map. Since all blocks x_i are Hermitian, we have

$$\begin{aligned}\mathcal{A}(X)^{(k)} &= \sum_{i=1}^{n} \langle A_i^{(k)}, x_i\rangle = \sum_{i=1}^{n} \langle P_{\Box\!\Box} B_i^{(k)}, x_i\rangle \\ &= \sum_{i=1}^{n} \mathrm{Re}\{\langle B_i^{(k)}, x_i\rangle\},\end{aligned} \tag{3.47}$$

with $\mathrm{Re}\{z\}$ denoting the real part of $z \in \mathbb{C}$. Since all entries of $B_i^{(k)}$ are i.i.d. complex normal random variables and x_i is Hermitian, $\mathrm{Re}\{\langle B_i^{(k)}, x_i\rangle\}$ are i.i.d. real random variables from the distribution $\mathcal{N}(0, \|x_i\|_{\mathrm{F}})$ for all i and k. We conclude that all entries $y_k = \mathcal{A}(X)^{(k)}$ of $\mathcal{A}(X)$ are Gaussian distributed with variance $\sigma^2 = \sum_i \|x_i\|_{\mathrm{F}}^2 = \|X\|_{\mathrm{F}}^2$ and have even moments $\mathbb{E}[{y_k}^{2t}] = 2^{-t} t! \binom{2t}{t} \sigma^{2t}$ [FR13, Corollary 7.7]. Correspondingly, the squared entries are sub-exponential random variables with mean $\mathbb{E}[y_k^2] = \sigma^2$. We denote the associated centred sub-exponential variable as

$$z_k := y_k^2 - \sigma^2. \tag{3.48}$$

The moments of z_k are bounded by

$$\mathbb{E}[|z_k|^t] \le 2^t \mathbb{E}[|y_k|^{2t}] = t! \binom{2t}{t} \sigma^{2t}, \tag{3.49}$$

where the first inequality follows from the triangle and Jensen's inequality. The binomial can be upper bounded using Stirling's formula [FR13, (C.13)] by $\binom{2t}{t} =$

$4^t r_t/\sqrt{\pi t}$ with $r_t \leq \mathrm{e}^{1/(24t)}$. Thus, for $t \geq 2$ we have $\mathbb{E}[|z_k|^t] \leq t! R^{t-2}\Sigma^2/2$ with $R = 4\sigma^2$ and $\Sigma^2 = \sqrt{2/\pi}\mathrm{e}^{1/48}16\sigma^4 \leq 0.815 \cdot 16\sigma^4$. Controlling the moments of z_k for $t \geq 2$, we can apply the Bernstein inequality [FR13, Theorem 7.30] and bound the probability that $\|\mathcal{A}(X)\|_{\ell_2}^2$ varies by more than $\Delta > 0$ from its expectation value

$$\begin{aligned}\mathbb{P}\left[\left|\frac{1}{m}\|\mathcal{A}(X)\|_{\ell_2}^2 - \|X\|_{\mathrm{F}}^2\right| \geq \Delta\right] &= \mathbb{P}\left[\left|\sum_{k=1}^m z_k\right| \geq m\Delta\right] \\ &\leq 2\exp\left[-\frac{m\Delta^2/2}{\Sigma^2 + R\Delta}\right] \\ &\leq 2\exp\left[\frac{-m\Delta^2}{32\|X\|_{\mathrm{F}}^4 + 8\|X\|_{\mathrm{F}}^2\Delta}\right].\end{aligned} \tag{3.50}$$

Let $\Delta = \delta\|X\|_{\mathrm{F}}^2$ for some $0 < \delta < 1$. Then we can rewrite the tail bound (3.50) as

$$\mathbb{P}\left[\left|\frac{1}{m}\|\mathcal{A}(X)\|_{\ell_2}^2 - \|X\|_{\mathrm{F}}^2\right| \geq \delta\|X\|_{\mathrm{F}}^2\right] \leq 2\exp\left[-\frac{m\delta^2}{C_\delta}\right] \tag{3.51}$$

with a constant $C_\delta \geq 40$. Hence, the condition

$$(1-\delta)\|X\|_{\mathrm{F}}^2 \leq \frac{1}{m}\|\mathcal{A}(X)\|_{\ell_2}^2 \leq (1+\delta)\|X\|_{\mathrm{F}}^2 \tag{3.52}$$

holds with probability at least $1 - 2\mathrm{e}^{-m\delta^2/C_\delta}$. □

Note that by the homogeneity of the RIP condition it suffices to restrict ourselves to normalized elements of $\hat{\Omega}_{s,r}^{n,d}$ in the proof of Theorem 30. In the following, we will therefore focus on the set

$$\bar{\Omega}_{s,r}^{n,d} := \{X \in \hat{\Omega}_{s,r}^{n,d} \mid \|X\|_{\mathrm{F}}^2 = 1\}. \tag{3.53}$$

To take a union bound over the set $\bar{\Omega}_{s,r}^{n,d}$ we need to bound the size of an ϵ-net that covers the set $\bar{\Omega}_{s,r}^{n,d}$. An ϵ-net $\mathcal{C}$ covering a set of matrices $\mathcal{M} \subset \mathbb{C}^{nd\times d}$ is a finite subset of $\mathcal{M}$ such that for all $X \in \mathcal{M}$ there exists $\bar{X} \in \mathcal{C}$ such that $\left\|X - \bar{X}\right\|_{\mathrm{F}} \leq \epsilon$. Our construction generalizes the construction of Ref. [SW19]. Therein, a covering net for the set of normalized block-wise low-rank matrices $\bar{\Omega}_{n,r}^{n,d}$ was derived. We summarize the statement given in Ref. [SW19] in the following lemma without giving a proof.

Lemma 32 ([SW19]). *For $\bar{\Omega}_{n,r}^{n,d}$ there exists an ϵ-covering net $\mathcal{C}_r^{n,d}$ with cardinality bounded by $(9/\epsilon)^{(2d+1)nr}$.*

The proof of Lemma 32 basically lifts the result of an ϵ-net for low-rank matrices of Ref. [CP11] to the set $\bar{\Omega}_{n,r}^{n,d}$ using the triangle inequality.

We can combine multiple ϵ-nets for $\bar{\Omega}_{s,r}^{s,d}$ to construct an ϵ-covering net for the set $\bar{\Omega}_{s,r}^{n,d}$ of block-sparse matrix vectors with low-rank blocks. The bound on the cardinality of the resulting ϵ-covering net is given in the following lemma:

Lemma 33 (Bound on the cardinality of a covering net). *For $\bar{\Omega}_{s,r}^{n,d}$ there exists an ϵ-covering net $\mathcal{C}_{s,r}^{n,d}$ of cardinality bounded by $\binom{n}{s}(9/\epsilon)^{(2d+1)sr}$. Furthermore, for each $X = [X_1, \dots, X_n] \in \bar{\Omega}_{s,r}^{n,d}$ there exists $\bar{X} = [\bar{X}_1, \dots, \bar{X}_n] \in \mathcal{C}_{s,r}^{n,d}$ such that $\left\|X - \bar{X}\right\|_{\mathrm{F}} \le \epsilon$ and $\left\|\bar{X}_k\right\|_{\mathrm{F}} = 0$ for all k for which $\left\|X_k\right\|_{\mathrm{F}} = 0$.*

Proof. Let $\Gamma \subset [n]$ with $|\Gamma| \le s$, i.e., the indices of the support of an s-sparse vector. The set

$$\bar{\Omega}_r^{\Gamma} := \left\{ \sum_{i\in\Gamma} \xi_i e_i \otimes x_i \;\middle|\; \xi_i \in \mathbb{R},\ x_i \in \mathcal{D}_r^{(d)}\ \forall i \right\} \subset \bar{\Omega}_{s,r}^{n,d} \tag{3.54}$$

shall consist of all elements of $\bar{\Omega}_{s,r}^{n,d}$ which have non-vanishing blocks only supported on Γ. To each element of $\bar{\Omega}_r^{\Gamma}$, we can associate an element of $\bar{\Omega}_{s,r}^{s,d}$ by omitting the vanishing blocks in the matrix vector and vice versa. By virtue of Lemma 32 we thus know that $\bar{\Omega}_r^{\Gamma}$ has a covering net $\mathcal{C}_r^{\Gamma}$ of cardinality bounded by $(9/\epsilon)^{(2d+1)sr}$.

We can decompose the entire set $\bar{\Omega}_{s,r}^{n,d}$ as

$$\bar{\Omega}_{s,r}^{n,d} = \bigcup_{\Gamma\subset[n],|\Gamma|\le s} \bar{\Omega}_r^{\Gamma}, \tag{3.55}$$

and thus, the set

$$\mathcal{C}_{s,r}^{n,d} = \bigcup_{\Gamma\subset[n],|\Gamma|\le s} \mathcal{C}_r^{\Gamma} \tag{3.56}$$

is an ϵ-covering net for $\bar{\Omega}_{s,r}^{n,d}$. The union is taken over $\binom{n}{s}$ different sets. Thus, the cardinality of $\mathcal{C}_{s,r}^{n,d}$ is upper bounded by $\binom{n}{s}(9/\epsilon)^{(2d+1)sr}$. The second statement follows by construction. □

We are now in the position to prove Theorem 30.

Proof of Theorem 30. The proof proceeds in two steps. First, we prove the RIP for elements of the ϵ-covering net $\mathcal{C}_{s,r}^{n,d}$ of $\bar{\Omega}_{s,r}^{n,d}$. To do so, we combine the concentration result of Lemma 31 and the union bound of Lemma 33 to establish uniform concentration. In a second step, following Ref. [SW19], we then use the definition of an ϵ-covering net to show that for elements $X \in \bar{\Omega}_{s,r}^{n,d}$ that are close enough to an element of the net, the RIP condition still holds.

Step 1: Taking the union bound over the ϵ-net $\mathcal{S}_{s,r}^{n,d}$ constructed in Lemma 33 and using the result of Lemma 31 in the form of (3.51) with constant $C_\delta \geq 40$ we get

$$\begin{aligned}\mathbb{P}\left(\max_{X\in\mathcal{S}_{s,r}^{n,d}}\left|\frac{1}{m}\|\mathcal{A}(X)\|_{\ell_2}^2 - \|X\|_{\mathrm{F}}^2\right| \quad \geq \delta/2\right)\\ \leq 2|\mathcal{S}_{s,r}^{n,d}|\mathrm{e}^{-m\delta^2/(4C_\delta)} \\ \leq 2\binom{n}{s}\left(\frac{9}{\epsilon}\right)^{(2d+1)sr} e^{-m\delta^2/(4C_\delta)}.\end{aligned} \tag{3.57}$$

The aim is to find a lower bound for the number of measurements m for which the probability (3.57) small. To this end, we rewrite

$$\begin{aligned}2\binom{n}{s}\left(\frac{9}{\epsilon}\right)^{(2d+1)sr} e^{-m\delta^2/(4C_\delta)}\\ \leq 2\exp\left[s\ln\frac{\mathrm{e}n}{s} + (2d+1)sr\ln\frac{9}{\epsilon} - \frac{m\delta^2}{4C_\delta}\right] \\ \leq \tau,\end{aligned} \tag{3.58}$$

using $\binom{n}{s} \leq \left(\frac{en}{s}\right)^s$ [FR13, Lemma C.5]. The latter inequality becomes true under the condition that

$$m \geq \frac{4C_\delta}{\delta^2}\left[s\ln\frac{\mathrm{e}n}{s} + (2d+1)sr\ln\frac{9}{\epsilon} + \ln\frac{2}{\tau}\right]. \tag{3.59}$$

Assuming that (3.59) holds, we have established the RIP condition for the ϵ-net $\mathcal{C}_{s,r}^{n,d}$, i.e., for all vectors $\overline{X} \in \mathcal{C}_{s,r}^{n,d}$ it holds that

$$(1-\delta/2)\left\|\overline{X}\right\|_{\mathrm{F}}^2 \leq \|\mathcal{A}(\overline{X})\|_{\ell_2}^2 \leq (1+\delta/2)\left\|\overline{X}\right\|_{\mathrm{F}}^2 \tag{3.60}$$

with probability at least $1-\tau$.

Step 2: Let us now transfer the RIP of $\mathcal{C}_{s,r}^{n,d}$ to the entire set $\bar{\Omega}_{s,r}^{n,d}$ while keeping the error under control. To this end, we choose the net parameter ϵ as $\frac{\delta}{4\sqrt{2}}$. By definition of an ϵ-net, for elements $X \in \bar{\Omega}_{s,r}^{n,d}$, there exists an element $\overline{X} \in \mathcal{C}_{s,r}^{n,d}$ such that

$$\left\|X - \overline{X}\right\|_{\mathrm{F}} \leq \frac{\delta}{4\sqrt{2}}. \tag{3.61}$$

To prove the RIP for the set $\bar{\Omega}_{s,r}^{n,d}$ we need to bound $\|\mathcal{A}(X)\|_{\mathrm{F}}$ from above and below.

We start with the upper bound, making use of Eq. (3.60):

$$\begin{aligned}\|\mathcal{A}(X)\|_{\ell_2} &\leq \|\mathcal{A}(\overline{X})\|_{\ell_2} + \|\mathcal{A}(X - \overline{X})\|_{\ell_2} \\ &\leq 1 + \frac{\delta}{2} + \|\mathcal{A}(X - \overline{X})\|_{\ell_2}.\end{aligned} \tag{3.62}$$

Now $\|\mathcal{A}(X - \overline{X})\|_{\ell_2}$ has to be bounded from above. We use that by the second statement of Lemma 33 the block supports of X and $\overline{X}$ coincide. Therefore, $X - \overline{X}$ has also s non-vanishing blocks that have rank of at most $2r$. We can, thus, decompose $X - \overline{X} = B + C$ in terms of orthogonal matrices $B, C \in \hat{\Omega}_{s,r}^{n,d}$ that obey $\langle B, C\rangle = 0$. In particular, B and C have the same block support as X. Let us define

$$\kappa_{s,r} := \sup_{X \in \bar{\Omega}_{s,r}^{n,d}} \|\mathcal{A}(X)\|_{\ell_2}. \tag{3.63}$$

Then we get using homogeneity

$$\begin{aligned}\|\mathcal{A}(X - \overline{X})\|_{\ell_2} &\leq \|\mathcal{A}(B)\|_{\ell_2} + \|\mathcal{A}(C)\|_{\ell_2} \\ &\leq \kappa_{s,r}(\|B\|_{\mathrm{F}} + \|C\|_{\mathrm{F}}) \leq \sqrt{2}\,\kappa_{s,r}\sqrt{\|B\|_{\mathrm{F}}^2 + \|C\|_{\mathrm{F}}^2} \\ &= \sqrt{2}\,\kappa_{s,r}\left\|X - \bar{X}\right\|_{\mathrm{F}},\end{aligned} \tag{3.64}$$

where the last step makes use of the orthogonality of B and C. Together with (3.61) it follows that

$$\|\mathcal{A}(X - \overline{X})\|_{\ell_2} \leq \frac{\delta \cdot \kappa_{s,r}}{4}. \tag{3.65}$$

It remains to derive an upper bound for $\kappa_{s,r}$. To this end, we use that, by definition, $\kappa_{s,r}$ is the best upper bound of the left-hand side of (3.62). Inserting (3.65) into the right-hand side of (3.62), we find the condition

$$\kappa_{s,r} \leq 1 + \frac{\delta}{2} + \frac{\delta \cdot \kappa_{s,r}}{4}. \tag{3.66}$$

Solving for $\kappa_{s,r}$, Eq. (3.66) implies for $0 < \delta < 1$

$$\kappa_{s,r} \leq \frac{1 + \delta/2}{1 - \delta/4} \leq 1 + \delta. \tag{3.67}$$

Altogether, this yields the desired upper bound

$$\|\mathcal{A}(X)\|_{\ell_2} \leq 1 + \frac{3}{4}\delta + \frac{\delta^2}{4} \leq 1 + \delta, \tag{3.68}$$

for $\delta < 1$. The lower bound is analogously obtained by combining the inequality

$$\begin{aligned} \|\mathcal{A}(X)\|_{\ell_2} &\geq \|\mathcal{A}(\overline{X})\|_{\ell_2} - \|\mathcal{A}(X - \overline{X})\|_{\ell_2} &(3.69)\\ &\geq 1 - \delta/2 - \|\mathcal{A}(X - \overline{X})\|_{\ell_2} &(3.70)\end{aligned}$$

with (3.65) (3.67) to arrive at

$$\|\mathcal{A}(X)\|_{\ell_2} \geq 1 - \delta/2 - \delta(1+\delta)/4 \geq 1 - \delta. \tag{3.71}$$

With the choice of ϵ, we can rewrite the condition (3.59) on m as

$$m \geq \frac{C}{\delta^2}\left[s \ln \frac{en}{s} + (2d+1)sr \ln \frac{c}{\delta} + \ln \frac{2}{\tau}\right] \tag{3.72}$$

with constants $C \geq 4C_\delta \geq 160$ and $c \geq 36\sqrt{2} \geq 51$. This completes the proof. □

3.5 Numerical results

The analytical results of the previous section provide worst-case bounds on the asymptotic scaling for a class of idealized, unstructured measurements. In order to benchmark and assess the non-asymptotic performance of compressed sensing algorithms in practice, however, numerical simulations are indispensable. In a first step we therefore perform numerical simulations for the idealized measurement model as given by random GUE matrices, comparing the performance of our algorithm to related established algorithms that do not entirely exploit the structure of the problem. In a second step, we compare the SDT algorithm 2 with standard CS tomography in a blind tomography setting involving measurements of Pauli correlators, cf. (3.3). To do so we randomly draw subsets of the possible Pauli measurements as possible calibrations $\mathcal{A}_i$ of the measurement apparatus. Finally, we demonstrate the feasibility of blind tomography under structure assumptions in the realistic measurement and calibration setting involving single-qubit coherent errors described in Section 3.1. To this end, we employ the Algorithm 3 that performs alternating constrained optimization. The algorithms and the scripts producing the plots have been implemented in Python and are available under Ref. [24].

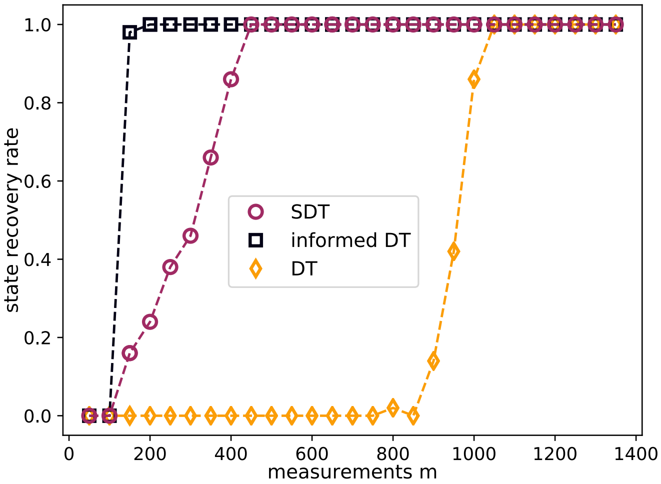


Figure 3.2: The figure displays the recovery rate for the SDT, DT and informed DT algorithm for different number of observables m for GUE measurements. Each point is averaged over 50 random measurement and signal instances with $r = 1$, $d = 16$, $n = 10$ and $s = 3$. A signal is considered successfully recovered if its Frobenius norm deviation from the original signal is smaller than 10^{-3}. One observes nearly coinciding recovery performances for the informed DT and the SDT algorithm. In comparison, the DT algorithm requires significantly more observables for recovery.

3.5.1 GUE measurements

The SDT algorithm goes beyond existing IHT algorithms for the de-mixing problem of low-rank matrices in that it additionally allows one to exploit a sparse mixture. We demonstrate that this yields a drastic and practically important improvement in the number of measurement required for the reconstruction.

To this end, we draw signal instances $X = \xi \otimes \rho$ at random from $\Omega^{n,d}_{s,r}$. We use four qubit pure states $\rho = |\psi\rangle\langle\psi|$ with $r = 1$ and $d = 16$, where $|\psi\rangle$ is drawn uniformly (Haar) random from the complex ℓ_2-norm sphere. The calibration vector $\xi \in \mathbb{R}^n$ with $n = 10$ has a support of size $s = 3$ drawn uniformly from the set of all $\binom{n}{s}$ possible supports. The non-vanishing entries of ξ are normal distributed with unit variance. The measurements are drawn at random from the GUE ensemble as defined above with a varying number of observables m.

The closest competitor to the SDT algorithm is the algorithm of Ref. [SW19]. The algorithm of Ref. [SW19] coincides with the special case of the SDT algorithm where we use the projection on to $\hat{\Omega}^{n,d}_{n,r}$ with $s = n$ ignoring the sparsity in the block structure. We will refer to this algorithm as the *DT algorithm*. We can also give the DT algorithm the 'unfair' advantage of restricting the problem to the correct block support of the signal from the beginning. We will refer to this variant as the *informed DT algorithm*.

Figure 3.2 shows the recovery rate for the SDT algorithm, the DT algorithm and its informed variant for different m. Each point is average over 50 random signal and measurement instances. We consider a signal as successfully recovered if the distance of the algorithm's output to the original signal is smaller than 10^{-3} in Frobenius norm. The algorithm terminated if either the stopping criterion (3.30) with $\gamma_{\text{break}} = 10^{-5}$ is met or after a maximal number of 600 iteration. We observe that if one of the algorithm successfully recovers a signal it typically meets the stopping criterion after less than 100 iterations.

The curves for all three algorithm in Figure 3.2 display a sharp phase transition from a regime where the number of measurement is too small to recover any signal to a regime of reliable recovery. While the phase transition for the SDT algorithm appears in a similar regime to the informed DT algorithm, the DT algorithm requires considerably more samples in order to recover the signal instances.

We conclude that the sparsity of the calibration parameters can be exploited by the SDT algorithm to considerably reduce the required number of measurements. Even more so, this does not require many more sampling points as compared to an algorithm which is given *a priori* knowledge which errors were present, that is, the block support of the signal. This shows that the SDT algorithm solves the de-mixing and blind tomography task in a highly efficient way and scalable. Finally, the number of possible erroneous measurements $\mathcal{A}_i$ can be scaled up at a very low cost in terms of required measurement settings.

3.5.2 Sub-sampled random Pauli measurements

For the application in characterizing quantum devices, it is key to compare the recovery performance of the SDT algorithm with standard low-rank quantum tomography algorithms. To this end note that the SDT algorithm restricted to $n, s = 1$ is also a state-of-the-art algorithm for standard low-rank state tomography without the on-the-fly calibration. Thus, we will make use of this implementation of conventional low-rank state tomography in the following.

We draw signal instances as before but using three-qubit states, $s \in \{3, 4\}$ and altering the model for the calibration parameter: We set the first entry of ξ to $\xi_0 = 1$. The support of the remaining entries is drawn uniformly at random. The

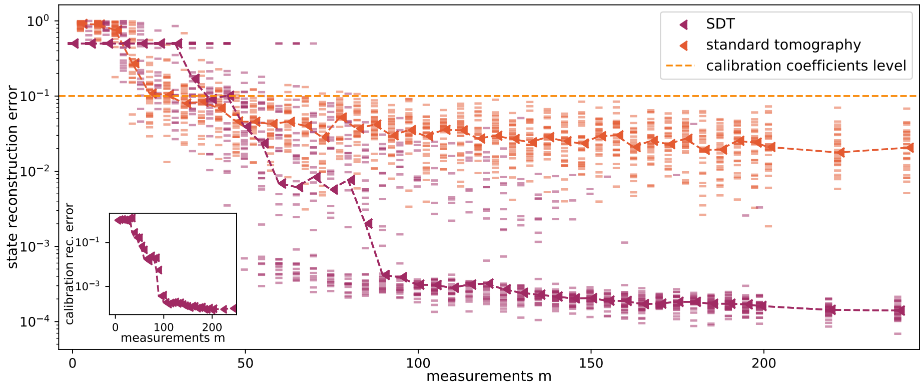


Figure 3.3: The figure displays the reconstruction error in trace distance (2.8) for the SDT compared to the standard tomography algorithm for different number of observables m for sub-sampled random Pauli measurements. Each point is averaged over 30 random measurement and signal instances with $r = 1$, $d = 8$, $n = 10$ and $s = 3$. The inline figure shows the mean ℓ_2-norm reconstruction error of the calibration coefficients for the SDT algorithm.

non-vanishing entries are then i.i.d. taken from the normal distribution rescaled by a factor of $1/10$. This mimics a setting where we have a dominant target measurement and a couple of small systematic deviation from a known set of candidates. The target measurements as well as the systematic deviations are uniformly sub-sampled Pauli observables. Thus, $\mathcal{A}_0$ till $\mathcal{A}_n$ have the form of (3.3) with differently i.i.d. selected Pauli observables uniformly selected from $\{\mathrm{Id}, X, Y, Z\}$. We simulate statistical noise using 10^8 samples per expectation value in order to realistically limit the resolution of the SDT algorithm.

We simultaneously perform recoveries with the SDT algorithm using the entire measurement matrix including the calibration measurement components and the SDT algorithm using only the target measurement $\mathcal{A}_0$ as in a conventional tomography setting.

The resulting trace distance of the state estimate, i.e., the trace-normalized first block of X, from the original ρ is shown for different number of measurements in Figure 3.3 and Figure 3.4 for different sparsity $s = 3$ and $s = 4$, respectively. The curves indicate the median over the depicted 30 sample points per value of m. The inline plot of boths Figures further show the ℓ_2-norm deviation of the reconstructed calibration parameters and the original ξ.

One observes that the conventional low-rank tomography becomes more accurate with an increasing number of measurement but is asymptotically still

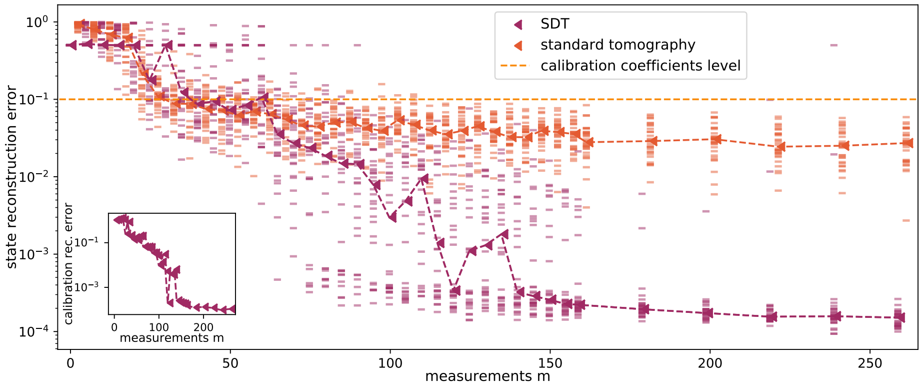


Figure 3.4: The figure displays the reconstruction error in trace distance (2.8) for the SDT compared to the standard tomography algorithm for different number of observables m for sub-sampled random Pauli measurements. Each point is averaged over 30 random measurement and signal instances with $r = 1$, $d = 8$, $n = 10$ and $s = 4$. The inline figure shows the mean ℓ_2-norm reconstruction error of the calibration coefficients for the SDT algorithm.

bounded from below by the systematic error induced by the calibration on the order of 10^{-1}. This agrees with the order of magnitude of variance of the calibration coefficients. In contrast, the SDT algorithm while performing slightly worse in a regime of insufficient measurements outperforms the conventional algorithm for a moderate number of samples and is ultimately only limited by the statistical noise. However, in the parameter regime under investigation there are even for large number of samples $m > 150$ a small number (well below 10%) of instance where SDT only reaches an accuracy comparable to standard tomography. In these instances we find that the support for the calibration measurement components was incorrectly identified. For $s = 4$ we furthermore observe one pathological instance of SDT for $m = 240$ that is worse in recovery than standard tomography is in this regime. For $s = 4$ the phase transition of SDT appears for a slightly larger values of m compared to $s = 3$. The curves for the reconstruction error of the quantum state approximately coincide with the curves for the error in the calibration parameter. We conclude that for a sufficient number of measurement settings, the SDT algorithm almost always performs a significantly more accurate state reconstruction and simultaneously extracts the calibration parameters. The precision is ultimately only limited by the statistical error in the estimation of the expectation values.

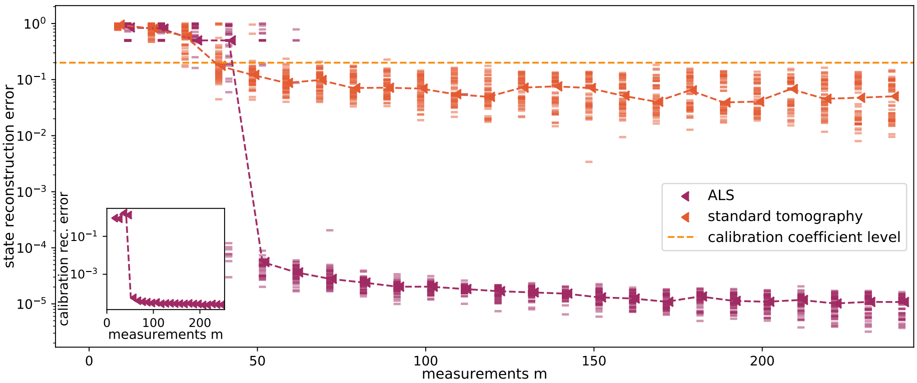


Figure 3.5: The figure displays the reconstruction error in trace distance (2.8) for the ALS compared to the standard tomography algorithm for different number of observables m for Pauli measurements with coherent single-qubit errors. Each point is averaged over 50 random measurement and signal instances with $r = 1$, $d = 16$, $n = 7$ and $s = 2$. The inline figure shows the mean ℓ_2-norm reconstruction error of the calibration coefficients for the SDT algorithm.

3.5.3 Pauli measurements with coherent single-qubit errors

We now come back to the concrete realistic scenario described in Section 3.1. There we derived the calibration measurement model originating from coherent errors in the gates that implement the single-qubit measurements.

For the numerical simulations, we draw a set of m Pauli observables uniformly at random as the target measurement. Subsequently, we introduce six calibration blocks such that every observable in the set $\{X, Y, Z\}$ is swapped with another Pauli observable in $\{X, Y, Z\}$ in a specific block. We generate data y for given states and calibration parameters using the linear calibration measurement model without noise as induced by finite statistics.

We find that in the parameter regimes that are easily amenable to numerical studies on desktop hardware the SDT algorithm is not capable of successfully reconstructing the states when the calibration parameters for the corrections are considerably smaller than the leading order measurement. To thoroughly understand this limitation, in the following, we briefly report the performance of the SDT algorithm on different sub-tasks related to the recovery problem.

First, we choose $d = 16$ and $n = s = 1$ such that only a single block, either the ideal measurement or one of the correction blocks, is used to generate the signal from a random pure state ($r = 1$). We observe that the SDT algorithm is able

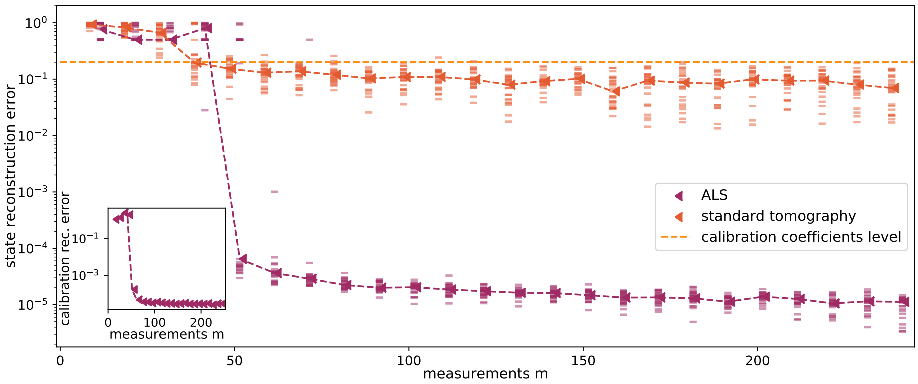


Figure 3.6: The figure displays the reconstruction error in trace distance (2.8) for the ALS compared to the standard tomography algorithm for different number of observables m for Pauli measurements with coherent single-qubit errors. Each point is averaged over 30 random measurement and signal instances with $r = 1$, $d = 16$, $n = 7$ and $s = 3$. The inline figure shows the mean ℓ_2-norm reconstruction error of the calibration coefficients for the SDT algorithm.

to recover the signals in this standard tomography problem. This indicates that also the calibration blocks individually allow for tomographic reconstruction of low-rank states. Second, the SDT algorithm can discriminate between different mixtures of the six correction blocks. To demonstrate this, we ignore the ideal measurement and employ only the correction blocks to generate the signal. We set the active calibration coefficients to one. Thus, $n = 6$, $s \leq n$ and $\xi_i = 1$ for i active. We observe that given a sufficient number of measurement settings the SDT algorithm correctly reconstructs pure states in this measurement setting. The same findings hold true if the target measurement is again considered as long as the active calibration coefficients are set to 1. We observe successful reconstructions of unit rank states for $n = 7$ and $s \in 1, 2, 3$.

A more natural setting however would typically have calibration coefficients that are considerably smaller than the ideal measurement. This justifies the linear expansion for the measurement model in the first place. If we choose, e.g., $\xi_i = 1/10$ for the indices i of active blocks, we were unable to identify a parameter regime on desktop hardware where the SDT algorithm can successfully recover the majority of instances of pure states. We observe that if the SDT algorithm settles on an objective variable with an incorrect block support in the first few iterations, it is not able to subsequently run into objective variables with a different block support in most instances.

Despite the negative result for the SDT algorithm in the most realistic setting, the general mindset to exploit structure (low-rankness) to allow quantum state tomography in a blind fashion is fruitful using a slightly different algorithmic strategy.

To this end, we use the constrained *alternating least square* (ALS) algorithm described in Section 3.3.4. We set the first calibration coefficient corresponding to the ideal measurement to one. The support of the remaining active calibration coefficients is drawn uniformly at random and their value are i.i.d. drawn from a shifted normal distribution with standard deviation 0.05 and mean value 0.2. We use Haar random pure states, $r = 1$ of a four-qubit system, $d = 16$, as the target states.

The algorithm is initialized with a Haar-randomly drawn pure state. We allow for a maximal number of 1000 iterations of the algorithm or terminate if the criterion (3.30) with $\gamma_{\text{break}} = 10^{-5}$ is met. Furthermore, if the stopping criterion is not met after 50 iterations, we re-initialize the algorithm with a new random pure state. We allowed for a maximal number of 10 or 20 re-initializations for $s = 2$ and $s = 3$, respectively. We observe that in case of successful recovery typically at most 3 re-initializations are required with most instances already correctly converging from the initial state.

As in the previous section, we compare the recovery performance of the ALS with the standard low-rank tomography algorithm. The trace-norm error and calibration error for different numbers of measurement settings for $s = 2$ and $s = 3$ are displayed in Figure 3.5 and 3.6, respectively. We observe that, as expected, the reconstruction error of standard low-rank tomography is again lower-bounded by a scale set by the magnitude of the calibration parameters. In contrast, with an only slightly larger number of measurement settings, the constrained ALS algorithm is capable of recovering the states and the calibration parameter with an accuracy that is improved by orders of magnitude and in the noiseless scenario only limited by the algorithms stopping criterion. Compared to recovery performance of the SDT algorithm we observe an even sharper phase transition to the regime of recovery.

3.6 Conclusions and outlook

In this chapter, we have shown that the natural assumption of low-rankness allows one to perform self-calibrating quantum state tomography. By relaxing the blind tomography problem to a sparse de-mixing problem, we have developed an efficient classical post-processing algorithm, the SDT algorithm, that is theoretically guaranteed to recover both the quantum state and the device calibration under a restricted isometry condition of the measurement model. We have demonstrated the necessity of relaxing the blind tomography problem within the framework of hard-thresholding algorithms by establishing the NP-hardness of the projection onto the set consisting of the outer products of vectors and fixed-rank matrices. Introducing a sparsity assumption on the calibration coefficients ensures that the reconstruction scheme can already be applied for fairly small system dimension. We have explicitly proven that a Gaussian random measurement model meets the required restricted isometry condition with a close-to-optimal measurement complexity in $O(s \ln n + drs)$. Furthermore, we have numerically demonstrated an improved performance of the SDT algorithm for random instances of measurement models compared to previously proposed non-sparse de-mixing algorithms and standard low-rank state tomography. While these generic measurement and calibration models enabled us to derive analytical guarantees, it is fair to argue that these models might at best capture some aspects of actual experimental implementations. A potential starting point for extending recovery guarantees to more realistic settings is the generalization of our results to random Pauli measurements as considered in Section 3.5 [Liu11] together with the coherence measures and structured measurement guarantees developed in the context of hierarchically spares signals [Spr+11; 12; 14; 13].

To complement our conceptually and rigorously minded work with a more pragmatic approach, we have additionally developed and implemented a structure-exploiting blind tomography algorithm based on alternating optimization. We have numerically demonstrated that the alternating algorithm is able to perform self-calibrating low-rank tomography in a measurement and calibration model that is well-motivated by gate implementations in ion traps. These numerical simulations indicate that the approach to the blind tomography problem developed here might be well-suited to improve tomographic diagnostics in current experiments. Ultimately, the recovery performance of the proposed algorithms has to be evaluated on experimental data.

Related work in signal processing. Recovery problems of the form (3.1) or the related de-mixing problem (3.2) also arise in other disciplines. For example, these problems appear in future mobile communication scenarios with the promise to yield much more scalable protocols with respect to the number of served devices [Wun+15]. Very concretely, our work can be directly applied in order to extend the *internet-of-things setup* described in Ref. [SW19] if one additionally wants to exploit the sporadic (sparse) user activity of machine-type messaging. Furthermore, our work identifies yet another set of hierarchical signal structures that allow for an efficient projection: It in this way also extends our work on compressed sensing with hierarchically sparse signals to low-rank matrices [12; 13; 14; 15; 16; 17; 18; 19; 20; 21].

4 Random Clifford designs and structured random ensembles

In this chapter, we take a look at random ensembles of unitaries and states that originate from uniform measures of subsets of the unitary groups. We derive results on the moments of random Clifford unitaries that will be essential for the proofs of the main theorem of Chapter 6 and collect multiple results that provide some further perspective on the topic.

In the preliminaries we have introduced a machinery to calculate the moments of random variables that are polynomials of Haar random unitaries. In particular, these results also apply to measures on the unitary group that can be shown to constitute a unitary k-design of an order k larger than the degree of the polynomial of interest. This is nothing but the definition of a unitary k-design. Especially, low-order k-designs with $k = 1$ till 4 feature prominently in quantum characterization: A state's orbit under a 1-design forms a POVM. In Chapter 5, we review that 2-designs give rise to particularly simple randomized benchmarking protocols. Shadow fidelity estimation [HKP20; 4] and filtered randomized benchmarking [9] is particularly efficient using a unitary 3-designs. And our approach to compressive randomized benchmarking tomography developed in Chapter 6 requires a unitary 4-design, at least at first sight.

These applications motivate finding examples of low-order unitary designs and studying exact and approximate constructions. The fruitful approach to identify exact unitary designs is to investigate the uniform measure with support only on a subgroup of the unitary group. For subgroups, the design properties can be established by comparing their irreducible decompositions of tensor power representations with the Schur-Weyl decomposition of the unitary group. A full classification of k-groups, i.e. finite groups that constitute a k-design, was recently derived and summarized by Bannai *et al.* in Ref. [Ban+20]. Combining this classification with a theorem about the universality of finitely generated subgroups by Sawicki and Karnas [SK17], yields a surprisingly simple picture: For the scalable application in quantum computing independent of the system's dimensions, we are interested in what we call a *finitely generated, scalable family of k-groups*. To be precise, consider a finite gate set $\mathcal{G} \subset \mathrm{SU}((\mathbb{C}^{\otimes q})^{\otimes r}$. For each $n \geq r$ it induces a subgroup $\mathcal{G}_n \subset \mathrm{SU}((\mathbb{C}^q)^{\otimes n}$ that is generated by acting with the elements of $\mathcal{G}$ on any r tensor components of $(\mathbb{C}^q)^{\otimes n}$. We call such a family $(\mathcal{G}_n)_{n\in\mathbb{N}}$ a *finitely generated, scalable family of k-groups* if there exists n_0 such that $\mathcal{G}_n$ is a k-group for all $n \geq n_0$. In fact, (i) the only finitely generated, scalable families of 2-groups are isomorphic to a subgroup of the Clifford group $\mathrm{Cl}(q^n)$ with prime q or are dense in $\mathrm{SU}(q^n)$.[1] The picture is even simpler for 3-designs. (ii) The only finitely generated, scalable family of 3-groups is isomorphic to the multi-qubit Clifford group $\mathrm{Cl}(2^n)$ (or dense in $\mathrm{SU}(q^n)$). Finally, (iii) there exist no finite unitary k-groups for $k \geq 4$ and dimension $d > 2$. We refer to Ref. [6, Section V] for a more careful formulation of the three statements and their arguments. Here we summarize that the Clifford group and its subgroups are ‘basically’ the only 2-groups and 3-group ‘within’ the special unitaries.[2]

In this sense, the Clifford group is as close as we get to a 4-design with a group structure. This limits the possibilities of implementing practical quantum computing applications that require 4-designs. Conversely, we might ask how far we can already get with Clifford t-designs, i.e. ensembles that reproduce the moments of the Clifford group instead of the moments of the full unitary group for $t \geq 4$. In Chapter 6, we show that compressive randomized benchmarking tomography already works with a Clifford 4-design. The main ingredient to the proof is an explicit formula for the integration of the fourth tensor power of the Clifford group. In the following section we derive this expression.

Additionally, we provide a broader perspective on the topic of Clifford designs in explaining two further results in this chapter: First, we ask how random quantum circuits consisting of an increasing number of local Clifford unitaries converge towards being random multi-qubit Clifford unitaries that constitute Clifford designs. Such a random circuit construction is of interest, inter alia, for more direct approaches to randomized benchmarking as explained in Chapter 5. Second, we answer a more conceptional question: Instead of asking if a Clifford 4-design suffices for a certain application of unitary 4-designs, we can ask the complementary question. How many non-Clifford resources do we need

[1] Note that we here refer (without definition) to the generalization of the qubit Clifford group defined for arbitrary prime-power dimension. And strictly speaking, we also assume that all $\mathcal{G}_n$ are either finite or infinite for the statement as presented here.

[2] For completeness, we mention that in specific dimensions other 2- and 3-groups exist. We refer to Ref. [Ban+20] for a complete list of such exceptional 2- and 3-groups existing only in fixed dimensions. Furthermore, there exist two types of 2-groups in dimensions not scaling as a prime power.

to break the barrier and construct higher approximate unitary k-designs? We show that perhaps surprisingly a number of non-Clifford unitaries that scales independently of the system size suffices to arrive at additive approximate designs with constant error. A phenomenon which we dub *quantum homeopathy*. This completes our conceptional study of Clifford designs.

In the final section of this chapter, we deviate from the core topic of this thesis, quantum device characterization, and take a look at a more fundamental question in quantum physics: How do equilibrium ensembles arise within the unitary time evolution of quantum mechanics? This discussion briefly illustrates further results arising from the study of the properties of local unitary or Clifford random quantum circuits.

4.1 The moment operator of Clifford four-designs

Compressed sensing applications where the measurements are generated by random unitaries typically require bounds of the fourth moments in the unitaries. To determine if a specific application already works with Clifford unitaries, we need a way to calculate the fourth-order moment operator of the Clifford group, i.e. an analogous result to Theorem 14 for $k = 4$. As Theorem 14 for the unitary group[3], the result for the Clifford group heavily relies on a characterization of the commutant of $\Delta^4_{\mathrm{Cl}(d)}$, i.e. Δ^4_d with domain restricted to the $\mathrm{Cl}(d)$. One characterization for the Clifford group was derived in Ref. [Zhu+16] and applies to multi-qubit dimensions $d = 2^n$. The work introduces the orthogonal projection

$$Q = \frac{1}{d^2} \sum_{k=1}^{d^2} W_k^{\otimes 4} \tag{4.1}$$

where $W_1, \dots, W_{d^2} \in \mathrm{L}\left(\mathbb{C}^d\right)$ are the multi-qubit Pauli matrices. In fact, the d^2-dimensional range of Q forms a particular stabilizer code. We denote by $Q^\perp = \mathrm{Id} - Q$ the orthogonal projection onto the complement of this stabilizer code. The orthogonal projection Q commutes with every $\pi_4(\sigma)$, $\sigma \in \mathfrak{S}_4$. Thus, Q acts trivially on the Specht modules S_λ in the Schur-Weyl decomposition (2.35). Following the notation conventions from Ref. [Zhu+16], we denote the subspace of the Weyl module W_λ that intersects with the range of Q by W_λ^+ and its dimension as D_λ^+. Analogously, the orthogonal complement of W_λ^+ shall

[3] The material of the remainder of the section has been published as Section A.2 of the supplemental material of Ref. [1]. It is altered to use a consistent notation.

be W_λ^- with dimension D_λ^-. We are now ready to state the main result of this section.

Theorem 34 (Integration over the Clifford group $\mathrm{Cl}(d)$)**.** *Let $A \in \mathrm{L}(V)$. Then,*

$$\begin{aligned}\mathcal{M}^{(4)}_{\mu_{\mathrm{Cl}(d)}}(A) = \frac{1}{4!} &\sum_{\lambda \vdash 4, l(\lambda) \leq d} d_\lambda \sum_{\sigma \in \mathfrak{S}_4} \\ &\times \left[\frac{1}{D_\lambda^+} \mathrm{Tr}(AQ\pi_4(\sigma^{-1}))Q + \frac{1}{D_\lambda^-} \mathrm{Tr}(AQ^\perp \pi_4(\sigma^{-1}))Q^\perp \right] \\ &\times \pi_4(\sigma) P_\lambda .\end{aligned} \tag{4.2}$$

To set up the proof we summarize the necessary results of Ref. [Zhu+16] in the following theorem:

Theorem 35 (Representation theory of the Clifford group [Zhu+16])**.** *Whenever $W_\lambda^\pm$ are non-trivial, the action of $\mathrm{Cl}(d) \times \mathfrak{S}_4$ on $(\mathbb{C}^d)^{\otimes 4}$ is multiplicity free and $(\mathbb{C}^d)^{\otimes 4}$ decomposes into irreducible components*

$$(\mathbb{C}^d)^{\otimes 4} \cong \bigoplus_{\lambda \vdash 4, l(\lambda) \leq d} (W_\lambda^+ \otimes S_\lambda) \oplus (W_\lambda^- \otimes S_\lambda), \tag{4.3}$$

on which $\mathrm{Cl}(d) \times \mathfrak{S}_4$ acts as $\Delta^\lambda_{\mathrm{Cl}(d)} \otimes \pi_4^\lambda$. The dimensions of W_λ^+ are of polynomials in d of degree 4 and the dimensions of W_λ^- are either vanishing or polynomials in d of degree 2.

From Theorem 35 we learn that an element of the commutant of the diagonal action of the Clifford group $\Delta^4_{\mathrm{Cl}(d)}$ can be written in the form

$$B = Q \bigoplus_{\lambda \vdash 4, l(\lambda) \leq d} (\mathrm{Id}_{D_\lambda} \otimes B_\lambda^+) + Q^\perp \bigoplus_{\lambda \vdash 4, l(\lambda) \leq d} (\mathrm{Id}_{D_\lambda} \otimes B_\lambda^-), \tag{4.4}$$

where $B_\lambda^\pm \in \mathrm{L}(S_\lambda)$ are linear operators acting on the Specht modules S_λ.

To expand elements of $\mathrm{Comm}(\Delta^4_{\mathrm{Cl}(d)})$, we define the map $\tilde{\Phi} : \mathrm{L}(V) \to \mathrm{L}(V)$, $\tilde{\Phi}(A) = \Phi(AQ)Q + \Phi(AQ^\perp)Q^\perp$ with Φ from (2.41). The map $\tilde{\Phi}$ has properties comparable to the map Φ, but is adapted to the diagonal representation of the Clifford group.

Lemma 36. *For all $A \in \mathrm{L}(V)$ and $B \in \mathrm{Comm}(\Delta^4_{\mathrm{Cl}(d)})$*

$$\begin{aligned}\tilde{\Phi}(A) =&\tilde{\Phi}(\mathcal{M}^{(4)}_{\mu_{\mathrm{Cl}(d)}}(A)), && (4.5)\\ \tilde{\Phi}(B) =&B\tilde{\Phi}(\mathrm{Id}), && (4.6)\\ \tilde{\Phi}(\mathrm{Id})^{-1} =&\frac{1}{4!}\sum_{\lambda\vdash 4, l(\lambda)\leq d} d_\lambda P_\lambda \left[\frac{1}{D^+_\lambda}Q + \frac{1}{D^-_\lambda}Q^\perp\right]. && (4.7)\end{aligned}$$

Proof.

1. Since $Q\pi_4(\sigma^{-1})$ and $Q^\perp\pi_4(\sigma^{-1})$ are in $\mathrm{Comm}(\Delta^4_{\mathrm{Cl}(d)})$ for all $\sigma \in \mathfrak{S}_4$, we can again apply Lemma 8 to get

$$\begin{aligned}\mathrm{Tr}(\mathcal{M}^{(4)}_{\mu_{\mathrm{Cl}(d)}}(A)Q\pi_4(\sigma^{-1})) &= \mathrm{Tr}(\mathcal{M}^{(4)}_{\mu_{\mathrm{Cl}(d)}}(AQ\pi_4(\sigma^{-1})))\\ &= \mathrm{Tr}(AQ\pi_4(\sigma^{-1}))\end{aligned}$$

and likewise for $Q^\perp$ instead of Q. Inserting this in the definition of $\tilde{\Phi}$ yields the first statement.

2. From the expansion of elements $B \in \mathrm{Comm}(\Delta^4_{\mathrm{Cl}(d)})$ in (4.4), we conclude that B can be expressed as $B = QB_1 + Q^\perp B_2$, where B_1 and B_2 are in the group ring $\mathbb{C}[\mathfrak{S}_4]$. Hence, it suffices to show the statement, $\tilde{\Phi}(B) = B\tilde{\Phi}(\mathrm{Id})$, for $B = Q\pi_4(\sigma)$ and $B = Q^\perp\pi_4(\sigma)$. In the first case, we find

$$\begin{aligned}\tilde{\Phi}(Q\pi_4(\sigma)) &= \Phi(Q\pi_4(\sigma))Q = \Phi(Q\,\mathrm{Id})Q\pi_4(\sigma)\\ &= \tilde{\Phi}(\mathrm{Id})Q\pi_4(\sigma)\,,\end{aligned} \tag{4.8}$$

where property (2.24) from Lemma 8 has been used in the second step. The proof of $Q^\perp$ is analogous.

3. Using the decomposition (4.3) of Theorem 35, we can calculate

$$\begin{aligned}\tilde{\Phi}(\mathrm{Id}) &= \sum_{\lambda\vdash 4, l(\lambda)\leq d}\sum_{\sigma\in\mathfrak{S}_4} \chi_{\pi_4}(\sigma^{-1})\pi_4(\sigma)\left[D^+_\lambda Q + D^-_\lambda Q^\perp_\lambda\right]\\ &= 4!\sum_\lambda \frac{1}{d_\lambda}P_\lambda\left[D^+_\lambda Q + D^-_\lambda Q^\perp\right],\end{aligned} \tag{4.9}$$

where the last line follows again from the expression (2.39) for the projectors. Inverting this expression yields the lemma's assertion.

□

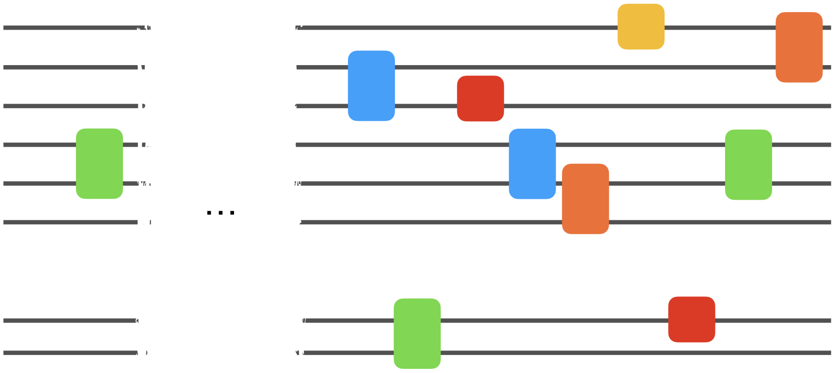


Figure 4.1: The figure depicts a local random Clifford Clifford circuit. At every step a gate from a set of generators of Cl(4) consisting of one- and two-local gates is applied on a random pair of *adjacent* qubits. The number of such layers consisting of a single gate is the depth of the circuit.

With these statements for the Clifford group at hand, we can proceed to prove Theorem 34.

Proof of Theorem 34. Equation (4.5) in Lemma 36 and 4.6 in Lemma 36 can be combined to conclude $\tilde{\Phi}(A) = \tilde{\Phi}(\mathcal{M}^{(4)}_{\mu_{\mathrm{Cl}(d)}}(A)) = \mathcal{M}^{(4)}_{\mu_{\mathrm{Cl}(d)}}(A)\tilde{\Phi}(\mathrm{Id})$ and, thus, $\mathcal{M}^{(4)}_{\mu_{\mathrm{Cl}(d)}}(A) = \tilde{\Phi}(A)\tilde{\Phi}(\mathrm{Id})^{-1}$. The expression for $\tilde{\Phi}(\mathrm{Id})^{-1}$ was derived in Lemma 36, Equation (4.7). Together with the definition of $\tilde{\Phi}$ the expression of the theorem follows after some simplification. □

4.2 Random Clifford designs via circuits

In this section, we will continue our study of Clifford designs.[4] A particular practically interesting question is how to implement a Clifford design of a certain order k in terms of a circuit of gates. One way is 'compiling' Clifford unitaries that are drawn uniformly at random from the Clifford group. An arbitrary Clifford unitary on n-qubit can be implemented using $O(n^2/\log(n))$ phase, Hadamard and CNOT gates [AG04]—the generators defined in (2.55). In this construction

[4] The section provides a short summary of one of the results of Ref. [6]. The work was done together with Jonas Haferkamp, who took the lead in the project, Felipe Montealegre-Mora, Markus Heinrich, Jens Eisert and David Gross. The thesis author developed the project idea together with Jonas Haferkamp and contributed to the conception, the analytical proofs and the write-up of the final manuscript.

the two-qubit CNOT gates act on arbitrary pairs of not necessarily adjacent qubits. Instead of compiling the gate, we can reverse the question and ask more directly: Assume we draw a random circuit by subsequently acting with a two-local Clifford gate randomly chosen from a set of generators of $\mathrm{Cl}(4)$ on a random pair of adjacent qubits in a one-dimensional chain again and again. See Figure 4.1 for an illustration. For which depth does the ensemble of such circuits approximate Clifford designs of a certain order?

We can formally formulate this question using the notion of a relative approximate Clifford design. A probability measure ν on $\mathrm{Cl}(d)$ is a *relative ϵ-approximate Clifford k-design* if

$$(1-\epsilon)\mathcal{M}^{(k)}_{\mu_{\mathrm{Cl}(d)}} \leq \mathcal{M}^{(k)}_{\nu} \leq (1+\epsilon)\mathcal{M}^{(k)}_{\mu_{\mathrm{Cl}(d)}}\,, \tag{4.10}$$

where we here write $A \leq B$ if $B - A$ is positive semi-definite. Let $G \subset \mathrm{Cl}(4)$ be a finite gate set that generates $\mathrm{Cl}(4)$ and is closed under inversion. For example, $G = \{H \otimes \mathbb{1}, S \otimes \mathbb{1}, S^3 \otimes \mathbb{1}, \mathrm{CNOT}\}$ is such a gate set, cmp. (2.55). Let $G_{i,j}$ be a set of embeddings of the elements of G into $\mathrm{Cl}(2^n)$ by acting non-trivially only on the ith and jth qubit, e.g. $G_{1,2} = G \otimes \mathbb{1}_{2^{n-2}}$. We denote by σ_G the probability measure on $\mathrm{Cl}(2^n)$ that has uniform support on the set $\bigcup_{i\in[n]} G_{i,i+1}$ with indices periodically identified. A *local random Clifford circuit of depth m* is a unitary drawn from the measure σ_G^{*m}, the m-fold convolution of the measure σ_G. In Ref. [6] we prove the following result.

Theorem 37. *Let $n \geq 12t$, then the ensemble of local, random Clifford circuit σ_G^{*m} of depth $m \in O(n \log^{-2} N(t) t^8 (2nt + \log(1/\epsilon)))$ constitutes a relative ϵ-approximate Clifford t-design.*

Thus, we observe that we also find a number of Clifford gates scaling quadratically with the number of qubits. Note that the results of Ref. [DLT02] implicitly give a scaling of $O(n^8)$. In contrast to the compiling scheme of Ref. [AG04; PMH08], however, in the circuit construction, we only have two-qubits gates acting on pairs of *adjacent* qubits.

In Chapter 6 we develop a process tomography scheme that employs random multiqubit Clifford unitaries. For such applications, Theorem 37 shows that one can directly use random circuits of local generators on adjacent qubits instead of implementing multi-qubit Clifford gates via compiling. Random circuits are, thus, a simpler and more direct experimental approach and avoid gates acting on distant qubits that can not natively be implemented on a certain architecture.

Refs. [BHH16b; BHH16a; BV10] developed a proof strategy for establishing that local unitary random circuits approximate *unitary* t-designs in depth $O(n^2 t^{10})$. The proof of Theorem 37 relies on a careful adaptation of the argument. In the following we sketch the argument, focusing on the deviations from the unitary case. For the complete proof, we refer to our Ref. [6].

Proof sketch for Theorem 37. As a first step, one relates the notion of an approximate Clifford t-design to the deviation of the moment-operators in spectral norm. Analogously to the result for the unitary group [BHH16b, Lemma 4 & 30], it holds that $\left\| \mathcal{M}^{(k)}_{\nu} - \mathcal{M}^{(k)}_{\mu_{\mathrm{Cl}(d)}} \right\|_{\infty} \leq \epsilon 2^{-2tn}$ implies that ν is a relative ϵ-approximate Clifford design [6, Lemma 5]. At the core of the argument connecting the operator norm with the deviation of the eigenvalues, is a lower bound on the minimal eigenvalue of $\mathcal{M}^{(k)}_{\mu_{\mathrm{Cl}(d)}}$. Since $\mathcal{M}^{(k)}_{\mu_{\mathrm{Cl}(d)}}$ is a projector, the eigenvalues are determined by dimensions of its range, i.e. the irreducible representations of $\Delta^k_{\mathrm{Cl}(d)}$. As a subgroup, the irreducible representations of $\Delta^k_{\mathrm{Cl}(d)}$ for the Clifford group arise as further decompositions of the irreducible representations of the tensor powers of the unitary group. An explicit example for $k = 4$ was given in Theorem 35. Thus, the irreducible representations are contained in the Weyl modules and have multiplicity spaces that contain the Specht modules. This allows us to reduce the bound on the eigenvalue $\Delta^k_{\mathrm{Cl}(d)}$ to a bound on the corresponding bound on the eigenvalues of $\Delta^k_{\mathrm{U}(d)}$. We refer to Ref. [6, Section VI.d] for the details of the argument.

In a second step, one realizes that the operator norm deviation of a single layer $\left\| \mathcal{M}^{(k)}_{\sigma_G} - \mathcal{M}^{(k)}_{\mu_{\mathrm{Cl}(d)}} \right\|_{\infty}$ can be recast as the spectral gap $\lambda_2(H_{n,t})$ of a family of frustration-free local Hamiltonians $H_{n,t}$. For such Hamiltonians the martingale method of Nachtergaele [Nac96] provides a bound of the spectral gap in terms of a spectral gap independent of n, $\lambda_2(H_{n,t}) \geq \lambda_2(H_{12t,t})/48$ for $n \geq 12t$. Instead of looking at the adjoint representation appearing in our construction, we can more abstractly consider the operator T_{σ_G} acting on the group algebra $L^2(\mathrm{Cl}(n))$ by convolution with σ_G. The spectrum of $H_{n,t}$ is then contained and in the spectrum of T_{σ_G} such that the gap to its second-largest eigenvalue $1 - \lambda_2(T_{\sigma_G})$ lower bounds $\lambda_2(H_{n,t})$. Using comparison techniques for random walks on groups [DS93, Corollary 1], $\lambda_2(T_{\sigma_G})$ is dominated by $1 - \eta/r_*^2$. Here, η is the lowest non-trivial probability mass of σ_G and r_*^2 is the minimal number of generators from G required to represent any elements of $\mathrm{Cl}(4)$. By construction we have $\eta = 1/|G|n$ and Ref. [AG04] establishes that $r_* = O(n^3/\log(n))$. Combining these bounds gives $\left\| \mathcal{M}^{(k)}_{\sigma_G} - \mathcal{M}^{(k)}_{\mu_{\mathrm{Cl}(d)}} \right\|_{\infty} \leq 1 - cn^{-1}t^{-8}\log^2(t)$ for some constant $c > 0$ depending only on the choice of G. See [6, Proposition 2] and its proof for more details. Interestingly, we find that the specifics of the Clifford group mainly enters the argument through the value of r_*.

Finally, observing $\left\| \mathcal{M}^{(k)}_{\sigma_G^{*m}} - \mathcal{M}^{(k)}_{\mu_{\mathrm{Cl}(d)}} \right\|_\infty \leq \left\| \mathcal{M}^{(k)}_{\sigma_G} - \mathcal{M}^{(k)}_{\mu_{\mathrm{Cl}(d)}} \right\|_\infty^m$ and tracing the scaling, yields Theorem 37. □

4.3 Quantum homeopathy

In the previous section, we have studied the complexity of random quantum circuits for approximating higher-order Clifford k-designs.[5] For $k \leq 3$ this construction also gives rise to unitary k-designs. In order to get unitary k-designs with $k > 3$, we need to resort to also use non-Clifford gates. Ref. [BHH16b] established that a local random circuit on n qubits with $O(n^2 k^{10})$ many two-local gates from $\mu_{\mathrm{U}(4)}$ is drawn from an approximate unitary k-designs. This motivates the question whether we can provide a more 'economical' construction of higher unitary k-designs using *mostly* Clifford gates. For example, in fault-tolerant quantum computing architectures Clifford unitaries can be more directly implemented while non-Clifford gates come with an overhead, e.g. of preparing so-called *magic states* [Got09; CTV17]. So how many non-Clifford gates do we actually need for an approximate unitary design with $k > 3$?

Some related observations in the literature indicate that the answer might be 'not so many'. Refs. [Zhu+16; GNW21] give a construction for *exact spherical* k-designs using only a system size-independent number of Clifford orbits. In Ref. [Zho+19] it was numerically observed that a single T gate in a random Clifford circuit drastically changes the entanglement spectrum.

Formally, we here use the notion of an (additive) ϵ-approximate design in diamond norm [HL09]. We say that a probability measure ν on $\mathrm{U}(d)$ is an *ϵ-approximate unitary k-design* if $\left\| \mathcal{M}^{(k)}_{\nu} - \mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}} \right\|_\diamond \leq \epsilon$. Note that this is a weaker notion than a relative ϵ-approximate design. Our circuit construction will make use of one non-Clifford gate $K \in \mathrm{U}(2)$ acting (multiple times) on the same qubit. We describe a K-layer by the probability measure μ_K on $\mathrm{U}(2^n)$

[5] The section provides a short summary of the main result of Ref. [6]. The work was done together with Jonas Haferkamp, who took the lead in the project, Felipe Montealegre-Mora, Markus Heinrich, Jens Eisert and David Gross. The thesis author developed the project idea together with Jonas Haferkamp and contributed to the conception, the analytical proofs and the write-up of the final manuscript.

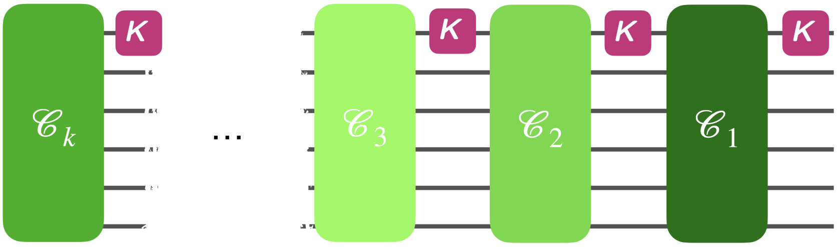


Figure 4.2: The figure depicts a K-interleaved Clifford circuit. A K-layer consisting of a single non-Clifford gate $K \in \mathrm{U}(2)$ acting on a fixed qubit is alternated with random multi-qubit Clifford gates.

with uniform support only on the set $\{K \otimes \mathbb{1}_{2^{n-1}}, K^\dagger \otimes \mathbb{1}_{2^{n-1}}, \mathbb{1}_{2^n}\}$. $K^\dagger$ and the identity are included for technical reasons but do not affect the physical implementation or the interpretation of the result. We define a *K-interleaved Clifford circuit of depth m* as a random circuit in $\mathrm{U}(2^n)$ drawn from the probability measure $\sigma_m = (\mu_{\mathrm{Cl}(2^n)} * \mu_K)^{*m}$. This means a K-interleaved Clifford circuit alternates between a uniformly random multi-qubit Clifford gates and K-layers acting only non-trivially on the first qubit. See Figure 4.2 for an illustration. This is arguably the most economic usage of a non-Clifford gate in a random circuit.

We prove the following result for K-interleaved Clifford circuits in Ref. [6].

Theorem 38 (Quantum homeopathy)**.** *Let $K \in U(2) \setminus \mathrm{Cl}(2)$. Suppose $k \geq C_1 \log^2(t)(t^4 + t\log(1/\epsilon))$, a K-interleaved Clifford circuit on n qubits of depth k is an additive ϵ-approximate t-design for all $n \geq C_2 t^2$ (and suitable constants $C_1, C_2 > 0$ only depending on K).*

The striking feature of this result is that we arrive at an additive ϵ-approximate unitary t-design with $O(t^4 \log^2(t))$ non-Clifford gates *independent of the system size* n. Considering the compilation of an n-qubit Clifford unitary of $O(n^2/\log(n))$ many local gates [AG04], our construction has an overall gate count in

$$O(n^2/\log n t^4 \log^2(t)) .$$

This is an improvement over the scaling $n^2 t^{10}$ of Ref. [BHH16b] in both n and t. The difference is even more striking when considering the Clifford gates as a 'free resource' as motivated from the perspective of fault-tolerant implementations.

We are tempted to refer to this phenomenon metaphorically as 'quantum homeopathy'. The concept of changing the ensemble properties of an infinitely large system using only a constant amount of non-trivial ingredients, summarizes the statement of Theorem 38 convincingly well. This is purely meant as an illustrative analogy and does of course not encompass any scientific statement about alternative health treatments.

We here only give the idea of the proof and refer to Ref. [6] for the complete and, in fact, lengthy proof. To derive Theorem 38 we have to carefully study the difference between the moment operators of tensor powers of Haar random unitaries and random Clifford unitaries as well as the overlap of this difference with a K-layer. With every application of a Clifford- and K-layer this difference is further suppressed. The difference between the moment operators of the Clifford group and the full unitary group is captured by the difference of the corresponding commutants. In Section 4.1 we have used the characterization of Ref. [Zhu+16] for the commutant of the fourth-order tensor representation of the Clifford group. Extending this result, Ref. [GNW21] derives an explicit construction of the commutant of higher-order tensor representations in terms of so-called *stochastic Lagrangian subspaces*. We here omit the precise definition [6, Definition 7] and only introduce the symbol $\Sigma_{k,k}$ for the set of Lagrangian subspaces. Further, each subspace $T \in \Sigma_{k,k}$ has a representation $r(T) \in \mathrm{L}((\mathbb{C}^{2^n})^{\otimes k})$ and $r(\Sigma_{k,k}) \supset \pi_k(\mathfrak{S}_k)$ the representation of the symmetric group acting by commuting the tensor powers, (2.28). We can, thus, identify $\mathfrak{S}_k$ with a subset of $\Sigma_{k,k}$. Ref. [GNW21] proves the following result:

Theorem 39 (Clifford commutant [GNW21])**.** *For $n \geq t-1$ it holds that*

$$\operatorname{comm}(\Delta_d^k(\mathrm{Cl}(2^n))) = \operatorname{span}\{r(T) \mid T \in \Sigma_{k,k}\} . \tag{4.11}$$

The theorem implies that the difference $\mathcal{M}^{(k)}_{\mu_{\mathrm{Cl}(d)}} - \mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}$ is the projector onto the space spanned by $r(\Sigma_{k,k} \setminus \mathfrak{S}_k)$. Interestingly, the dimension of the space turns out to be independent of the system size. The proof of Theorem 38 explicitly constructs this projector using a Gram-Schmidt orthogonalization of the elements $r(\Sigma_{k,k} \setminus \mathfrak{S}_k)$. In terms of the orthogonal basis one can then bound all contributions to the deviation of the moment operator of the K-interleaved Clifford circuit σ_m. One finds that $\left\|\mathcal{M}^{(k)}_{\sigma_m} - \mathcal{M}^{(k)}_{\mu_{\mathrm{U}(d)}}\right\|_\diamond \leq 2^{O(k^4)+k\log m}(1 + 2^{O(k^2)-n})^{5m}\eta^{m-1}$, where η is the supremum of the moment operator of the K-layer restricted to $r(\Sigma_{k,k} \setminus \mathfrak{S}_k)$. This supremum can be bounded as $\eta \leq 1 - c\log^2(t)$ with a constant c depending only on the gate K using the seminal result on bounds for restricted spectral gaps by Vajú [Var13] and a novel bound on the overlap of the Haar moment operator with the elements in $r(\Sigma_{k,k} \setminus \mathfrak{S}_k)$. The latter bound is derived as Lemma 13 in Ref. [6] and might become useful in other related applications. We here content ourselves with this level of details.

4.4 *Viewpoint:* Locally random matrix-product states and entanglement ergodic systems

We conclude this chapter with an illustration of further applications of random ensembles of unitaries and quantum states beyond the fields of quantum computing and quantum characterization.[6] Random matrix theory has found abundant application in the study of interacting many-body systems [ABD15]. Similar to what motivates the study of random quantum circuits, it is often infeasible to derive rigorous results on the properties of a specific instance of a many-body system but the study of the properties of a random ensemble of such systems is more fruitful. Such arguments allow one to 'at least' make statements about *typical* or *generic* situations encountered in such systems.

A fundamental question is the explanation of 'equilibration'. We observe a world that is mostly described by statistical ensembles in equilibrium. How does this world arise via unitary time evolution? How do *closed* complex many-body systems relax to statistical equilibrium ensembles within the framework of unitary quantum mechanics? While the question itself stands from the beginnings of the theory of quantum mechanics, its mathematical rigorous study is still an ongoing endeavour. Excitingly, the theoretical endeavour is nowadays accompanied by the novel possibilities of quantum technologies to actually realize models of many-body quantum systems in well-controlled experiments, e.g. [BDZ08; BDN12; Sch+12; Tro+12; Bra+15; Sch+15].

Formally, one can describe the equilibration of a closed system in terms of the expectation values of bounded observables. Let $\rho(t)$ denote the time-evolved state of the system that started initially in the state $\rho(0)$. Let A be an observable with expectation value $\langle A(t)\rangle_\rho = \mathrm{Tr}[\rho(t)A]$. We say a system is equilibrated if

[6] This section presents results derived in Refs. [7; 8]. Ref. [7] was authored together with Henrik Wilming, who first-authored the work, Marcel Goihl and Jens Eisert. Ref. [8] originated in a joint project with Jonas Haferkamp, the lead-author of the manuscript, Christian Bertoni and Jens Eisert. In both works the thesis author did central contributions to all parts of the work including the conception, proofs, presentation and write-up.

$\langle A(t)\rangle_\rho$ exhibits only small fluctuations around a static value for most times. We denote the infinite time-average of the fluctuations as

$$\Delta A_\rho^\infty := \lim_{T\to\infty} \frac{1}{T} \int_0^t \left|\langle A(t)\rangle_\rho - A_\rho^\infty\right|^2 \mathrm{d}t, \tag{4.12}$$

where the infinite time-average of the expectation value of A is

$$A_\rho^\infty := \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathrm{Tr}[\rho(t)A]\mathrm{d}t = \mathrm{Tr}[\omega A] \tag{4.13}$$

and the state ω is the infinite time-average of $\rho(t)$. It is now well understood that the dynamical build-up of entanglement leads to equilibration [Tas98; Rei08; Lin+09; Gol+10; Sho11; Rei12; RK12; SF12; MRA13; GE16; EFG15; BR17]. In particular, for Hamiltonians with non-degenerate energy differences the fluctuations can be bounded in terms of the Rényi-2 entropy as

$$\Delta A_\rho^\infty \leq \|A\|_\infty^2 \,\mathrm{e}^{-S_2(\omega)}\,, \tag{4.14}$$

with the Rényi-α entropies defined as $S_\alpha(\rho) = \frac{1}{1-\alpha}\log \mathrm{Tr}[\rho^\alpha]$ [Sho11]. Note that via Proposition 2 the characterization of equilibration in terms of bounded observables implies that the reduced density matrices of the system similarly equilibrate in trace-distance.

A large Rényi-2 entropy of the time-averaged state is a sufficient condition for equilibration. Rigorous arguments that a physical system meets this condition when starting out in a natural initial state however have been lacking [FBC17; Gal+18]. Ref. [FBC17] establishes that the Rényi-2 entropy of states with finite correlation length scales poly-logarithmically with the system size, insufficient to explain equilibration to exponential precision in the system size.

To remedy the situation, we proposed the notion of *entanglement ergodicity* in Ref. [7]—a supposedly *weak condition* on the entanglement present in the eigenstates of a system but strong enough to ensure equilibration to exponential precision in the system size even for initial product state.

Let us formally state the definition for entanglement ergodicity. We consider families of systems of local Hamiltonians

$$H_\Lambda = \sum_x h_x \tag{4.15}$$

on a regular ν-dimensional lattice Λ indexed by the lattice cardinality $N = |\Lambda|$. We define a sequence of increasing (in N) systems as *entanglement-ergodic* if there exist $N_0 \in \mathbb{N}$ and a Lipschitz continuous function $g : \mathbb{R} \to [0, \infty)$ that is positive in the interior of the energy spectrum such that for all $N \geq N_0$ the system fulfils the following *weak entropic volume law*: For every eigenstate $|e\rangle$, where e is its energy density, there exist a subsystem A such that the reduced state $\rho_A = \mathrm{Tr}_{A^c}(|e\rangle\langle e|)$ has Rényi-2 entropy:

$$S_2(\rho_A) \geq g(e)N\,. \tag{4.16}$$

The main result of Ref. [7] is now to prove that entanglement-ergodic systems in fact equilibrate when they start out in a product state. A Hamiltonian H_λ of the form (4.15) is called strictly local and uniformly bounded if each h_x has non-trivial support on at most l subsystems and $\|h_x\|_\infty \leq h$ with constants l, h independent of the system size.

Theorem 40 ([7, Theorem 2])**.** *Consider an entanglement-ergodic system with strictly local, uniformly bounded Hamiltonian. Then for any energy density $e > 0$ there exists $k(e), N_0(e) > 0$ such that for all system-sizes $N > N_0(e)$ and for all product-states $|\Psi\rangle$ with energy density e, we have*

$$S_\alpha(\omega) \geq k(e)N, \tag{4.17}$$

where ω is the time-averaged state of the system when initialized in $|\Psi\rangle$.

By (4.14), Theorem 40 has the immediate consequence that the systems fulfilling the theorem's assumptions and having non-degenerate energy gaps equilibrate to exponentially precision in the sense that $\Delta A_\rho^\infty \leq \|A\|_\infty^2 \,\mathrm{e}^{-k(e)N}$ for all observables A.

The proof of the theorem starts from the observation that states that have rank one between a subsystem A and its complement have an overlap with pure states that is exponentially small in the Rényi-α entropies with $\alpha > 1$ of the state reduced to the subsystem A [7, Lemma 4]. This observation is combined with a result on energy densities of product states for strictly local, uniformly bounded Hamiltonians derived by Anshu [Ans16]. We refer to Ref. [7] for further details.

Our work has reduced the question of equilibration from natural initial states, product states, to a weak entropic volume law of the eigenstates. While it is plausible that many quantum systems might be entanglement-ergodic, making a rigorous statement about the entropy scaling of every eigenstate is still challenging, to say the least. To collect some examples of states exhibiting a weak-volume law,

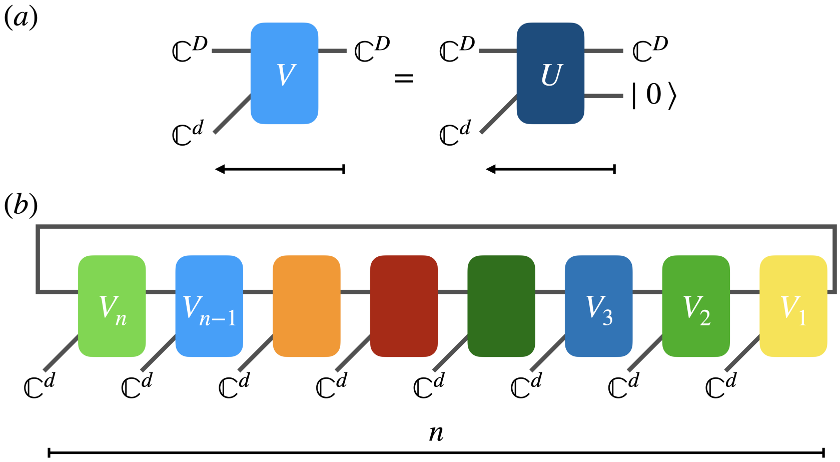


Figure 4.3: (a) A random isometry $V : \mathbb{C}^D \to \mathbb{C}^{D\times d}$ can be constructed by acting with a Haar random unitary $U \in \mathrm{U}(Dd)$ partially on a normalized vector $|0\rangle \in \mathbb{C}^d$. (b) A random disordered MPS (rMPS) drawn from the measure $\mu_{d,n,D}$ is defined by the contraction of n i.i.d. random isometries $V_k : \mathbb{C}^D \to \mathbb{C}^{D\times d}$ along the spaces of dimension D. The resulting tensor is a vector in $\mathbb{C}^{nd} \cong (\mathbb{C}^d)^{\otimes n}$.

one way forward is to study 'natural' random ensembles of states. Ref. [RW20] showed that generic translationally invariant matrix product state (MPS) fulfil a weak volume law for a subregion A that is every kth side. In Ref. [8] we studied the case of random disordered MPSs. An MPS of bond-dimension D on $(\mathbb{C}^d)^{\otimes n}$ (with periodic boundary conditions) is described by a collection of $n \cdot d$ matrices $((A^{(k)}_{i_k} \in \mathbb{C}^{D\times D})_{i_k\in[d]})_{k\in[n]}$ to be the pure state

$$|\psi\rangle = \sum_{i_1,\dots,i_n\in[d]} \mathrm{Tr}[A^{(1)}_{i_1} A^{(2)}_{i_2} \cdots A^{(n)}_{i_n}]\, |i_1,\dots,i_n\rangle\,, \tag{4.18}$$

where $\{|i_1,\dots,i_n\rangle\}$ is a local basis of $(\mathbb{C}^d)^{\otimes n}$ [FNW92; Per+07]. For each k, we can regard the collection of matrices $(A^{(k)}_{i_k} \in \mathbb{C}^{D\times D})_{i_k\in[d]}$ as a 3-tensor $A^{(k)} \in \mathbb{C}^{D\times D\times d}$, the so-called *tensor core* of the kth site. The tensor cores can furthermore be identified with homeomorphisms from $\mathbb{C}^D \otimes \mathbb{C}^d \to \mathbb{C}^D$ as $A^{(k)} = \sum_{l,m\in[D],i\in[d]} A^{(k)}_{l,m,i}\, |l\rangle\langle m|\, \langle i|$ after choosing bases.

A random disordered MPS (rMPS) is defined by choosing all tensor cores i.i.d. as random isometries from $\mathbb{C}^{D\times d} \to \mathbb{C}^D$ from the unitarily invariant measure [Gar+10]. The unitarily invariant measure on the isometries is simply induced by the Haar measure on the unitary group. Concretely, let $|0\rangle$ be a fixed state in $\mathbb{C}^d$ and $U \in \mathrm{U}(Dd)$ a Haar random unitary. We can set $V = U\mathbb{1} \otimes |0\rangle$ to arrive at a random isometry from $\mathbb{C}^{D\times d} \to \mathbb{C}^D$. See also Figure 4.3 for an illustration. We denote the measure on $(\mathbb{C}^d)^{\otimes n}$ defining rMPSs with bond dimension D as $\mu_{d,n,D}$.

Note that $\mu_{d,n,D}$ takes values on vectors that are not exactly normalized in ℓ_2 norm. Their squared norm, however, is ϵ-close to one with probability of at least $1-\epsilon^{-2}d^{-n}$ [8, Lemma 1]. MPSs are well-known to approximate the ground states of gapped one-dimensional local Hamiltonians capturing finite-ranged interactions [SPC11]. Therefore, disordered rMPSs can be seen as arising from disordered parent Hamiltonians and are typical representative of one-dimensional quantum phases of matter. This does not directly suffice to relate rMPSs to also higher-excited eigenstate of an ensemble of Hamiltonians. Interestingly, the eigenstates of systems exhibiting the phenomenon of many-body localization, that equilibrate but fail to thermalize, are expected to be well-approximated by matrix-product states capturing the finite-correlation length even at higher energies [BN13; Fri+15].

This motivates the study of the entanglement ergodicity of rMPSs. In Ref. [8] we prove that rMPSs in fact fulfil a weak entropic volume law:

Theorem 41 ([8, Theorem 2])**.** *Suppose that n is divisible by $k \in \mathbb{N}$. Let A be subsystem that excludes every kth site. For $|\psi\rangle$ an rMPS drawn from $\mu_{d,n,D}$, it holds that*

$$\mathbb{P}\left[S_2(\mathrm{Tr}_{A^c}[\,|\psi\rangle\langle\psi|\,]) \geq \Omega\left(\frac{n}{k}\right)\right] \geq 1 - \mathrm{e}^{\Omega(n/k)} \,. \tag{4.19}$$

The proof makes use of techniques for calculating the moments of circuits of independent Haar random unitaries by mappings to partition functions of classical statistical models [Hun19; NVH18]. We here only sketch the principle idea of this toolbox. A lower bound on the Rényi-2 entropy, directly follows from an upper bound of the state's purity. The purity of the reduced state arising from an rMPS is a specific contraction of the second moment operators of the Haar random unitaries defining the tensor cores. By Proposition 12 the moment operators can be expressed in terms of the projections onto the symmetric and anti-symmetric subspace of $\mathfrak{S}_2$ or, equivalently, by as a sum over the permutation in $\mathfrak{S}_2$ as

$$\mathcal{M}^{(2)}_{\mu_{\mathrm{U}(d)}}(A) = \sum_{\sigma,\tau\in\mathfrak{S}_2} \mathrm{Wg}(\sigma^{-1}\tau, d)\pi_2(\sigma)\,\mathrm{Tr}[\pi_2^\dagger(\tau)A]\,, \tag{4.20}$$

with the Weingarten function $\mathrm{Wg}(\tau, d) = (d^2-1)^{-1}(-d)^{-\delta_{\tau,\mathbb{F}}}$. In the contraction we can therefore replace every moment operator by an input operator and an output operator and sum over all combinations of choosing $\mathbb{1}$ and $\mathbb{F}$ for each operator. For every moment operator we multiply a weight of $(d^2-1)^{-1}$. For the terms where the input and output operators differ, we additionally multiply with $-1/d$. In this way, we can see the entire contraction as a partition function of a one dimensional spin-$1/2$ system with a $-1/d$ interaction penalizing spin-alignment. The partition function can be bounded with combinatoric arguments and making use of an entanglement area law [ECP10]. Markov's inequality then implies the concentration result. We refer to Ref. [8] for the detailed proof.

The same proof techniques allow us to calculate other second-order polynomials in the entries of an rMPS. We can make use of this to derive another result in the context of equilibration. With entanglement ergodicity we introduced a weak property on the eigenstates to show equilibration already from all product states. Changing the perspective, one can also aim at deriving typicality arguments for ensembles of initial states. Then, one has to argue that a typical initial state $\rho(0)$ evolves to a state ω that has a large Rényi-2 entropy more or less independent of the actually Hamiltonian of the system that governs the time-evolution. Such a result was derived, e.g. in Ref. [HH19] for random product states. Similarly, in Ref. [8] we prove the following statement for rMPSs.

Theorem 42 ([6, Theorem 1])**.** *Let H be a Hamiltonian with non-degenerate spectrum and spectral gap. Let $|\psi\rangle \sim \mu_{d,n,D}$ (and normalized). For all observables A, it holds with probability $1-\mathrm{e}^{-c_1\alpha(d,D)n}$ that*

$$S_2 \geq c_2\alpha(d,D)n, \tag{4.21}$$

with constants $c_1, c_2 \geq 0$ and $\alpha(d,D) \geq 0$ not depending on n.

Again by (4.14), this theorem has the immediate consequence that a system with non-degenerate energy gaps initialized in a disordered rMPS equilibrates to exponentially precision with exponentially high probability in the system size.

In two different readings, we have shown that already *locally random* ensembles of states have generically suitable properties to ensure equilibration. Ultimately however, it is still an important open question to establish entanglement ergodicity or a similarly weak condition that ensures equilibration directly for the ensemble of eigenstates of 'natural' Hamiltonians.

5 Randomized benchmarking – Estimating average gate fidelities

In Chapter 3 we studied the problem of reconstructing quantum states while being partially agnostic about the calibration of the measurement apparatus. We now turn to the task of identifying quantum channels commonly referred to as quantum process tomography. The most important desiderata of semi-device-dependence in quantum process tomography is captured by the notion of robustness against errors in the state preparation and measurement (SPAM). In Chapter 6 we study how tomographic information about quantum process and in particular unitary gates can be extracted from the relative average gate fidelities, (2.68), with respect to Clifford unitaries. The main motivation for these tomographic schemes is that in principle this type of data can be efficiently extracted by so-called RB protocols that exhibit robustness against SPAM errors. To provide the necessary context for the *compressive randomized-benchmarking tomography* scheme devised in the following chapter, we now take a closer look at RB protocols. We begin with an introduction of the foundational primitives of RB and its variants of *interleaved* RB that aim at extracting relative average-gate fidelities. This pedagogical introduction was originally written for the tutorial on quantum certification [4]. To highlight the general scope of RB, we afterwards briefly sketch the results of our work, Ref. [9]. Therein we develop the mathematical tools for deriving theoretical guarantees for virtually all RB protocols in a unified framework. We conclude the chapter with new general results on the sampling complexity of extracting multiple average gate fidelities by-passing the complications arising from RB.

5.1 Randomized benchmarking

Prepare-and-measure schemes for estimating measures of quality of quantum processes fail in the presence of sizeable SPAM errors.[1] In the context of digital

[1] Prepare-and-measure schemes …inversion of the sequence.] This review section is also published in Ref. [4] with minor modifications.

quantum computing, the sensitivity to SPAM errors is dramatically reduced by so-called *RB protocols* [EAŻ05a; Lév+07; Kni+08; Dan+09; MGE11]. These protocols can extract certain quantitative measures of a quantum process associated to a *quantum gate set*. The process can be, for example, a certain gate, an error channel or an error map associated to the deviation of a quantum gate set from its ideal implementation. While still concerned with the physical layer of a quantum device, randomized benchmarking protocols already make explicit use of a gate layer, the abstraction at the heart of digital quantum computing.

Randomized benchmarking comprises a large zoo of different protocols. Therefore, we begin with a fairly general description. The principle idea to achieve the SPAM(-error) robustness is the following: After preparing an input state, one applies the quantum process under scrutiny multiple times in sequences of different length before performing a measurement. Thereby, the effect of the process on the measurement is attenuated with increasing sequences length. At the same time errors in the state preparation and measurements enter the measured quantities only linearly and are independent of the sequence length. In this way, fitting the attained signals for different sequence lengths with functions depending on the length reveals properties of the quantum process disentangled from the SPAM errors.

A prototypical RB protocol implements this rough idea for a digital quantum computer as follows. Let $\mathsf{G} \subset \mathrm{U}(d)$ be a subgroup of unitary operations and $\phi : \mathsf{G} \to \mathbb{L}(\mathbb{C}^d)$ be their implementation on a quantum computer. In simple RB protocols $\phi(g)$ just models the faulty implementation of $\mathcal{G}$ on the actual device. More generally, the targeted implementation of the protocol can also include, e.g., a non-uniform sampling over the group or the implementation of another fixed gate after $\mathcal{G}$. Also in these cases ϕ is the faulty version of the targeted implementation. Note that the assumption of the existence of such a map ϕ already encodes assumptions on the quantum device and its noise process: The map ϕ might model the compilation into elementary gates, effects and imperfections of the physical control and noise. All these steps are not allowed to depend on the gate sequence the gate is part of, the overall time that evolves during the protocol, or other external variables. With these ingredients we can state a prototypical RB protocol, see Figure 5.1 for an illustration.

Protocol 1 (Prototypical RB). Let $\mathsf{G} \subset \mathrm{U}(d)$ be a subgroup, $\rho \in \mathcal{D}(\mathbb{C}^d)$ an initial state, and $\mathsf{M} = \{M, \mathbb{1} - M\} \subset \mathrm{Pos}(\mathbb{C}^d)$ a measurement. Furthermore, let $\mathfrak{M} \subset \mathbb{N}$ be a set of sequence lengths.

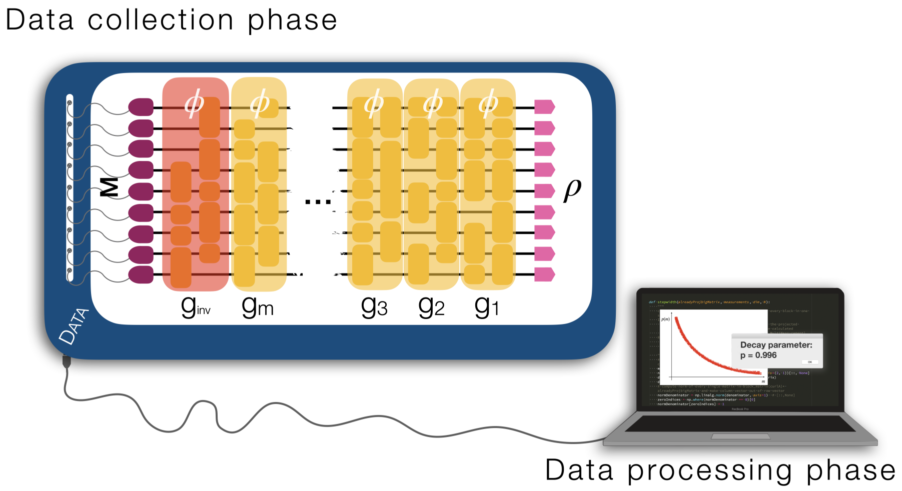


Figure 5.1: Illustration of a prototypical RB protocol. After the preparation of an initial state, one applies a random sequence of unitaries $g = (g_1, \ldots, g_m)$ succeeded by an inversion gate and final measurement of M. This experiment is repeated for different sequences and different sequence lengths m. In the classical postprocessing, the decay parameter of resulting empirical estimates for different sequence lengths m are extracted and reported as the RB parameters.

For every sequence length $m \in \mathfrak{M}$, we do the following estimation procedure multiple times:

Draw a sequence $g = (g_1, \ldots, g_m)$ of m group elements chosen i.i.d. uniformly at random. For the sequence calculate the inverse elements $g_{\text{inv}} = g_1^{-1} g_2^{-1} \cdots g_m^{-1}$.

For each sequence preform the following experiment:

- Prepare ρ
- Apply $S_g = \phi(g_{\text{inv}})\phi(g_m)\ldots\phi(g_2)\phi(g_1)$, i.e. the sequence of implementations of g followed by the implementation of g_{inv}, to ρ.
- Perform the measurement M.

Multiple repetitions of the experiment yield an estimator $\hat{p}_g$ for the probabilities

$$p_g(m) = \operatorname{Tr}[M S_g \rho] \tag{5.1}$$

Repeating these steps for different random sequences, we can calculate an estimator $\hat{p}(m)$ for

$$p(m) = \mathbb{E}_{g_1}\mathbb{E}_{g_2}\cdots\mathbb{E}_{g_m} p_{(g_1,g_2,\ldots g_m)}(m). \tag{5.2}$$

Post-processing: extract the decay parameters of the data $\mathfrak{M} \to [0,1], m \mapsto \hat{p}(m)$ and report as the RB parameters.

* * *

More generally, RB protocols might go beyond Protocol 1 in various ways: for example, by calculating the inverse of a sequence only up to specific gates, using a different measure than the uniform measure for drawing the group elements of the sequence, or performing a measurement POVM with multiple outputs or measurements adapted to the sequence. In addition, the post-processing might combine different RB data series in order to get simpler decay signatures.

The first step in the theoretical analysis of RB protocols is to establish the fitting model of the RB data $p(m)$. Ideally, $p(m)$ is well-approximated by a single exponential decay. Subsequently, the RB decay parameters can in certain settings be connected to the average gate fidelity of a noise process effecting the implementation map, as we now discuss. The data model of most RB protocols can be understood as estimating the m-fold self-convolution of the implementation map [MPF21]. More precisely, for $\phi, \psi : \mathsf{G} \to \mathbb{L}(\mathbb{C}^d)$ we can define a convolution operation as

$$\phi * \psi(g) = \mathbb{E}_{\tilde{g}}\phi(g\tilde{g}^{-1})\psi(\tilde{g}). \tag{5.3}$$

Note that this definition naturally generalizes, e.g., the discrete circular convolution on vectors in $\mathbb{C}^n$, which can be seen as an operation on functions on the finite group $(\mathbb{Z}_n, +) \to \mathbb{C}$. With the convolution (5.3), we can rewrite the averages of the RB sequences as

$$\begin{aligned}\mathbb{E}_g S_g &= \mathbb{E}_{g_1,g_2,\ldots,g_m}\phi(g_1^{-1}g_2^{-1}\cdots g_m^{-1})\phi(g_m)\cdots\phi(g_2)\phi(g_1)\\ &= \mathbb{E}_{h_1,h_2,\ldots,h_m}\phi(h_m^{-1})\phi(h_m h_{m-1}^{-1})\cdots\phi(h_2 h_1^{-1})\phi(h_1)\\ &= \phi^{*(m+1)}(\mathrm{id}),\end{aligned} \tag{5.4}$$

where the replacements $h_1 = g_1$ and $h_j = g_j h_{j-1}$ for $j \in \{2,\ldots,m\}$ have been made the second equality, id denotes the identity element of G and ϕ^{*k} denotes the k-fold convolution of ϕ with itself. In expectation the RB data $p(m)$ is thus a contraction defined by M and ρ of the $(m+1)$-fold self-convolution of ϕ evaluated at the identity element.

In the simplest instance of an RB protocol one can directly calculate this expression: namely, when G is a unitary 2-design, the targeted implementation is simply the action of G as quantum gates, and the noise in ϕ can be modelled by a single gate-independent quantum channel $\Lambda \in \mathrm{CPT}(\mathbb{C}^d)$. Denoting by $\mathcal{G}$ the (adjoint) action of g as the unitary channel $X \mapsto \mathcal{G}(X) = gXg^\dagger$, we have the noise model

$$\phi(g) = \Lambda \circ \mathcal{G} \,. \tag{5.5}$$

With this ansatz for ϕ we can calculate that

$$\mathbb{E}_{g \in \mathsf{G}^m} S_g = \phi^{*(m+1)}(\mathrm{id}) = \Lambda \left[\mathbb{E}_{g \in \mathsf{G}} \mathcal{G}^\dagger \Lambda \mathcal{G} \right]^m . \tag{5.6}$$

The operator $\mathrm{tw}_\mu : \mathbb{L}(\mathbb{C}^d) \to \mathbb{L}(\mathbb{C}^d)$, $\mathcal{X} \mapsto \mathbb{E}_{U \sim \mu}[\mathcal{U}\mathcal{X}\mathcal{U}^\dagger]$ appearing in (5.6) is the so-called *(channel) twirling map* and appears in different contexts in quantum information. If we write out the twirling map with the individual unitaries it reads

$$\mathrm{tw}_\mu(\mathcal{X}) = (\rho \mapsto \mathbb{E}_{U\sim\mu}[U\mathcal{X}(U^\dagger \rho U)U^\dagger]\,). \tag{5.7}$$

It becomes apparent that tw_μ is related to second moment operator $\mathcal{M}_\mu^{(2)}$, (2.16), by simple vector space isomorphisms. Recall that for a unitary 2-design μ Proposition 12 gives us an explicit description of $\mathcal{M}_\mu^{(2)}$. We can simply track the isomorphism to derive the following convenient expression.

Theorem 43 (Twirling of channels [Nie02; EAŻ05a]). *Let $\mathcal{X} \in \mathbb{L}(\mathbb{C}^d)$ be trace-preserving and μ be a unitary 2-design. Then*

$$\mathrm{tw}_\mu(\mathcal{X}) = \mathcal{D}_{p(\mathcal{X})} \,, \tag{5.8}$$

where $\mathcal{D}_p$ is the depolarizing channel (2.60) *and $p(\mathcal{X})$ is the effective depolarizing parameter defined in* (2.79).

Proof. First we note that any map $\mathcal{X} \in \mathbb{L}(\mathbb{C}^d)$ is uniquely determined by $(\mathcal{X} \otimes \mathrm{id})(\mathbb{F})$, which is a similar construction as the Choi-Jamiołkowski isomorphism. This isomorphism is given by $\mathrm{Tr}_{2,3}[(\mathcal{X} \otimes \mathrm{id})(\mathbb{F}) \otimes A] = \mathcal{X}(A)$, but its explicit form is not needed. Hence, we can make the isomorphisms between the twirling map tw_μ and the second moment operator $\mathcal{M}^{(2)}_\mu$ from (2.16) explicit by writing

$$\begin{aligned}
&(\mathrm{tw}_\mu(\mathcal{X}) \otimes \mathrm{id})(\mathbb{F}) \\
&= \mathbb{E}_{U\sim\mu}\left[(U\otimes \mathbb{1})\mathcal{X}\otimes \mathrm{id}\left((U^\dagger \otimes \mathbb{1})\mathbb{F}(U\otimes \mathbb{1})\right)(U^\dagger\otimes \mathbb{1})\right] \\
&= \mathbb{E}_{U\sim\mu}\left[(U\otimes \mathbb{1})\mathcal{X}\otimes \mathrm{id}\left((\mathbb{1} \otimes U)\mathbb{F}(\mathbb{1}\otimes U^\dagger)\right)(U^\dagger\otimes \mathbb{1})\right] \\
&= \mathbb{E}_{U\sim\mu}\left[(U\otimes U)\mathcal{X}\otimes \mathrm{id}\left(\mathbb{F}\right)(U^\dagger\otimes U^\dagger)\right] \\
&= \mathcal{M}^{(2)}_\mu(\mathcal{X}\otimes \mathrm{id}(\mathbb{F}))\,.
\end{aligned} \tag{5.9}$$

For μ a unitary 2-design, $\mathcal{M}^{(2)}_\mu$ takes the value of the moment operator of the Haar measure. Schur-Weyl duality, Theorem 10, tells us that

$$\mathcal{M}^{(2)}_\mu(\mathcal{X}\otimes \mathrm{id}(\mathbb{F})) \in \mathrm{span}\{\mathbb{1},\mathbb{F}\}\,. \tag{5.10}$$

Observing that $\mathcal{D}_0 \otimes \mathrm{id}(\mathbb{F}) = \mathbb{1}/d$ and trivially $\mathcal{D}_1 \otimes \mathrm{id}(\mathbb{F}) = \mathbb{F}$, we conclude that

$$\mathrm{tw}_\mu(\mathcal{X}) \in \mathrm{span}\{\mathcal{D}_0,\mathcal{D}_1\}\,. \tag{5.11}$$

Furthermore, one quickly checks that if $\mathcal{X}$ is trace-preserving so is $\mathrm{tw}_\mu(\mathcal{X})$. Hence, $\mathrm{tw}_\mu(\mathcal{X})$ is an affine combination of $\mathcal{D}_0$ and $\mathcal{D}_1$. Thus, $\mathrm{tw}_\mu(\mathcal{X}) = \mathcal{D}_p$ holds for some $p \in \mathbb{C}$, and it remains to determine p. One way forward is a straight-forward calculation using the expressions for the coefficients provided by Proposition 13. A shortcut is to calculate the effective depolarization of both sides. Due to the unitary invariance of $\mu_{\mathbb{S}(\mathbb{C}^d)}$, it follows from (2.69) that $F_{\mathrm{avg}}(\mathcal{X}) = F_{\mathrm{avg}}(\mathrm{tw}(\mathcal{X}))$ and correspondingly for the affinely related effective depolarization parameter $p(\mathcal{X}) = p(\mathrm{tw}(\mathcal{X}))$. Combined with $p(\mathcal{D}_p) = p$, (2.80), yields the theorem's assertion. □

Theorem 43 allows us to explicitly calculate the RB data model from (5.6). To this end, a short calculation reveals that $\mathcal{D}_p^m = \mathcal{D}_{p^m}$. With this we find the RB data model to be

$$\begin{aligned}
p(m) &= \mathrm{Tr}[\tilde{M}\Lambda\mathcal{D}_{p(\Lambda)^m}(\tilde{\rho})] \\
&= p(\Lambda)^m\,\mathrm{Tr}[\tilde{M}\Lambda(\tilde{\rho})] + (1-p(\Lambda)^m)\,\mathrm{Tr}[\tilde{M}\Lambda(\mathbb{1}/d)] \\
&= p(\Lambda)^m\,\mathrm{Tr}[\tilde{M}\Lambda(\tilde{\rho}-\mathbb{1}/d)] + \mathrm{Tr}[\tilde{M}\Lambda(\mathbb{1}/d)]\,,
\end{aligned} \tag{5.12}$$

with $\tilde{M}$ and $\tilde{\rho}$ denoting the potentially faulty implementation of the measurement M and initial state ρ.

If we define the so-called *SPAM constants*

$$A \coloneqq \mathrm{Tr}[\tilde{E}\Lambda(\tilde{\rho} - \mathbb{1}/d)] \qquad \text{and} \qquad B \coloneqq \mathrm{Tr}[\tilde{E}\Lambda(\mathbb{1}/d)], \tag{5.13}$$

Equation (5.12) yields the simple RB fitting model

$$p(m) = A\,p^m + B\,. \tag{5.14}$$

Thus, fitting a single exponential decay to the estimator $\hat{p}(m)$ yields estimates $\hat{p}$, $\hat{A}$ and $\hat{B}$ for the model parameters p, A and B. In particular, the estimated RB decay parameter $\hat{p}$ is an estimator for the effective depolarizing parameter $p(\Lambda)$ of the error channel Λ. Recall that the effective depolarizing parameter is affinely related to the average gate fidelity (2.69) via (2.79). From the RB decay parameter, we thus equivalently obtain an estimate for the average gate fidelity of the noise channel Λ as

$$\hat{F}_{\mathrm{avg}} = \left(1 - \frac{1}{d}\right)\hat{p} + \frac{1}{d}. \tag{5.15}$$

The resulting estimate of the average gate fidelity (2.69) is robust against SPAM errors, which only enter the SPAM constants A and B.

Deriving rigorous performance guarantees for the RB estimator $\hat{p}$ is involved: it requires the analysis of confidence region of the estimator $\hat{p}_g(m)$ of the probability (5.1) that is a random variable of the quantum measurement statistics and $\hat{p}(m)$ obtained by the subsampling of the sequences g. Furthermore, the error of these estimators for each m enters the errors of the fidelity estimator via the exponential fitting procedure. This step depends on the choice of algorithm and the estimated sequence lengths. Using the fact that $\hat{p}(m)$ is the mean estimator of a bounded random variable, one can use Hoeffding's inequality to derive confidence intervals for an overall sampling complexity that is independent of the number of qubits in the regime of high fidelity. Such bounds however are prohibitively large for practical implementations. A refined analysis by Wallman and Flammia [WF14] derived tighter bounds for short sequences and small number of qubits. However, bounds that are practical and scalable in the number of qubits require a careful analysis of the variance of the estimator $\hat{p}_g(m)$ over the choice of the random sequences. For G being the Clifford group, Helsen *et al.* [Hel+19a] work out explicit variance bounds for the estimator $\hat{p}_g(m)$ and derive sampling complexities for $\hat{p}(m)$ that are practical, independent of the number of qubits and scale favourable with the sequence length. To this end, they employed a refined representation theoretical analysis of the commutant of the 4-th order diagonal action of the Clifford group [HWW18; Zhu+16] in order to calculate the corresponding moment operator; an endeavour that is complicated by the fact that the Clifford group itself is not a unitary 4-design—the same complications arising in proving the recovery guarantee in Chapter 6. A rigorous analysis of a simplified fitting procedure was derived in Ref. [Har+19]. Therein (again using trivial bounds on the variance) the authors show that a ratio estimator for the infidelity $r = 1 - p$ that employs the estimates of $p(m)$ for two different sequence length has multiplicative error using an efficient number of samples again in the regime of high fidelity.

All of these performance guarantees indicate that in principle RB protocols can be efficiently scalable in the number of qubits. To ensure also an efficient classical pre-processing of the prototypical RB protocol it is important to have an efficiently tractable group structure so that the inverse of the gate sequence can be computed. For the essential example of the Clifford group, the Gottesman-Knill theorem, see e.g. Ref. [NC10], allows the efficient computation of the inverse of a sequence $g_m \cdots g_2 g_1$ in polynomial time (w.r.t. the number of qubits). Furthermore, since the Clifford group is a unitary 3-design [Web16; Zhu17], it meets the requirement of Theorem 43. For this reason the presented analysis applies to the Clifford group under the assumption of gate-independent noise.

It is natural to ask of additional examples of groups that constitute a unitary 2-design and are covered by the presented analysis without modifications. But as discussed in the introduction of Chapter 4, it has been established that these two requirements are already suprisingly restrictive. If one requires a family of 2-groups that can be constructed for an arbitrary number of qubits, one is left with subgroups of the Clifford group or $\mathrm{SU}(d)$ itself as the only examples [Ban+20; SK17; 6].

We provide more details how the analysis of the prototypical RB protocol can be generalized in the subsequent section. Now, we want to discuss another variant of RB that is of particularly interest for estimating the input data for our approach towards quntum process tomography.

Interleaved randomized benchmarking

The prototypical RB protocol estimates the effective depolarizing parameter or the average gate fidelity of the average error channel of a gate *set*. In contrast, *interleaved RB* protocols [Mag+12] allow one to extract the effective depolarizing parameter of *individual* gates from a group with respect to their ideal implementation provided the noise is sufficiently incoherent.

In an interleaved RB protocol one performs in addition to the standard RB protocol a modified version, where the random sequences are interleaved with the specific target gate. The second experiment yields estimates for the effective depolarization parameter of the error channel associated to the group concatenated with the error channel of the individual target gate. Under certain assumptions the effective depolarization parameter of the implementation of the target gate can be estimated from the decay parameters of both RB protocols.

Protocol 2 (Interleaved RB)**.** For $\mathsf{G} \subset \mathrm{U}(d)$ and a target gate $g_T \in \mathsf{G}$,

1. follow Protocol 1,
2. follow Protocol 1 but modify the sequences to be

$$g = (g_1, g_T, g_2, g_T, g_3, \dots, g_T, g_m), \tag{5.16}$$

where g_T is the target gate and $g_i \in \mathsf{G}$ for $i \in [m]$ are drawn uniformly at random. The inverse g_{inv} is also calculated w.r.t. the modified sequence g.

The output of the protocol are the decay parameters of both experiments.

For the analysis we again consider a 'mostly' gate-independent noise model and assume that G is a unitary 2-design. In the noise model we assume that the same noise channel $\Lambda \in \mathrm{CPT}(\mathcal{H})$ follows the ideal implementation of all gates but the target gate, i.e.,

$$\phi(g) = \Lambda \circ \mathcal{G} \tag{5.17}$$

for all $g \in \mathsf{G} \setminus \{g_T\}$. The first step of the protocol is the unmodified RB protocol. If we neglect that ϕ deviates from the form (5.17) on g_T, we can apply the analysis of the previous section for gate-independent noise and conclude that the protocol outputs and estimator for the effective depolarizing constant $p(\Lambda)$. For example, for a large group it is plausible to neglect the contribution of the noise associated to the g_T gate to the group average. It remains to analyse the second protocol. In analogy to (5.3) we can in general rewrite

$$\begin{aligned}
&\mathbb{E}_{g_1,\dots,g_m} S_g \\
&\quad = \mathbb{E}_{g_1,\dots,g_m} \phi(g_1^{-1} g_T^{-1} g_2^{-1} g_T^{-1} \dots g_m^{-1}) \phi(g_m) \phi(g_T) \cdots \phi(g_2) \phi(g_T) \phi(g_1) \\
&\quad = \mathbb{E}_{g_1,\dots,g_m} \phi(g_m^{-1}) \cdots \phi(g_3 g_2^{-1} g_T^{-1}) \phi(g_T) \phi(g_2 g_1^{-1} g_T^{-1}) \phi(g_T) \phi(g_1),
\end{aligned}$$

by substituting g_i with $g_i g_{i-1}^{-1} g_T^{-1}$ for all $i > 1$. Inserting the noise model (5.17) yields

$$\mathbb{E}_{g_1,\dots,g_m} S_g = \Lambda \left[\mathbb{E}_{g \in \mathsf{G}} \, \mathcal{G}^\dagger \mathcal{G}_T^\dagger \phi(g_T) \Lambda \mathcal{G} \right]^m . \tag{5.18}$$

This is the same expression as (5.6) with Λ replaced by $\mathcal{G}_T^\dagger \phi(g_T)\Lambda$. Hence, applying the same arguments as in the analysis of the standard RB protocol for unitary 2-designs yields a single-exponential fitting model with decay parameter estimating the effective depolarizing parameter $p(\mathcal{G}_T^\dagger \phi(g_T)\Lambda)$. The second part of the interleaved RB protocol, thus, returns an estimate of the effective depolarizing parameter or equivalently, via (5.15), of the fidelity of the error map $\mathcal{G}_T^\dagger \phi(g_T)$ of the target gate $\mathcal{G}_T$ concatenated with the error channel Λ.

From $p(\Lambda)$ and $p(\mathcal{G}_T^\dagger \phi(g_T))$ it is indeed possible to infer $p(\mathcal{G}_T^\dagger \phi(g_T))$. In meaningful practical regimes this however requires additional control with the unitarity of Λ [CWE19]: For sequences of unitary channels the infidelity of their composition can scale quadratically in the sequence length in leading order. In contrast, highly non-unitary channels will feature a close to linear scaling in the sequence length. Thus, using the unitarity one can derive bounds for fidelity measures of composite channels that exploit the linear scaling. We simply state the required bound without proof for interleaved RB:

Theorem 44 (Composite channel bound [CWE19]). *For any two quantum channels $\mathcal{X}, \mathcal{Y}$ it holds that*

$$\left| p(\mathcal{X}) - \frac{p(\mathcal{X}\mathcal{Y})p(\mathcal{Y})}{u(\mathcal{Y})} \right| \leq \sqrt{1 - \frac{p(\mathcal{Y})^2}{u(\mathcal{Y})}} \sqrt{1 - \frac{p(\mathcal{X}\mathcal{Y})^2}{u(\mathcal{Y})}} \tag{5.19}$$

With an estimate for the unitarity $\hat{u}(\Lambda)$, Theorem 44 allows for the estimation of the effective depolarizing constant and thus the average gate fidelity of the target gate by

$$\hat{F}_{\mathrm{avg}}(\phi(g_T), \mathcal{G}_T) = \frac{d-1}{d} \frac{\hat{p}(\mathcal{G}_T^\dagger \phi(g_T))\hat{p}(\Lambda)}{\hat{u}(\Lambda)} + \frac{1}{d} \tag{5.20}$$

up to a systematic error that is given by evaluating the right-hand side of (5.19). The systematic error is small in the regime where $u(\Lambda) \approx p(\Lambda)^2$ which is the case if Λ is decoherent. The unitarity of Λ can be estimated using variants of the RB protocol itself developed in Refs. [Wal+15; DHW19] or potentially following the proposal put forward in Section 6.5.

Alternatively, one can just assume that the error is sufficiently incoherent, i.e. that $|1 - p(\Lambda)^2/u(\Lambda)| \leq \epsilon$. Conditioned on this external belief, one obtains the simpler estimator

$$\hat{F}_{\mathrm{avg}}(\phi(g_T), \mathcal{G}_T) = \frac{d-1}{d} \frac{\hat{p}(\mathcal{G}_T^\dagger \phi(g_T))}{\hat{p}(\Lambda)} + \frac{1}{d} \tag{5.21}$$

that comes with a systematic error that is controlled in ϵ. Thereby, interleaved RB can be used to arrive at average-performance certificates of individual quantum gates.

We have already seen that for interleaved RB controlling the unitarity is helpful in deriving tighter error bounds. In addition, estimating the unitarity can also yield relevant worst-case performance bounds in terms of the average gate fidelities using Theorem 22.

Interleaved RB was proposed in Refs. [Mag+12; Gae+12] and demonstrated in practice. Already standard RB provides a trivial bound for individual gates of the group by simply attributing the average error to a single gate. In the original proposal of interleaved RB, the analysis does not allow for rigorous certificates that go significantly beyond this trivial bound for few qubits [CWE19]. A general bound by Kimmel *et al.* [Kim+14], was considerably refined using the unitarity by Carignan-Dugas *et al.* [CWE19]. Thereby it was established that if the error channel is sufficiently incoherent interleaved RB yields rigorous certificates for individual gates with reasonable error bars. There exist multiple variants of the interleaved RB scheme [Erh+19; She+16; HF17; Cha+17]. Another class of interleaved RB was introduced in Ref. [OWE19]. Here, the average gate fidelity of individual gates is inferred from measurements of random sequences of gates that are drawn from the symmetry group of the gate. The individual gates are not part of the group itself and are also not included in the inversion of the sequence.

A variant of interleaved RB arises when one does not implement the interleaved target gate g_T in the experimental sequence but still calculate the inverse with respect to the interleaved sequence (5.16) that includes g_T. Again for G a unitary 2-design and gate-independent noise $\phi(g) = \Lambda \circ \mathcal{G}$, we read-off from (5.18) that for the modified RB protocol it holds that

$$\mathbb{E}_{g_1,\dots,g_m} S_g = \Lambda [\mathbb{E}_{g \in \mathsf{G}} \mathcal{G}^\dagger \mathcal{G}_T^\dagger \Lambda \mathcal{G}]^m \,. \tag{5.22}$$

Hence, by the same argument that yielded (5.15), the output RB data is described by a single exponential decay with decay parameter $p(\mathcal{G}_T^\dagger \Lambda)$. Using the affine relation (5.15), we can thus infer relative average gate fidelities $F_{\mathrm{avg}}(\Lambda, \mathcal{G}_T)$ of the gate set's error channel with specific target gates $\mathcal{G}_T$ via a 'standard' RB experiment with a modified classical post-processing—interleaving 'in post'.

In Chapter 6 we study identification protocols for an unknown quantum channel $\mathcal{X}$ that use relative average gate fidelities $F_{\mathrm{avg}}(\mathcal{C}_i, \mathcal{X})$ as their input data, where $\mathcal{C}_i$ are Clifford gates. The motivation for studying such schemes is that these relative average gate fidelities can be estimated via RB experiments. In this way our identification protocols inherit the SPAM-robustness of the RB protocol rendering them semi-device dependent. Note that the magnitude of the relative average gate fidelities for tomographic protocols typically scales as d^{-1}, i.e. inversely proportionally to the Hilbert space dimension. Therefore one also expects a sampling complexity that scales exponentially in the number of qubits in order to be able to resolve the decay constants of the fast-decaying exponentials. Thereby the RB experiments are expected to become significantly more challenging than standard RB. At the same time quantum process tomography unavoidably involves a sampling complexity scaling polynomially in d—without making a more restrictive structure assumption than unit Kraus rank. Thus, also RB tomography is in any case more ressource-intense than standard RB.

The idea of combining different relative average gate fidelities obtained by interleaved RB schemes to acquire tomographic information about the error channel was proposed by Kimmel *et al.* in Ref. [Kim+14]. Kimmel *et al.* formulate a similar modified interleaved RB protocol as above that instead of interleaving in the inversion performs the inverse target gate after every error channel invocation. For completeness, we mention that for Pauli channels tomographic information can be efficiently obtained performing a character RB protocol on multiple qubits simultaneously [HFW20; FW20; HYF20; SSS21].

5.2 A general framework for randomized benchmarking

The exposition to RB in the previous section took a narrow perspective in order to be brief in the derivation of the results. In this section, we provide a more comprehensive review on the larger body of work on RB protocols and present the general guarantees that we derived together with collaborators in Ref. [9].

Randomized benchmarking was originally developed in a series of work focusing on the unitary group and Clifford gates [EAŻ05a; Lév+07; Kni+08; Dan+09; MGE11].[2] The early analyses used the gate-independent noise model (5.5), which we also assumed in the previous section. In many applications this is however a questionable assumption. After first perturbative approaches to derive the RB signal model under gate-dependent noise by Magesan *et al.* [MGE11; MGE12]

[2] Randomized benchmarking . . . tensor copies of the Clifford group] is taken from Ref. [4].

and Proctor *et al.* [Pro+19], Wallman rigorously derived the fitting model for unitary 2-designs in Ref. [Wal18].

Using the elegant description of the RB data as the m-fold convolution of the implementation map, recently proposed by Merkel *et al.* [MPF21], one can abstractly understand the result as follows: as the standard discrete circular convolution, the convolution operator of maps on a group can be turned into a (matrix) multiplication using a Fourier transform. This abstract Fourier transform for functions on the group is defined to be a function on the irreducible representations of the group. In the case of RB, this function is matrix-valued, and we observe matrix powers of the Fourier transforms for every irreducible representation superimposed by a linear map. For every irreducible representation, for sufficiently large m, the matrix powers are proportional to the mth power of the largest eigenvalue of the matrix-valued Fourier transform. Contributions from other eigenvalues are suppressed. In this sense RB is akin to the power method of numerical linear algebra but in Fourier space. A rigorous analysis requires to perturbatively bound the contribution of the sub-dominant eigenvalues. For unitary 2-groups the adjoint representation decomposes into two irreducible representations the trace representation and the unital part of the quantum channel. For close to trace-preserving maps the trace representation will only contribute a very slow decay, i.e. a constant contribution to the fit model, and the RB decay parameter is the dominant eigenvalue of the unital representation. Wallman [Wal18] derived norm bounds for the contribution of sub-dominant eigenvalues and showed that the contribution is exponentially suppressed with the sequence length. Furthermore, Wallman showed that there is a gauge choice of the gate set such that the decay parameter can be connected to the average gate fidelity of the average error channel over the gate set. For qubits this gauge was demonstrated to yield a physical gate set by Carignan-Dugas *et al.* [Car+18]. The physicality of this gauge is, however, in general not guaranteed, and we give a counter example in Ref. [9]. As discussed by Proctor *et al.* [Pro+17], this complicates the interpretation of the RB decay rates as related to average fidelities that have a clear physical interpretation.

While the Clifford gates are definitely a prominent use case in the benchmarking of digital quantum computers and will be also in the focus of our tomographic methods, more flexible RB protocols require analysing groups that are not a unitary 2-design. Randomized benchmarking protocols for other groups were developed in Refs. [Gam+12; CWE15; Cro+16; Has+18; BE18; FH18; CW15; Hel+19b]. These protocols, for example, allow inclusion of the T-gate in the gate set [CWE15] or characterization of leakage between qubit registers by using tensor copies of the Clifford group [Gam+12].

For the different RB protocols analytical guarantees were derived under varying sets of assumptions and with different standard of rigor. Together with collaborators we remedied this situation in Ref. [9]. We formulized required preconditions that ensure the proper functioning of RB protocols in general. At the core of this framework for analysing RB protocols is the following theorem establishing the functional form of the output of RB experiments.

Theorem 45 (RB data form)**.** *Consider an RB experiment with sequence length m, and gates drawn uniformly from a group G. We assume that the action of the group is implemented through a reference representation $\omega(g)$ that irreducibly decomposes as $\omega(g) = \bigotimes_{\lambda\in\Lambda} \sigma_\lambda^{\otimes n_\lambda}(g)$. Denote the corresponding (noisy) actual implementation $\phi(g)$. We assume that*

$$\mathbb{E}_{g\in\mathsf{G}} \left\|\omega(g) - \phi(g)\right\|_\diamond \leq \delta \leq \frac{1}{9}\,. \tag{5.23}$$

Then, the output data $p(m)$ of the RB experiment obeys the relation

$$\left| p(m) - \sum_{\lambda\in\Lambda} \mathrm{Tr}(A_\lambda M_\lambda)^m \right| \leq O(\delta^m)\,. \tag{5.24}$$

Here A_λ and M_λ are $n_\lambda \times n_\lambda$ matrices, with M_λ depending only on ϕ.

The theorem establishes that under the assumption that the actual experimental implementation is close to an ideal implementation on average over the group, the output data is described by a sum of (matrix) exponential decays, one for each irreducible representation of the reference representation. The proof of this theorem combines the matrix Fourier picture introduced in Ref. [MPF21] together with a careful analysis of the perturbation of the invariant subspaces of non-normal matrices. We refer to Ref. [9] for further details.

In Ref. [9] we also demonstrate that the formulation of Theorem 45 is general enough to reduce specific interleaved and non-uniform RB protocols to the same formulation and derive corollaries that establish the functional form for these protocols. Non-uniform RB is another practically important variation of RB where one does not draw the gates from the uniform but another distribution over the group [Kni+08; FH18; Boo+19; Pro+19]. For example, drawing the sequences randomly from the generating gates of the group, reduces the required sequence lengths [Pro+19].[3] It is an interesting open problem to extend and combine the framework of Ref. [9] with convergence guarantees of random circuits, such as the results discussed in Chapter 4.

[3] For example, drawing …was proposed in Ref. [Com+17]] is taken from Ref. [4] with minor modifications.

As the adjoint representation of other groups typically decomposes into multiple irreducible representation, RB data is expected to feature multiple decays in general. In order to isolate the different decays, variants of RB have been developed. These either rely on directly preparing a state that has high overlap with only one irreducible representation or cleverly combining data from different RB experiments to achieve the same effect. Many of these techniques can be understood as variants of the character benchmarking protocol developed by Helsen *et al.* [Hel+19b]. Character benchmarking uses inversions of the RB sequence not to the identity but randomly drawn gates from the group. In the classical post-processing, data from sequences of different end gates are linearly combined weighted according to the character formulas. Thereby, the data is projected onto the irreducible representation of the respective character and can be subsequently fitted by a single decay. This however comes with the disadvantage of a sampling complexity that is inversely proportional to the dimension of the irreducible representation. As worked out in Ref. [9] the inversion step and the decrease in sampling complexity can be avoided by using a larger measurement frame at the end of the RB sequence. The 'interleaving in post' protocol we sketched above makes use of the same strategy for the data selection. A flexible post-processing scheme for general RB type data and performance guarantees for fitting multiple decays are derived in Ref. [9].

It has been realized early in the development of RB that also other quantities can be measured by variants of the RB protocols. These include the unitarity [Wal+15; DHW19], measures for the losses, leakage, addressability and cross-talk [Gam+12; WBE15; WBE16]. Furthermore, RB of operations on the logical level of an error correcting quantum architecture was proposed in Ref. [Com+17].

5.3 Achievable sample complexity for relative average gate fidelities

Our approaches to channel tomography in Chapter 6 focuses on using relative average fidelities. One of the main motivations is that these data can be in principle extracted in RB experiments. Theoretical guarantees for the complete channel tomography protocol including an RB data aquisition is beyond the scope of the current thesis and ongoing work. To still work out the required resources of estimating multiple relative average fidelities with a given precision, we here analyse the sampling complexity of more direct measurement approaches of these quantities. Besides being of independent interest, these results enable us to at least to some degree discuss overall sampling complexities in the next chapter.

Importantly, we find that average fidelities with respect to unitary designs have a more favourable scaling in their error than a worst-case bound using the individual scaling indicates. In the following, we assume that $\mathcal{X} \in \mathrm{CPT}(\mathbb{C}^d)$ is a quantum channel. Via Proposition 18 and (2.67), the relative average fidelity $F_{\mathrm{avg}}(\mathcal{C}, \mathcal{X})$ is affinely related to $\langle \mathcal{C}, \mathcal{X} \rangle = d^2 \langle \mathfrak{J}(\mathcal{C}), \mathfrak{J}(\mathcal{X}) \rangle$. To suppress dimensional factors in the expression, we write in the following $(\mathcal{C}, \mathcal{X}) \coloneqq \langle \mathfrak{J}(\mathcal{C}), \mathfrak{J}(\mathcal{X}) \rangle$. Note that the proportionality constant between $F_{\mathrm{avg}}(\mathcal{C}, \mathcal{X})$ and $(\mathcal{C}, \mathcal{X})$ is in $O(1)$. Therefore, we aim at estimating the vector $y \in \mathbb{R}^m$ with entries $y_i \coloneqq (\mathcal{C}_i, \mathcal{X})$ for a set of unitary gates $\{\mathcal{C}_i\}_{i=1}^m$ with unitaries that constitute a unitary 1-design.

Independent projective measurements. We first consider the setting where every fidelity is directly estimated using a projective measurement. We assume that we have access to m_s copies of a quantum system in the state $\mathfrak{J}(\mathcal{X})$ and measure the dichotomic projector-valued measure (PVM) $\{\Pi_i = \mathfrak{J}(\mathcal{C}_i), \mathrm{Id} - \Pi_{\mathcal{C}}\}$. We count the frequency f_i of recording the outcome associated to Π_i in m_s repetitions. The frequency f_i, a binomial random variable with $p_i = (\mathcal{C}_i, \mathcal{X})$, directly yields an estimator $\hat{y}_i = f_i/m_s$ for y_i as $\mathbb{E}[f_i] = m_s p_i$ and $\mathrm{Var}[f_i] = m_s p_i (1 - p_i)$. Let us first look at the sampling complexity of the fidelity estimators for each $\mathcal{C}_i$ individually. Hoeffding's inequality implies that $\hat{y}_i$ is an ϵ-accurate estimator with confidence δ for y_i provided that $m_s \geq \frac{1}{2}\epsilon^{-2} \log(2/\delta)$. Let $y \in \mathbb{R}^m$ be a vector with entries y_i and $\hat{y}$ the corresponding estimator. Then, the Hoeffding bound and the union bound imply that $\|\hat{y} - y\|_{\ell_\infty} \leq \epsilon$ for $m_s \geq \frac{1}{2}\epsilon^{-2} \log[2m/\delta]$. We conclude that the total number of samples $M = m m_s$ for an ϵ-accurate estimate of $y \in \mathbb{R}^m$ in ℓ_2 norm is $M \geq \frac{1}{2} m^{3/2} \epsilon^{-2} \log(2m/\delta)$.

So far we have regarded the different entries of y as being unrelated. We can improve on the result of the Hoeffding bound by controlling the concentration of the estimator with the variance. To this end, we exploit the fact that the cumulative variance of the vector y with $\mathcal{C}_i$s distributed uniformly on the unitary group is of $O(md^{-2})$. More precisely, we make use of the following lemma.

Lemma 46. *Let $\{\mathcal{C}_i\}_{i=1}^m \subset \mathbb{L}(\mathbb{C}^d)$ be a unitary 1-design. Then, for all Hermiticity-preserving $\mathcal{X} \in \mathbb{L}(\mathbb{C}^d)$*

$$\frac{1}{m} \sum_{i=1}^{m} (\mathcal{C}_i, \mathcal{X}) = \frac{\mathrm{Tr}[\mathcal{X}(\mathbb{1}/d)]}{d^2} \,. \tag{5.25}$$

In particular, $\mathrm{Tr}[\mathcal{X}(\mathrm{Id}\,/d)] = 1$ for $\mathcal{X} \in \mathrm{CPT}(\mathbb{C}^d)$.

Proof. Let $\{\,|m\rangle\}_{m=1}^{d}$ denote a basis. Using the explicit form of the Choi matrix (2.64), (2.61), the 1-design assumption, and the expression for the moment-operator acting on unit-rank states of Lemma 13, we calculate

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^{m}\mathfrak{J}(\mathcal{C}_i) &= \frac{1}{d}\sum_{m,n=1}^{d}\mathcal{M}_{\mu}^{(1)}(\,|m\rangle\langle n\,|) \otimes |m\rangle\langle n\,| \\ &= \frac{1}{d^2}\sum_{m,n}\delta_{m,n}\mathbb{1}\otimes|m\rangle\langle n\,| = \frac{1}{d^2}\mathbb{1}_{d^2}.\end{aligned} \tag{5.26}$$

Thus,

$$\frac{1}{m}\sum_{m}(\mathcal{C}_i,\mathcal{X}) = \frac{1}{d^2}\langle\mathbb{1}_{d^2},\mathfrak{J}(\mathcal{X})\rangle = \frac{1}{d^2}\,\mathrm{Tr}[\mathfrak{J}(\mathcal{X})] = \frac{1}{d^2}\,\mathrm{Tr}[\mathcal{X}(\mathbb{1}/d)]\,. \tag{5.27}$$

□

Note the expression can also be directly read-off from Lemma 55 that we derive in the next chapter.

For the mean estimator $\hat{y}$, defined above via the frequency, the variance is insufficient to control tail bounds with high confidence. Replacing the mean estimator by a median of mean estimator, however, allows us to achieve exponentially high confidence.

Theorem 47 (Median of means for random vectors [LM19]). *Let $\mu_1,\ldots,\mu_m \in \mathbb{R}^d$ be i.i.d. random vectors with mean μ and co-variance matrix Σ and denote by $S_k := \frac{1}{k}\sum_{i=1}^{k}\mu_i$ the empirical mean from k i.i.d. samples. Take l such empirical means $S_{k,j}$, $j \in [l]$, that are (i.i.d.) copies of S_k and set*

$$\hat{\mu} = \mathrm{median}\{S_{k,1},\ldots,S_{k,l}\}\,. \tag{5.28}$$

Then, for $\delta \in (0,1)$ and $k = \lceil 8\log(1/\delta)\rceil$ and $m = kl$ with probability of at least $1-\delta$,

$$\|\hat{\mu}-\mu\|_{\ell_2} \leq \sqrt{\frac{32\,\mathrm{Tr}[\Sigma]\log(d/\delta)}{n}}\,. \tag{5.29}$$

The theorem follows as a corollary from [LM19, Theorem 2], see discussion on page 21 of Ref. [LM19]. The theorem uses an element-wise median of mean estimator. Even sharper concentration results can be proven for more sophisticated estimators [LM19].

Let $\mu \in \{0,1\}^m$ be a random vector recording the output in entry μ_i of measuring once the PVM with projector Π_i. Set again $p_i = \langle \mathcal{C}_i, \mathcal{X} \rangle = y_i$. The random vector μ has mean y and co-variance matrix $\Sigma_{ij} = \delta_{ij} p_i(1-p_i)$. By Lemma 46 $\mathrm{Tr}[\Sigma] \leq \sum_i p_i = m/d^2$. Repeat each PVM measurement m_s times and calculate the entry-wise median of mean estimator $\hat{\mu}$ with blocking as in Theorem 47. Then, Theorem 47 establishes that $\hat{\mu}$ is an ϵ-accurate estimator of y in ℓ_2-norm with confidence δ provided that the total number of samples $M = m m_s$ fulfils

$$M \geq 32 \frac{m^2}{d^2} \frac{1}{\epsilon^2} \log \frac{m}{\delta} \,. \tag{5.30}$$

Simultaneous projective measurements. For completeness we notice that this sampling complexity can be further improved by simultaneously measuring all fidelities. For $\{\mathcal{C}_i\}_{i=1}^m$ a unitary 1-design, Lemma 46 indicates that we can define a POVM with elements $\Pi_i = \frac{d^2}{m} \mathfrak{J}(\mathcal{C}_i)$ for $i \in [m]$ and $\sum_{i=1}^m \Pi_i = \mathrm{Id}_{d^2}$. We encode the outcome i of a single measurement with this POVM as the vector $(\delta_{ij})_{j=1}^m$. Thus, the measurement realizes a random vector $\mu \in \{0,1\}^m$ with entries following a single trial multinomial distribution, $\mathbb{E}[\mu_i] = p_i = \frac{d^2}{m}(\mathcal{C}_i, \mathcal{X})$ and co-variance matrix $\Sigma_{i,i} = p_i(1-p_i)$ for $i \in [m]$ and $\Sigma_{i,j} = -p_i p_j$ for $i \neq j$. For each measurement outcome, we calculate the random variable $\tilde{\mu} = \frac{m}{d^2}\mu$ that has mean y and its covariance matrix has trace $\mathrm{Tr}[\tilde{\Sigma}] \leq \frac{m^2}{d^4} \sum_i p_i = \frac{m^2}{d^4}$. Let $\hat{y}$ be the entry-wise median of mean estimator of M (i.i.d.) copies of y. Again by Theorem 47, $\hat{y}$ is an ϵ-accurate estimator for y in ℓ_2-norm with confidence δ provided that the total number of samples M fulfils

$$M \geq 32 \frac{m^2}{d^4} \frac{1}{\epsilon^2} \log \frac{m}{\delta} \,. \tag{5.31}$$

Designing RB protocols that SPAM-error robustly achieve the scaling derived in this section by direct ancillar-based measurements is subject of ongoing research.

6 Compressive randomized benchmarking tomography

Building on the merits of RB, outlined in the previous chapter, provides a promising route to devise semi-device-dependent tomographic schemes for quantum processes. The idea of Kimmel *et al.* [Kim+14] is to use multiple average gate fidelities of a unital quantum process with respect to different Clifford gates as the input to a reconstruction algorithm. This motivates the question of how many of such average fidelities are required to extract such tomographic information?[1]

In Section 6.1, we here first provide a more general characterization for the set of unitaries whose average gate fidelities allow for a stable reconstruction of unital quantum channels, Proposition 48. As already argued in the introduction, the most important use-case of process tomography is the characterization of coherent errors, i.e. the reconstruction of unitary processes. Simply counting the degrees of freedoms of an arbitrary unital process acting on a d-dimensional Hilbert space in comparison to a unitary channel indicates a square-root reduction from d^4 to d^2 in the information gain of the compressive reconstruction task. The main result of this chapter, Theorem 52 of Section 6.2, establishes that this square-root improvement in the number of required average gate fidelities with respect to Clifford gates is indeed achievable for a constraint least-square optimization. We furthermore derive the corresponding sampling complexity and prove optimality of our result using the results of Section 5.3. To this end, we derive an information theoretical lower-bound, Theorem 71. Finally, we establish a new characterization of the unitarity (2.91) in terms of the variance over average gate fidelities, Theorem 78.

[1] The results of this chapter were reported in the letter [1]. Most of the technical material that we present was previously made available as supplemental material to Ref. [1]. We here significantly reorganized the material of the letter and the supplemental material for a linear presentation. The work was conducted in close collaboration and with extensive support by Richard Kueng, Shelby Kimmel, Yi-Kai Liu, David Gross, Jens Eisert, and Martin Kliesch. Collectively, we acknowledge helpful discussions with Mateus Araújo, Steven T. Flammia, Christian Krumnow, Robin Harper, and Michał Horodecki.

6.1 Unital quantum channels

In this chapter, we take average gate fidelities (AGFs), $F_{\mathrm{avg}}(\mathcal{X},\mathcal{U})$ as defined in (2.69), with respect to different unitary quantum channels $\mathcal{U}$ as the input data to reconstruct a quantum process $\mathcal{X}$. The first question to ask is what quantum channels can in principle be reconstructed from this type of data. A first answer to this question can be found in Ref. [MW09] that provides a detailed analysis of the geometry of unital channels. There, it was shown that a quantum channel is unital if and only if it can be written as an affine combination of unitary gates.[2] *Affine* here means that the expansion coefficients sum to 1. Unlike *convex* combinations, they are, however, not restrict to being non-negative. Ref. [Kim+14] then showed that for many-qubit systems (i.e., $d = 2^n$), any unital and trace-preserving map is fully characterized by its AGFs (2.69) with respect to the Clifford group. The Clifford group constitutes a particularly important family of unitary gates that are featured prominently in state-of-the-art quantum architectures. Motivated by the result for Clifford gates, one can ask more generally: What are the subsets of unitary gates that span the set of unital and trace-preserving maps?

A general answer to this question can be given using the notion of unitary t-designs that we introduced in Definition 1. Recall that a unitary t-designs [Dan+09; GAE07] (and their state cousins, spherical t-designs [DGS77; Ren+04], respectively) are discrete subsets of the unitary group $\mathrm{U}(d)$ (resp., complex unit sphere) that are evenly distributed in the sense that their average reproduces the Haar (resp., uniform) measure over the full unitary group (resp., complex unit sphere) up to the tth moment. The multiqubit Clifford group forms a *unitary* 3*-design* [Zhu17; Web16; KG15]. For spherical designs, a close connection between informational completeness for quantum state estimation and the notion of a 2-design has been established in Ref. [Ren+04]; see also Refs. [Sco06; App05; GKK15]. A similar result holds for quantum process estimation, and provides an answer to the question of properly conditioned spanning sets. Indeed, the following result is essentially due to Ref. [Sco08]. Below we here give a concise proof in the form of the slightly more general Theorem 51.

Proposition 48 (Informational completeness and unitary designs). *Let $\{\mathcal{U}_k\}_{k=1}^N$ be the gate set of a unitary 2-design, represented as channels. Every unital and trace-preserving map $\mathcal{X}$ can be written as an affine combination $\mathcal{X} = \frac{1}{N}\sum_{k=1}^N c_k(\mathcal{X})\mathcal{U}_k$*

[2] There, it was shown …for unital channels.] is based on the result summary in the main text of Ref. [1] with verbatim adoptions.

of the $\mathcal{U}_k$'s. The coefficients are given by $c_k(\mathcal{X}) = C\, F_{\mathrm{avg}}(\mathcal{U}_k, \mathcal{X}) - \frac{C}{d} + 1$, where $C = d(d+1)(d^2-1)$.

Hence, every unital and trace-preserving map is uniquely determined by the AGFs with respect to an arbitrary unitary 2-design. Clifford gates are a particularly prominent gate set with this 2-design feature. However, its cardinality scales superpolynomially in the dimension d. For explicit characterizations, this is far from optimal. In certain dimensions there exist subgroups of the Clifford group with cardinality proportional to d^4 that also form a 2-design [Cha04; GAE07]—see also the discussion at the beginning of Chapter 4. More generally, order of $d^4 \log(d)$ Clifford gates drawn independent and identically distributed (i.i.d.) from the uniform distribution are an approximate 2-design [ABW09]. From Proposition 48, we expect that such randomly generated approximate 2-designs and the local circuit construction of Theorem 37 can be used for approximate reconstruction schemes for unital channels.

We here give an instructive proof of Proposition 48 and show that the linear span of the *unital* channels coincides with the linear span of the *unitary* ones, even if one restricts to the unitaries from a unitary 2-design.[3] We also link this finding to AGFs. On the way, we establish the simple formula of Proposition 48 that allows for the reconstruction of unital and trace-preserving maps from measured AGFs with respect to an arbitrary unitary 2-design, e.g. Clifford gates.

Recall that $\mathrm{L}_{\overline{\mathrm{u,tp}}}(H_d)$ denotes the linear hull of unital and trace-preserving maps acting on the space H_d of hermitian operators on a d-dimensional complex Hilbert space. At the heart of the reconstruction formula is the observation that every unitary 2-design constitutes a Parseval frame for $\mathrm{L}_{\overline{\mathrm{u,tp}}}(H_d)$. More abstractly, this observation stems from the general fact that irreducible unitary representations form Parseval frames on the space of endomorphisms of their representation space. For this reason it is instructive, to derive the connection explicitly in the 'natural' representation-theoretic language. We begin with formalizing the connection between irreducible representations and Parseval frames.

Lemma 49 (Irreps form a Parseval frame)**.** *Let $R : G \to \mathrm{L}(V)$ be an irreducible unitary representation of a group G. Then the set $\{\sqrt{\dim V}\, R(g)\}_{g\in G}$ forms a Parseval frame for the space $\mathrm{L}(V)$ equipped with the Hilbert-Schmidt inner product*

[3] We here give an instructive . . .] This section is Section F of the Supplemental material of Ref. [1]

$A, B \mapsto \mathrm{Tr}[A^\dagger B]$, *in the sense that*

$$T_G(A) := \dim(V) \int_G R(g)\,\mathrm{Tr}[R(g)^\dagger A]\, d\mu(g) = A \tag{6.1}$$

for all $A \in \mathrm{L}(V)$.

Proof. Since $\mathrm{L}(V)$ is generated as an algebra by $\{R(g)\}_{g\in G}$ (see e.g. [FH91, Proposition 3.29]), it suffices to show the statement for $A = R(g)$ with $g \in G$. Due to the invariance of the Haar measure, the map T_G is covariant in the sense that $T_G(R(g)B) = R(g)T_G(B)$ for all $B \in \mathrm{L}(V)$. In particular, for $B = \mathrm{Id}$, we thus get $T_G(R(g)\,\mathrm{Id}) = R(g)T_G(\mathrm{Id})$. With $\chi(g) = \mathrm{Tr}\,R(g)$ the character of the representation, we have

$$T_G(\mathrm{Id}) = \dim(V) \int_G R(g)\bar{\chi}(g)\, d\mu(g) = \mathrm{Id} \tag{6.2}$$

from the well-known expression for projection onto a representation space in terms of the character, see e.g. Ref. [FH91, Chapter 2.4]. Thus, we have established that $S_R(R(g)) = R(g)$ for all $g \in G$. □

Applying this lemma to unitary channels, we can derive the following expression for the orthogonal projection onto the linear hull of unital and trace-preserving maps.

Theorem 50. *Let* $\{\mathcal{U}_k\}_{k=1}^N$ *be a unitary* 2*-design. The orthogonal projection onto the linear hull of unital and trace-preserving maps* $\mathrm{L}_{\overline{\mathrm{u,tp}}}(H_d)$ *is give by*

$$P_{\overline{\mathrm{u,tp}}}(\mathcal{X}) = \frac{1}{N}\sum_{k=1}^{N} c_{\mathcal{U}_k}(\mathcal{X})\,\mathcal{U}_k \tag{6.3}$$

with coefficients

$$c_{\mathcal{U}}(\mathcal{X}) = C\,F_{\mathrm{avg}}(\mathcal{U},\mathcal{X}) - \frac{1}{d}\left(\frac{C}{d} - 1\right)\mathrm{Tr}(\mathcal{X}(\mathrm{Id}))\,, \tag{6.4}$$

where $C := d(d+1)(d^2-1)$.

Proof. Throughout the proof, we denote the unitary channel representing the unitary $U \in U(d)$ on space of Hermitian operators H_d by $\mathcal{U} : \rho \mapsto U\rho U^\dagger$. The vector space H_d is a direct sum of the space $\mathcal{K}_0$ of trace-less hermitian matrices, and of $\mathcal{K}_1 = \{z \operatorname{Id}\}_{z\in\mathbb{C}}$. The group of unitary channels acts trivially on $\mathcal{K}_1$, and irreducibly on $\mathcal{K}_0$. In particular, $\mathcal{U}$ is "block-diagonal" $\mathcal{U} = \mathcal{U}_0 \oplus 1$ with respect to this decomposition, where $\mathcal{U}_0 \in \mathrm{L}(\mathcal{K}_0)$ is the irreducible (d^2-1)-dimensional block. More generally, the projection of a map $\mathcal{X}$ onto the linear hull of unital and trace-preserving maps $\mathrm{L}_{\overline{\text{u,tp}}}(H_d)$ is of the form $\mathcal{X}_0 \oplus x_1$. The map $\mathcal{X}_0 \oplus x_1$ is trace-preserving and unital if and only if $x_1 = \operatorname{Tr}(\mathcal{X}(\operatorname{Id}/d)) = 1$. For the map $\mathcal{X} \in \mathrm{L}(H_d)$ we have

$$\operatorname{Tr}[\mathcal{U}^\dagger \mathcal{X}] = \operatorname{Tr}[\mathcal{U}_0^\dagger \mathcal{X}_0] + x_1. \tag{6.5}$$

Using this formula, Lemma 49 for the choice $V = \mathcal{K}_0$, and the fact that a group integral over a non-trivial irrep vanishes[4], we find

$$\begin{aligned}
&(d^2-1)\int_{U(d)} \mathcal{U} \operatorname{Tr}\left[\mathcal{U}^\dagger \mathcal{X}\right] d\mu(U) \\
&\quad= (d^2-1)\int_{U(d)} (\mathcal{U}_0 \oplus 1)(\operatorname{Tr}[\mathcal{U}_0^\dagger \mathcal{X}_0] + x_1)\, d\mu(U) \\
&\quad= (d^2-1)\int_{U(d)} \mathcal{U}_0(\operatorname{Tr}[\mathcal{U}_0^\dagger \mathcal{X}_0] + x_1)\, d\mu(U) \\
&\qquad\oplus (d^2-1)\int_{U(d)} (\operatorname{Tr}[\mathcal{U}_0^\dagger \mathcal{X}_0] + x_1)\, d\mu(U) \\
&\quad= \mathcal{X}_0 \oplus (d^2-1)x_1\,.
\end{aligned} \tag{6.6}$$

Hence, for $\mathcal{X} \in \mathrm{L}_{\overline{\text{u,tp}}}(H_d)$ we obtain the completeness relation

$$\int_{U(d)} \mathcal{U} \left((d^2-1)\operatorname{Tr}[\mathcal{U}^\dagger \mathcal{X}] + \frac{2-d^2}{d} \operatorname{Tr}[\mathcal{X}(\operatorname{Id})] \right) d\mu(U) = \mathcal{X}\,. \tag{6.7}$$

For $\mathcal{X}$ in the ortho-complement of $\mathrm{L}_{\overline{\text{u,tp}}}(H_d)$ the left-hand side of (6.7) vanishes. The expression, thus, defines the orthogonal projection $P_{\overline{\text{u,tp}}}$ onto $\mathrm{L}_{\overline{\text{u,tp}}}(H_d)$. The projection can be re-expressed in terms of the AGF. With the help of Proposition 18,

$$\operatorname{Tr}[\mathcal{U}^\dagger \mathcal{X}] = d(d+1)\, F_{\text{avg}}(\mathcal{U}, \mathcal{X}) - \operatorname{Tr}(\mathcal{X}(\operatorname{Id}))\,. \tag{6.8}$$

Hence,

$$P_{\overline{\text{u,tp}}}(\mathcal{X}) = \int_{U(d)} c_{\mathcal{U}}(\mathcal{X})\mathcal{U}\, d\mu(U), \tag{6.9}$$

[4] More explicitly, for $X \in H_d$ we can calculate $\int_{U(d)} \mathcal{U}(X)\, d\mu(U) = \mathcal{M}^{(1)}_{\mu_{U(d)}}(X) = \mathbb{O} \oplus 1$ using Theorem 14.

with expansion coefficients

$$\begin{aligned} c_{\mathcal{U}}(\mathcal{X}) &= d(d+1)(d^2-1)\, F_{\text{avg}}(\mathcal{U},\mathcal{X}) \\ &\quad - \frac{1}{d}\big((d+1)(d^2-1)-1\big)\, \operatorname{Tr}(\mathcal{X}(\mathrm{Id})) \\ &= C\, F_{\text{avg}}(\mathcal{U},\mathcal{X}) - \frac{1}{d}\left(\frac{C}{d}-1\right) \operatorname{Tr}(\mathcal{X}(\mathrm{Id})). \end{aligned}$$

Since the integrand in (6.9) is linear in $U^{\otimes 2} \otimes \bar{U}^{\otimes 2}$, the completeness relation continues to hold if the Haar integral is replaced by the average

$$\frac{1}{N}\sum_{k=1}^{N} c_{\mathcal{U}_k}(\mathcal{X})\,\mathcal{U}_k = P_{\overline{\text{u,tp}}}(\mathcal{X}) \tag{6.10}$$

over any unitary 2-design $\{\mathcal{U}_k\}_{k=1}^N$. □

In the proof, we have used that linear hull of the unital and trace-preserving maps $\mathrm{L}_{\overline{\text{u,tp}}}(H_d)$ is given by the space of block diagonal matrices $\mathrm{L}(\mathcal{K}_0) \oplus \mathrm{L}(\mathcal{K}_1)$. If $\mathcal{X}$ is not unital and trace-preserving, the image $\mathcal{X}_{\overline{\text{u,tp}}}$ will thus be equal to $\mathcal{X}$, with the off-diagonal blocks set to zero. In particular, the two-norm deviation of a map $\mathcal{X}$ from its projection onto $\mathrm{L}_{\overline{\text{u,tp}}}(H_d)$ is given by

$$\|\mathcal{X} - P_{\overline{\text{u,tp}}}(\mathcal{X})\|^2 = \frac{1}{d^3}\left(\|\mathcal{X}(\mathrm{Id})\|_2^2 + \|\mathcal{X}^\dagger(\mathrm{Id})\|_2^2 - \frac{2}{d}\operatorname{Tr}\left(\mathcal{X}(\mathrm{Id})\right)^2\right). \tag{6.11}$$

Based on the arguments used to establish Theorem 50, we can derive the following variant, which includes a converse statement.

Theorem 51 (Informational completeness and unitary designs). *Let $\{\mathcal{U}_k\}_{k=1}^N$ be a set of unitary channels. Then the following are equivalent:*

(i) Every unital and trace-preserving map $\mathcal{X}$ can be written as an affine combination $\mathcal{X} = \frac{1}{N}\sum_{k=1}^N c_k(\mathcal{X})\mathcal{U}_k$ of the $\mathcal{U}_k$, with coefficients given by $c_k(\mathcal{X}) = C\, F_{\text{avg}}(\mathcal{U}_k,\mathcal{X}) - \frac{C}{d} + 1$, where $C = d(d+1)(d^2-1)$.

(ii) The set $\{U_k\}_{k=1}^N$ forms a unitary 2-design.

Proof. To show that (ii) implies (i) we apply Theorem 50. From (6.7) we can read of that

$$\frac{1}{N}\sum_{k=1}^{N} c_k(\mathcal{X}) = \mathrm{Tr}[\mathcal{X}(\mathrm{Id}\,/d)] = 1. \tag{6.12}$$

Thus, the linear expansion of $\mathcal{X}$ in terms of the unitary 2-design is affine.

It remains to establish the converse statement. Let $\{\mathcal{U}_k\}_{k=1}^N$ be a set of unitary channels fulfilling

$$\frac{1}{N}\sum_{k=1}^{N}\mathcal{U}_k\left((d^2-1)\,\mathrm{Tr}[\mathcal{U}_k^\dagger\mathcal{X}] + 2 - d^2\right) = \mathcal{X} \tag{6.13}$$

for all $\mathcal{X} \in \mathrm{L}_{\mathrm{u,tp}}(H_d)$. A handy criterion for verifying that $\{\mathcal{U}_k\}_{k=1}^N$ is a unitary 2-design can be formulated in terms of its frame potential

$$P = \frac{1}{N^2}\sum_{k,k'=1}^{N} |\,\mathrm{Tr}(U_k^\dagger U_{k'})|^4, \tag{6.14}$$

where again U_k is the unitary matrix defining the unitary channel $\mathcal{U}_k$. A set of unitary gates is a unitary 2-design if and only if $P = 2$ [GAE07, Theorem 2]. In fact, Equation (6.13) allows one to calculate the frame potential as follows. Inserting $\mathcal{X} = 0 \oplus 1$ (the depolarizing channel), we find that

$$\frac{1}{N}\sum_{k=1}^{N}\mathcal{U}_k = 0 \oplus 1. \tag{6.15}$$

Note that this implies that the set $\{\mathcal{U}_k\}_{k=1}^N$ constitutes a unitary 1-design. Therefore, (6.13) takes the form

$$\frac{1}{N}\sum_{k=1}^{N}\mathcal{U}_k(d^2-1)\,\mathrm{Tr}[\mathcal{U}_k^\dagger\mathcal{X}] + 0 \oplus (2-d^2) = \mathcal{X} \tag{6.16}$$

for all $\mathcal{X} \in \mathrm{L}_{\mathrm{u,tp}}(H_d)$. Let the left-hand side of (6.16) define a linear operator $F : \mathcal{X} \mapsto F(\mathcal{X})$. Then (6.16) implies

$$\begin{aligned}\frac{1}{N}\sum_{k'=1}^{N}\mathrm{Tr}[\mathcal{U}_{k'}^\dagger F(\mathcal{U}_{k'})] &= \frac{d^2-1}{N^2}\sum_{k,k'=1}^{N} |\,\mathrm{Tr}(U_{k'}^\dagger U_k)|^4 + 2 - d^2\\ &= d^2 \end{aligned} \tag{6.17}$$

and hence

$$\frac{1}{N^2}\sum_{k,k'=1}^{N} |\operatorname{Tr}(U_{k'}^\dagger U_k)|^4 = 2. \tag{6.18}$$

This completes the proof. □

Note that for quantum channels, the affine expansion is *almost* convex in the sense that $c_k(\mathcal{X}) \geq 2 - d^2/N \geq -1/d^2$.

6.2 Unitary channels

Our main result for compressive Clifford RB tomography, focuses on the particular task of reconstructing multiqubit unital channels that are close to being unitary, i.e., well approximated by a channel of Kraus rank equal to 1.[5] We have already encountered techniques for low-rank matrix reconstruction in the task of blind tomography in Chapter 3. We will here again leverage compressed sensing results [FHB01; Gro+10; Gro11; Fla+12; Kab+16; BKD14] in order to exploit the low Kraus-rank and reduce the number of AGFs required to uniquely reconstruct an unknown unitary gate. We first give a more concise summary of our technical results. A considerably more refined version is provided together with the proof below.

The setting is as follows: Suppose we are given a list of m AGFs

$$f_i = F_{\mathrm{avg}}(\mathcal{C}_i, \mathcal{X}) + \epsilon_i \tag{6.19}$$

between the unknown unitary gate $\mathcal{X}$ and Clifford gates $\mathcal{C}_i$. The gates $\mathcal{C}_i$ are chosen uniformly at random. The estimates f_i are possibly corrupted by additive noise ϵ_i. In order to reconstruct $\mathcal{X}$ from these observations, we propose to perform a least-squares fit over the set of unital quantum channels, i.e.

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{m} [F_{\mathrm{avg}}(\mathcal{C}_i, \mathcal{Z}) - f_i]^2 \\ \text{subject to} \quad & \mathcal{Z} \text{ is a unital quantum channel.} \end{aligned} \tag{6.20}$$

[5] Our main result ...rank-1 measurements.] is based on the corresponding result summary in the main text of Ref. [1] with verbatim adoption.

We emphasize that this is an efficiently (in d) solvable convex optimization problem. The feasible set is convex since it is the intersection of an affine subspace (unital and trace-preserving maps) and a convex cone (completely positive maps).

Valid for multiqubit gates ($d = 2^n$), our main result states that this reconstruction procedure is guaranteed to succeed with exponentially high probability, provided that the number m of AGFs is proportional (up to a $\log(d)$-factor) to the number of degrees of freedom in a general unitary gate. The error of the reconstructed channel is measured with the Frobenius norm in Choi representation $\|\cdot\|$.

Theorem 52 (Recovery guarantee for unitary gates)**.**
Fix the dimension $d = 2^n$. Then,

$$m \geq c d^2 \log(d) \tag{6.21}$$

noisy AGFs with randomly chosen Clifford gates suffice with high probability (of at least $1 - \mathrm{e}^{-\gamma m}$) to reconstruct any *unitary quantum channel $\mathcal{X}$ via (6.20). This reconstruction is stable in the sense that the minimizer $\mathcal{Z}^\sharp$ of (6.20) is guaranteed to obey*

$$\left\|\mathcal{Z}^\sharp - \mathcal{X}\right\| \leq \tilde{C}\frac{d^2}{\sqrt{m}}\|\epsilon\|_{\ell_2}\,. \tag{6.22}$$

The constants $\tilde{C}, c, \gamma > 0$ are independent of d.

The theorem considers the case of exactly unitary gates. A more general version—Theorem 61—shows that the result can be extended to cover approximately unitary channels. Furthermore, the general version treats also an optimization of the ℓ_1 norm instead of the ℓ_2 norm in (6.20), resulting in a slightly stronger error bound. Before jumping into the proof, we briefly highlight the implications of the theorem.

Equation (6.22) shows the protocol's inherent stability to additive noise. This stability combined with the robustness of randomized benchmarking against SPAM errors results in an estimation procedure that is potentially more resource intensive but considerably less susceptible to experimental imperfections and systematic errors than many other reconstruction protocols [FL11; Fla+12; Kli+19].

The main feature of the theorem is that it achieves the desired quadratic improvement (up to a log-factor) over the minimal number of AGFs required for a direct reconstruction via linear inversion for the case of noiseless measurements.

It does not directly give rise to an overall sampling complexity, say in terms of the number of required channel invocations. To this end, one needs to consider the number of samples required to obtain the AGFs and to suppress the effect of the measurement noise ϵ in the reconstruction error (6.22). For randomized benchmarking setups, a fair accounting of all involved errors is still beyond the scope of the current work. But in order to build evidence that the scaling of the noise term in our reconstruction error (6.22) is essentially optimal, we consider the conceptually simpler measurement setting where the channel's Choi state is measured directly. The average fidelities can then be measured, e.g., with the POVMs described in Section 5.3. In Section 6.3, we prove upper and lower bounds to the minimum number of channel uses sufficient for a reconstruction via algorithm (6.20) with reconstruction error (6.22) bounded by $\varepsilon_{\text{rec}} > 0$. This number of channel uses scales as $d^4/\varepsilon_{\text{rec}}^2$ up to log-factors when the fidelities are measured in separate experiments akin to interleaved RB experiments. In order to establish a lower bound, we extend information theoretic arguments from Ref. [Fla+12] to rank-1 measurements.

Establishing the recovery guarantee

The AGFs can be interpreted[6] as expectation values of certain observables, which are unit rank projectors onto directions that correspond to elements of the Clifford group. In contrast, most previous work on tomography via compressed sensing features observables that have full rank, e.g., tensor products of Pauli operators. Since we now want to utilize observables that have unit rank, a different approach is needed. One approach [KL17] is to use strong results from low-rank matrix reconstruction and phase retrieval [CSV13; CL14; GKK15; KRT15; Kab+16]. These methods [KRT15; Kab+16] require measurements that look sufficiently random and unstructured, in that their fourth-order moments are close to those of the uniform Haar measure. The multiqubit Clifford group, however, does constitute a 3-design, but not a 4-design. In Ref. [KL17], this discrepancy is partially remedied by imposing additional constraints (a "nonspikiness condition"; see also Ref. [KL18]) on the unitary channels to be reconstructed. In turn, their result also required these constraints to be included in the algorithmic reconstruction which renders the algorithm impractical.[7] Moreover, important

[6] The AGFs can be interpreted …for the task at hand.] is taken from the main text of Ref. [1].

[7] The cardinality of the Clifford group grows superpolynomially in the Hilbert space dimension d. Therefore, the non-spikiness with respect to the Clifford group quickly corresponds to a demanding number of constraints. In fact, about 10^8 and 10^{13} constraints are already required for 3-qubits and 4-qubits, respectively.

classes of channels, e.g. Pauli channels, do not satisfy this condition in general. Here, we overcome these issues by appealing to recent works that fully characterize the fourth moments of the Clifford group [Zhu+16; HWW18]. In order to apply these results, we developed an integration formula for fourth moments over the Clifford group in Chapter 4. This formula is analogous to the integration over the unitary group know as Collins's calculus with Weingarten functions [Col03]. Equipped with this new representation-theoretic technique, we now show that the deviation of the Clifford group from a unitary 4-design is—in a precise sense—mild enough for the task at hand.

To this end, recall the following notation:[8] It will again be convenient to work with the differently normalized version of the Hilbert-Schmidt inner product on $\mathrm{L}(H_d)$ that coincides with the Hilbert-Schmidt inner-product on the normalized Choi states :

$$(\mathcal{X}, \mathcal{Y}) := \frac{1}{d^2}\langle \mathcal{X}, \mathcal{Y}\rangle = \langle \mathfrak{J}(\mathcal{X}), \mathfrak{J}(\mathcal{Y})\rangle \,. \tag{6.23}$$

Most of our results feature the corresponding Hilbert-Schmidt norm $\|\cdot\|$ on $\mathrm{L}(H_d)$ (the Frobenius in Choi representation). This norm is naturally induced on by the average gate fidelity (AGF) via Theorem 18.

In the centre of the proof of the recovery guarantee are the moments of the following random variable that encodes the measurement data: For a map $\mathcal{T} : H_d \to H_d$ we define the random variable

$$S_{\mathcal{T}} = d^2(\mathcal{T}, \mathcal{U})\,, \tag{6.24}$$

where $\mathcal{U}$ is a unitary channel $\mathcal{U}(X) = UXU^\dagger$ with U either chosen uniformly at random from the full unitary group $\mathrm{U}(d)$, or the Clifford group $\mathrm{Cl}(d)$, depending on the context. The main technical ingredients for the proofs of the recovery guarantees are an expression for the second and fourth moment of $S_{\mathcal{T}}$. To this end, we make use of the integration formula for the first four moments over the Clifford group of Chapters 2 and 4. We then derive an explicit expression for the second moment of $S_{\mathcal{T}}$ and an upper bound on the fourth moment of $S_{\mathcal{T}}$ in the following sections. These bounds are essential prerequisites for applying strong techniques from low-rank matrix reconstruction to subsequently prove our recovery guarantee, Theorem 52, for unitary gates.

[8] To this end, …directly follows.] The remainder of this section has been published as supplemental material to Ref. [1]. We made modifications to unify the notation with the rest of the monograph.

The second moment. The main result of this section is the following expression for the second moment of $S_{\mathcal{T}}$ defined in (6.24). We shall use this statement multiple times in the proofs of our main results.

Lemma 53 (The 2-nd moment for $\mathrm{U}(d)$)**.** *Let $\mathcal{T} : H_d \to H_d$ be a map. Then*

$$\begin{aligned}&\mathbb{E}_{U\sim\mu_{\mathrm{U}(d)}}[S_{\mathcal{T}}^2]\\ &\quad= \frac{1}{d^2-1}\left\{d^2\left\|\mathcal{T}\right\|^2 + \operatorname{Tr}(\mathcal{T}(\mathrm{Id}))^2 - \frac{1}{d}\left(\left\|\mathcal{T}(\mathrm{Id})\right\|_{\mathrm{F}}^2 + \left\|\mathcal{T}^\dagger(\mathrm{Id})\right\|_{\mathrm{F}}^2\right)\right\},\end{aligned} \tag{6.25}$$

for $S_{\mathcal{T}}$ defined in (6.24).

For a trace-annihilating and Id-annihilating map, one arrives at a much simpler expression. A (hermicity-preserving) map $\mathcal{T} \in L(H_d)$ is *trace-annihilating* and Id-*annihilating* if $\mathcal{T}(\mathrm{Id}) = 0$ and $\operatorname{Tr}[\mathcal{T}(X)] = 0$ for all $X \in H_d$, respectively. We denote the set of trace- and Id-annihilating maps by $\mathrm{V}_{\mathrm{u,tp,0}} \subset \mathrm{L}(H_d)$.

Corollary 54 (Expression for trace- and Id-annihilating maps)**.** *Let $\mathcal{T} \in \mathrm{V}_{\mathrm{u,tp,0}}$. Then the second moment of $S_{\mathcal{T}}$ is*

$$\mathbb{E}_{U\sim\mu_{\mathrm{U}(d)}}[S_{\mathcal{T}}^2] = \frac{d^2}{d^2-1}\left\|\mathcal{T}\right\|^2. \tag{6.26}$$

Proof. This follows directly from Lemma 53 and the observation that $\mathcal{T}$ being trace-annihilating translates to $\operatorname{Tr}(\mathcal{T}(\mathrm{Id}))) = 0$ and $\left\|\mathcal{T}^\dagger(\mathrm{Id})\right\|_{\mathrm{F}} = 0$ and $\mathcal{T}$ being Id-annihilating further requires $\left\|\mathcal{T}(\mathrm{Id})\right\|_{\mathrm{F}} = 0$. □

Before proving Lemma 53, we derive a general expression for the kth moment of $S_{\mathcal{T}}$. To this end, recall that by Choi's theorem an endomorphism $\mathcal{T}$ of H_d (i.e. a hermiticity preserving map) can be decomposed as

$$\mathcal{T}(X) = \sum_{i=1}^{r} \lambda_i T_i X T_i^\dagger, \tag{6.27}$$

where $\lambda_i \in \mathbb{R}$ and $T_1, \dots, T_r$ are linear operators with unit Frobenius norm. In this decomposition, the random variable $S_{\mathcal{T}}$ from (6.24), with $\mathcal{U}(X) = UXU^\dagger$ takes the form

$$S_{\mathcal{T}} = d^2(\mathcal{T}, \mathcal{U}) = \sum_{i=1}^{r} \lambda_i |\operatorname{Tr}(U^\dagger T_i)|^2 \tag{6.28}$$

and its kth moment can be expressed as follows:

Lemma 55 (kth moment of $S_{\mathcal{T}}$)**.** *For $k \in \mathbb{N}$ and T_i defined by (6.27) we have*

$$\begin{aligned}&\mathbb{E}_{U\sim\mu_{\mathrm{U}(d)}}[S_{\mathcal{T}}^k]\\&=\sum_{i_1,\dots,i_k=1}^{r}\lambda_{i_1}\cdots\lambda_{i_k}\frac{1}{k!}\sum_{\tau\in\mathfrak{S}_k}\sum_{\lambda\vdash k,\,l(\lambda)\le d}\frac{d_\lambda}{D_\lambda}\operatorname{Tr}\left[\bigotimes_{j=1}^{k}T_{i_{\tau(j)}}^{\dagger}P_\lambda\bigotimes_{j=1}^{k}T_{i_j}\right].\end{aligned}\tag{6.29}$$

Proof. We can rewrite the kth unitary moment of $S_{\mathcal{T}}$ as

$$\begin{aligned}&\mathbb{E}_{U\sim\mu_{\mathrm{U}(d)}}[S_{\mathcal{T}}^k]\\&=\mathbb{E}_U\sum_{i_1,\dots,i_k=1}^{r}\lambda_{i_1}\cdots\lambda_{i_k}|\operatorname{Tr}(U^\dagger T_{i_1})|^2\cdots|\operatorname{Tr}(U^\dagger T_{i_k})|^2\\&=\mathbb{E}_U\sum_{i_1,\dots,i_k=1}^{r}\lambda_{i_1}\cdots\lambda_{i_k}\operatorname{Tr}\left[\bigotimes_{j=1}^{k}T_{i_j}^{\dagger}\,U^{\otimes k}\right]\operatorname{Tr}\left[U^{\dagger\otimes k}\bigotimes_{j=1}^{k}T_{i_j}\right]\\&=\sum_{i_1,\dots,i_k=1}^{r}\lambda_{i_1}\cdots\lambda_{i_k}\sum_{m,n=1}^{d^k}\langle m|\bigotimes_{j=1}^{k}T_{i_j}^{\dagger}T_{\Delta_d^k}(|m\rangle\langle n|)\bigotimes_{j=1}^{k}T_{i_j}\,|n\rangle\end{aligned}\tag{6.30}$$

where in the last line we evaluated the trace in an orthonormal basis $\{\,|m\rangle\mid m\in\{1,\dots,d^k\}\}$ for $(\mathbb{C}^d)^{\otimes k}$. Using the expression for $T_{\Delta_d^k}=\mathcal{M}_{\mu_{\mathrm{U}(d)}}^{(k)}$ of Theorem 14 we get

$$\begin{aligned}&\mathbb{E}_{U\sim\mu_{\mathrm{U}(d)}}[S_{\mathcal{T}}^k]\\&=\sum_{i_1,\dots,i_k=1}^{r}\lambda_{i_1}\cdots\lambda_{i_k}\frac{1}{k!}\sum_{\tau\in\mathfrak{S}_k}\sum_{\lambda\vdash k,\,l(\lambda)\le d}\frac{d_\lambda}{D_\lambda}\\&\qquad\times\operatorname{Tr}\left[\pi_k(\tau)\bigotimes_{j=1}^{k}T_{i_j}^{\dagger}\pi_k(\tau^{-1})P_\lambda\bigotimes_{j=1}^{k}T_{i_j}\right]\\&=\sum_{i_1,\dots,i_k=1}^{r}\lambda_{i_1}\cdots\lambda_{i_k}\frac{1}{k!}\sum_{\tau\in\mathfrak{S}_k}\sum_{\lambda\vdash k,\,l(\lambda)\le d}\frac{d_\lambda}{D_\lambda}\operatorname{Tr}\left[\bigotimes_{j=1}^{k}T_{i_{\tau(j)}}^{\dagger}P_\lambda\bigotimes_{j=1}^{k}T_{i_j}\right].\end{aligned}\tag{6.31}$$

□

Proof of Lemma 53. We evaluate the expression of Lemma 55 for the case $k=2$. To this end recall that the irreducible representations of $\mathfrak{S}_2$ are the symmetric

($\ydiagram{2}$) and antisymmetric representation ($\ydiagram{1,1}$). The central projections are given by $P_{\ydiagram{2}} = \frac{1}{2}(1+\mathbb{F})$ and $P_{\ydiagram{1,1}} = \frac{1}{2}(1-\mathbb{F})$, see Lemma 9, where $\mathbb{F}$ is the bipartite flip operator $\mathbb{F} : (\mathbb{C}^d)^{\otimes 2} \to (\mathbb{C}^d)^{\otimes 2}$, $|x\rangle \otimes |y\rangle \mapsto |y\rangle \otimes |x\rangle$. The dimensions are $d_{\ydiagram{2}} = d_{\ydiagram{1,1}} = 1$, $D_{\ydiagram{2}} = \frac{d(d-1)}{2}$ and $D_{\ydiagram{1,1}} = \frac{d(d+1)}{2}$. For $A, B \in H_d^{\otimes 2}$ we introduce the following shorthand notation

$$\Gamma_{AB} \coloneqq \sum_{i,j}^{r} \lambda_i \lambda_j \operatorname{Tr}\left[A(T_i^\dagger \otimes T_j^\dagger) B (T_i \otimes T_j)\right]. \tag{6.32}$$

Rearranging the terms in the first statement of the Lemma 55 then yields

$$\begin{aligned}
&\mathbb{E}_{U\sim\mu_{\mathrm{U}(d)}}[S_{\mathcal{T}}^2] \\
&\quad= \frac{1}{4}\left\{\left[\frac{1}{D_{\ydiagram{2}}} + \frac{1}{D_{\ydiagram{1,1}}}\right][\Gamma_{\mathrm{Id}\,\mathrm{Id}} + \Gamma_{\mathbb{F}\mathbb{F}}] + \left[\frac{1}{D_{\ydiagram{2}}} - \frac{1}{D_{\ydiagram{1,1}}}\right][\Gamma_{\mathbb{F}\,\mathrm{Id}} + \Gamma_{\mathrm{Id}\,\mathbb{F}}]\right\} \\
&\quad= \frac{1}{d^2-1}\left\{\Gamma_{\mathrm{Id}\,\mathrm{Id}} + \Gamma_{\mathbb{F}\mathbb{F}} - \frac{1}{d}\left(\Gamma_{\mathrm{Id}\,\mathbb{F}} + \Gamma_{\mathbb{F}\,\mathrm{Id}}\right)\right\}.
\end{aligned} \tag{6.33}$$

The four Γ-terms can be evaluated explicitly. For the first term, we obtain

$$\begin{aligned}
\Gamma_{\mathrm{Id}\,\mathrm{Id}} &= \sum_{i,j=1}^{r} \lambda_i \lambda_j \|T_i\|_{\mathrm{F}}^2 \|T_j\|_{\mathrm{F}}^2 \\
&= \left(\sum_i \lambda_i \operatorname{Tr}(T_i \,\mathrm{Id}\, T_i^\dagger)\right)^2 = \operatorname{Tr}(\mathcal{T}(\mathrm{Id}))^2.
\end{aligned} \tag{6.34}$$

The second terms reads

$$\Gamma_{\mathbb{F}\mathbb{F}} = \sum_{i,j=1}^{r} \lambda_i \lambda_j |\operatorname{Tr}(T_i^\dagger T_j)|^2 = d^2 \|\mathcal{T}\|^2 \tag{6.35}$$

and the third term can be written as

$$\Gamma_{\mathbb{F}\,\mathrm{Id}} = \sum_{i,j=1}^{r} \lambda_i \lambda_j \operatorname{Tr}\left(T_i^\dagger T_i T_j^\dagger T_j\right) = \left\|\mathcal{T}^\dagger(\mathrm{Id})\right\|_{\mathrm{F}}^2. \tag{6.36}$$

Moreover, a computation that closely resembles this reformulation yields $\Gamma_{\mathrm{Id}\,\mathbb{F}} = \|\mathcal{T}(\mathrm{Id})\|_{\mathrm{F}}^2$ and the claim follows. □

A fourth moment bound. The main result of this section is an upper bound for the fourth moment of $S_{\mathcal{T}}$ when $\mathcal{U}$ is a Clifford operation drawn uniformly at random. To gain some intuition, let us first derive an upper bound on the fourth moment taken with respect to the full unitary group. Note that a similar bound for the full unitary group has already been derived in Ref. [KL17].

Lemma 56 (4th moment bound for $\mathrm{U}(d)$)**.** *Let $\mathcal{T} : H_d \to H_d$ be a map. Then for $S_{\mathcal{T}}$ defined in* (6.24)

$$\mathbb{E}_{U\sim\mu_{\mathrm{U}(d)}}[S_{\mathcal{T}}^4] \leq C\,\|\mathfrak{J}(\mathcal{T})\|_1^4 \tag{6.37}$$

with some constant $C > \frac{1}{3}$ independent of the dimension d.

Proof. Applying Cauchy-Schwarz to an individual summand on the right-hand side of Lemma 55 yields for all k

$$\begin{aligned}\left|\mathrm{Tr}\left[\bigotimes_{j=1}^{k} T_{i_{\tau(j)}}^\dagger P_\lambda \bigotimes_{j=1}^{k} T_{i_j}\right]\right| &\leq \left\|P_\lambda \bigotimes_{j=1}^{k} T_{i_{\tau(j)}}\right\|_{\mathrm{F}} \left\|P_\lambda \bigotimes_{j=1}^{k} T_{i_j}\right\|_{\mathrm{F}} \\ &\leq \left\|\bigotimes_{j=1}^{k} T_{i_{\tau(j)}}\right\|_{\mathrm{F}} \left\|\bigotimes_{j=1}^{k} T_{i_j}\right\|_{\mathrm{F}} \\ &= \prod_{j=1}^{k} \|T_{i_j}\|_{\mathrm{F}}^2 ,\end{aligned} \tag{6.38}$$

which is independent of the permutation $\tau \in \mathfrak{S}_k$. We may therefore conclude

$$\mathbb{E}_{U\sim\mu_{\mathrm{U}(d)}}[S_{\mathcal{T}}^k] \leq \sum_{i_1,\dots,i_k=1}^{r} \prod_{j=1}^{k} |\lambda_{i_j}|\,\|T_{i_j}\|_{\mathrm{F}}^2 \sum_{\lambda\vdash k,\ l(\lambda)\leq d} \frac{d_\lambda}{D_\lambda}. \tag{6.39}$$

From Theorem 35 we observe that for $k = 4$

$$\sum_{\lambda\vdash 4,\ l(\lambda)\leq d} \frac{d_\lambda}{D_\lambda} \leq \frac{C}{d^4} \tag{6.40}$$

for some constant $C > \frac{1}{3}$ independent of d. Thus, (6.39) implies the desired bound. □

In an analogous way we can derive a sufficient bound on the fourth moment of $S_{\mathcal{T}}$ when the average is performed over the Clifford group. The result will be stated in Lemma 60. To get the correct dimensional pre-factors in the bound, we have to rely on particular properties of the projection Q of (4.1) appearing in the representation theory of the fourth order diagonal action of the Clifford group in Theorem 34. The following technical result takes care of this issue.

Lemma 57 (Properties of the projection Q)**.** *For $\{T_l\}_{l=1}^r \subset \mathrm{L}(\mathbb{C}^d)$ and Q defined in* (4.1)

$$\left\| Q \bigotimes_{j=1}^4 T_{i_j} Q \right\|_{\mathrm{F}} \leq \frac{1}{d} \prod_{j=1}^4 \|T_{i_j}\|_{\mathrm{F}} \,. \tag{6.41}$$

This bound is tight. In fact, one can show that it is saturated if all T_i's are chosen to be the same stabilizer state. The proof of Lemma 57 requires two other properties of the multi-qubit Pauli matrices $W_1, \dots, W_{d^2}$. The first property is summarized by the following lemma.

Lemma 58 (Magnitude of multi-qubit Pauli matrices)**.** *For $A, B \in \mathrm{L}(\mathbb{C}^d)$,*

$$\mathrm{Tr}(W_j A W_k B) \leq \|A\|_{\mathrm{F}} \|B\|_{\mathrm{F}} \tag{6.42}$$

for all $j, k \in \{1, \dots, d^2\}$.

Proof. This statement follows directly from the Cauchy-Schwarz inequality and the unitary invariance of the Frobenius norm:

$$\begin{aligned} \mathrm{Tr}\left(W_j A W_k B\right) &= \left(B^\dagger, W_j A W_k\right) \\ &\leq \|B^\dagger\|_2 \|W_j A W_k\|_2 = \|B\|_2 \|A\|_2 . \end{aligned} \tag{6.43}$$

□

The second property is that the two multi-qubit flip operator $\mathbb{F}$ can be expanded in terms of tensor products of Pauli matrices.

Lemma 59 (Multi-qubit flip operator in terms of Pauli matrices)**.**

$$\mathbb{F} = \frac{1}{d} \sum_{i=1}^{d^2} W_i^{\otimes 2} . \tag{6.44}$$

Proof. The re-normalised Pauli matrices form an orthonormal basis of H_d:

$$X = \frac{1}{d}\sum_{k=1}^{d} W_k \operatorname{Tr}(W_k X) \quad \forall X \in H(\mathbb{C}^n). \tag{6.45}$$

We can extend this to a basis of $H_d^{\otimes 2}$ by considering all possible tensor products of Pauli matrices. Expanding the flip operator in this basis yields

$$\begin{aligned}\mathbb{F} &= \frac{1}{d^2}\sum_{k,l=1}^{d^2} W_k \otimes W_l \operatorname{Tr}(\mathbb{F} W_k \otimes W_l)\\ &= \frac{1}{d^2}\sum_{k,l=1}^{d^2} W_k \otimes W_l d\delta_{k,l} = \frac{1}{d}\sum_{k=1}^{d^2} W_k^{\otimes 2}\end{aligned}$$

as claimed. □

We are now equipped to prove Lemma 57.

Proof of Lemma 57. We start by inserting the definition of Q, Equation (4.1). Fixing w.l.o.g. an order of the indices, we obtain

$$\operatorname{Tr}\left[Q\bigotimes_{j=1}^{4} T_j Q \bigotimes_{j=1}^{4} T_j^\dagger\right] = \frac{1}{d^4}\sum_{k,l=1}^{d^2}\prod_{j=1}^{4}\operatorname{Tr}\left[W_k T_j W_l T_j^\dagger\right] \tag{6.46}$$

$$= \frac{1}{d^4}\sum_{k,l=1}^{d^2} c_{k,l}(T_1)c_{k,l}(T_2)c_{k,l}(T_3)c_{k,l}(T_4), \tag{6.47}$$

where we defined $c_{k,l}(T_j) := \operatorname{Tr}(W_k T_j W_l T_j^\dagger) \in \mathbb{C}$. These numbers obey

$$\begin{aligned}\overline{c_{k,l}(T_j)} =& \overline{\operatorname{Tr}\left(W_k T_j W_l T_j^\dagger\right)} = \operatorname{Tr}\left(\left(W_k T_j W_l T_j^\dagger\right)^\dagger\right)\\ =& \operatorname{Tr}\left(T_j W_l^\dagger T_j^\dagger W_k\right) = c_{k,l}(T_j^\dagger).\end{aligned} \tag{6.48}$$

In addition, Lemma 58 implies

$$|c_{k,l}(T_j)|^2 = \left|\operatorname{Tr}\left(W_k T_j W_l T_j^\dagger\right)\right|^2 \leq \|T_j\|_2^4. \tag{6.49}$$

Equation (6.47) can be viewed as a complex-valued inner product between two d^2-dimensional vectors indexed by k and l. This expression can be upper bounded by the Cauchy-Schwarz inequality:

$$\begin{aligned}
&\frac{1}{d^4}\sum_{k,l=1}^{d^2} c_{k,l}(T_1)c_{k,l}(T_2)c_{k,l}(T_3)c_{k,l}(T_4)\\
&=\frac{1}{d^4}\sum_{k,l=1}^{d^2} \overline{c_{k,l}(T_1^\dagger)c_{k,l}(T_2^\dagger)}c_{k,l}(T_3)c_{k,l}(T_4)\\
&\leq \frac{1}{d^2}\sqrt{\frac{1}{d^2}\sum_{k,l}\left|c_{k,l}(T_1^\dagger)c_{k,l}(T_2^\dagger)\right|^2}\sqrt{\frac{1}{d^2}\sum_{k,l}|c_{k,l}(T_3)c_{k,l}(T_4)|^2}.
\end{aligned}\tag{6.50}$$

The first square-root can be bounded in the following way

$$\begin{aligned}
&\sqrt{\frac{1}{d^2}\sum_{k,l}|c_{k,l}(T_3)c_{k,l}(T_4)|^2}\\
&\leq \sqrt{\|T_1^\dagger\|_2^4\frac{1}{d^2}\sum_{k,l}c_{k,l}(T_2^\dagger)}\\
&=\|T_1\|_2^2\sqrt{\frac{1}{d^2}\sum_{k,l}\operatorname{Tr}\left(W_kT_2^\dagger W_lT_2\right)^2}\\
&=\|T_1\|_2^2\sqrt{\operatorname{Tr}\left(\frac{1}{d}\sum_k W_k^{\otimes 2}(T_2^\dagger)^{\otimes 2}\frac{1}{d}\sum_l W_l^{\otimes 2}T_2^{\otimes 2}\right)}\\
&=\|T_1\|_2^2\sqrt{\operatorname{Tr}\left(\mathbb{F}\,(T_2^\dagger)^{\otimes 2}\,\mathbb{F}\,T_2^{\otimes 2}\right)}\\
&=\|T_1\|_2^2\sqrt{\operatorname{Tr}\left(T_2^\dagger T_2\right)^2}=\|T_1\|_2^2\|T_2\|_2^2.
\end{aligned}\tag{6.51}$$

Here, we have applied the magnitude bound (6.49) for $c_{k,l}(T_1^\dagger)$ in the second line and applied Lemma 59. The second square root can be bounded in a complete analogous fashion, i.e.

$$\sqrt{\frac{1}{d^2}\sum_{k,l}|c_{k,l}(T_3)c_{k,l}(T_4)|^2}\leq\|T_3\|_2^2\|T_4\|_2^2.\tag{6.52}$$

Inserting both bounds into (6.50) yields the desired claim. □

Having established Lemma 57, we will now state the bound on the fourth moment of $S_{\mathcal{T}}$ when the average is performed over the Clifford group.

Lemma 60 (4th moment bound for $\mathrm{Cl}(d)$)**.** *Let $\mathcal{T} : H_d \to H_d$ be a map. For $S_{\mathcal{T}}$ defined in* (6.24)*, it holds*

$$\mathbb{E}_{U\sim\mu_{\mathrm{Cl}(d)}}[S_{\mathcal{T}}^4] \leq C\,\|\mathfrak{J}(\mathcal{T})\|_1^4\,, \tag{6.53}$$

where $\|\cdot\|_1$ denotes the trace (or nuclear) norm and the constant $C > 0$ is independent of d.

Proof. As for the unitary group, we can rewrite the kth moment of $S_{\mathcal{T}}$ for the Clifford group as

$$\begin{aligned}&\mathbb{E}_{U\sim\mu_{\mathrm{Cl}(d)}}[S_{\mathcal{T}}^k]\\ &= \sum_{i_1,\ldots,i_k=1}^{r} \lambda_{i_1}\cdots\lambda_{i_k} \sum_{m,n=1}^{d^k} \langle m|\bigotimes_{j=1}^{k} T_{i_j}^{\dagger} T_{\Delta^k_{\mathrm{Cl}(d)}}(|m\rangle\langle n|)\bigotimes_{j=1}^{k} T_{i_j}\,|n\rangle\end{aligned} \tag{6.54}$$

using a basis $\{|m\rangle \mid m \in \{1,\ldots,d^k\}\}$ for $(\mathbb{C}^d)^{\otimes k}$. The expression for $T_{\Delta^4_{\mathrm{Cl}(d)}}$ with $k = 4$ was derived in Theorem 34. It implies that

$$\begin{aligned}\mathbb{E}_{U\sim\mu_{\mathrm{Cl}(d)}}[S_{\mathcal{T}}^4] = &\sum_{i_1,\ldots,i_k=1}^{r} \lambda_{i_1}\cdots\lambda_{i_k}\frac{1}{4!}\sum_{\tau\in\mathfrak{S}_k}\sum_{\lambda\vdash k,\ l(\lambda)\leq d} d_\lambda\\ &\times\left\{\frac{1}{D_\lambda^+}\mathrm{Tr}\left[Q\bigotimes_{j=1}^{4} T_{i_{\tau(j)}}^{\dagger} QP_\lambda\bigotimes_{j=1}^{4} T_{i_j}\right]\right.\\ &\left.+\frac{1}{D_\lambda^-}\mathrm{Tr}\left[Q^{\perp}\bigotimes_{j=1}^{4} T_{i_{\tau(j)}}^{\dagger} Q^{\perp}P_\lambda\bigotimes_{j=1}^{4} T_{i_j}\right]\right\}.\end{aligned} \tag{6.55}$$

We may bound the first trace term by

$$\begin{aligned}&\left|\mathrm{Tr}\left[Q\bigotimes_{j=1}^{4} T_{i_{\tau(j)}}^{\dagger} QP_\lambda\bigotimes_{j=1}^{4} T_{i_j}\right]\right| \leq \left\|P_\lambda Q\bigotimes_{j=1}^{4} T_{i_{\tau(j)}} Q\right\|_{\mathrm{F}} \left\|P_\lambda Q\bigotimes_{j=1}^{4} T_{i_j} Q\right\|_{\mathrm{F}}\\ &\leq \left\|Q\bigotimes_{j=1}^{4} T_{i_{\tau(j)}} Q\right\|_{\mathrm{F}} \left\|Q\bigotimes_{j=1}^{4} T_{i_j} Q\right\|_{\mathrm{F}} \leq \frac{1}{d^2}\prod_{j=1}^{4}\|T_{i_j}\|_{\mathrm{F}}^2\,,\end{aligned} \tag{6.56}$$

where we have used Cauchy-Schwarz and applied Lemma 57 in the last line. For the second trace term a looser bound suffices:

$$\left\| Q^{\perp} \bigotimes_{j=1}^{k} T_{i_{\tau(j)}} Q^{\perp} \right\|_{\mathrm{F}} \leq \prod_{j=1}^{k} \|T_{i_j}\|_{\mathrm{F}} \tag{6.57}$$

for all $\tau \in \mathfrak{S}_4$. This follows directly from Cauchy-Schwarz. Altogether we conclude that

$$\begin{aligned} &\mathbb{E}_{U \sim \mu_{\mathrm{Cl}(d)}}[S_{\mathcal{T}}^4] \\ &\leq \sum_{i_1,\dots,i_4=1}^{r} \prod_{j=1}^{4} |\lambda_{i_j}| \, \|T_{i_j}\|_{\mathrm{F}}^2 \sum_{\lambda \vdash k,\ l(\lambda)\leq d} d_\lambda \left[\frac{1}{d^2 D_\lambda^+} + \frac{1}{D_\lambda^-} \right] \\ &\leq C \, \|\mathfrak{J}(\mathcal{T})\|_1^4 \end{aligned} \tag{6.58}$$

with some constant $C > 0$ independent of d. The last step follows from the dimensions given in Theorem 35. □

Proof of Theorem 52 (recovery guarantee). We have now established the main auxiliary results for deriving the recovery guarantee, Theorem 52 or more precisely a considerably refined statement. We consider the following measurements: For a map $\mathcal{X} \in \mathrm{L}(H_d)$ the measurement outcomes $f \in \mathbb{R}^m$ are given by

$$\begin{aligned} f_i &= F_{\mathrm{avg}}(\mathcal{C}_i, \mathcal{X}) + \epsilon_i \\ &= \frac{1}{d+1} \left[d\,(\mathcal{C}_i, \mathcal{X}) + \frac{1}{d} \operatorname{Tr}(\mathcal{X}^\dagger(\mathrm{Id})) \right] + \epsilon_i, \end{aligned} \tag{6.59}$$

where $\mathcal{C}_i$ are random Clifford channels and $\epsilon \in \mathbb{R}^m$ accounts for additional additive noise.

To make use of the proof techniques developed for low rank matrix reconstruction [KRT15; Kab+16], we will in the following work in the Choi representation of channels. This has the advantage, that the Kraus rank directly translates to the familiar matrix rank. Recall the following results and definitions from the preliminaries: The Choi matrix of a map is positive semi-definite if and only if the map is completely positive. We denote the cone of positive semidefinite matrices by Pos_{d^2}. A channel $\mathcal{X}$ is trace-preserving and unital if and only if both partial traces of the Choi matrix yield the maximally mixed state, i.e. $\mathrm{Tr}_1(\mathfrak{J}(\mathcal{X})) = \mathrm{Tr}_2(\mathfrak{J}(\mathcal{X})) = \mathrm{Id}/d$. We will denote the set of Choi matrices that correspond to channels in $\mathrm{L}_{\mathrm{u,tp}}(H_d)$ by $\mathfrak{J}(\mathrm{L}_{\mathrm{u,tp}})$ dropping the H_d for convenience. Furthermore, we define $\mathfrak{J}(\mathrm{V}_{\mathrm{u,tp},0})$ as the set of Choi matrices corresponding to trace- and identity-annihilating channels, i.e., both partial traces of operators in $\mathfrak{J}(\mathrm{V}_{\mathrm{u,tp},0})$ vanish. Moreover, by Equation (6.23), the inner product $(\cdot,\cdot)$ on $\mathrm{L}_{\mathrm{u,tp}}(H_d)$ coincides with the Hilbert-Schmidt inner product of the corresponding Choi states. Adhering to this correspondence, we slightly abuse notation and use $(\mathcal{X},\mathcal{Y})$ and $(\mathfrak{J}(\mathcal{X}),\mathfrak{J}(\mathcal{Y}))$ interchangeably.

To formalize the robustness of our reconstruction we need to introduce the following notation. For a Hermitian matrix $Z \in H_d$ let λ be the largest eigenvalue with an eigenvector v. We write $Z|_1 = \lambda \left|v\right\rangle\!\left\langle v\right|$ for the best unit rank approximation to Z and $Z|_c := Z - Z|_1$ denotes the corresponding "tail".

In terms of the Choi matrix of $\mathcal{X}$ the measurement outcomes $f \in \mathbb{R}^m$ read

$$f_i = \frac{1}{d+1}\left[d\left(\mathfrak{J}(\mathcal{C}_i), \mathfrak{J}(\mathcal{X})\right) + \mathrm{Tr}(\mathfrak{J}(\mathcal{X}))\right] + \epsilon_i, \tag{6.60}$$

The underlying linear measurement map $\mathcal{A} : H_{d^2} \to \mathbb{R}^m$ is given by

$$\mathcal{A}_i(X) = \frac{1}{d+1}\left[d(\mathfrak{J}(\mathcal{C}_i), X) + \mathrm{Tr}(X)\right]. \tag{6.61}$$

Since unital and trace preserving maps $\mathcal{X}$ have trace normalized Choi matrices the second trace-term of the measurement map is just a constant shift. We also define the set of measurement matrices $\{A_i\}_{i=1}^{b}$ that encode the measurement map as $\mathcal{A}_i(X) = (A_i, X)$: $A_i = \frac{d}{d+1}\left[\mathfrak{J}(\mathcal{C}_i) + \mathrm{Id}/d\right]$, where each $\mathcal{C}_i$ is a gate that is chosen uniformly at random from the multi-qubit Clifford group. In the Choi representation, we want to consider the optimization problem

$$\begin{aligned} \underset{Z}{\text{minimize}} \quad & \|\mathcal{A}(Z) - f\|_{\ell_q} \\ \text{subject to} \quad & Z \in \mathfrak{J}(\mathrm{L}_{\mathrm{u,tp}}) \cap \mathrm{Pos}_{d^2}, \end{aligned} \tag{6.62}$$

where we allow the minimization of an arbitrary ℓ_q norm. The optimization problem (6.20) is equivalent to (6.62) for $q = 2$. We are interested in using the optimization procedure (6.62) for the recovery of unitary quantum channels. In this section, we will derive the following recovery guarantee:

Theorem 61 (Recovery guarantee). *Let $\mathcal{A} : H_{d^2} \to \mathbb{R}^m$ be the measurement map (6.61) with*

$$m \geq c d^2 \log(d). \tag{6.63}$$

Then, for all $X \in \mathfrak{J}(L_{\mathrm{u,tp}}) \cap \mathrm{Pos}_{d^2}$ given noisy observations $f = \mathcal{A}(X) + \epsilon \in \mathbb{R}^m$, the minimizer $Z^\sharp$ of the optimization problem (6.62) fulfils for $p \in \{1, 2\}$

$$\left\|Z^\sharp - X\right\|_p \leq \tilde{C}_1 \left\|X|_c\right\|_1 + 2\tilde{C}_2 d^2 m^{-1/q} \left\|\epsilon\right\|_{\ell_q} \tag{6.64}$$

with probability at least $1 - \mathrm{e}^{-c_f m}$ *over the random measurements. The constants* $\tilde{C}_1, \tilde{C}_2, c, c_f > 0$ *only depend on each other.*

The recovery guarantee of Theorem 52 is the special case of Theorem 61 for $q = 2$ and $p = 2$ restricted to measurements of a unitary quantum channel. In contrast, the more general formulation of Theorem 61 allows for a violation of the unit rank assumption. The first term in (6.64) is meant to absorb violations of this assumption into the error bound. We note in passing that the choice of $p = 1$ actually yields a tighter bound compared to $p = 2$.

More generally, one can ask for a recovery guarantee if the measured map X can not be guaranteed to be unital or trace preserving. From (6.11) one observes that as long as the map X is trace normalized the measured AGFs are identical to the average fidelities of the projection $X_{\text{u,tp}}$ of X onto the affine space of unital and trace-preserving maps. But since $X_{\text{u,tp}}$ is not necessarily positive, it is not straight-forward to apply Theorem 61 to $X_{\text{u,tp}}$. We expect the reconstruction algorithm to recover the trace-preserving and unital part of an arbitrary map. The reconstruction error (6.64) is expected to additionally feature a term proportional to the distance of X to the intersection of $\mathfrak{J}(L_{\text{u,tp}})$ with the cone Pos_{d^2} of positive semi-definite matrices.

Another way to proceed is to use a trace-norm minimization subject to unitality, trace-preservation and the data constraints $\|\mathcal{A}(Z) - f\|_{\ell_q} < \eta$. The derivation of Theorem 61 readily yields a recovery guarantee for the trace-norm minimization that is essentially identical to Theorem 61. See Ref. [Kab+16] for details on the argument. The main difference is that such a recovery guarantee does not need to assume complete positivity of the map that is to be reconstructed. Correspondingly, the result of the trace-norm minimization is not guaranteed to be positive semi-definite. This implies that the robustness of this algorithm against violations of the unitality and trace-preservation is different compared to (6.62). For example, the AGFs of a not necessarily unital or trace-preserving map $\mathcal{X}$ to unitary gates coincide with the AGFs of its unital and trace-preserving part $\mathcal{X}_{\text{u,tp}}$ as long as X is still normalized in trace norm. This is a consequence of (6.11). Thus, a trace-norm minimization will reconstruct $X_{\text{u,tp}}$ up to an error given by $\left\|\mathfrak{J}(\mathcal{X}_{\text{u,tp}})|_c\right\|_1$ and noise. We leave a more extensive study of the robustness of the discussed reconstruction algorithms against violations of this particular model assumption to future work.

The proof of the recovery guarantee relies on establishing the so-called *nullspace property (NSP)* for the measurement map $\mathcal{A}$. We refer to Ref. [FR13] for a history of the term. The NSP ensures injectivity, i.e. informational completeness, of the measurement map $\mathcal{A}$ restricted to the matrices that should be recovered. Informally, for our purposes, a measurement map $\mathcal{A} : H_{d^2} \to \mathbb{R}^m$ obeys the NSP if no unit rank matrix in $\mathfrak{J}(\mathrm{V}_{\mathrm{u,tp},0})$ is in the kernel (nullspace) of $\mathcal{A}$.

Definition 4 (Robust NSP, Definition 3.1 in Ref. [Kab+16])**.** $\mathcal{A} : H_{d^2} \to \mathbb{R}^m$ satisfies the nullspace property (NSP) with respect to ℓ_q with constant $\tau > 0$ if for all $X \in \mathfrak{J}(\mathrm{V}_{\mathrm{u,tp},0})$

$$\|X|_1\|_{\mathrm{F}} \leq \frac{1}{2}\,\|X|_c\|_1 + \tau\|\mathcal{A}(X)\|_{\ell_q}. \tag{6.65}$$

The factor $1/2$ in front of the first term of (6.65) is only one possible choice. In fact, one can instead introduce a constant with value in $(0,1)$. The constants appearing in Theorem 61 then depend on the specific value of the pre-factor. In particular, the different choices of the pre-factor in the definition of the NSP result in different trade-offs between the constant c that appears in the sampling complexity and the constant $\tilde{C}_1$ that decorates the model-mismatch term in the reconstruction error. For the simplicity, we leave these dependencies implicit.

The main consequence of the NSP that we require is captured by the following reformulation of Theorem 12 of [Kab+16].

Theorem 62. *Fix $p \in \{1,2\}$ and let $\mathcal{A} : H_{d^2} \to \mathbb{R}^m$ satisfy the NSP with constant $\tau > 0$. Then, for all $Y, Z \in \mathfrak{J}(\mathrm{L}_{\mathrm{u,tp}})$*

$$\|Z - Y\|_p \leq \frac{9}{2}\left[\|Z\|_1 - \|Y\|_1 + 2\,\|Y|_c\|_1\right] + 7\tau\,\|\mathcal{A}(Z-Y)\|_{\ell_q}\,. \tag{6.66}$$

In fact, the measurement $\mathcal{A}$ of (6.60) obeys the NSP. More precisely:

Lemma 63. *Let $\mathcal{A} : H_{d^2} \to \mathbb{R}^m$ be the measurement map defined in (6.61) with $m \geq cd^2\log(d)$. Then $\mathcal{A}$ obeys the NSP with constant $\tau = C^{-1}d(d+1)m^{-1/q}$ with probability of at least $1 - \mathrm{e}^{-c_f m}$. The constants $C, c, c_f > 0$ only depend on each other.*

The proof of Lemma 63 is developed in the subsequent section.

Proof of Theorem 61. With the requirements of Lemma 63 we can apply Theorem 62 and set $Z = Z^\sharp$, the reconstructed result of the algorithm, as well as $Y = X$. The theorem's statement then reads

$$\left\| Z^\sharp - X \right\|_p \leq 9 \left\| X|_c \right\|_1 + 7\tau \left\| \mathcal{A}(Z^\sharp - X) \right\|_{\ell_q} , \tag{6.67}$$

because $\|X\|_1 = \|Z\|_1 = 1$ is true for arbitrary Choi matrices of (trace-preserving) quantum channels. The second term is dominated by

$$\left\| \mathcal{A}(Z^\sharp - X) \right\|_{\ell_q} \leq \left[\left\| \mathcal{A}(X - Z^\sharp) + \epsilon \right\|_{\ell_q} + \|\epsilon\|_{\ell_q} \right] \leq 2 \|\epsilon\|_{\ell_q} , \tag{6.68}$$

where the last step follows from $Z^\sharp$ being the minimizer of (6.62). Thus, we can replace it by any point in the feasible set including X on the right-hand side of the first line. Inserting (6.68) and the NSP constants of Lemma 63 into (6.67) the assertion of the theorem follows. □

In the remainder of this section, we will establish the NSP for our measurement matrix $\mathcal{A}$ as summarized in Lemma 63.

Establishing the null space property. To prove Lemma 63 at the end of this section we start with deriving a criterion for the NSP following the approach taken in Refs. [Kab+16; Kli+19].

Lemma 64. *A map $\mathcal{A} : H_{d^2} \to \mathbb{R}^m$ obeys the null space property with respect to ℓ_q norm with constant $\tau > 0$ if*

$$\inf_{X \in \mathbf{\Omega}} \|\mathcal{A}(X)\|_{\ell_1} \geq \frac{m^{1-1/q}}{\tau} \tag{6.69}$$

with $\mathbf{\Omega} := \{ Z \in \mathfrak{J}(\mathrm{V}_{\mathrm{u,tp},0}) \mid \|Z|_1\|_{\mathrm{F}} \geq \frac{1}{2} \|Z|_c\|_1 , \|Z\|_{\mathrm{F}} = 1 \}$.

Proof. For matrices X with the property $\|X|_1\|_{\mathrm{F}} \leq \frac{1}{2} \|X|_c\|_1$ the NSP condition (6.65) is satisfied independently of the map $\mathcal{A}$. Hence, to establish the NSP for a specific map $\mathcal{A}$ it suffices to show that the condition (6.65) holds for all $X \in \mathbf{\Omega} = \{ Z \in \mathfrak{J}(\mathrm{V}_{\mathrm{u,tp},0}) \mid \|Z|_1\|_{\mathrm{F}} \geq \frac{1}{2} \|Z|_c\|_1 , \|Z\|_{\mathrm{F}} = 1 \}$. The additional assumption of $\|Z\|_{\mathrm{F}} = 1$ is no restriction since both sides of (6.65) are absolutely homogeneous functions of the same degree. By definition, for all $X \in \mathbf{\Omega}$ we have $\|X|_1\|_{\mathrm{F}} \leq \|X\|_{\mathrm{F}} \leq 1$. Therefore, for $X \in \mathbf{\Omega}$

$$\|\mathcal{A}(X)\|_{\ell_q} \geq \frac{1}{\tau} \tag{6.70}$$

implies the NSP condition (6.65). Using the norm inequality $\|x\|_{\ell_q} \geq m^{1/q-1} \|x\|_{\ell_1}$ yields the criterion of the lemma. □

Recall that every rank-r matrix X obeys $\|X\|_1^2 / \|X\|_F^2 \leq r$. This motivates thinking of the matrices of $\boldsymbol{\Omega}$ as having *effective unit rank* since the norm ratio bounded in $\mathcal{O}(1)$. More precisely, the following statement holds:

Lemma 65 (Ratio of 1 and 2 norms)**.** *Every matrix $X \in \boldsymbol{\Omega}$ has* effective unit rank *in the following sense:*

$$\frac{\|X\|_1^2}{\|X\|_2^2} \leq 9. \tag{6.71}$$

Proof. From $\|X|_1\|_2 \leq 1$ and the definition of $\boldsymbol{\Omega}$ it follows that $\|X|_1\|_2 + \frac{1}{2}\|X|_1\|_1 \leq \frac{3}{2}$. Hence $\frac{1}{2}\|X|_1\|_2 + \|X|_1\|_1 \leq 3$. Therefore, we have that $\|X\|_1 \leq \|X|_1\|_1 + \|X|_c\|_1 \leq 3$ from which the assertion follows, because every $X \in \boldsymbol{\Omega}$ has unit Frobenius norm. □

In summary, we want to prove a lower bound on the ℓ_q norm of the measurement outcomes for trace- and identity-annihilating channels with effective unit Kraus rank. The proof uses Mendelson's small ball method. See Ref. [Kli+19, Lemma 9] for details of the method as it is stated here, which is a slight generalization of Tropp's formulation [Tro15] of the original method developed in Refs. [Men15; KM15]. Mendelson's proof strategy requires multiple ingredients. These necessary ingredients will become obvious from the following theorem, which can be found in Ref. [Tro15] and lies at the heart of the small ball method.

Theorem 66 (Mendelson's small ball method)**.** *Suppose that $\mathcal{A}$ contains m measurements of the form $f_k = \mathrm{Tr}[A_k X]$ where each A_k is an independent copy of a random matrix A. Fix $E \subseteq \mathfrak{J}(\mathrm{V}_{\mathrm{u,tp},0})$ and $\xi > 0$ and define*

$$W_m(E; A) := \mathbb{E}\left[\sup_{Z \in E} \mathrm{Tr}\,(ZH)\right], \quad H = \frac{1}{\sqrt{m}} \sum_{k=1}^{m} \epsilon_k A_k, \tag{6.72}$$

$$Q_\xi(E; A) := \inf_{Z \in E} \mathbb{P}\left[|\mathrm{Tr}\,[AZ]| \geq \xi\right], \tag{6.73}$$

where the ϵ_k's are i.i.d. Rademacher random variables, i.e. are uniformly distributed in $\{-1, 1\}$. Then, with probability of at least $1 - \mathrm{e}^{-2t^2}$, where $t \geq 0$,

$$\inf_{Z \in E} \|\mathcal{A}(Z)\|_{\ell_1} \geq \sqrt{m}\left(\xi\sqrt{m} Q_{2\xi}(E; A) - 2W_m(E; A) - \xi t\right).$$

A lower bound of $\|\mathcal{A}(X)\|_{\ell_1}$ thus requires two main ingredients: 1.) a lower bound on the so-called *mean empirical width* $W_m(E;A)$ and 2.) an upper bound on the so-called *marginal tail function* $Q_{2\xi}(E;A)$. We will derive those bounds for $E = \mathbf{\Omega}$ and our measurement map $\mathcal{A}$ at hand.

Bound on the mean empirical width. With a different normalization the following statement is derived in Ref. [KL17].

Lemma 67. *Fix $d = 2^n$ and suppose that the measurement matrices are given by $A_i = \frac{d}{d+1}\left[\mathfrak{J}(\mathcal{C}_i) + \mathrm{Id}/d\right]$ with a gate $\mathcal{C}_i$ chosen uniformly from the Clifford group for all i. Also, assume that $m \geq d^2 \log(d)$. Then*

$$W_m(\mathbf{\Omega}, A) \leq \frac{24}{d+1}\sqrt{\log(d)}. \tag{6.74}$$

The proof is analogous to the one in Refs. [KRT15; KL17; Kli+19]. In order to adjust the normalization we provide a short summary.

Proof. For $Z \in \mathbf{\Omega}$ it holds that

$$(A_i, Z) = \frac{d}{d+1}(\mathfrak{J}(\mathcal{C}_i), Z). \tag{6.75}$$

The constant shift by the identity matrix does not appear hear since every $Z \in \mathbf{\Omega}$ is trace-less. Thus, we can set $H = \frac{d}{\sqrt{m}(d+1)}\sum_{i=1}^m \epsilon_i \mathfrak{J}(C_i)$. Applying Hölder's inequality for Schatten norms to the definition of the mean empirical width yields

$$W_m(\mathbf{\Omega}, A) \leq \sup_{Z\in\mathbf{\Omega}} \|Z\|_1 \mathbb{E}\|H\|_\infty \leq 3\,\mathbb{E}\|H\|_\infty, \tag{6.76}$$

where we have used the effective unit rank of Z, Lemma 65. Also, the ϵ_i's in the definition of H form a Rademacher sequence. The non-commutative Khintchine inequality, see e.g [Ver12, Eq. (5.18)], can be used to bound this sequence

$$\mathbb{E}_{\epsilon_i, C_i}\|H\|_\infty \leq \frac{d}{d+1}\sqrt{\frac{2\log(2d^2)}{m}\mathbb{E}_{C_i}\left\|\sum_{i=1}^m \mathfrak{J}(C_i)^2\right\|_\infty} \tag{6.77}$$

and $\mathfrak{J}(C_i)^2 = \mathfrak{J}(C_i)$ further simplifies the remaining expression. Moreover, $\mathbb{E}\left[\mathfrak{J}(C_i)\right] = \frac{1}{d^2}\mathbb{I}$, $\|\mathfrak{J}(C_i)\|_\infty = 1$ and a Matrix Chernoff inequality for expectations (with parameter $\theta = 1$), see e.g. [Tro12, Theorem 5.1.1] implies

$$\mathbb{E}_{C_i}\left\|\sum_{i=1}^m \mathfrak{J}(C_i)\right\|_\infty \leq (\mathrm{e}-1)\frac{m}{d^2} + \log(d^2) \leq 4\frac{m}{d^2}, \tag{6.78}$$

where the second inequality follows from the assumption $m \geq d^2 \log(d)$. Inserting this bound into (6.77) yields

$$\mathbb{E}_{\epsilon_i, C_i} \|H\|_\infty \leq \frac{d}{d+1} \sqrt{\frac{8 \log(2d^2)}{d^2}} \tag{6.79}$$

and the claim follows from combining this estimate with the bound (6.76) and $\log(2d^2) \leq 4 \log(d)$. □

Bound on the marginal tail function. Here, we establish an anti-concentration bound to the marginal tail function. The precise result is summarized in the following statement.

Lemma 68. *Suppose the random variable $A \in H_d$ is given by $A = \frac{d}{d+1} [\mathfrak{J}(\mathcal{C}) + \mathrm{Id}/d]$, where $\mathcal{C}$ is a Clifford channel drawn uniformly from the Clifford-group $\mathrm{Cl}(d)$. For $0 \leq \xi \leq \frac{1}{d(d+1)}$ it holds that*

$$Q_\xi(\boldsymbol{\Omega}, A) \geq \frac{1}{\hat{C}} \left(1 - d^2(d+1)^2 \xi^2\right)^2, \tag{6.80}$$

where $\hat{C}$ is the constant from Lemma 69.

This statement follows from applying the Paley-Zygmund inequality to the non-negative random variable $S_\mathcal{T}^2$ defined in (6.24). For this purpose, we will make use of the bounds on the second and fourth moment of $S_\mathcal{T}$. In particular, we establish the following relation between the second and fourth moment of $S_\mathcal{T}$. This is one of the technical core result of this work.

Lemma 69. *Let $\mathcal{T} \in V_{\mathrm{u,tp,0}}$ be a map with $\mathfrak{J}(\mathcal{T})$ of effective unit rank, i.e. $\|\mathfrak{J}(\mathcal{T})\|_\mathrm{F}^2 \leq c \|\mathfrak{J}(\mathcal{T})\|_1^2$ with some constant $c > 0$, then*

$$\mathbb{E}_{U \sim \mu_{\mathrm{Cl}(d)}} [S_\mathcal{T}^4] \leq \hat{C}\, \mathbb{E}_{U \sim \mu_{\mathrm{Cl}(d)}} [S_\mathcal{T}^2]^2 \tag{6.81}$$

for some constant $\hat{C}$ independent of the dimension d.

Proof. Since the Clifford group is a unitary 3-design [Zhu17; Web16], Corollary 54 implies

$$\mathbb{E}_{U \sim \mu_{\mathrm{Cl}(d)}} [S_\mathcal{T}^2] \geq \|\mathfrak{J}(\mathcal{T})\|_\mathrm{F}^2 . \tag{6.82}$$

Furthermore, the effective unit rank assumption, $\|\mathfrak{J}(\mathcal{T})\|_1^2 \leq c\,\|\mathfrak{J}(\mathcal{T})\|_{\mathrm{F}}^2$, together with Lemma 60 yields for the fourth moment

$$\mathbb{E}_{U\sim\mu_{Cl(d)}}[S_{\mathcal{T}}^4] \leq \hat{C}\,\|\mathfrak{J}(\mathcal{T})\|_{\mathrm{F}}^4 \tag{6.83}$$

for some constant $\hat{C} = cC > 0$ independent of d. Combining these two equations, the statement of the proposition follows. □

Note that with the help of Lemma 56 one arrives at the same conclusion for the moments of $S_{\mathcal{T}}$ when the average is taken over the unitary group. This reproduces the previous technical core result of Ref. [KL17].

Proof of Lemma 68. In the following we always understand by $\mathcal{T}$ the map in $\mathrm{L}(H_d)$ with Choi matrix $T = \mathfrak{J}(\mathcal{T})$. In terms of the random variable $S_{\mathcal{T}} = d^2\ \mathrm{Tr}[T\,\mathfrak{J}(\mathcal{C})]$, (6.24), the marginal tail function can be expressed as

$$Q_\xi(\mathbf{\Omega}, A) = \inf_{T\in\mathbf{\Omega}} \mathbb{P}\left[\frac{|S_{\mathcal{T}}|}{d(d+1)} \geq \xi\right]. \tag{6.84}$$

Here we again used that every $Z \in \mathbf{\Omega}$ is trace-less. Consequently, the shift by the identity matrix in the measurements A_i vanishes. Using Lemma 69, the theorem follows by a straight-forward application of the Paley-Zygmund inequality,

$$\begin{aligned}
&\inf_{T\in\mathbf{\Omega}} \mathbb{P}\left[\frac{1}{d(d+1)}|S_{\mathcal{T}}| \geq \xi\right] \\
&\quad= \inf_{T\in\mathbf{\Omega}} \mathbb{P}\left[\frac{1}{d^2(d+1)^2} S_{\mathcal{T}}^2 \geq \frac{\mathbb{E}[S_{\mathcal{T}}^2]}{d^2(d+1)^2}\,\tilde{\xi}^2\right] \\
&\quad\geq (1-\tilde{\xi}^2)^2 \frac{\mathbb{E}[S_{\mathcal{T}}^2]^2}{\mathbb{E}[S_{\mathcal{T}}^4]} \geq \frac{1}{\hat{C}}(1-\tilde{\xi}^2)^2,
\end{aligned} \tag{6.85}$$

where $\hat{C} > 0$ and $\tilde{\xi} = \frac{d(d+1)}{\sqrt{\mathbb{E}[S_{\mathcal{T}}^2]}}\,\xi$ is required to fulfil $\tilde{\xi} \in [0,1]$. According to Corollary 54 and the normalization of $T \in \mathbf{\Omega}$ we have $\tilde{\xi} = \frac{d(d+1)\,\xi}{\|T\|_{\mathrm{F}}} = d(d+1)\,\xi$. □

Completing the proof of Lemma 63. We are finally in position to deliver the proof for the NSP of $\mathcal{A}$. With the bounds on the mean empirical width, Lemma 67, and the marginal tail function, Lemma 68, Mendelson's small ball method, Theorem 66, yields the following lemma:

Lemma 70. *Suppose that $\mathcal{A}$ contains*

$$m \geq m_0 = c\, d^2 \log(d) \tag{6.86}$$

measurements of the form $f_k = \operatorname{Tr}[A_k X]$ where each $A_k = \frac{d}{d+1}\left[\mathfrak{J}(\mathcal{C}_i) + \operatorname{Id}/d\right]$ is given by an independent and uniformly random Clifford unitary channel $\mathcal{C}_i$. Fix $\boldsymbol{\Omega} \subset \mathfrak{J}(\mathrm{V}_{\mathrm{u,tp},0})$ as defined in Lemma 64. Then

$$\inf_{Z\in\boldsymbol{\Omega}} \|\mathcal{A}(Z)\|_{\ell_1} \geq C\frac{m}{d(d+1)} \tag{6.87}$$

with probability at least $1-\mathrm{e}^{-c_f m}$ over the random measurements. The constants $C, c, c_f > 0$ only depend on each other.

Proof. Combining the Lemmas 66, 67, and 68 yields with probability at least $1 - \mathrm{e}^{-2t^2}$ that

$$\begin{aligned}
&\inf_{\mathcal{Z}\in\boldsymbol{\Omega}} \|\mathcal{A}(\mathcal{Z})\|_{\ell_1}\\
&\geq \sqrt{m}\left(\frac{\xi\sqrt{m}}{\hat{C}}\left(1-(d(d+1)\xi)^2\right)^2 - \frac{48}{d+1}\sqrt{\log(d)} - \xi t\right)\\
&\geq \frac{\sqrt{m}}{d+1}\left(c_1\frac{\sqrt{m}}{d} - 48\sqrt{\log(d)} - \frac{t}{2d}\right)
\end{aligned} \tag{6.88}$$

where we have chosen $\xi = \frac{1}{2d(d+1)}$. The statement follows from the scaling (6.86) of m. □

From Lemma 70 and Lemma 64 the assertion of Lemma 63 directly follows.

6.3 Sample optimality in the number of channel uses

The compressed sensing recovery guarantees, Theorem 52 and Theorem 61, focus on the minimal number of AGFs m that are required for the reconstruction of a unital and trace-preserving quantum channel using the reconstruction procedure (6.20) and (6.62), respectively.[9] This can be regarded as the number of

[9] The compressed sensing recovery guarantees ...] Most parts of the section are published in the supplemental material of Ref. [1]. Compared to the published material the section contains a more direct argument for the sample achievable sampling complexity based on the results of Section 5.3.

measurement settings. But already the measurement of single fidelities up to some desired additive error will require a certain number of repetitions of some experiment. Therefore, to quantify the total measurement effort a more relevant figure of merit is the minimum number of channel uses M required for taking all the data used in a reconstruction.[10]

We will now show that the equivalent algorithms (6.20) and (6.62) reach an optimal parametric scaling of the required number of channel uses in a simplified measurement setting. To this end, we first combine the sample complexity results of Section 5.3 with our recovery strategy to provide an upper bound on the number of channel uses required for the reconstruction of a unitary gate up to a constant error. We find that in specific cases, the sampling complexity scales as $O(d^4)$ in d up-to logarithmic factors when one measures the fidelities individually in separate experiments. When we measure the fidelities simultaneously with a single POVM, we can achieve a scaling of $O(d^2)$ up-to logarithmic overhead.

To argue that the scaling of $O(d^4)$ is indeed optimal, we then derive a lower bound on the number of channel uses required by any POVM measurement scheme that individually determines the AGFs with Clifford gates and any subsequent reconstruction protocol that only relies on these AGFs. The latter proof will follow the strategy of Ref. [Fla+12, Section III]. We find that the lower bound matches the upper bound up to log-factors. Hence, we have build evidence for the optimality of the our recovery guarantee.

In order to obtain an optimality result we consider a measurement setting that is arguably simpler than the one in randomized benchmarking and more basic from a theoretical perspective. We consider a unitary channel $\mathcal{U}$ given by a unitary $U \in \mathrm{U}(d)$ and measurements given by Clifford channels $\mathcal{C}_i$ with $C_i \in \mathrm{Cl}(d)$. Via Proposition 18, the AGFs $F_{\mathrm{avg}}(\mathcal{C}_i, \mathcal{X})$ are determined by

$$f_i = \langle \mathfrak{J}(\mathcal{C}_i), \mathfrak{J}(\mathcal{X}) \rangle = \frac{1}{d^2} |\operatorname{Tr}[C_i U]|^2 \,. \tag{6.89}$$

In this section, we consider $U/\sqrt{d}$ as a pure state vector in $\mathbb{C}^d \otimes \mathbb{C}^d$, i.e., as the state vector corresponding to the Choi state of the channel $\mathcal{U}$. This state can be prepared by applying the operation U to one half of a maximally entangled state.

[10] Note that if we'd use an RB experiment for the measurement every sequence might additionally involve multiple channel invocations introducing an additional polynomial factor in the maximal sequence length but independent of the system dimension. Since we are not explicitly concerned with RB sequences, we will not make a distinction between the sample complexity and the number of channel invocations.

6.3.1 An upper bound from direct POVM measurements

To derive an overall achievable sampling complexity of the tomography scheme, we combine the recovery guarantee, Theorem 52, with the sample complexity results for estimating the vector f in ℓ_2 norm derived in Section 5.3. The latter was derived under the assumption that the set $\{\mathcal{C}_i\}_{i=1}^{m}$ defining f_i constitutes a unitary 1-design. The Pauli group of cardinality $O(d^2)$ constitute a unitary 1-design. We can always trivially construct unitary 1-designs of a multiple size by combining, e.g. differently rotated copies of a unitary 1-design. A randomly drawn subset of the Clifford group of cardinality $O(d^2 \log d)$ is unlikely to constitute an exact unitary 1-design, or even an approximate unitary 1-design with a suitably small additive error $O(d^{-2})$ in the expression of Lemma 46. But since such designs exist, we might accidentally arrive at an exact unitary 1-design with a non-vanishing probability. In this case, if we want a reconstruction error in Theorem 52 of size ϵ_{rec}, we need to estimate f with accuracy $\sqrt{m}\epsilon_{\text{rec}}/d^2$ in ℓ_2 norm. Using separate POVM measurements, this requires according to (5.30) a number of samples in $O(md^2\epsilon_{\text{rec}}^{-2} \log m)$. For $m \geq \tilde{m} \in O(d^2 \log d)$, we find an overall sampling complexity in $O(d^4 \log(d)\epsilon_{\text{rec}}^{-2})$. For the simultaneous POVM measurement, we analogously get a sampling complexity scaling as d^2 in d up-to log factors.

6.3.2 Information theoretic lower bound

We now derive a lower bound on the number of channel uses that holds in a general POVM framework focusing on the case where the f_i are measured separately. Up to log-factors, it has the same dimensional scaling as the upper bound derived in the last paragraph. We extend the arguments of Ref. [Fla+12, Section III] to prove a lower bound on the number of channel uses required for QPT of unitary channels from measurement values of the form (6.89). We consider each of these values to be an expectation value in a binary POVM measurement setting given by the unit rank projector $\mathfrak{J}(\mathcal{C}_i)$ are applied to the Choi state $\mathfrak{J}(\mathcal{U})$. Then we are in the situation of [Fla+12, Section 3], which proves a lower bound for the *minimax risk*—a prominent figure of merit for statistical estimators.

Let us summarize this setting. We denote by $\mathcal{D} \subset H_d$ the set of density matrices and by $\mathcal{M}$ the set of all two-outcome positive-operator-valued measurements (POVMs), each of them given by a projector $\pi \in H_d$. Next, we assume that we measure M copies of an unknown state $\rho \in \mathcal{D}$ sequentially. By Y_i we denote the binary random variable that is given by choosing the ith measurement $\pi_i \in \mathcal{M}$ and measuring ρ. These are mapped to an estimate $\hat{\rho}(Y_1, \dots, Y_M) \in H_d$. Any such estimation protocol is specified by the estimator function $\hat{\rho}$ and a set of functions $\{\Pi_i\}_{i\in[M]}$ that correspond to the measurement choices, where $\Pi_i(Y_1, \dots Y_{i-1}) \in \mathcal{M}$, i.e., the ith measurement choice Π_i only depends on previous measurement outcomes. Let $\varepsilon > 0$ be the maximum trace distance error we like to tolerate between the estimation $\hat{\rho}$ and ρ. Then the *minimax risk* is defined as

$$R^*(M, \varepsilon) := \inf_{\substack{\hat{\rho} \\ \Pi_1, \dots, \Pi_M}} \sup_{\rho \in \mathcal{D}} \mathbb{P}\left[\|\hat{\rho}(Y) - \rho\|_1 > \varepsilon\right], \tag{6.90}$$

where we denote by Y the vector consisting of all random variables Y_i. An estimation protocol $(\hat{\rho}, \{\Pi_i\}_{i\in[M]})$ minimizing the minimax risk has the smallest possible worst-case probability over the set of quantum states.

The following theorem provides a lower bound on the minimax risk for the estimation of the Choi matrix of a unitary gate from unit rank measurements.

Theorem 71 (Lower bound, unit rank measurements). *Fix a set $\mathcal{M}$ of rank-1 measurements. For $\varepsilon > 0$ the minimax risk (6.90) of measurements of M copies is bounded as*

$$R^*(M, \varepsilon) \geq 1 - c_1 \frac{\log(d)\log(|\mathcal{M}|)}{d^4(1-\varepsilon/2)^2} M - \frac{c_2}{d^2(1-\varepsilon^2)}, \tag{6.91}$$

where c_1 and c_2 are absolute constants.

Before providing a proof for this theorem let us work out its consequences. If the measurements project onto Clifford unitaries, we get the following lower bound on the minimax risk.

Corollary 72 (Lower bound, Clifford group). *Let $\varepsilon > 0$ and consider measurements of the form (6.89) given by Clifford group unitaries on M copies. Then the minimax risk (6.90) is bounded as*

$$R^*(M, \varepsilon) \geq 1 - c_3 \frac{\log(d)^3}{d^4(1-\varepsilon/2)^2} M - \frac{c_2}{d^2(1-\varepsilon^2)}, \tag{6.92}$$

where c_3 and c_2 are absolute constants.

Proof. The cardinality of the n-qubit Clifford group ($d = 2^n$) is bounded as

$$|\operatorname{Cl}(d)| = 2^{n^2+2n} \prod_{j=1}^{n} (4^j - 1) < 2^{2n^2+4n} \tag{6.93}$$

[Cal+98]. This implies that in case of our Clifford group measurements we have $\log(|\mathcal{M}|) < 2\log(d)^2 + 4\log(d)$. □

In every meaningful measurement and reconstruction scheme the minimax risk needs to be small. The corollary implies that, in the case of Clifford unitaries, the number of copies M need to scale with the dimension as

$$M \in \Omega\left(\frac{d^4}{\log(d)^3}\right) , \tag{6.94}$$

where we have assumed $\varepsilon > 0$ to be small. This establishes a lower bound on the number of channel uses that every POVM measurement and reconstruction scheme requires for a guaranteed successful recovery of unitary channels from AGFs with respect to Clifford unitaries.

From the argument as it is presented here it is not possible to extract the optimal parametric dependence of the number of channel uses M on the desired reconstruction error ε. For quantum state tomography such bounds were derived in Ref. [Haa+17] by extending the argument of Ref. [Fla+12] and constructing different ε-packing nets. By adapting the ε-packing net constructions of Ref. [Haa+17] to unitary gates one might be able to derive an optimal parametric dependence of M on ε. But it is not obvious how one can incorporate the restriction of the measurements to unit rank in the argument of Ref. [Haa+17]. We leave this task to future work.

In the remainder of this section we prove Theorem 71. The proof proceeds in two steps. At first we derive a more general bound on the minimax risk, Lemma 73, that follows mainly from combining Fano's inequality with the data processing inequality, see e.g. [CT12]. This is a slight generalization of Lemma 1 of Ref. [Fla+12] adjusted to the situation where the outcome probabilities of the POVM measurements do not necessarily concentrate around the value $1/2$. Lemma 73 assumes the existence of an ε-packing net for the set of unitary gates whose measurement outcomes are in a small interval to establish a lower bound on the minimax risk. Hence, in order to complete the proof, we have to establish the existence of a suitable packing net, Lemma 77, in a second step. Combining the general bound of Lemma 73 and the existence of the packing net of Lemma 77, the proof of Theorem 71 follows.

We begin with the general information theoretic bound on the minimax risk.

Lemma 73 (Lower bound to the minimax risk). *Let $\varepsilon > 0$ and $0 < \alpha < \beta \leq 1/2$. Assume that there are states $\rho_1, \dots, \rho_s \in \mathrm{Pos}_D$ and orthogonal projectors $\pi_1, \dots, \pi_n \in \mathrm{Pos}_D$ such that*

$$\left\|\rho_i - \rho_j\right\|_1 \geq \varepsilon \tag{6.95}$$

$$\mathrm{Tr}[\pi_k \rho_i] \in [\alpha, \beta] \tag{6.96}$$

for all $i \neq j \in [s]$ and $k \in [n]$. Then the minimax risk (6.90) of M single measurements is bounded as

$$R^*(M, \varepsilon) \geq 1 - \frac{M(h(\beta) - h(\alpha)) + 1}{\log(s)}\,, \tag{6.97}$$

where h denotes the binary entropy.

Proof. We start by following the proof of [Fla+12, Lemma 1]: Let X be the random variable uniformly distributed over $[s]$ and let $Y_1, \dots, Y_M$ be the random variables describing the M single POVM measurements performed on ρ_X. Consider any estimator $\hat{\rho}$ of the state ρ_X from the measurements Y and define

$$\hat{X}(Y) := \operatorname*{arg\,min}_{i \in [s]} \left\|\hat{\rho}(Y) - \rho_i\right\|_1 \,. \tag{6.98}$$

Then, for all $i \in [s]$,

$$\mathbb{P}[\|\hat{\rho}(Y) - \rho_i\|_1 \geq \varepsilon] \geq \mathbb{P}[\hat{X}(Y) \neq X]. \tag{6.99}$$

Following Ref. [Fla+12], we combine Fano's inequality and the data processing inequality for the mutual information $I(X;Z) = H(X) - H(X|Z)$, where H denotes the entropy and conditional entropy, to obtain

$$\mathbb{P}[\hat{X}(Y) \neq X] \geq \frac{H(X|\hat{X}(Y)) - 1}{\log(s)} \geq 1 - \frac{I(X;Y) + 1}{\log(s)}\,. \tag{6.100}$$

Now we start deviating from Ref. [Fla+12]. We use that $I(X;Y) = I(Y;X)$, the chain rule, and the definition of the conditional entropy to obtain

$$\begin{aligned}
\mathbb{P}[\hat{X}(Y) \neq X] &\geq 1 - \frac{H(Y) - H(Y|X) + 1}{\log(s)} \\
&= 1 - \frac{1}{\log(s)}\Bigg(\sum_{j=1}^{M}\Big\{H(Y_j|Y_{j-1}, \dots, Y_1) \\
&\qquad\qquad - \frac{1}{s}\sum_{i=1}^{s} H(Y_j|Y_{j-1}, \dots, Y_1, X = i)\Big\} + 1\Bigg).
\end{aligned}$$

Now we use that $H(Y_j|Y_{j-1},\dots,Y_1,X=i) \geq h(\alpha)$ and $H(Y_j|Y_{j-1},\dots,Y_1) \leq h(\beta)$, where h is the binary entropy, to arrive at

$$\begin{aligned}\mathbb{P}[\hat{X}(Y) \neq X] &\geq 1 - \frac{M(h(\beta)-h(\alpha))+1}{\log(s)} \\ &\geq 1 - \frac{M(h(\beta)-h(\alpha))+1}{\log(s)}.\end{aligned}$$

□

To apply Lemma 73 we need to proof the existence of an ε-packing net $\{\rho_i\}_{i=1}^{s}$ consisting of unitary quantum gates with the properties (6.95) and (6.96). The construction of such a suitable ε-packing net will use the fact that the modulus of the trace of a Haar random unitary matrix is a sub-gaussian random variable. This can be viewed as a non-asymptotic version of a classic result by Diaconis and Shahshahani [DS94]: the trace of a Haar random unitary matrix in $\mathrm{U}(d)$ is a complex Gaussian random variable in the limit of infinitely large dimensions d.

The trace of Haar random unitaries is sub-gaussian. The statement follows from the fact that the moments of the modulus of the trace of a Haar random unitary are dominated by the moments of a Gaussian variable.

Proposition 74. *For all $d, k \in \mathbb{Z}_+$*

$$\mathbb{E}_{U\sim\mu_{\mathrm{U}(d)}}\left[|\operatorname{Tr}[U]|^{2k}\right] \leq k!, \tag{6.101}$$

with equality if $k \leq d$.

Proof. Denote by $S := |\operatorname{Tr}(U)|^2$ the random variable with $U \in \mathrm{U}(d)$ drawn from the Haar measure. Let $\{|n\rangle\}_{n=1}^{d^k}$ be an orthonormal basis of $(\mathbb{C}^d)^{\otimes k}$. The kth moment of S is given by

$$\mathbb{E}[S^k] = \sum_{n,m=1}^{d^k} \langle n| U^{\otimes k} |n\rangle \langle m| (U^\dagger)^{\otimes k} |m\rangle . \tag{6.102}$$

Applying Theorem 14, we get

$$\mathbb{E}[S^k] = \frac{1}{k!} \sum_{n,m=1}^{d^k} \sum_{\tau \in \mathfrak{S}_k} \sum_{\lambda \vdash k, l(\lambda) \leq d} \frac{d_\lambda}{D_\lambda} \langle m| \pi_k(\tau) |n\rangle \langle n| \pi_k(\tau^{-1}) P_\lambda |m\rangle \tag{6.103}$$

$$= \frac{1}{k!} \sum_{\tau \in \mathfrak{S}_k} \sum_{\lambda \vdash k, l(\lambda) \leq d} \frac{d_\lambda}{D_\lambda} \operatorname{Tr}(\pi_k(\tau)\pi_k(\tau^{-1}) P_\lambda) \tag{6.104}$$

$$= \sum_{\lambda \vdash k, l(\lambda) \leq d} \frac{d_\lambda}{D_\lambda} \operatorname{Tr}(P_\lambda). \tag{6.105}$$

Since $\operatorname{Tr}(P_\lambda) = d_\lambda D_\lambda$, we conclude

$$\mathbb{E}[S^k] = \sum_{\lambda \vdash k, l(\lambda) \leq d} d_\lambda^2 \leq \sum_{\lambda \vdash k} d_\lambda^2 = k!\,. \tag{6.106}$$

The last equality can be seen from the orthogonality relation of the characters of the symmetric group, see e.g. Ref. [FH91, Chapter 2] for more details. Note that the second inequality is saturated in the case where $k \leq d$ since in this case the restriction $l(\lambda) \leq d$ is automatically fulfilled. □

As a simple implication of the previous lemma is that the random variable $S = |\operatorname{Tr}(U)|^2$ has sub-exponential tail decay.

Lemma 75. *Let S be a real-valued random variable that obeys $\mathbb{E}\left[|S|^k\right] \leq k!$ for all $k \in \mathbb{N}$. Then, the right tail of X decays at least sub-exponentially. For any $t \geq 0$,*

$$\mathbb{P}\left[S \geq t\right] \leq \mathrm{e}^{-\kappa t + 2},$$

with $\kappa = 1 - \frac{1}{2\mathrm{e}}$.

This is a consequence of a standard result in probability theory that can be found in many textbooks, e.g. [Ver12] and [FR13, Section 7.2]. We present a short proof here in order to be self-contained.

Proof. We use Markov's inequality, Proposition 74, and Stirling's bound $k! \leq \mathrm{e}\sqrt{k}\, k^k\, \mathrm{e}^{-k}$ to obtain for any $k \in \mathbb{N}$

$$\mathbb{P}[S \geq k] \leq \frac{\mathbb{E}[|S|^k]}{k^k} \leq \frac{k!}{k^k} \leq \mathrm{e}\sqrt{k}\mathrm{e}^{-k}. \tag{6.107}$$

In order to prove the tail bound, we choose $t \geq 0$ arbitrary and let k be the largest integer that is smaller or equal to t ($k = \lfloor t \rfloor$). Then

$$\Pr\left[S \geq t\right] \leq \Pr\left[S \geq k\right] \leq \mathrm{e}\sqrt{k}\mathrm{e}^{-k} \leq \mathrm{e}^{-\kappa k+1} \leq \mathrm{e}^{-\kappa t+1+\kappa}.$$

Here, we have used $\sqrt{k}\mathrm{e}^{-k} \leq \mathrm{e}^{-\kappa k}$ and $t \leq k+1$. □

Random variables with sub-gaussian tail decay—*sub-gaussian random variables* —are closely related to random variables with sub-exponential tail decay: X is sub-gaussian if and only if X^2 is sub-exponential.

Thus, Proposition 74 highlights that the trace of a Haar-random unitary is a sub-gaussian random variable. This is the aforementioned generalization of the classical result by Diaconis and Shashahani.

A packing net with concentrated measurements. The proof of existence of an ε-packing net to apply Lemma 73 uses a probabilistic argument as in Ref. [Fla+12]. Here, the strategy is the following: We assume we are already given an ε-packing net of a size $s-1$ that satisfies the desired concentration condition (6.96). We then show that a Haar random unitary gate also fulfils the concentration condition and is ε-separated from the rest of the net with strictly positive probability. Consequently, if one can be lucky to randomly arrive at a suitable ε-packing net of size s in this way then it must also exist.

We start by deriving an anti-concentration result for the Choi matrix $\mathfrak{J}(\mathcal{U})$ of a unitary channel given by a Haar random unitary U in $\mathrm{U}(d)$.

Lemma 76. *Let $\mathcal{V}$ be a unitary gate. For all $\varepsilon > 0$*

$$\mathbb{P}_{U\sim\mu_{\mathrm{U}(d)}}\left[\|\mathfrak{J}(\mathcal{U}) - \mathfrak{J}(\mathcal{V})\|_1 \leq \varepsilon\right] \leq \mathrm{e}^{-\kappa d^2(1-\varepsilon/2)^2+2} \tag{6.108}$$

with $\kappa > 0$ being the constant from Lemma 75.

Proof. Due to the unitary invariance of the trace norm and the Haar measure, it suffices to show the statement for $\mathcal{V} = \mathrm{Id}$. For a unitary channel with Choi-matrix $\mathfrak{J}(\mathcal{U}) = d^{-1}\operatorname{vec}(U)\operatorname{vec}(U^\dagger)^t$ and Kraus-operator $U \in \mathrm{U}(d)$ we have

$$\|\mathfrak{J}(\mathcal{U}) - \mathfrak{J}(\mathrm{Id})\|_1 = 2\sqrt{1 - \frac{1}{d^2}|\operatorname{Tr}(U)|^2} \geq 2\left(1 - \frac{1}{d}|\operatorname{Tr}(U)|\right). \tag{6.109}$$

For the first equation we calculate the set eigenvalues of $\mathfrak{J}(\mathcal{U}) - \mathfrak{J}(\mathrm{Id})$, which is $\{\pm\sqrt{1-d^{-2}|\operatorname{Tr}(U)|^2}\}$. Introducing the random variable $S_U := |\operatorname{Tr}(U)|^2$, we can rewrite the probability as

$$\begin{aligned}\mathbb{P}[\|\mathfrak{J}(\mathcal{U}) - \mathfrak{J}(\mathrm{Id})\|_1 \leq \varepsilon] &\leq \mathbb{P}\left[2\left(1 - \frac{1}{d}\sqrt{S_U}\right) \leq \varepsilon\right] \\ &= \mathbb{P}\left[S_U \geq d^2\left(1 - \frac{\varepsilon}{2}\right)^2\right]. \qquad (6.110)\end{aligned}$$

From Lemma 75 we know that

$$\mathbb{P}\left[S_U \geq d^2\left(1 - \frac{\varepsilon}{2}\right)^2\right] \leq \mathrm{e}^{-\kappa d^2(1-\varepsilon/2)^2+2} \qquad (6.111)$$

from which the assertion follows. □

The anti-concentration result of Lemma 76 implies the existence of a large ε-packing net $\mathcal{N}_\varepsilon$ of unitary quantum channels. The desired concentration of the measurement outcomes can be established using Lemma 75. In summary we arrive at the following assertion:

Lemma 77 (Packing net with concentrated measurements). *Let $0 < \varepsilon < 1/2$, $\kappa = 1 - \frac{1}{2\mathrm{e}}$, and $C_1, \dots, C_K \in \mathrm{U}(d)$. Then, for any number $s < \frac{1}{2}\mathrm{e}^{\kappa(1-\varepsilon/2)^2 d^2 - 2}$, there exist $U_1, \dots, U_s \in \mathrm{U}(d)$ such that for all $i, j \in [s]$ with $i \neq j$ and for all $k \in [K]$*

$$\|\mathfrak{J}(\mathcal{U}_i) - \mathfrak{J}(\mathcal{U}_j)\|_1 \geq \varepsilon\,, \qquad (6.112)$$

$$\frac{1}{d^2}|\operatorname{Tr}[C_k^\dagger U_i]|^2 \leq \frac{\log(2K) + 2}{\kappa d^2}\,. \qquad (6.113)$$

Proof. As outlined above the existence of the described ε-packing net follows inductively from the fact that if one adds a Haar random unitary gate $\mathcal{U}$ to an ε-packing $\tilde{\mathcal{N}}_\varepsilon$ of size $s-1$ that already fulfils all requirements of the lemma the resulting set $\tilde{\mathcal{N}}_\varepsilon \cup \{\mathcal{U}\}$ has still a strictly positive probability to be an ε-packing net with the desired concentration property (6.113).

We start with bounding the probability that the resulting set $\tilde{\mathcal{N}}_\varepsilon \cup \{\mathcal{U}\}$ fails to be an ε-packing net. Let us denote the probability that a Haar random $\mathcal{U}$ is not ε-separated from $\tilde{\mathcal{N}}_\varepsilon$ by $\bar{p}_\varepsilon$. In other words, $\bar{p}_\varepsilon$ is the probability that there exists $\mathcal{V} \in \tilde{\mathcal{N}}_\varepsilon$ with $\|\mathfrak{J}(\mathcal{U}) - \mathfrak{J}(\mathcal{V})\|_1 \leq \varepsilon$. Taking the union bound for all $\mathcal{V} \in \tilde{\mathcal{N}}_\varepsilon$, Lemma 76 implies that

$$\bar{p}_\varepsilon \leq s\mathrm{e}^{-\kappa d^2(1-\epsilon/2)^2+2} \qquad (6.114)$$

with $\kappa = 1 - \frac{1}{2\mathrm{e}}$. Thus, for $s < \frac{1}{2}\mathrm{e}^{-\kappa d^2(1-\epsilon/2)^2+2}$ we ensure that $\bar{p}_\varepsilon < \frac{1}{2}$. We now also have to upper-bound the probability $\bar{p}_c$ of $\mathcal{U}$ not having a concentration property

$$\frac{1}{d^2}|\operatorname{Tr}[C_k^\dagger U_i]|^2 \leq \beta \tag{6.115}$$

with respect to K different unitaries $C_1, \dots, C_K$. Using the unitary invariance of the Haar measure and taking the union bound, the tail-bound for the squared modulus of the trace of a Haar random unitary, Lemma 75, yields

$$\bar{p}_c \leq K\mathrm{e}^{-\kappa\beta d^2+2} \tag{6.116}$$

for $\beta \geq 2$. In order for $\bar{p}_c$ to be at most $1/2$, we need that

$$\beta \geq \frac{\log(2K)+2}{\kappa d^2}. \tag{6.117}$$

In summary, we have established that $\bar{p}_\varepsilon + \bar{p}_c < 1$ as long as $s < \frac{1}{2}\mathrm{e}^{-\kappa d^2(1-\epsilon/2)^2+2}$ and the achievable concentration is $\beta \geq (\log(2K)+2)/(\kappa d^2)$. Hence, in this parameter regime there always exist at least one additional unitary gate extending the ε-packing net. Inductively this proves the existence assertion of the lemma. □

Having established a suitable ε-packing net, we can now apply Lemma 73 to derive the lower bound on the minimax-risk for the recovery of unitary gates from unit rank measurements of Theorem 71, the main result of this section.

Proof of Theorem 71. We will apply Lemma 73 with $\alpha = 0$ and

$$\beta = \frac{\log(2|\mathcal{M}|)+2}{\kappa d^2}, \tag{6.118}$$

and we use that $h(\beta) \leq 2\beta\log(1/\beta)$ for $\beta \leq 1/2$. Combining the Lemmas 73 and 77 we obtain

$$\begin{aligned} R^*(M,\varepsilon) &\geq 1 - \frac{Mh(c/d^2)+1}{(\kappa(1-\varepsilon/2)^2 d^2+2)/\log(2)-2} && (6.119)\\ &\geq 1 - \frac{2\frac{\log(2|\mathcal{M}|)+2}{\kappa d^2}\log\left(\frac{\log(2|\mathcal{M}|)+2}{\kappa d^2}\right)M+1}{d^2(\kappa(1-\varepsilon/2)^2 d^2+2)/\log(2)-2}, && (6.120)\end{aligned}$$

where, in Lemma 77 we have chosen s to be the strict upper bound minus one. Finally, we simplify the bound by choosing large enough constants c_1 and c_2. □

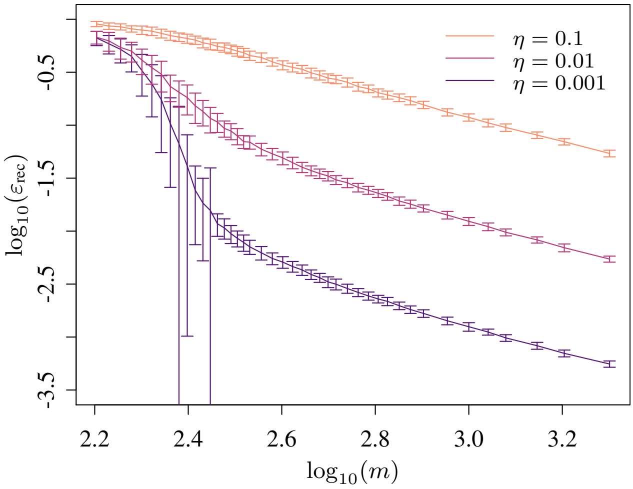


Figure 6.1: Reconstruction of a Haar-random 3-qubit channel using the optimization (6.20): The plots show the dependence of the observed average reconstruction error $\varepsilon_{\text{rec}} := \left\| \mathcal{Z}^\sharp - \mathcal{X} \right\|$ on the number of AGFs m for different noise strengths $\eta := \|\epsilon\|_{\ell_2}$. The error bars denote the observed standard deviation. The averages are taken over 100 samples of random i.i.d. measurements and channels (nonuniform). The MATLAB code and data used to create these plots can be found on GitHub [25].

6.4 Numerical demonstrations

Finally, with numerical simulations, we briefly explore the practical feasibility of the reconstruction procedure (6.20) and discuss some of its subtleties. The reconstruction algorithm (6.20) can be implemented with standard optimization packages. The Matlab code for our numerical experiments can be found on GitHub [25].[11] Let $\mathcal{X}$ denote a unitary quantum channel. Given measurements f_i from (6.60) with Clifford unitaries C_i we approximately recover $\mathcal{X}$ using the semi-definite program (SDP) (6.62) with $q = 2$. In the numerical experiments we draw a three-qubit unitary channel $\mathcal{X}$ uniformly at random, the m Clifford unitaries for the measurements uniformly at random, and the noise $\epsilon \in \mathbb{R}^m$ uniformly from a sphere with radius η, i.e., $\|\epsilon\|_{\ell_2} = \eta$. Then we solve the SDP using Matlab, CVX and SDPT3. The resulting average reconstruction error is plotted against the number of measurement settings m and the noise strength η in Figure 6.1 and Figure 6.2 (left), respectively. As a comparison we run simulations for Haar random unitary measurements, see Figure 6.2 (right). We find that the measurements based on random Clifford unitaries perform equally well as mea-

[11] Finally, with numerical simulations, …] This section has been published in the supplemental material of Ref. [1] with minor modifications.

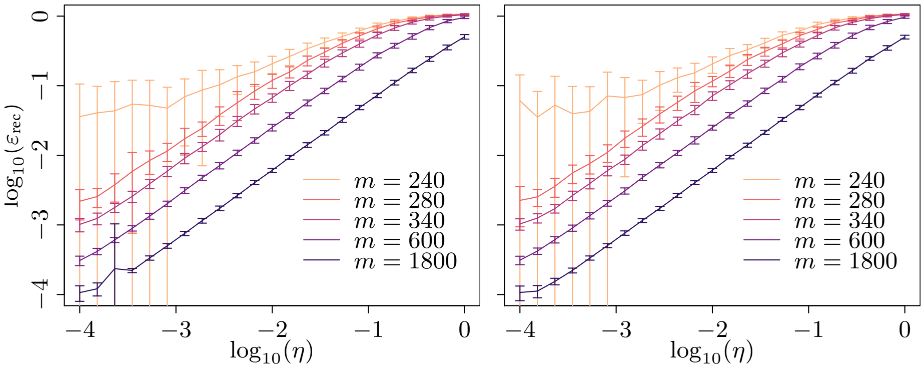


Figure 6.2: Comparison of the reconstruction (6.20) from AGFs (6.19) with random Clifford unitaries (left) and Haar random unitaries (right). The plots show the dependence of the observed average reconstruction error $\varepsilon_{\mathrm{rec}} := \left\|\mathcal{Z}^\sharp - \mathcal{X}\right\|$, on the noise strength $\eta := \|\epsilon\|_{\ell_2}$ for 3 qubits and different numbers of AGFs m. The error bars denote the observed standard deviation. The averages are taken over 100 samples of random i.i.d. measurements and channels (non-uniform). The Matlab code and data used to create these plots can be found on GitHub [25].

surements based on Haar random unitaries. This is in agreement with a similar observation made for the noiseless case by two of the authors in Ref. [KL17].

We observed that sometimes the SDP solver cannot find a solution. We also tested the use of Mosek instead of SDPT3. We find that the Mosek solver is faster but has more problems finding the correct solution. For the cases where the SDP solver does not exit with status “solved” we relax the machine precision on the equality constraints in the SDP (6.62) and change the measurement noise by a machine precision amount. More explicitly, for an integer $j \geq 0$ we try to solve

$$\begin{aligned} \underset{Z}{\text{minimize}} \quad & \|\mathcal{A}(Z) - f\|_{\ell_2} \\ \text{subject to} \quad & Z \geq 0, \\ & \left\|\mathrm{Tr}_1(Z) - \frac{\mathrm{Id}}{d}\right\|_{\mathrm{F}} \leq 10^j \,\mathrm{eps}, \\ & \left\|\mathrm{Tr}_2(Z) - \frac{\mathrm{Id}}{d}\right\|_{\mathrm{F}} \leq 10^j \,\mathrm{eps} \end{aligned} \tag{6.121}$$

where eps denotes the machine precision and Tr_1 and Tr_2 the partial traces on $\mathrm{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$. We successively try to solve these SDPs for $j = 0, 1, 2, \dots, 6$. Moreover, we change the measurement noise ϵ to $\epsilon' + \zeta$ in each of these trials, where each $\zeta_i = \mathrm{eps} \cdot g_i$ with $g_i \sim \mathcal{N}(0,1)$ is an independent normally distributed random number. For the Clifford type measurement (Figures 6.1 and 6.2 left) a total of 24 400 channels were reconstructed and j was increased 1 865 many times in total. For the Haar random measurement unitaries (Figure 6.2 left) a total of 12 900 channels were reconstructed and j was increased 950 times. So, we observed that with a probability of $\sim 7.5\%$ the SDP solver cannot solve the given SDP with machine precision constraints.

Some error bars in the plots in Figures 6.1 and 6.2 might seem quite large, which we would like to comment on. Note that in compressed sensing it is typical to have a phase-transition from having no recovery for too small numbers of measurements m to having a recovery with very high probability once m exceeds a certain threshold. This phase transition region becomes smeared out if the noise strength $\|\epsilon\|_{\ell_2}$ is increased. For those m in the phase transition region the reconstruction errors are expected to fluctuate a lot, which we observe in the plots.

The slope of the linear part of plots $\varepsilon_{\mathrm{rec}}(m)$ in Figure 6.1 is roughly $\delta\varepsilon_{\mathrm{rec}}(m)/\delta m \approx -1.3$. This means that the reconstruction error scales like $\varepsilon_{\mathrm{rec}}(m) \sim m^{-1.3}$, which is better than Theorem 52 suggests. The reason for this discrepancy is that the theorem also bounds systematic errors and even adversarial noise ϵ whereas in the numerics we have drawn ϵ_i uniformly from a sphere, i.e., ϵ_i are i.i.d. up to a rescaling.

6.5 Unitarity

The final result of this chapter addresses the *unitarity* of a quantum channel.[12] Introduced by Wallman *et al.* [Wal+15], the unitarity is a measure for the coherence of a (noise) channel $\mathcal{E}$. Recall from the preliminares, Equation (2.91), that it is defined to be the average purity of the output states of a slightly altered[13]channel $\mathcal{E}'$

$$u(\mathcal{E}) = \int \mathrm{d}\psi \, \mathrm{Tr}\left(\mathcal{E}'\left(|\psi\rangle\!\langle\psi|\right)^\dagger \mathcal{E}'\left(|\psi\rangle\!\langle\psi|\right)\right) \tag{6.122}$$

that flags the absence of trace preservation and unitality. The unitarity can be estimated efficiently by using techniques similar to randomized benchmarking [Fen+16]. It is also an important figure of merit when one aims to compare the

[12] The final result of … e.g. in Ref. [MBE11].] Is based on the discussion of the unitarity characterization in the main text of Ref. [1].

[13] $\mathcal{E}'$ is defined so that $\mathcal{E}'(\mathrm{Id}) = 0$ and $\mathcal{E}'(X) = \mathcal{E}(X) - \mathrm{Tr}(\mathcal{E}(X)/\sqrt{d}\,\mathrm{Id}$ for all traceless X.

AGF of a noisy gate implementation to its diamond distance [Kue+16; Wal15]—a task that is important for certifying fault-tolerance capabilities of quantum devices.

Although useful, the existing definition of the unitarity (6.122) is arguably not very intuitive. Here, we try to (partially) amend this situation by providing a simple statistical interpretation:

Theorem 78 (Operational interpretation of unitarity)**.** *Let $\{\mathcal{U}_k\}_{k=1}^{N}$ be the gate set of a unitary 2-design. Then, for all Hermicity-preserving maps $\mathcal{X}$*

$$\mathrm{Var}\left[F_{\mathrm{avg}}\left(\mathcal{U}_k, \mathcal{X}\right)\right] = \frac{u(\mathcal{X})}{d^2(d+1)^2}, \tag{6.123}$$

where the variance is computed with respect to $\mathcal{U}_k$ drawn randomly from the unitary 2-design.

Note that the variance is taken with respect to unitaries drawn from the unitary 2-design and not the variance of the average fidelity with respect to the input state as calculated, e.g. in Ref. [MBE11].

Before providing the proof of the theorem, we allow ourselves to speculate about possible applications for Theorem 78.[14] A direct estimation procedure for the unitarity has been proposed in Ref. [Wal+15] and refined in Ref. [DHW19]. Inspired by randomized benchmarking, this procedure is robust towards SPAM errors, but has other drawbacks: Estimating the purity of outcome states directly is challenging, because the operator square function is not linear. Although Wallman *et al.* have found ways around this issue, their approaches are not yet completely satisfactory.

We suggest an alternative approach based on Theorem 78. It might be conceivable that techniques like importance sampling could be employed to efficiently estimate this variance—and thus the unitarity—from "few" samples. The fourth moment bounds computed here could potentially serve as bounds on the "variance of this variance" and help control the convergence. One major appeal to this strategy is that the same type of data relative average fidelities could then be used for multiple diagnostic tasks, without the necessity to perform different experimental setups.

[14] Before providing the proof ...] The remainder of the section is based on Section G of the supplemental material of Ref. [1].

It remains to prove the theorem. The proof is most naturally phrased by decomposing the linear hull of unital and trace preserving maps $\mathrm{L}_{\overline{\mathrm{u,tp}}}(H_d)$ into endomorphism acting on the spaces that carry irreducible representations of the unitary channels. In the proof of Theorem 50 we have explicitly seen that the projection of any map $\mathcal{X}$ onto $\mathrm{L}_{\overline{\mathrm{u,tp}}}(H_d)$ has the block-diagonal structure:

$$P_{\overline{\mathrm{u,tp}}}(\mathcal{X}) = \mathcal{X}_0 \oplus x_1,$$

where $x_1 = \mathrm{Tr}\left(\mathcal{X}\left(\mathrm{Id}/d\right)\right)$. For channels that are already unital and trace preserving, this projection acts as the identity and $x_1 = 1$. Particular examples of this class are unitary channels $\mathcal{U} = \mathcal{U}_0 \oplus 1$ and the depolarizing channel $\mathcal{D} = \mathbb{O} \oplus 1$ acting as $\mathcal{D}(X) = \frac{\mathrm{Tr}(X)}{d}\,\mathrm{Id}$ on $X \in H_d$. Unitary channels are also special in the sense that they are normalized with respect to the inner products defined in (6.23):

$$d^2 = \mathrm{Tr}\left[\mathcal{U}^\dagger \mathcal{U}\right] = (\mathcal{L}(\mathcal{U}), \mathcal{L}(\mathcal{U})) = d^2\left(\mathcal{U}, \mathcal{U}\right).$$

In fact, unitary channels are the only maps with this property (provided that we also adhere to our convention of normalizing maps with respect to the trace norm of the Choi matrix). Combining this feature with the “block diagonal” structure of unitary channels yields

$$d^2 = \mathrm{Tr}\left[\mathcal{U}^\dagger \mathcal{U}\right] = \mathrm{Tr}\left[\mathcal{U}_0^\dagger \oplus 1\, \mathcal{U}_0 \oplus 1\right] = 1 + \mathrm{Tr}\left[\mathcal{U}_0^\dagger \mathcal{U}_0\right].$$

This computation implies that a map $\mathcal{X}$ is unitary if and only if

$$u(\mathcal{X}) := \frac{\mathrm{Tr}\left[\mathcal{X}_0^\dagger \mathcal{X}_0\right]}{d^2 - 1}$$

equals one. Otherwise, the *unitarity* $u(\mathcal{X}) \in [0, 1]$ is strictly smaller. For instance, $u(\mathcal{D}) = 0$ for the depolarizing channel. This definition of the unitarity is equivalent to the one presented in (6.122), see [Wal+15, Proposition 1]. The argument outlined above succinctly summarizes the main motivation for this figure of merit: it captures the coherence of a noise channel $\mathcal{X}$.

Equipped with this characterization of the unitarity, we can now give the proof for the interpretation of the unitarity as the variance of the AGF with respect to a unitary 2-design.

Proof of Theorem 78. The unitarity $u(\mathcal{X})$ may be expressed as

$$\frac{\mathrm{Tr}\left[\mathcal{X}_0^\dagger \mathcal{X}_0\right]}{d^2 - 1} = \frac{\mathrm{Tr}\left[\left(\mathcal{X}_0 \oplus (d^2 - 1)x_1\right)^\dagger \mathcal{X}\right]}{d^2 - 1} - x_1^2. \tag{6.124}$$

Equation (6.15) allows us to rewrite x_1 as an average over a unitary 1-design $\{\mathcal{U}_k\}_{k=1}^N$:

$$x_1 = \mathrm{Tr}\left[(\mathbb{O}\oplus 1)^\dagger\,\mathcal{X}\right] = \frac{1}{N}\sum_{k=1}^{N}\mathrm{Tr}\left[\mathcal{U}_k^\dagger\mathcal{X}\right] = \mathbb{E}\,\mathrm{Tr}\left[\mathcal{U}^\dagger\mathcal{X}\right]$$

Let us now assume that the set $\{\mathcal{U}_k\}_{k=1}^N$ is also a 2-design. Then, (6.6) implies

$$\frac{(\mathcal{X}_0\oplus(d^2-1)x_1)^\dagger}{d^2-1} = \sum_{k=1}^{n}\mathcal{U}_k^\dagger\overline{\mathrm{Tr}\left[\mathcal{U}_k^\dagger\mathcal{X}\right]} = \mathbb{E}\,\mathcal{U}^\dagger\,\mathrm{Tr}\left[\mathcal{X}^\dagger\mathcal{U}\right]$$

Inserting both expressions into (6.124) yields

$$\begin{aligned} u(\mathcal{X}) =& \mathrm{Tr}\left[\mathcal{X}^\dagger\mathbb{E}\,\mathcal{U}\,\mathrm{Tr}\left[\mathcal{U}^\dagger\mathcal{X}\right]\right] - \left(\mathbb{E}\,\mathrm{Tr}\left[\mathcal{X}^\dagger\mathcal{U}\right]\right)^2 \\ =&\mathbb{E}\,\left|\mathrm{Tr}\left[\mathcal{X}^\dagger\mathcal{U}\right]\right|^2 - \left(\mathbb{E}\,\mathrm{Tr}\left[\mathcal{X}^\dagger\mathcal{U}\right]\right)^2 \\ =&\mathrm{Var}\left[\mathrm{Tr}\left[\mathcal{X}^\dagger\mathcal{U}\right]\right], \end{aligned}$$

where we have used linearity of the expectation value and the fact that the random variable $\mathrm{Tr}\left[\mathcal{X}^\dagger\mathcal{U}\right]$ is real-valued. Finally, we employ the relation between $\mathrm{Tr}\left[\mathcal{U}^\dagger\mathcal{X}\right]$ and $F_{\mathrm{avg}}(\mathcal{U},\mathcal{X})$ presented in (6.8) to conclude

$$\begin{aligned} u(\mathcal{X}) =&\mathrm{Var}\left[\mathrm{Tr}\left[\mathcal{U}^\dagger\mathcal{X}\right]\right] \\ =&\mathrm{Var}\left[d(d+1)\,F_{\mathrm{avg}}(\mathcal{U},\mathcal{X}) - \mathrm{Tr}(\mathcal{X}(\mathrm{Id}))\right] \\ =&(d(d+1))^2\,\mathrm{Var}\left[F_{\mathrm{avg}}(\mathcal{U},\mathcal{X})\right], \end{aligned}$$

because variances are invariant under constant shifts and depend quadratically on scaling factors. This establishes Theorem 78. □

6.6 Conclusion and outlook

In this chapter, we addressed the crucial task of characterizing quantum channels. We do so by relying on AGFs of the quantum channel of interest with simple-to-implement Clifford gates.[15] More specifically, we start by noting that (i) the unital part of any quantum channel can be written in terms of a unitary

[15] In this chapter, we addressed the …] The conclusions are taken from the conclusions section of Ref. [1] with slight modifications.

2-design with expansion coefficients given by AGFs. As a consequence, for certain Hilbert space dimensions d, the unital part can be reconstructed from d^4 AGFs with Clifford-group operations by a straightforward and stable expansion formula. (ii) As the main result, we prove for the case of unitary gates that the reconstruction can be practically done using only essentially order of d^2 random AGFs with Clifford gates. In a simplified measurement setting, we prove that this also leads to a resource optimal scaling in terms of the total number of channel invocations required to estimate the AGFs up to a precision of ϵ. For the proof, we derived—in Chapter 4—a formula for the integration of fourth moments over the Clifford group, which is similar to Collins's calculus with Weingarten functions. This integration formula might also be useful for other purposes. (iii) We prove that the unitarity of a quantum channel, which is a measure for the coherence of noise [Wal+15], has a simple statistical interpretation: It corresponds to the variance of the AGF with unitaries sampled from a unitary 2-design.

The focus of this work is on the reconstruction of quantum gates. Here, the assumption of unitarity considerably simplifies the representation-theoretic effort for establishing the fourth moment bounds required for applying strong existing proof techniques from low-rank matrix recovery. These extend naturally to higher Kraus ranks, and we leave this generalization to future work. Existing results [KZG16b; KZG16a] indicate that the deviation of the Clifford group from a unitary 4-design may become more pronounced when the rank of the states or channels in question increases. This may lead to a sub-optimal rank scaling of the required number of observations m. In fact, a straightforward extension of Theorem 52 to the Kraus rank-r case already yields a recovery guarantee with a scaling of $m \sim r^5 d^2 \log(d)$.

Practically, it is important to explore how the identification protocol behaves when applied to data obtained from interleaved randomized benchmarking experiments. Such numerical studies would further allow for a comparison to other established schemes such as GST, for which no theoretical guarantees exist. In Ref. [Kim+14], the authors show how to use interleaved randomized benchmarking experiments to measure the AGF between a known Clifford gate and the combined process of an unknown gate concatenated with the average Clifford error process. In order to obtain tomographic information about the isolated unknown gate, the authors had to do a linear inversion of the average Clifford error. However, in most cases, we expect the average Clifford error to be close to a depolarizing channel which has very high rank. Thus, building on our intuition obtained for quantum states [Rio+17] and using our techniques, we could obtain a low-rank approximation to the combined unknown gate and average

Clifford error, which under the assumption of a high-rank Clifford error, would naturally pick out the coherent part of the unknown gate.

7 Hamiltonian identification for analogue simulation

In this chapter, we turn to the identification problem of a Hamiltonian that is implemented in a analogue simulation of time trace data. For a quantum dynamics analogue simulator such identifications task are arguably the most important diagnostic primitives as they assess the quality and identify the deviations from the device's specification—Hamiltonian identification assess the quality of the application layer of a analogue simulator.

Interestingly, the level of control of modern quantum analogue simulators permits one to assess the quality of the Hamiltonian implementation in a bottom-up approach: starting with very simple Hamiltonians and, subsequently, increasing the complexity of the tasks. As we will see in this chapter, already the identification of very simple Hamiltonians—non-interacting, quadratic bosonic Hamiltonians—gives rise to a rich signal processing task, and presents severe challenges in terms of uncontrolled effects in the state-preparation and measurement phase in practice. We here report on the precise identification of non-interacting, quadratic bosonic Hamiltonians that were experimentally realized on a superconducting-qubit analogue quantum simulator. In the setup the state-preparation and the measurement phase include ramping phases between the idle frequencies of the qubits, that realize the excitations and couplings, to the Hamiltonian parameter. This introduces a particular form of SPAM errors that, if not accounted for, severely limit the precision of the identification task. We devise an identification method that is custom-tailored to the problem at hand and fully exploits its structure in order to achieve the required level of semi-device dependence for a precise identification. Furthermore, by comparing the identified Hamiltonians with their control targets for families of Hamiltonians we provide an *analogue simulator benchmark* for the components of super-conducting processors.

We begin with reporting the experimental results and subsequently give a more detailed account of the method.

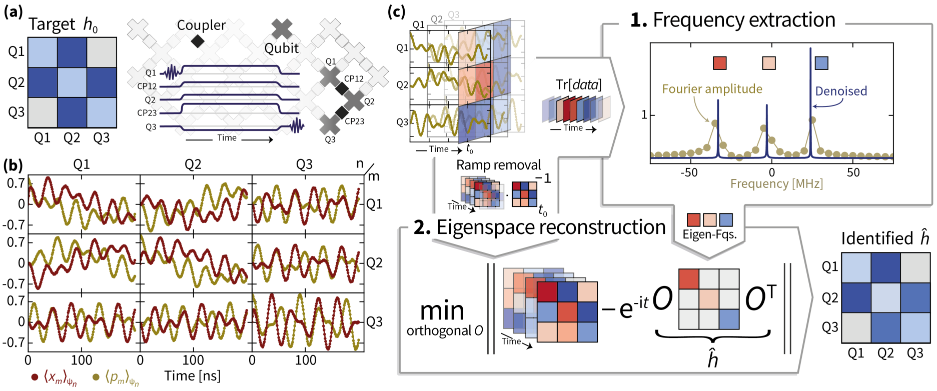


Figure 7.1: **Outline of the experiment and identification algorithm. (a)** The time evolution under a target Hamiltonian h_0 is implemented on an part of the Google Sycamore chip (gray) using the pulse sequence depicted in the middle. **(b)** The expected value of canonical coordinates x_m and p_m for each qubit m over time is estimated from measurements using different ψ_n as input states. **(c)** The data shown in (b) for each time t_0 can be interpreted as a (complex-valued) matrix with entries indexed by measured and initial excited qubit, m and n. The identification algorithm proceeds in two steps: 1. From the time-dependent matrix trace of the data, the Hamiltonian eigenfrequencies are extracted using a super-resolving, denoising algorithm. The blue line indicates the denoised, high-resolution signal as 'seen' by the algorithm. 2. After removing the ramp using the tomographic estimate provided by a the matrix at a fixed time, the Hamiltonian eigenspaces are reconstructed using a non-convex optimization algorithm over the orthogonal group. From the extracted frequencies and reconstructed eigenspaces, we can calculate the identified Hamiltonian $\hat{h}$ that describes the measured time evolution.

7.1 Experimental results

Setup. We characterize the Hamiltonian[1] governing analogue dynamics of a Google Sycamore chip, which consists of a two-dimensional array of nearest-neighbour coupled superconducting qubits. Each physical qubit is a non-linear

[1] We characterize …improving analogue quantum devices.] The remainder of this chapter focusing on Hamiltonian identification has been published as the preprint Ref. [3]. The work has been conducted together with Dominik Hangleiter and under supervision of Jens Eisert. It contains equal contributions of Dominik Hangleiter and the thesis' author to all parts of the project. The experimental data was taken by Pedram Roushan at Google Quantum AI. The hardware used for this experiment was developed by the Google Quantum AI hardware team, under the direction of Anthony Megrant, Julian Kelly and Yu Chen. We further acknowledge contributions from Charles Neill, Kostyantyn Kechedzhi, and Alexander Korotkov to the calibration procedure used in this analogue approach. We would like to thank Christian Krumnow, Benjamin Chiaro, Alireza Seif, Markus Heinrich, and Juani Bermejo-Vega for fruitful discussions in early stages of the project.

oscillator with bosonic excitations (microwave photons) [Car+20]. Using the rotating-wave approximation the dynamics governing the excitations of the qubits in the rotating frame can be described by the Bose-Hubbard Hamiltonian

$$H_{\mathrm{BH}} = \sum_i \left(\mu_i a_i^\dagger a_i + \eta_i a_i^\dagger a_i^\dagger a_i a_i \right) - \sum_{i \neq j} J_{i,j} a_i^\dagger a_j \,, \tag{7.1}$$

where $a_i^\dagger$ and a_i denote bosonic creation and annihilation operators at site i, respectively, μ_i is the on-site potential, $J_{i,j}$ is the hopping rate between nearest neighbour qubits, and η_i is the on-site interaction strength. The qubit frequency, the nearest-neighbour coupling between them, and the non-linearity set μ, J, and η, respectively. We are able to tune μ and J on nanoseconds timescales, while η is fixed.

Here, we focus on the specific task of identifying the values μ_i and $J_{i,j}$. The corresponding non-interacting Hamiltonian acting on N modes can be conveniently parametrized as

$$H(h) = - \sum_{i,j=1}^{N} h_{i,j} a_i^\dagger a_j \tag{7.2}$$

with an $N \times N$ Hermitian parameter matrix h with entries $h_{i,j}$, which is composed of the on-site chemical potentials μ_i on its diagonal and the hopping energies $J_{i,j}$ for $i \neq j$. The identification of the non-interacting part $H(h)$ of H_{BH} can be viewed as a first step in a hierarchical procedure for characterizing dynamical quantum simulations with tunable interactions and numbers of particles.

The non-interacting part $H(h)$ of the Hamiltonian H_{BH} can be inferred when initially preparing a state where only a single qubit is excited with a single photon. For initial states with a single excitation, the interaction term vanishes, hence effectively $\eta = 0$. Consequently, only the two lowest energy level of the non-linear oscillators enter the dynamic. Therefore, referring to them as qubits (two-level systems) is precise. Specifically, we identify the parameters $h_{i,j}$ from dynamical data of the following form. We initialize the system in $|\psi_n\rangle = (\mathrm{id} + a_n^\dagger) |0\rangle^{\otimes N} / \sqrt{2}$ and measure the canonical coordinates $x_m = (a_m + a_m^\dagger)/2$ and $p_m = (a_m - a_m^\dagger)/(2\mathrm{i})$ for all combinations of $m, n = 1, \dots, N$. In terms of the qubit architecture, this amounts to local Pauli-X and Pauli-Y basis measurements, respectively. We combine the statistical averages over multiple measurements to obtain an empirical estimator for $\langle a_m(t) \rangle_{\psi_n} = \langle x_m(t) \rangle_{\psi_n} + \mathrm{i} \langle p_m(t) \rangle_{\psi_n}$.

For particle-number preserving dynamics, this data is of the form

$$\langle a_m(t) \rangle_{\psi_n} = \frac{1}{2} \exp(-\mathrm{i}th)_{m,n} \,. \tag{7.3}$$

It therefore directly provides estimates of the entries of the time-evolution unitary at time t in the single-particle sector of the bosonic Fock space.

In Figure 7.1, we show an overview of the experimental procedure, and the different steps of the Hamiltonian identification algorithm. Every experiment uses a few coupled qubits, from the larger array of qubits on the device (Figure 7.1(a)). On those qubits, the goal is to implement the time-evolution with targeted Hamiltonian parameters h_0, which are subject to connectivity constraints imposed by the couplings of the qubits. To achieve this, we perform the following pulse sequence to collect dynamical data of the form (7.3). Before the start of the sequence, the qubits are at frequencies (of the $|0\rangle$ to $|1\rangle$ transition) that could be a few hundred MHz apart from each other. In the beginning, all qubits are in their ground state $|0\rangle$. To prepare the initial state, a $\pi/2$-pulse is applied to one of the qubits, resulting in its Bloch vector moving to the equator. Then ramping pulses are applied to all qubits to bring them to the desired detuning around a common rendezvous frequency (6500 MHz in this work). At the same time, pulses are applied to the couplers to set them to the desired value (20 MHz in this work). The pulses are held at the target values for time t, corresponding to the evolution time of the experiment. Subsequently, the couplers are ramped back to zero coupling and the qubits back to their initial frequency, where $\langle x_m(t) \rangle$ and $\langle p_m(t) \rangle$ on the desired qubit m is measured. The initial and final pulse ramping take place over a finite time of 2–3 ns, and therefore give rise to a non-trivial effect on the dynamics, which we take into account in the identification procedure. The experimental data (Figure 7.1(b)) on N qubits are $N \times N$ time-series estimates of $\langle a_m(t) \rangle_{\psi_n}$ for $t = 0, 1, \ldots, T$ ns and all pairs $n, m = 1, \ldots, N$. Given those data, the identification task amounts to identifying the 'best' coefficient matrix h, describing the time-sequence of snapshots of the single-particle unitary matrix $\frac{1}{2} \exp(-\mathrm{i}th)$.

Identification method. We can identify the generator h of the unitary in two steps (Figure 7.1(c)), making use of the eigendecomposition of the Hamiltonian (see the Methods section below). In the first step, the time-dependent part of the identification problem is solved, namely, identifying the Hamiltonian eigenvalues (eigenfrequencies). In the second step, given the eigenvalues, the eigenbasis for the Hamiltonian of h is determined. In order to make the identification method noise-robust, we furthermore exploit structural constraints of the model.

First, the Hamiltonian has a non-degenerate spectrum such that the time-series data has a time-independent, sparse frequency spectrum with exactly N distinct contributions. Second, the Fourier coefficients of the data have an explicit form as the outer product of the orthogonal eigenvectors of the Hamiltonian. Third, the Hamiltonian parameter matrix has an a priori known sparse support due to the experimental connectivity constraints. These structural constraints are not respected by various sources of incoherent noise, including particle loss and finite shot noise, and also coherent noise such as the accumulation of diagonal phases before the measurement. Thereby, an identification protocol that takes these constraints into account is intrinsically robust against various imperfections.

To identify the sparse frequencies from the experimental data, in the first step of the algorithm, we use a super-resolution and denoising algorithm (ESPRIT) [RPK86; Fan16]. Achieving the high precision in this step is crucial for identifying the eigenvectors in the presence of noise. To identify the eigenbasis, in the second step, we perform least-square optimization of the time-series data under the orthonormality constraint with a gradient descent algorithm on the manifold structure of the orthogonal group [AEK09]. Here, we incorporate the connectivity constraint on the coefficient matrix h by making use of regularization techniques [BG11].

Robustness against ramp errors. The initial and final ramping pulses result in a time-independent, approximately unitary transformation at the beginning and end of the time series. It is important to stress that such ramping pulses are expected to be generic in a wide range of experimental implementations of dynamical analogue quantum simulations. We can model the effect of such state preparation and measurement (SPAM) errors via linear maps S and M, respectively. This alters our model of the ideal data (7.3) to

$$\langle a_m(t)\rangle_{\psi_n} = \frac{1}{2}(M \cdot \exp(-\mathrm{i}th) \cdot S)_{m,n}. \tag{7.4}$$

While for the frequency identification such time-independent errors ‘only’ deteriorate the signal-to-noise ratio, for the identification of the eigenvectors of h it is crucial to take the effects of non-trivial S and M into account. Given the details of the ramping procedure, we expect that the deviation of the initial map S from the identity will be significantly larger than that of the final map M and we provide evidence for this in the subsequent method section. In particular, the final map will be dominated by phase accumulation on the diagonal.

We make the identification robust to arbitrary initial maps S and diagonal final maps M: By pre-processing the data, we can robustly remove the initial map S. In the post-processing, we can remove diagonal final maps M in the eigenspace identification. Given the identified Hamiltonian $\hat{h}$, we can obtain estimates $\hat{S}$ of S and $\hat{D}_M$ of the orthogonal and diagonal part of M (see Methods).

Error sources. There are two main remaining sources of error that affect the Hamiltonian identification. First, the estimate $\hat{h}$ has a statistical error due to the finite number of measurements in estimating the expectation values. We estimate the statistical error of the identified Hamiltonian $\hat{h}$ via parametric bootstrapping. Second, any final map M with a non-diagonal orthogonal part will alter the eigenbasis of the Hamiltonian that describes the dynamical data after removal of the initial map S giving rise to a systematic error. We estimate the magnitude of the resulting error on the identified Hamiltonian $\hat{h}$ using a simple model for the ramping phase.

Notice that while the statistical error determines the predictive power of the identified Hamiltonian $\hat{h}$, the systematic error does not necessarily: A future experiment with the same ramping pulses will be best described by the identification results $\hat{h}$, $\hat{S}$ and $\hat{D}_M$ up to the statistical error. The systematic error then captures the deviation of the identified Hamiltonian $\hat{h}$ from the Hamiltonian implemented during the time-evolution phase of the experiment within our model.

Results. We implement different Hamiltonians with a fixed overall hopping strength $J_{i,j} = 20\,\mathrm{MHz}$ and site-dependent local potentials μ_i on subsets of qubits and take time-series data as described above. Specifically, we choose the local potentials quasi-randomly $\mu_q = 20\cos(2\pi qb)\,\mathrm{MHz}$, for $q = 1, \ldots, N$, where b is a number between zero and one. In one dimension, this choice corresponds to implementing the Harper Hamiltonian, which exhibits characteristic 'Hofstadter butterfly' frequency spectra as a function of the dimensionless magnetic flux b [Hof76].

In Figure 7.2, we illustrate the properties of a single Hamiltonian identification instance in terms of both how well the simulated time evolution fits the experimental data (a, d, e) and how it compares to the targeted Hamiltonian (b) and SPAM (c). We measure all deviations in terms of the ℓ_2 norm, defined for a matrix A as $\|A\|_{\ell_2} = (\sum_{i,j} |A_{i,j}|^2)^{1/2}$ (Frobenius norm). We also define a metric of

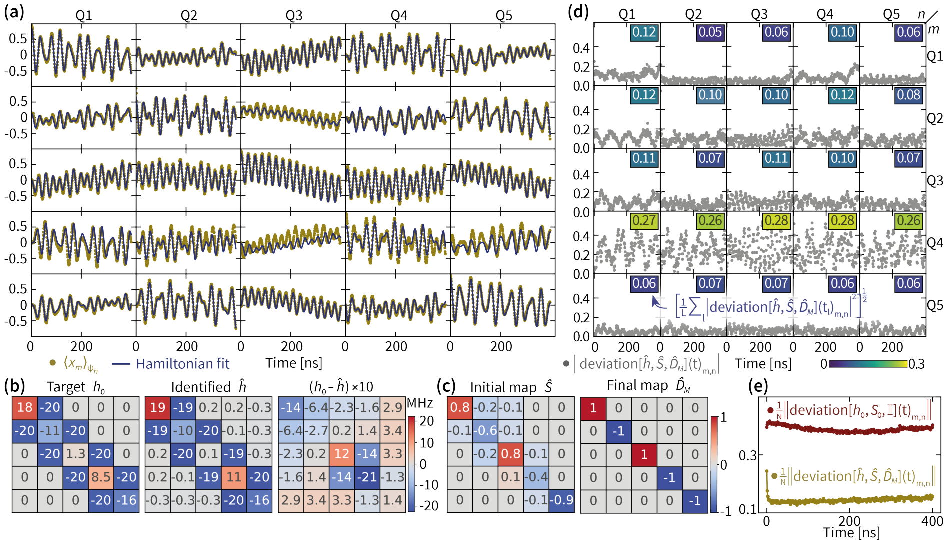


Figure 7.2: **A single Hamiltonian recovery of a 5-mode Hamiltonian and the corresponding time domain data. (a)** The full experimental time-series data $\langle x_m(t)\rangle_{\psi_n}$ for $m, n = 1, \dots, 5$ and the best fit of those data in terms of our model $\frac{1}{2}(M \exp(-ith)S)_{m,n}$ for a diagonal and orthogonal M and linear map S (solid lines). **(b)** The target Hamiltonian matrix h_0, the identified Hamiltonian $\hat{h}$, and the deviation between them. The error of each diagonal entry is $\pm(0.16+1.22)$ MHz and of each off-diagonal entry $\pm(0.12 + 0.60)$ MHz and comprises the statistical and the systematic error, respectively. The analogue implementation accuracy $\mathcal{E}_{\text{analogue}}(\hat{h}, h_0)$ is $0.73 \pm (0.07+0.37)$ MHz, and 0.32 ± 0.00 MHz for the eigenfrequencies. The analogue implementation accuracy $\mathcal{E}_{\text{analogue}}(\hat{S}, \text{id})$ of the identified initial map is $0.61 \pm (0.00+0.61)$. **(c)** The real part of the initial map $\hat{S}$ and the diagonal orthogonal part $\hat{D}_M$ of the final map M, inferred from the data using the identified Hamiltonian $\hat{h}$. **(d)** Absolute value of the time-domain deviation of the fit from the full experimental data for each time series, given by $\mathsf{deviation}[\hat{h}, \hat{S}, \hat{D}_M]_{m,n} := \langle a_m(t)\rangle_{\psi_n} - \frac{1}{2}\hat{D}_M \exp(-it\hat{h})\hat{S}$. The insets represent the root-mean-square deviation of the Hamiltonian fit from the experimental data per time series, averaged over the evolution time for each matrix entry (m, n), resulting in an entry-wise summarized quality of fit. We find a total root-mean-square deviation of the fit of 0.14. **(e)** Instantaneous root-mean-square deviation of the identified Hamiltonian $\hat{h}$, initial map $\hat{S}$ and final map $\hat{D}_M$ and of the target Hamiltonian h_0 with initial map fit S_0 from the experimental data averaged over the distinct time series.

analogue implementation accuracy of the identified Hamiltonian $\hat{h}$ with respect to the targeted Hamiltonian h_0 as

$$\mathcal{E}_{\text{analogue}}(\hat{h}, h_0) := \frac{1}{N} \left\| \hat{h} - h_0 \right\|_{\ell_2} . \tag{7.5}$$

Likewise, we can quantify the implementation accuracy of the identified initial map $\hat{S}$ as $\mathcal{E}_{\text{analogue}}(\hat{S}, \mathrm{id})$, and of the identified eigenfrequencies $\mathrm{eig}(\hat{h})$ as $\mathcal{E}_{\text{analogue}}(\mathrm{eig}(\hat{h}), \mathrm{eig}(h_0))$.

Notice that the implementation accuracy of the frequencies in the data from the targeted Hamiltonian eigenfrequencies give a lower bound to the overall implementation accuracy of the identified Hamiltonian. This is because the ℓ_2 norm used in the definition (7.5) of $\mathcal{E}_{\text{analogue}}$ is unitarily invariant and any deviation in the eigenbasis, which we identify in the second step of our algorithm, will tend to add up with the frequency deviation.

We find that most entries of the identified Hamiltonian deviate from the target Hamiltonian by less than 0.5 MHz with a few entries deviating by around 1–2 MHz. The overall implementation accuracy is less than 1 MHz. The error of the identification method is dominated by the systematic error due to the final ramping phase that is around 1 MHz for the individual entries. Small long-range couplings exceeding the statistical error are necessary to fit the data well even when penalizing those entries via regularization. These entries are rooted in the effective rotation by the final ramping before the measurement and within the estimated systematic error.

The fit deviation from the data (Figure 7.2(e)) exhibits a prominent decrease within the first few nanoseconds of the time evolution. This indicates that the time evolution significantly differs during the initial phase of the experiment as compared to the main phase of the experiment, which we can attribute to the initial pulse ramping of the experiment. The identified initial map describing this ramping (Figure 7.2(c)) is approximately band-diagonal and deviates from being unitary, indicating fluctuations of the effective ramps between different experiments.

We find a significantly larger time-averaged real-time error (Figure 7.2(d)) in all data series $\langle a_m \rangle_{\psi_n}$ in which Q_4 was measured, indicating a measurement error on Q_4. We also observe significant deviation between the parameters of the target and identified Hamiltonian in qubits Q_3 and Q_4 and the coupler between them. Since the deviation of the eigenfrequencies is much smaller than of the full Hamiltonian, we attribute those errors also to a non-diagonal part of the final ramping phase at those qubits that leads to a rotated eigenbasis.

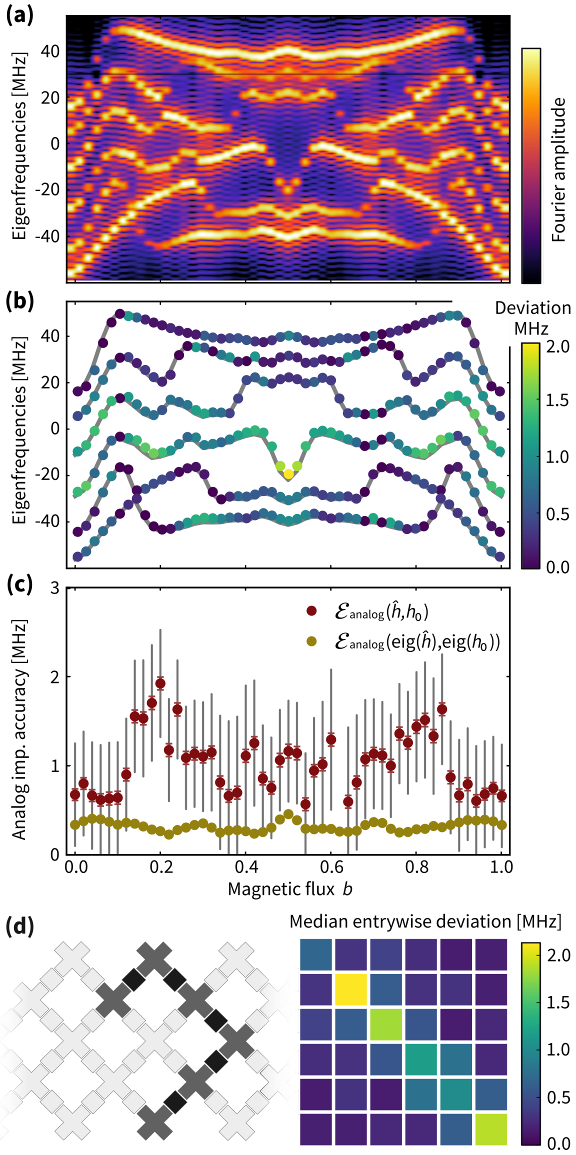


Figure 7.3: **Comparing frequency and full identification errors.** **(a)** In an $N = 6$ subset of connected qubits, by varying b from 0 to 1, we implement 51 different Hamiltonians. The plot shows the Fourier transform of the time domain data. **(b)** The extracted eigenfrequencies (denoised peaks in panel (a)) are shown as colored dots, where the assigned color is indicative of the deviation between targeted eigenfrequencies (gray lines) and the identified ones from position of the peaks. **(c)** Analog implementation accuracy $\mathcal{E}_{\text{analogue}}(\hat{h}, h_0)$ of the identified Hamiltonian (dark red) compared to the implementation accuracy $\mathcal{E}_{\text{analogue}}(\text{eig}(\hat{h}), \text{eig}(h_0))$ of the identified frequencies (golden). Colored (gray) error bars quantify the statistical (systematic) error. **(d)** Layout of the six qubits on the Sycamore processor and median of the entry-wise absolute-value deviation of the Hamiltonian matrix entries from their targeted values across the ensemble of 51 different values of $b \in [0, 1]$.

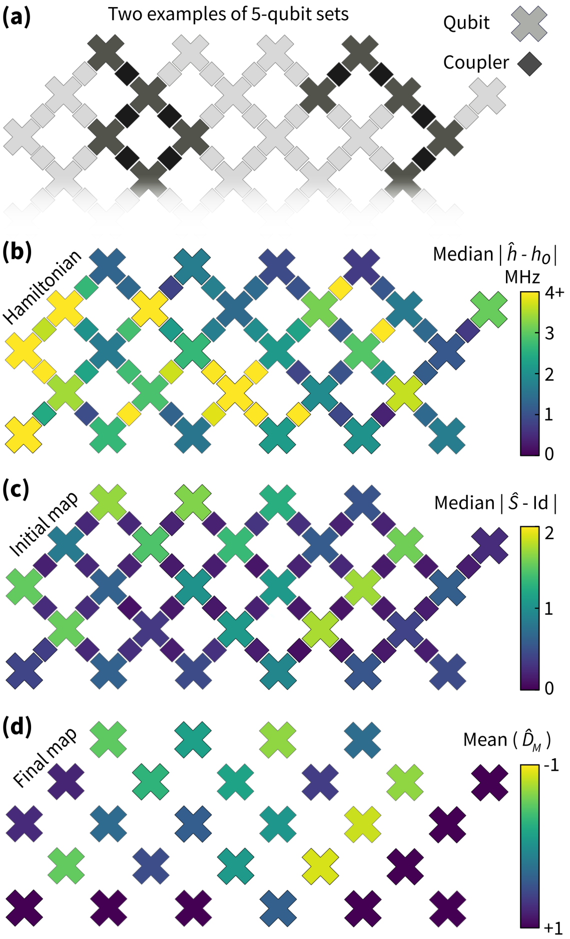


Figure 7.4: **Error map of Hamiltonian implementation across the Sycamore processor.** Over the grid of 27 qubits, we randomly choose subsets of connected qubits and couplers of size $N = 5$. On each subset we implement two Hamiltonians with $b = 0, 0.5$ and run the identification algorithm. Two instances are shown in panel **(a)**. For each subset, we compute the deviation of the identified Hamiltonian and initial map from their respective target and assign it to each qubit or coupler involved. Due to overlap of subsets, each qubit or coupler has been involved in at least 5 different choices of subsets. Panels **(b)** and **(c)** show the median deviation for the Hamiltonian and initial map implementations, respectively. Panel **(d)** shows the mean of the sign flips in the identified (diagonal ± 1) final map for each qubit.

In Figure 7.3, we summarize multiple identification data of this type to benchmark the overall performance of a fixed set of qubits. In panel (a) we show the measured Fourier domain data for 51 different values of the magnetic flux $b \in [0, 1]$. In panel (b) we plot the deviation of the frequencies identified from the data. Most implemented frequencies deviate by less than 1 MHz from their targets. While the Hamiltonian implementation accuracies (Figure 7.3(c)) are up to a factor of four larger than the corresponding frequency accuracies, this decrease is mostly within our estimated systematic error induced by the final ramping. Importantly, the frequency identification is robust against systematic measurement errors. This indicates that the final ramping is the limiting factor for an even more accurate implementation and identification of the Hamiltonians.

In Figure 7.3(d) we show the median of the entry-wise deviation of the identified Hamiltonian from its target over all magnetic flux values. Thereby, the ensemble of Hamiltonians defines an overall error benchmark. This benchmark can be associated to the individual constituents of the quantum processor, namely, the qubits, corresponding to diagonal entries of the Hamiltonian deviation, and the couplers, corresponding to the first off-diagonal matrix entries of the deviation.

We use this benchmark over an ensemble of two flux values to assess a 27-qubit array of superconducting qubits. To do so, we repeat the analysis reported in Figure 7.3 for 5-qubit dynamics on different subsets of qubits and extract average errors of the individual qubits and couplers involved in the dynamics, both in terms of the identified Hamiltonian and the initial and final maps. Summarized in Figure 7.4, we find significant variation in the implementation accuracy of different couplers and qubits. While for some qubits the effects of the initial and final maps are negligible, for others they indicate the potential of a significant implementation error. From a practical point of view, such diagnostic data allows to maximally exploit the chip's accuracy for small-scale analogue simulation experiments. Let us note that within the accuracy of our method the overall benchmark for the qubits and couplers for 5-qubit dynamics agrees with that of 3- and 4-qubit dynamics.

7.2 Methods

After summarizing the experimental finding, we here give a more detailed account of the method used, in particular, its robustness and the estimation of systematic errors.

7.2.1 Experimental details

We use the Sycamore quantum processor composed of quantum systems arranged in a two-dimensional array. This processor consists of gmon qubits (transmons with tunable coupling) with frequencies ranging from 5 to 7 GHz. These frequencies are chosen to mitigate a variety of error mechanisms such as two-level defects. Our coupler design allows us to quickly tune the qubit–qubit coupling from 0 to 40+ MHz. The chip is connected to a superconducting circuit board and cooled down to below 20 mK in a dilution refrigerator. Each qubit has a microwave control line used to drive an excitation and a flux control line to tune the frequency. The processor is connected through filters to room-temperature electronics that synthesize the control signals. We execute single-qubit gates by driving 25 ns microwave pulses resonant with the qubit transition frequency.

The qubits are connected to a resonator that is used to read out the state of the qubit. The state of all qubits can be read simultaneously by using a frequency-multiplexing. Initial device calibration is performed using ‘Optimus’ [Kel+18] where calibration experiments are represented as nodes in a graph.

7.2.2 Details of the identification algorithm

Succinctly written, our data model is given by

$$y_{m,n}[l] = \langle a_m(t_l) \rangle_{\psi_n} = \frac{1}{2}(M \cdot \exp(-\mathrm{i} t_l h) \cdot S)_{m,n}, \tag{7.6}$$

where $m, n = 1, \ldots, N$ label the distinct time series, $l = 1, \ldots, L$ labels the time stamps of the L data points per time series. The matrices S and M are arbitrary linear maps that capture the state preparation and measurement stage as effected by the ramping of the eigenfrequencies of the qubits and couplers to their interacting value and back (see Figure 7.1). In the experiment, we empirically estimate each such expectation value with 1000 single shots.

Our mindset for solving the identification problem is based on the eigendecomposition $h = \sum_{k=1}^{N} \lambda_k \, |v_k\rangle\langle v_k|$ of the coefficient matrix h in terms of eigenvectors $|v_k\rangle$ and eigenvalues λ_k. We can write the data (7.6) in matrix form as

$$y[l] = \frac{1}{2} \exp(-\mathrm{i} t_l h) \cdot = \frac{1}{2} \sum_{k=1}^{N} \mathrm{e}^{-\mathrm{i} t_l \lambda_k} \, |v_k\rangle\langle v_k| \,, \tag{7.7}$$

where we have dropped S and M for the time being. This decomposition suggests a simple procedure to identify the Hamiltonian using Fourier data analysis. From the matrix-valued time series data $y[l]$ (7.7), we identify the Hamiltonian coefficient matrix h in two steps. First, we determine the eigenfrequencies of h. Second, we recover the eigenbasis of h. To achieve those identification tasks with the largest possible robustness and accuracy, it is key to exploit all available structure at hand.

Step 1: Frequency extraction. In order to estimate the spectrum to the best available accuracy levels, we exploit that the signal is *sparse* in Fourier space. Exploiting this structure allows to substantially denoise the signal and achieve a *super-resolution* beyond the Nyquist limit [CF13; CF14]. We achieve this with an adaptation of the ESPRIT algorithm proposed in Ref. [RPK86] and analysed in its functioning in Refs. [Fan16; LLF20].

To achieve optimal robustness of the recovery, we superimpose different time series $y_{m,n}$. Specifically, we compute the trace of the data matrix (for $S = M = \mathrm{id}$) as

$$F[l] := \mathrm{Tr}[y[l]] = \sum_{m=1}^{N} y_{m,m}[l] = \frac{1}{2} \sum_{k=1}^{N} \mathrm{e}^{-\mathrm{i} t_l \lambda_k} \,. \tag{7.8}$$

We find $F[l]$ to be an *equally weighted* linear combination of sinusoids. Thus, using $F[l]$ instead of the separate time series ensures an optimal signal-to-noise ratio for the simultaneous extraction of all frequencies. In a future companion work [Rot+22] we describe in detail how ESPRIT [Fan16] recovers the desired spectral values robustly with super-resolution from such data. Suffice it here to say that the key idea is to set up a Hankel matrix and identify the dominant n-dimensional subspace of its range.

Step 2: Eigenspace identification. To identify the eigenspaces of the Hamiltonian, we use the eigenfrequencies found in Step 1 to fix the oscillating part of the dynamics in Eq. (7.7). What remains is the problem of finding the eigenspaces $|v_k\rangle\langle v_k|$ from the data. The problem is a non-convex inverse quadratic problem, subject to the constraint that the resulting Hamiltonian matrix respects the connectivity constraint of the superconducting architecture. Formally, we denote the a priori known support set of the Hamiltonian matrix as Ω, so that this constraint amounts to $h_{\overline{\Omega}} = 0$. We can cast this problem into the form of a least-square optimization problem

$$\begin{aligned} \underset{\{|v_k\rangle\}}{\text{minimise}} \quad & \sum_{l=1}^{L} \left\| y[l] - \sum_k \mathrm{e}^{-\mathrm{i}\lambda_k t_l} \, |v_k\rangle\langle v_k| \right\|_{\ell_2}^2 , \\ \text{subject to} \quad & \langle v_m | v_n \rangle = \delta_{m,n}, \ \left(\sum_k \lambda_k \, |v_k\rangle\langle v_k| \right)_{\bar{\Omega}} = 0, \end{aligned} \tag{7.9}$$

equipped with non-convex constraints enforcing the orthogonality and the quadratic constraint restricting the support. In order to approximately enforce the support constraint, we make use of regularization [BG11]. It turns out that this can be best achieved by adding a term [10, App. A]

$$\mu \left\| \left(\sum_k \lambda_k \, |v_k\rangle\langle v_k| \right)_{\bar{\Omega}} \right\|_{\ell_2} \tag{7.10}$$

to the objective function (7.9), where $\mu > 0$ is a parameter weighting the violation of the support constraint. We then solve the resulting minimization problem by using a conjugate gradient descent on the manifold of the orthogonal group [EAS98; AEK09], see also the recent work [Luc+21; LKF21; 2] for the use of geometric optimization for quantum characterization.

Without the support constraint this gives rise to an optimization algorithm that converges well. A detailed study will be published in Ref. [Rot+22]. However, the regularization term makes the optimization landscape rugged as it introduces an entry-wise constraint that is skew to the orthogonal manifold. This is why we consecutively ramp up μ until the algorithm does not converge any more in order to find the Hamiltonian that best approximates the support constraint while at the same time being a proper solution of the optimization problem. For example, for the data in Figure 7.2 the value of μ is 121. In order to avoid that we identify a Hamiltonian from a local minimum of the rugged landscape, we only accept Hamiltonians which achieve a total fit of the experimental data within a 5% margin of the fit quality of the unregularized recovery problem, and use the Hamiltonian recovered without the regularization otherwise.

7.2.3 Robustness to state preparation and measurement errors

The experimental design requires a ramping phase of the qubit and coupler frequencies from their idle location to the desired target Hamiltonian and back for the measurement. In effect, the data model (7.6) includes time-independent linear maps M and S that are applied at the beginning and end of the Hamiltonian time-evolution. The maps affect both frequency extraction and the eigenspace identification. For the frequency extraction, the Fourier coefficients of the trace signal $F[l]$ become $\langle v_k| SM |v_k\rangle$. While the frequencies remain unchanged the Fourier coefficients now deviate significantly from unity, significantly impairing the noise-robustness of the frequency identification. More severely, the eigenspace reconstruction yields a rotated eigenbasis if M and S cause an orthogonal rotation. We can however remove either S or M by appropriately pre-processing the data.

Ramp removal via pre-processing. To remove the initial map S from the data we apply the pseudoinverse $(\cdot)^+$ of the data $y[l_0]$ at a fixed time t_{l_0} to the entire (time-dependent) data series in matrix form. For arbitrary S and $M = \mathrm{id}$ this gives rise to

$$y^{(l_0)}[l] = y[l](y[l_0])^+ = \sum_{k=1}^{N} \mathrm{e}^{-\mathrm{i}\lambda_k(t_l - t_{l_0})} |v_k\rangle\langle v_k| \,. \tag{7.11}$$

The caveat of this approach is that the noise that altered $y[l_0]$ is now present in every entry of the new data series $y^{(l_0)}$. We can remedy this effect by using several time points $y[l_0]$ for the inversion. We compute the concatenation of data series for different choices of l_0, e.g., for every s data points $0, s, 2s, \ldots, \lfloor L/s \rfloor s$ giving rise to new data $y_{\text{total, s}} = (y^{(0)}, y^{(s)}, y^{(2s)}, \ldots, y^{(\lfloor L/s \rfloor s)}) \in \mathbb{C}^{\lfloor L/s \rfloor L}$. If the data suffers from drift errors, it is also beneficial to restrict each data series $y^{(l_0)}$ to entries $y^{(l_0)}[\kappa]$ with $\kappa \in [l_0 - w, l_0 + w]$, i.e., the entries in a window of size w around l_0. In practice, we use $s = 1$ and $w = 50$. We note that due to the symmetry of the problem, replacing the right multiplication with $(y[l_0])^+$ in the pre-processing by a left multiplication yields the analogue algorithm that removes errors in the measurement M instead of the state preparation error S.

Given an estimate for the Hamiltonian $\hat{h}$, we can estimate the initial map S via

$$\hat{S} = \frac{2}{L} \sum_{l=1}^{L} \exp[\mathrm{i} t_l \hat{h}] y[l] \,. \tag{7.12}$$

Systematic measurement error. The pre-processing step allows us to either remove the initial map or the final map from the data. Removing the initial map still leaves us with the final map M as a source of systematic error. The final pulse ramping is in fact one of the major sources of systematic error in the experimental application of the identification method. When allowing for an arbitrary initial map, it is intrinsically impossible to identify the final map at the same time. This is because any perturbation of the measurement basis induced by a final pulse ramping effectively changes the Hamiltonian eigenbasis. However, due to the constraints of the problem, the eigenspace identification algorithm is only susceptible to a real orthogonal rotation at the end of the time series.

We now make this explicit. Including the final ramp M from our data model (7.6), the description Eq. (7.11) of the data, obtained after removing the initial ramp S in the pre-processing, becomes

$$y^{(t_0)}[l] = M \exp[-\mathrm{i}(t_l - t_0)h]M^{-1}\,. \tag{7.13}$$

Taking the trace of $y^{(t_0)}[l]$ yields $F[l] = \frac{1}{2}\sum_{k=1}^{N} \mathrm{e}^{-t_l\lambda_k}$ so that the data for the frequency extraction step is unaltered by M in expectation. Consequently, we only observe an effect of M in the eigenspace identification.

To understand the effect of M on the identification, we observe that the eigenbasis of h is constrained to be real and orthogonal. Correspondingly, only the real, orthogonal part of M will rotate the identified eigenspaces. We write $M = O_M(\mathrm{Id} + X)$, where $O_M \in \arg\min_{O\in O(N)} \|M - O\|_{\ell_2}$. Let us assume that X is small in operator norm so that we can drop higher order terms in the following. Then the identification algorithm—assuming convergence to the global minimum—returns the Hamiltonian estimate

$$\hat{h} = O_M h O_M^T. \tag{7.14}$$

The result of the identification algorithm will therefore be rotated by the orthogonal part of M compared to the 'actual Hamiltonian' h.

To see the effect on the estimate $\hat{S}$ of the initial map S, let us first observe that using $\hat{h}$ to predict time series data results in

$$y_{\hat{h}}(t) = \frac{1}{2} O_M \exp\{-\mathrm{i}th\} O_M^T \hat{S}\,. \tag{7.15}$$

This in turn yields an estimate $\hat{S}$ for S via linear inversion for every t of $\hat{S}_t = O_M \mathrm{e}^{\mathrm{i}th} O_M^T M \mathrm{e}^{-\mathrm{i}th} S = O_M S + O_M \mathrm{e}^{\mathrm{i}th} X \mathrm{e}^{-\mathrm{i}th} S$. Taking the average over $t \in$

$\{t_l\}_{l=1}^L$ gives

$$\hat{S} = O_M(\mathrm{Id} + \overline{X})S \tag{7.16}$$

with $\overline{X} = \frac{1}{L}\sum_{l=1}^L \mathrm{e}^{\mathrm{i}t_l h} X \mathrm{e}^{-\mathrm{i}t_l h}$. Due to unitary invariance, the ℓ_2 norm of $\overline{X}$ is controlled by $\left\|\overline{X}\right\|_{\ell_2} \leq \|X\|_{\ell_2} = \|M - O_M\|_{\ell_2} \leq \|M - \mathrm{Id}\|_{\ell_2}$. We find that the estimate $\hat{S}$ is affected both by the orthogonal part O_M and the average of the non-orthogonal part over the time evolution.

Finally, let us consider the effect on the quality of fit of the data with $\hat{h}$. The deviation of the data predicted with $\hat{h}$ from the expectation value of the actual data is thus

$$y(t) - y_{\hat{h}}(t) = \frac{1}{2} O_m \left[\mathrm{e}^{-\mathrm{i}th}\overline{X} - X\mathrm{e}^{-\mathrm{i}th}\right] S\,. \tag{7.17}$$

We have $\left\|y(t) - y_{\hat{h}}(t)\right\|_{\ell_2} \leq \|X\|_{\ell_2}\|S\|_{\ell_2}$. In particular, the deviation of the data from the Hamiltonian fit is only produced by the non-orthogonal part of M, i.e. $O_M X$, and vanishes for orthogonal M where $X = 0$. The identification method itself, however, is robust against these non-orthogonal deviations.

We conclude that neither the identification algorithm nor the deviation in the data can distinguish between a real, orthogonal rotation before the measurement and the time evolution under the correspondingly rotated Hamiltonian. Indeed, any future data with the same final ramping will be correctly predicted by the identified Hamiltonian $\hat{h}$ and initial map $\hat{S}$.

Removing diagonal unitary M in post-processing. However, using a (mild) assumption on the accuracy of the actual implemented Hamiltonian h with respect to its target h_0, we can identify and correct for the orthogonal part of diagonal ramps in $\hat{h}$ and $\hat{S}$. Indeed, we expect M to be approximately diagonal such that its main contribution is to add a complex phase to the signal for each measured qubit. This is because effectively ramping the couplers out of the relevant frequency spectrum can be achieved much more rapidly than ramping the qubits back to their idling frequencies.

Importantly, for $M = \mathrm{diag}(\mathrm{e}^{\mathrm{i}\delta_1}, \ldots, \mathrm{e}^{\mathrm{i}\delta_N})$ a diagonal unitary, its projection onto real, orthogonal matrices is given by $O_M = \mathrm{diag}(s_1, \ldots, s_n)$ with $s_i = +1$ for $|\delta_i| \leq \pi/2$ and $s_i = -1$ for $|\delta_i| > \pi/2$. If we exclude the possibility to miss a target Hamiltonian in the implementation by a phase of more than $\pi/2$ on a qubit, we can attribute an observed sign-flip in the identified Hamiltonian compared to the target Hamiltonian to the final map. Thus, by optimizing our recovery error over all possible *diagonal* real, orthogonal matrices D_M (i.e., diagonal ± 1-matrices) we can correct for a systematic error caused by an arbitrary diagonal unitary M. We can identify the optimal such matrix $\hat{D}_M$ and thereby also determine the diagonal phases of M that are larger than $\pi/2$ in absolute value. Given the identification results $\hat{h}'$ and $\hat{S}'$ before post-processing, we then obtain the final identified Hamiltonian $\hat{h} = \hat{D}_M \hat{h}' \hat{D}_M$ and initial map $\hat{S} = \hat{D}_M \hat{S}'$.

To summarize: Via pre- and post-processing we arrive at an identification method that is robust against arbitrary errors in the state-preparation (initial ramping) and diagonal errors in the measurement (final ramping).

Imbalance between initial and final ramping phase. As explained above, the pre-processing step allow us to remove either the initial map S or final map M from the data. Subsequently, we can treat the diagonal phase accumulation of the remaining map in the post-processing but its non-diagonal, real orthogonal part remains a systematic error. A priori it is unclear which one of the two maps is more beneficial to remove in order to reduce the systematic error.

We have already treated the initial and final ramping phases on a different footing, however. The reason for this is rooted in the specifics of the ramping of the couplers compared to the qubits. The couplers need to be ramped from their idle frequencies to provide the desired target frequencies of 20 MHz. This is why we expect the timescale of the initial ramping to be mainly determined by the couplers, namely the delay until they arrive around the target frequency and the time it takes to stabilize at the target frequency. In contrast, the final ramping map becomes effectively diagonal as soon as the couplers are again out of the MHz regime. We therefore expect that the initial map has a sizeable non-diagonal orthogonal component, whereas the final map is approximately diagonal.

Therefore, it is advantageous to remove the initial map in the pre-processing, and correct for the diagonal phases of the final map in the post-processing. To confirm this and to build trust in the theoretical considerations above, in Figure 7.5 we compare the properties of the identification result when we remove the initial map in the pre-processing with the corresponding result when we remove the final map instead. We observe that the deviation of the orthogonal part $\hat{O}_S$ from its projection $\hat{D}_S$ to diagonal orthogonal matrices is much larger for the identified initial map $\hat{S}$ than the corresponding deviation for the final map (Figure 7.5(a)). Moreover, both the root-mean-square fit of the data (Figure 7.5(c)) and the analogue implementation accuracy of the identified Hamiltonian with its target (Figure 7.5(b)) are significantly improved when removing the initial ramp as compared to removing the final ramp. This indicates that S induces

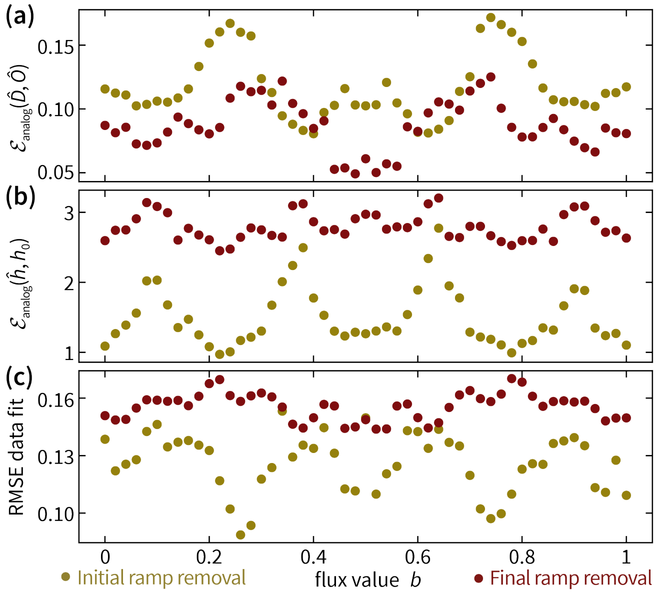


Figure 7.5: **Initial ramp removal versus final ramp removal.** We identify Hamiltonians of a set of 5-qubit Hamiltonians with Hofstadter butterfly potentials $\mu_q = 20\cos(2\pi qb)$ MHz for qubits $q = 1, \ldots, 5$ and flux value b in without regularization. **(a)** Deviation of the orthogonal part $\hat{O}_S$ ($\hat{O}_M$) of the identified initial map $\hat{S}$ (final map $\hat{M}$) from the closest diagonal orthogonal matrix $\hat{D}_S$ ($\hat{D}_M$). **(b)** Analog implementation accuracy of the corresponding identified Hamiltonians $\hat{h}_S$ ($\hat{h}_M$). **(c)** Total root-mean-square deviation of the time series data from the Hamiltonian fit.

a larger systematic error than M. Correspondingly, it is indeed more advantageous to remove the initial map in the pre-processing and fit the final map with a diagonal orthogonal matrix validating the approach taken here.

7.2.4 Error estimation

Finally, we estimate the error on the Hamiltonian $\hat{h}$ and initial map $\hat{S}$ identified via the robust identification method including pre- and post-processing. This error comprises two contributions. First, it has a systematic contribution which is due to the non-diagonal and orthogonal part O_M of the final map M. Second, it has a statistical contribution due to the estimation of the expectation values (7.6) from finite statistics. The two contributions to the error play different roles for the predictive power of the identified Hamiltonian. The identified Hamiltonian can be regarded as the 'actual' Hamiltonian during the time evolution as described in the frame rotated by the orthogonal part of the systematic measurement error. For experiments that include the same ramping phases, the identified Hamiltonian $\hat{h}$ together with the initial map $\hat{S}$ and the diagonal orthogonal final map $\hat{D}_M$ is thus expected to predict the time-evolution up the statistical error. Thus, in such settings, the systematic error of the identified Hamiltonian does not affect its predictive power.

Final ramp effect estimation. To estimate the magnitude of the systematic error that is induced by an orthogonal, non-diagonal part O_M of the final pulse ramping M, we use an idealized model of the final ramping phase with a fixed speed v and a maximal ramping time τ. Let $M = \mathcal{T} \exp\{-i \int_0^\tau H(t) dt\}$, where $\mathcal{T}$ denotes the time-ordering operator. We set $H(t) = T_{h_m}(h_0 + \operatorname{sign}(h_m - h_0)vt)$ with parameter matrix h_m at the end off the ramp pulse and the thresholding operator acting entry-wise as $T_{h_m}(x) = \max\{\min\{x, -|h_m|\}, |h_m|\}$. The thresholding ensures that the entries of $H(t)$ stay equal to those of h_m once they reach their final value. For τ we use the minimal time at which all entries of $H(t)$ are equal to h_m plus a small additional wait time. We assume that the matrix after the ramp pules h_m is a diagonal matrix with frequencies corresponding to the idling frequencies of the qubits.

This leaves us with the ramping speed as an unknown parameter in our model. We estimate the ramping speed to be lower bounded by 150 MHz/ns and a nonzero wait time of 0.1ns and experimentally build trust in this assumption below. We numerically evaluate the time-dependent integral to arrive at an estimate $\bar{M}$ for M. The projection of $\bar{M}$ onto the real, orthogonal matrices yields an estimate $\bar{O}_M$ for the induced rotation in the Hamiltonian reconstruction (7.14). Given $\bar{O}_M$ we rotate the identified Hamiltonian $\hat{h}'$ before the post-processing stage, obtaining $\bar{h} = \bar{O}_M^T \hat{h}' \bar{O}_M$. We then calculate the entry-wise deviation of $\bar{h}$ after post-processing from the identified Hamiltonian $\hat{h}$. The maximal deviation over all diagonal entries and over all off-diagonal entries is used as an estimate for the systematic error for the respective entries of $\hat{h}$ caused by the final pulse ramp. Analogously, we obtain a systematic error for the overall analogue implementation accuracy. The resulting systematic error strongly depends on the parameters of h_m and h_0, and is compatible with the deviation of the analogue implementation accuracy of the identified Hamiltonian from that of the identified frequencies.

We find that the phases of the qubits in the simulated $\bar{O}_M$ are rarely larger than

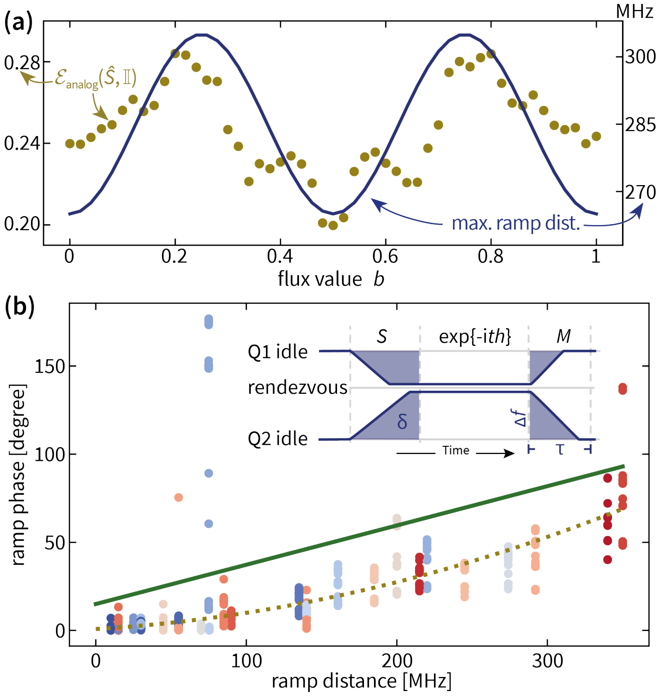


Figure 7.6: **Validating the ramp model.** **(a)** Distance of the identified initial map before post-processing $\hat{S}'$ from the identity for the 5-qubit butterfly data of Figure 7.5 (golden dots) and maximum ramp distance $\max_i |(h_0 - h_m)_{i,i}|$ (solid line) for each flux value $b \in [0, 1]$. **(b)** Phases of the diagonal entries of $\hat{S}$ for various instances of diagonal Hamiltonians as a function of the corresponding ramping distances from their idle frequency to the common rendezvous frequency, which is given by $|(h_m)_{i,i}|$. Different measurements of the same qubits are depicted in the same colour. The solid green line is a linear upper bound to the ramp phases, yielding an upper bound to the total ramp time of 0.6 ns. The dotted, golden line is a quadratic fit of the phase data, excluding outliers above $140°$. *Inset illustration of the ramp model.* The qubits initially at frequencies Q1 and Q2 are ramped to the common rendezvous frequency of 6500 MHz giving rise to an initial map S, where they evolve under the Hamiltonian h for time t until they are ramped back to their idle frequencies, giving rise to a final map M. The shaded areas show the total acquired phase δ during the ramp phases.

$\pi/2$. In comparison, we observe frequent phase flips in the data. With the assumed minimal ramping speed we are therefore unable to explain all sign flips that are observed for certain qubits. As a consequence, the ramping speed of these qubits must either be particularly slower, or further effects that result in diagonal phase accumulation at the end of the time evolution play a role in the experiment. But as explained above a diagonal unitary rotation as caused by a ramping of the qubits after the couplers are already effectively 0, can only cause sign flips of entire rows and columns and can, thus, be corrected for in the post-processing.

Empirical validation of ramp model and parameter estimation. Our model for estimating the systematic error induced by the final ramping phase implies that the deviation of the initial and final ramp from the identity transformation depends on the ramp distance, that is, the absolute value of the entries of $h - h_m$. Indeed, the maximal ramp distance is expected to set the timescale of the ramp phase and, thus, determines magnitude of the ramping effect in the data. In Figure 7.6(a) we validate that, indeed, the deviation of the identified initial map $\hat{S}$ is proportional to the ramp distance $\max_{i,j} |(h - h_m)_{i,j}|$.

Our model of the final ramp phase depends on a lower bound on the ramping speed of the qubits. We estimate the parameter by implementing the Hamiltonian with zero entries and reconstructing it with our identification method. In the rotating frame of the idle frequencies of the qubits, we effectively observe a diagonal Hamiltonian with eigenfrequencies that are the difference between the common rendezvous frequency (6500 MHz) and the idle frequencies. Since no couplers are involved, both the corresponding final and initial ramping maps are diagonal and contribute a complex phase to the data which is proportional to the combined surface area underneath the ramps, see the inset of Figure 7.6(b).

Since the Hamiltonian is itself diagonal, it commutes with the final diagonal unitary so that effectively the data can be described as $y_{\text{diagonal}}(t) = \exp(-\mathrm{i}th)MS$. We can thus determine the phase directly as the phases of the diagonal entries of our estimate $\hat{S}$ of the initial map. For the ramp model with a fixed ramp speed we expect the total (diagonal) phase $\delta = \Delta f^2$ to scale quadratically in the ramp distance Δf. For ramping up to the final value within a fixed time τ, we expect the total phase to scale linearly as $\delta = \Delta f \cdot \tau$. Since for every Hamiltonian instance, only one of the entries will reach its final value in the optimal time, and all others effectively see a fixed ramp time, we expect the dependence for our model to be a linear combination between a quadratic and a linear behaviour.

This expectation is confirmed in Figure 7.6, where we show the identified ramp phases as a function of the ramp distance from the idle frequencies of individual qubits to their common rendezvous frequency. From the linear upper bound to the acquired phases as a function of the travelled qubit distances, we can further estimate an upper bound of 0.6 ns to the total ramp time from the slope and a non-zero additional wait time due to the non-zero offset. Since the maximum ramp distance of all data sets is at least 90 MHz (excluding the outliers at 75 MHz) this builds trust in our estimate for a minimal ramping speed of $150\,\mathrm{MHz/ns}$. At this ramping speed, the offset of 15° corresponds to an additional wait time of $0.1\,\mathrm{ns}$. We also observe individual qubits that exhibit considerably larger phases for relatively short ramping distance indicating that their ramping is slower.

Statistical error: Bootstrapping. Let us now turn to estimating the statistical error of the identification result. This error is due to the finite measurement statistics which introduces an error to the empirical estimates of the expectation value governing the data (7.3). We estimate the size of the induced error to the Hamiltonian estimate that is returned by the identification method via parametric bootstrapping. To this end, we simulate time series data with finite statistical noise according to the model (7.6) with $M = \mathrm{id}$ using the identified Hamiltonian $\hat{h}$ and a Haar-random unitary for the initial state-preparation error S. We then run the Hamiltonian identification method without the regularization on 10^5 instances of such synthetic data. As the statistical error of the entry we use the 0.99-quantile (99% confidence level) of the absolute deviation of each entry in the Hamiltonians, obtained from the synthetic data, from the corresponding entry of the identified Hamiltonian, used to generate the data. We observe that the statistical errors of the entries are of comparable size and only report the maximal statistical error over all entries.

We also calculate 0.99-quantile of the deviation of the synthetically identified Hamiltonian $\hat{h}_{\mathrm{bt}}$ from the originally identified Hamiltonian $\hat{h}$ in terms of the analogue implementation accuracy $\mathcal{E}_{\text{analogue}}(\hat{h}_{\mathrm{bt}}, \hat{h})$, Eq. (7.5), and likewise for the eigenfrequencies. This is used as the statistical error estimate for the overall analogue implementation accuracy.

Omitting the regularization in the identification method reduces the computational complexity of the bootstrapping and produces more well-behaved empirical distributions of the deviation error. At the same time the regularization is expected to improve the estimate and, thus, the statistical error obtained in this way is expected to dominate the statistical error of the regularized identification method.

7.3 Summary and outlook

We have implemented the analogue simulation of the time-evolution of non-interacting bosonic Hamiltonians with tunable parameters for up to 6 qubit lattice sites. A custom-tailored identification method allows us for the first time to robustly recover the implemented Hamiltonian that governs the time-evolution. Thereby, we diagnose the deviation from the target Hamiltonian and assess the precision of the implementation. We achieve sub-MHz accuracy of the Hamiltonian parameters compared to their targeted values in most implementations. Combining the average performance measures over ensembles of Hamiltonians we associate benchmarks to the components of the superconducting qubit chip that quantify the performance of the hardware for the analogue task at hand and provide specific diagnostic information. Within our Hamiltonian identification framework, we are able to identify SPAM errors due to parameter ramp phases as a severe limitation of the architecture. Importantly, such ramp phases are present in *any* analogue quantum simulation of quenched dynamics. Our results show that minimizing those is crucial for precisely implementing a Hamiltonian.

The experimental and computational effort of the identification method scales efficiently in the number of modes of the Hamiltonian. Generalizing our two-step approach developed here, we expect a polynomial scaling with the dimension of the diagnosed particle sector and therefore remain efficient for diagnosing two-, three- and four-body interactions, thus allowing to build trust in the correct implementation of interacting Hamiltonian dynamics as a whole—in future work.

From a broader perspective, with this work, we hope to substantially contribute to the development of a machinery for precisely characterizing and thereby improving analogue quantum devices. As we have seen here, considerations of semi device dependence come with different twists for analogue simulators than for digital quantum computing devices providing new specific challenges.

* * *

7.4 *Viewpoint:* Easing the Monte-Carlo sign-problem

So far in this chapter, we devised and experimentally tested a practical, semi-device-dependent identification method for the Hamiltonian of an analogue simulation. At the heart of the method is the exploitation of structural constraint such as the orthogonality of the Hamiltonian eigenbasis. We focused on the dynamics of single particle excitations under a non-interacting Hamiltonians. This puts us in the comfortable position that the exact calculation of the ideal time evolution in the respective symmetry sector is efficient on a classical computer. The dynamics is arguably boring from a many-body physics point of view. For an analogue simulator, however, we regard it as an inevitable first step for building trust in its correct function.

Nonetheless, the main motivation of analogue simulators of course stems from their potential to efficiently perform simulations of quantum systems that are intractable on a classical computing device. But how exactly does the simulation of a quantum system become intractable in the numerical practice? We conclude the chapter with a discussion on the possibilities and limitation of classical simulations of quantum systems from the perspective of a key approach in classical quantum simulations, quantum Monte-Carlo (QMC) [Hir+82; Tro+03; Pol12; Tro+10]. Within the context of this chapter, this final discussion provides a complementary perspective on the merits of analogue simulations. At the same time it allows us to briefly present our contribution to the question. Besides the topical relation, our work on QMC is also methodically related to our approach to Hamiltonian identification. The work on QMC centres around the practical optimization over bases choices, and we use the same optimization algorithms for the orthogonal group as for the Hamiltonian identification.

Quantum Monte-Carlo (QMC) techniques are the central tool to perform numerical simulations of many-body quantum physics in order to study their equilibrium physics.[2] Our identification method focused on the real time simulation of Hamiltonian dynamics. In contrast, QMC techniques are typically used for efficiently calculating expectation values of observables in the Hamiltonian's ground and thermal state [Hir+82; Tro+03; Pol12; Tro+10], in other words, imaginary time evolution. To this end, Monte Carlo techniques probabilistically estimate thermal averages after expressing them in a suitable basis expansion. For

[2] Quantum Monte-Carlo (QMC) …] The remainder of this section summarizes our work published in Ref. [10]. The work has been conducted in collaboration with Dominik Hangleiter, Daniel Nagai, and Jens Eisert. The author of this thesis made central contributions to all analytical and numerical results of the work from the conception and calculation to the write-up.

example, consider the expectation value of an observable O in the Gibbs state of a system at inverse temperature β with Hamiltonian H

$$\langle O\rangle_{\beta,H} = Z_{\beta,H}^{-1}\operatorname{Tr}[O\exp\{-\beta H\}] = \sum_{\lambda} p(\lambda)f(\lambda) \tag{7.18}$$

with the partition function $Z_{\beta,H}$ as the normalizing factor. In QMC we want to rewrite $\langle O\rangle_{\beta,H}$ as the expected value of a function $f : \Lambda \to \mathbb{R}$ on a random variable taking values in Λ with probability mass function p

$$\langle O\rangle_{\beta,H} = \sum_{\lambda\in\Lambda} p(\lambda)f(\lambda)\,. \tag{7.19}$$

Having a method to generate samples from Λ according to p, we can estimate $\langle O\rangle_{\beta,H}$ by evaluating f on the samples and, e.g., using an empirical mean estimator. The complexity of estimating $\langle O\rangle_{\beta,H}$ now manifests itself in two ways: (1) The complexity of generating a single sample from p, (2) the sampling complexity of the empirical estimators as, e.g., controlled by its variance.

The first obstacle can already arise in the classical variants of Monte Carlo methods, where the Hamiltonian is always diagonal. For example, one observes exponentially long mixing times of Markov Chains that were set up to generate the samples for certain Hamiltonian models. But even when samples can be efficiently generated, in the quantum variant, one can encounter a so-called *sign problem*. Here, expanding the relevant quantities in terms of a basis, gives only rise to a quasi-probability distribution (normalized but non-positive) not a non-negative probability distribution. One can still proceed by defining a proxy probability distribution and an accordingly modified estimator that reproduces the original expected value. But the modified estimator typically exhibit an exponentially increased sampling complexity and hence run-time of the estimation procedure.

The sign problem manifests itself after expanding the Hamiltonian in a basis [HS92; Has15]. In fact, with a change of basis, one can *cure* the sign problem in certain settings by exploiting specific properties of the physical system [WHZ03; OH14; LJY15; LJY16; Nak98; ADP16; Hon+16; Wes+17]. A meaningful notion for the severity of the sign problems, thus, needs to consider the equivalence class of a Hamiltonian under basis transformations. Crucially, performing the basis transformations needs to be computationally tractable for the Hamiltonian. For example, in its eigenbasis a Hamiltonian does not cause a sign problem. Diagonalizing the Hamiltonian, however, is typically at least as hard as the original estimation problem. Tractable orbits of a local Hamiltonian can for example be generated by local Hadamard, Clifford or unitary transformations or, more generally, with quasi-local circuits which are efficiently computable [Has15], such as short circuits and matrix product unitaries [Cir+17; Sah+18]. The intrinsic sign-problem of the Hamiltonian is then a property of its orbit under a suitable subgroup of the unitary group.

A sufficient condition for avoiding quasi-probability distributions in the estimation problem is *stoquasticity* [Bra+08]: A Hamiltonian matrix is called *stoquastic* if it only has non-positive off-diagonal entries. Recently, it has been shown that deciding if a stoquastic Hamiltonian exists in the orbit of an *arbitrary 2-local* Hamiltonian under on-site basis transformation is NP-complete [MLH19; Kla+20]. For *strictly 2-local* Hamiltonians on the contrary, an efficient algorithm is given in Refs. [KT19; Kla+20].

From a practical perspective, however, aiming at stoquasticity might be unnecessarily ambitious. After all, the sign problem manifests itself as an increase in the variance of estimators in practice. Thus, instead of *exactly curing*, it already suffices to improve the variance by a computationally tractable basis change—in the best case yielding a polynomial sample complexity in the system size. Following this strategy, we developed a generally applicable, systematic framework for *easing* the sign problem in Ref. [10]. We therein both demonstrate the practical feasibility of the approach, and formally establish its limitations. In the following we summarize our findings.

The variance of the QMC estimators can be written as the inverse of the expected value of the signs of the quasi-probability distribution with respect to the proxy distribution—the so-called *average sign*. Ironically, computing the average sign suffers itself from the very same sign problem as the original estimation task. Hence, directly attempting to optimize the average sign is in general not efficient. Stoquasticity instead is defined directly as a property of the Hamiltonian matrix in a basis. In analogy, Ref. [10] introduces an efficiently computable measure of *approximate stoquasticity* for a Hamiltonian matrix. For a Hamiltonian matrix H we measure its distance from the set of stoquastic Hamiltonians in terms of *the sum of all its positive matrix entries*, $\nu_1(H) = \|H_\neg\|_{\ell_1}$. Here, $H_\neg$ denote the non-stoquastic part of the Hamiltonian defined by $(H_\neg)_{i,j} = h_{i,j}$ for $h_{i,j} > 0$ and $i \neq j$, and zero otherwise. The measure can be efficiently calculated for arbitrary local Hamiltonians on bounded-degree graphs. We further show that the measure can be efficiently estimated up-to inverse polynomial error for two-local Hamiltonian on any graph. Note that the ℓ_1 norm of the measure furthermore provides a natural regularizer promoting a sparse representation in the mindset of compressed sensing [MA15; Tho+15; DLA19].

For easing the sign problem it is now important to relate the measure of approximate stoquasticity to the sample complexity of the QMC estimation. Interestingly, we find that general analytic bounds on the variance in terms of approximate stoquasticity measures are not possible. In the context of the word-line QMC method [LB00], we can explicitly give examples of Hamiltonians with large positive off-diagonal but unit average sign and conversely also Hamiltonians arbitrarily close to being stoquastic but with nearly vanishing average sign [10]. The examples exploit the combinatoric structure of how the entries of very sparse Hamiltonians can appear in the polynomial expansion of the partition function when fine-tuning the parameters of the Monte-Carlo algorithm. As such the examples are rather pathological. Ref. [10] collects analytic and numerical evidence that for generic two-local Hamiltonians the average sign actually scales exponentially in the measure ν_1.

This serves as a motivation to actually try to ease the sign problem by optimizing the measure ν_1 over a suitable set of basis transformation. One of the simplest settings to practically test the approach are translationally invariant Hamiltonians on a quasi-one-dimensional geometry, such as anti-ferromagnetic Heisenberg Hamiltonians on a ladder geometry [MK04; DR96; Tak96]. We consider translationally invariant on-site orthogonal basis transformations $O \in \mathrm{O}(d)$, $H \mapsto O^{\otimes n} H (O^T)^{\otimes n}$. The optimization of ν_1 over $\mathrm{O}(d)$ can be performed using the manifold optimization algorithm that we also used for Hamiltonian learning [AEK09]. Ref. [10] shows numerically that the algorithm is able to find the basis in which 'curable' random two-locals Hamiltonians are stoquastic and can improve the stoquasticity measure for frustrated anti-ferromagnetic Heisenberg Hamiltonians on different ladder geometries. Figure 7.7 shows the corresponding improvement in the average sign after optimization for one of the ladder models and different parameters of the Heisenberg Hamiltonian. The source code of the numerical simulations is available as Ref. [26].

Our findings demonstrate already with simple basis transformation, one can significantly ease the sign problem. We therefore expect that in particular the optimization over more general ansatz classes can yield practically relevant improvements of the sampling complexity for QMC estimation problems that go beyond the toy models that we studied.

From a more fundamental perspective, one can ask what the fundamental limitations of this optimization approach for easing the sign problem are. Formally, we can formalize the optimization problem for the non-stoquasticity measure as a decision problem and study its computational complexity:

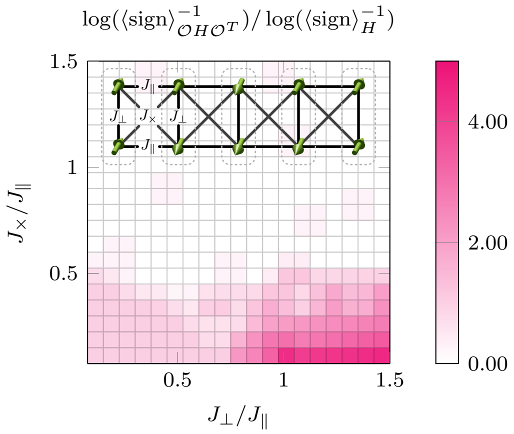


Figure 7.7: Improvement of the inverse average sign after optimizing the non-stoquasticity measure ν_1 for a frustrated ladder model depicted in the inset. The plot's coloring shows the ratio of the logarithm of the inverse sign after and before optimization. The axis vary the coupling parameters of the model on the ladder model. The ratio illustrates the achievable improvement in the required number of samples in a Monte-Carlo estimation of the partition function. The average sign was computed via exact diagonalization. The figure is taken from [10, Figure 2], see also the reference for further information on the model and numerical simulation.

Definition 5 (SignEasing). Given an n-qubit Hamiltonian H, constants $B > A \geq 0$ with $B - A \geq 1/\mathsf{poly}(n)$, and a set of allowed unitary transformations $\mathcal{U}$, decide which of the following is the case:

$$\text{YES}: \quad \exists U \in \mathcal{U} : \nu_1(UHU^\dagger) \leq A, \text{ or} \quad \text{NO}: \quad \forall U \in \mathcal{U} : \nu_1(UHU^\dagger) \geq B\,.$$

In Ref. [10] we establish the following theorem for the complexity of SignEasing for strictly 2-local (XYZ) Hamiltonians under particularly simple transformations $\mathcal{U}$.

Theorem 79 (Complexity of SignEasing [10, Theorem 2]). SignEasing *is* NP-*complete for 2-local (XYZ) Hamiltonians under both on-site orthogonal Clifford transformations, and on-site general orthogonal transformations.*

The proof encodes a promise version of the MAXCUT-problem into SignEasing. We refer to the supplemental material of our Ref. [10] for details.

Interestingly, we find that SignEasing is NP-complete exactly in the setting where Klassen *et al.* [KT19] established that one can efficiently decide if a stoquastic

Satisfiability	Stoquasticity	Complexity	Ref.
MAX2SAT	Easing strictly 2-local Hamiltonians	NP-complete	[10]
3SAT	Curing 2+1-local Hamiltonians	NP-complete	[MLH19; Kla+20]
2SAT	Curing strictly 2-local Hamiltonians	in P	[KT19; Kla+20]

Table 7.1: The complexity of curing and easing of local Hamiltonians in analogy to satisfiability problems.

Hamiltonian exist in the orbit. In other word, for strictly 2-local Hamiltonians curing is efficient if possible. Furthermore, for Hamiltonians further including one-local terms already curing is NP-complete [MLH19; Kla+20]. But—as our result shows—if curing is not possible, it is NP-complete to optimally ease the sign problem. The complexity of mitigating the sign problem turns out to be analogous to the hardness of satisfiability problem, where 2SAT is in P but 3SAT and MAX2SAT is NP-complete. We summarize this state of affairs in Table 7.1.

In summary one way how the simulation of quantum systems can become intractable is via the appearance of signs in their classical description. Exploring efficient and systematic ways to modify this prescription, here in terms of direct bases transformations, in order to ease these sign problems is a promising route to extend the range of classical simulation. At the same time, we established that such approaches ultimately are in the worst case only shifting the hardness of the problem. In general, it can already be NP-complete to even optimize an approximate stoquasticity measure of the Hamiltonian.

From the perspective of the quantum technologies the potential to push classical simulations further and further in specific settings constitute a big challenge. Quantum devices improve in scale and accuracy—a technological endeavour that requires a flexible, practical and theoretically well-understood toolbox of characterization techniques. Beyond their technological motivation such characterization techniques often touch upon the fundamental questions of what and how we can practically (efficiently) learn about nature in the quantum realm. At the same time we understand better and better how to exploit structure in order to improve the scaling of classical algorithms. We have here encountered a couple of such structure-exploiting classical algorithmic paradigms in the context of quantum characterization. These are of course much more generally applicable and in fact regularly applied in many engineering and scientific disciplines. Even though many of these techniques do not yield a polynomial scaling in the worst-case, they perform exceptionally well in many practical settings. As of today, the boundary between problems amenable to future quantum technologies but not to purely classical computation and simulation devices is still to be determined for practical applications.

Acronyms

AGF	average gate fidelity	32
ALS	alternating least square	89
NISQ	noisy and intermediate scale quantum	15
POVM	positive operator valued measure	41
PVM	projector-valued measure	144
CB	completely boundedness	66
CP	completely positive	59
CPT	completely positive and trace-preserving	59
CNOT	controlled NOT	58
IHT	iterative hard thresholding	72
MPS	matrix product state	125
MUBs	mutually unbiased bases	59
NSP	nullspace property	168
SDT	sparse de-mixing thresholding	86
SIC	symmetric, informationally complete	59
SPAM	state preparation and measurement	129
RB	randomized benchmarking	31
RIP	restricted isometry property	90
rMPS	random disordered MPS	125
QMC	quantum Monte-Carlo	219
QCVV	quantum characterization, verification and validation	16
GST	gate set tomography	30

Bibliography

[AA11] S. Aaronson and A. Arkhipov. “The computational complexity of linear optics”. In: *STOC’11: Proc. 43rd Ann. ACM Symp. Theor. Comput.* ACM. 2011, 333–342. arXiv: 1011.3245 [quant-ph].

[ABD15] G. Akemann, J. Baik, and P. Di Francesco, eds. *The Oxford Handbook of Random Matrix Theory.* Oxford [u.a.]: Oxford Univ. Press, 2015.

[ABW09] A. Ambainis, J. Bouda, and A. Winter. “Nonmalleable encryption of quantum information”. *J. Math. Phys.* **50** (2009), 042106. arXiv: 0808.0353 [quant-ph].

[Aci+07] A. Acin, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani. “Device-Independent Security of Quantum Cryptography against Collective Attacks”. *Phys. Rev. Lett.* **98** (2007), 230501.

[Ací+18] A. Acín, I. Bloch, H. Buhrman, T. Calarco, C. Eichler, J. Eisert, D. Esteve, N. Gisin, S. J. Glaser, F. Jelezko, S. Kuhr, M. Lewenstein, M. F. Riedel, P. O. Schmidt, R. Thew, A. Wallraff, I. Walmsley, and F. K. Wilhelm. “The quantum technologies roadmap: a European community view”. *New J. Phys.* **20** (2018), 080201. arXiv: 1712.03773 [quant-ph].

[ADP16] F. Alet, K. Damle, and S. Pujari. “Sign-Problem-Free Monte Carlo Simulation of Certain Frustrated Quantum Magnets”. *Phys. Rev. Lett.* **117** (2016), 197203. arXiv: 1511.01586 [cond-mat.str-el].

[AE07] A. Ambainis and J. Emerson. “Quantum t-designs: t-wise Independence in the Quantum World”. In: *Computational Complexity, 2007. CCC ’07. Twenty-Second Annual IEEE Conference on.* 2007, 129–140. arXiv: quant-ph/0701126 [quant-ph].

[AEK09] T. Abrudan, J. Eriksson, and V. Koivunen. “Conjugate gradient algorithm for optimization under unitary matrix constraint”. *Signal Process.* **89** (2009), 1704–1714.

[AG04] S. Aaronson and D. Gottesman. "Improved simulation of stabilizer circuits". *Phys. Rev. A* **70** (2004), 052328. arXiv: quant-ph/0406196 [quant-ph].

[AMS09] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds.* Princeton University Press, 2009.

[Ans16] A. Anshu. "Concentration bounds for quantum states with finite correlation length on quantum spin lattice systems". *New J. Phys.* **18** (2016), 083011. arXiv: 1508.07873 [quant-ph].

[App05] D. M. Appleby. "Symmetric informationally complete-positive operator valued measures and the extended Clifford group". *J. Math. Phys.* **46** (2005), 052107. eprint: quant-ph/0412001.

[ARR14] A. Ahmed, B. Recht, and J. Romberg. "Blind deconvolution using convex programming". *IEEE Trans. Inf. Th.* **60** (2014), 1711–1732.

[Aru+19] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Land huis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis. "Quantum supremacy using a programmable superconducting processor". *Nature* **574** (2019), 505–510. arXiv: 1910.11333 [quant-ph].

[Bai+20] E. Bairey, C. Guo, D. Poletti, N. H. Lindner, and I. Arad. "Learning the Dynamics of Open Quantum Systems from Their Steady States". *New J. Phys.* **22** (2020), 032001.

[BAL19] E. Bairey, I. Arad, and N. H. Lindner. "Learning a Local Hamiltonian from Local Measurements". *Phys. Rev. Lett.* **122** (2019), 020504.

[Ban+13] A. S. Bandeira, E. Dobriban, D. G. Mixon, and W. F. Sawin. “Certifying the restricted isometry property is hard”. *IEEE Trans. Inf. Th.* **59** (2013), 3448–3450.

[Ban+20] E. Bannai, G. Navarro, N. Rizo, and P. H. Tiep. “Unitary t-groups”. *J. Math. Soc. Japan* **72** (2020), 909–921.

[Bar+10] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. “Model-based compressive sensing”. *IEEE Trans. Inf. Th.* **56** (2010), 1982–2001. arXiv: 0808.3572 [cs.IT].

[Bar+12] S. Barz, E. Kashefi, A. Broadbent, J. F. Fitzsimons, A. Zeilinger, and P. Walther. “Demonstration of Blind Quantum Computing”. *Science* **335** (2012), 303–308.

[Bar+16] D. Barredo, S. de Leseleuc, V. Lienhard, T. Lahaye, and A. Browaeys. “An atom-by atom assembler of defect-free arbitrary 2d atomic arrays”. *Science* **354** (2016), 1021.

[BB19] M. Brennan and G. Bresler. “Optimal average-case reductions to sparse PCA: From weak assumptions to strong hardness”. In: *32nd Annual Conference on Learning Theory.* Vol. 99. Proceedings of Machine Learning Research. 2019. arXiv: 1902.07380 [cs.CC].

[BD08] T. Blumensath and M. E. Davies. “Iterative Thresholding for Sparse Approximations”. *Journal of Fourier Analysis and Applications* **14** (2008), 629–654.

[BD09] T. Blumensath and M. E. Davies. “Sampling theorems for signals from the union of finite-dimensional linear subspaces”. *IEEE Trans. Inf. Theory* **55** (2009), 1872–1882.

[BDN12] I. Bloch, J. Dalibard, and S. Nascimbène. “Quantum simulations with ultracold quantum gases”. *Nat. Phys.* **8** (2012), 267–276.

[BDZ08] I. Bloch, J. Dalibard, and W. Zwerger. “Many-body physics with ultracold gases”. *Rev. Mod. Phys.* **80** (2008), 885–964.

[BE18] W. G. Brown and B. Eastin. “Randomized benchmarking with restricted gate sets”. *Phys. Rev. A* **97** (2018), 062323. arXiv: 1801.04042 [quant-ph].

[Ber+17] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletic, and M. D. Lukin. “Probing many-body dynamics on a 51-atom quantum simulator”. *Nature* **551** (2017), 579–584.

[Ber+18] J. Bermejo-Vega, D. Hangleiter, M. Schwarz, R. Raussendorf, and J. Eisert. “Architectures for Quantum Simulation Showing a Quantum Speedup”. *Phys. Rev. X* **8** (2018), 021010. arXiv: 1703.00466 [quant-ph].

[BFK18] A. Bouland, J. F. Fitzsimons, and D. E. Koh. “Complexity Classification of Conjugated Clifford Circuits”. In: *33rd Computational Complexity Conference (CCC 2018)*. Ed. by R. A. Servedio. Vol. 102. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018, 21:1–21:25. arXiv: 1709.01805 [quant-ph].

[BG11] P. Bühlmann and S. V. D. Geer. *Statistics for high-dimensional data*. Vol. 9. Springer Series in Statistics. Berlin: Springer, 2011.

[BH13] F. G. S. L. Brandao and M. Horodecki. “Exponential quantum speedups are generic”. *Q. Inf. Comp.* **13** (2013), 0901. arXiv: 1010.3654 [quant-ph].

[Bha97] R. Bhatia. *Matrix analysis*. Graduate texts in mathematics. New York; Heidelberg [u.a.]: Springer, 1997, XI, 347 S.

[BHH16a] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki. “Efficient Quantum Pseudorandomness”. *Phys. Rev. Lett.* **116** (2016).

[BHH16b] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki. “Local random quantum circuits are approximate polynomial-designs”. *Comm. Math. Phys.* **346** (2016), 397–434. arXiv: 1208.0692 [quant-ph].

[BJS10] M. J. Bremner, R. Jozsa, and D. J. Shepherd. “Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy”. *Proc. Roy. Soc.* **467** (2010), 2126. arXiv: 1005.1407 [quant-ph].

[BKD14] C. H. Baldwin, A. Kalev, and I. H. Deutsch. “Quantum process tomography of unitary and near-unitary maps”. *Phys. Rev. A* **90** (2014), 012110. eprint: arXiv:1404.2877.

[Blu+13] R. J. Blume-Kohout, J. K. Gamble, E. Nielsen, J. A. Mizrahi, J. D. Sterk, and P. L. W. Maunz. “Robust self-consistent closed-form tomography of quantum logic gates on a trapped ion qubit.” *Nat. Phys.* (2013).

[Blu+17] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz. “Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography”. *Nat. Commun.* **8** (2017), 14485. arXiv: 1605.07674 [quant-ph].

[BM09] D. Burgarth and K. Maruyama. “Indirect Hamiltonian Identification through a Small Gateway”. *New J. Phys.* **11** (2009), 103019.

[BMN09] D. Burgarth, K. Maruyama, and F. Nori. “Coupling Strength Estimation for Spin Chains despite Restricted Access”. *Phys. Rev. A* **79** (2009), 020305.

[BMN11] D. Burgarth, K. Maruyama, and F. Nori. “Indirect Quantum Tomography of Quadratic Hamiltonians”. *New J. Phys.* **13** (2011), 013019.

[BMS17] M. J. Bremner, A. Montanaro, and D. J. Shepherd. “Achieving quantum supremacy with sparse and noisy commuting quantum computations”. *Quantum* **1** (2017), 8. arXiv: 1610.01808 [quant-ph].

[BMZ16] A. Bouland, L. Mančinska, and X. Zhang. “Complexity Classification of Two-Qubit Commuting Hamiltonians”. In: *31st Conference on Computational Complexity (CCC 2016).* Ed. by R. Raz. Vol. 50. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016, 28:1–28:33. arXiv: 1602.04145 [quant-ph].

[BN13] B. Bauer and C. Nayak. “Area laws in a many-body localized state and its implications for topological order”. *Journal of Statistical Mechanics: Theory and Experiment* **2013** (2013), P09005.

[Boi+18] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven. “Characterizing Quantum Supremacy in Near-Term Devices”. *Nat. Phys.* **14** (2018), 595–600. arXiv: 1608.00263 [quant-ph].

[Boo+19] K. Boone, A. Carignan-Dugas, J. J. Wallman, and J. Emerson. “Randomized benchmarking under different gate sets”. *Phys. Rev. A* **99** (2019), 032329. arXiv: 1811.01920 [quant-ph].

[Bou+18] A. Bouland, B. Fefferman, C. Nirkhe, and U. Vazirani. “Quantum Supremacy and the Complexity of Random Circuit Sampling”. *Nat. Phys.* **15** (2018), 159–163. arXiv: 1803.04402 [quant-ph].

[BR12] R. Blatt and C. F. Roos. “Quantum simulations with trapped ions”. *Nat. Phys.* **8** (2012), 277.

[BR13a] Q. Berthet and P. Rigollet. “Complexity theoretic lower bounds for sparse principal component detection”. In: *Conference on Learning Theory.* 2013, 1046–1066.

[BR13b] Q. Berthet and P. Rigollet. “Optimal detection of sparse principal components in high dimension”. *Ann. Statist.* **41** (2013), 1780–1815.

[BR17] B. N. Balz and P. Reimann. “Typical Relaxation of Isolated Many-Body Systems Which Do Not Thermalize”. *Phys. Rev. Lett.* **118** (2017), 190601.

[Bra+08] S. Bravyi, D. P. DiVincenzo, R. I. Oliveira, and B. M. Terhal. “The Complexity of Stoquastic Local Hamiltonian Problems”. *Quant. Inf. Comp.* **8** (2008), 0361. arXiv: quant-ph/0606140.

[Bra+12] A. M. Brańczyk, D. H. Mahler, L. A. Rozema, A. Darabi, A. M. Steinberg, and D. F. V. James. “Self-calibrating quantum state tomography”. *New J. Phys.* **14** (2012), 085003.

[Bra+15] S. Braun, M. Friesdorf, S. S. Hodgman, M. Schreiber, J. P. p. Ronzheimer, A. Riera, M. del Rey, I. Bloch, J. Eisert, and U. Schneider. “Emergence of coherence and the dynamics of quantum phase transitions”. *Proc. Natl. Acad. Sci.* **112** (2015), 3641–3646.

[BSH21] P. Bienias, A. Seif, and M. Hafezi. “Meta Hamiltonian learning”. 2021. arXiv: 2104.04453 [quant-ph].

[BV10] W. G. Brown and L. Viola. “Convergence rates for arbitrary statistical moments of random quantum circuits”. *Phys. Rev. Lett.* **104** (2010), 250501. arXiv: 0910.0913 [quant-ph].

[BW08] R. Blatt and D. Wineland. “Entangled states of trapped atomic ions”. *Nature* **453** (2008), 1008–1015.

[BY12] D. Burgarth and K. Yuasa. “Quantum System Identification”. *Phys. Rev. Lett.* **108** (2012), 080502.

[Cal+98] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane. “Quantum error correction via codes over GF(4)”. *IEEE Trans. Inf. Theory* **44** (1998), 1369–1387. arXiv: quant-ph/9608006.

[Car+18] A. Carignan-Dugas, K. Boone, J. J. Wallman, and J. Emerson. “From randomized benchmarking experiments to gate-set circuit fidelity: how to interpret randomized benchmarking decay parameters”. *New J. Phys.* **20** (2018), 092001. arXiv: 1804.01122 [quant-ph].

[Car+20] I. Carusotto, A. A. Houck, A. J. Kollar, P. Roushan, D. I. Schuster, and J. Simon. “Photonic materials in circuit quantum electrodynamics”. *Nat. Phys.* **16** (2020), 268–279.

[CC18] E. Chertkov and B. K. Clark. “Computational Inverse Method for Constructing Spaces of Quantum Models from Wave Functions”. *Phys. Rev. X* **8** (2018).

[CDH06] J. H. Cole, S. J. Devitt, and L. C. L. Hollenberg. "Precision Characterization of Two-Qubit Hamiltonians via Entanglement Mapping". *J. Phys. A: Math. Gen.* **39** (2006), 14649–14658.

[CEW12] T. S. Cubitt, J. Eisert, and M. M. Wolf. "Extracting Dynamical Equations from Experimental Data is NP Hard". *Phys. Rev. Lett.* **108** (2012), 120503.

[CF13] E. J. Candès and C. Fernandez-Granda. "Super-resolution from noisy data". *J. Fourier An. App.* **19** (2013), 1229–1254.

[CF14] E. J. Candès and C. Fernandez-Granda. "Towards a mathematical theory of super-resolution". *Comm. Pure App. Math.* **67** (2014), 906–956.

[Cha+17] T. Chasseur, D. M. Reich, C. P. Koch, and F. K. Wilhelm. "Hybrid benchmarking of arbitrary quantum gates". *Phys. Rev. A* **95** (2017), 062335. arXiv: 1606.03927 [quant-ph].

[Cha04] H. F. Chau. "Unconditionally secure key distribution in higher dimensions by depolarization". *IEEE Trans. Inf. Theory* **51** (2004), 1451–1468.

[Che+14] Y. Chen, C. Neill, P. Roushan, N. Leung, M. Fang, R. Barends, J. Kelly, B. Campbell, Z. Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, A. Megrant, J. Y. Mutus, P. J. J. O'Malley, C. M. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, M. R. Geller, A. N. Cleland, and J. M. Martinis. "Qubit Architecture with High Coherence and Fast Tunable Coupling". *Phys. Rev. Lett.* **113** (2014), 220502–220502.

[Che+21a] L. Che, C. Wei, Y. Huang, D. Zhao, S. Xue, X. Nie, J. Li, D. Lu, and T. Xin. "Learning Quantum Hamiltonians from Single-Qubit Measurements". *Phys. Rev. Research* **3** (2021), 023246.

[Che+21b] X. Chen, Y. Li, Z. Wu, R. Liu, Z. Li, and H. Zhou. "Experimental Realization of Hamiltonian Tomography by Quantum Quenches". *Phys. Rev. A* **103** (2021), 042429.

[Cho+16] J.-y. Choi, S. Hild, J. Zeiher, P. Schauß, A. Rubio-Abadal, T. Yefsah, V. Khemani, D. A. Huse, I. Bloch, and C. Gross. "Exploring the many-body localization transition in two dimensions". *Science* **352** (2016), 1547–1552. arXiv: 1604.04178 [cond-mat.quant-gas].

[Cho75] M.-D. Choi. "Completely positive linear maps on complex matrices". *Lin. Alg. App.* **10** (1975), 285–290.

[Cir+17] J. I. Cirac, D. Perez-Garcia, N. Schuch, and F. Verstraete. "Matrix product unitaries: structure, symmetries, and topological invariants". *J. Stat. Mech.* **2017** (2017), 083105. arXiv: 1703.09188 [cond-mat.str-el].

[CL14] E. Candès and X. Li. "Solving quadratic equations via PhaseLift when there are about as many equations as unknowns". *Found. Comput. Math.* **14** (2014), 1017–1026. arXiv: 1208.6247 [cs.IT].

[CM87] J. P. Crutchfield and B. S. McNamara. "Equations of motion from a data series". *Complex Systems* **1** (1987), 417–452.

[CN16] B. Collins and I. Nechita. "Random matrix techniques in quantum information theory". *J. Math. Phys.* **57** (2016), 015215.

[CN97] I. L. Chuang and M. A. Nielsen. "Prescription for experimental determination of the dynamics of a quantum black box". *J. Mod. Opt.* **44** (1997), 2455–2467. eprint: quant-ph/9610001.

[COB20] P. Cerfontaine, R. Otten, and H. Bluhm. "Self-Consistent Calibration of Quantum-Gate Sets". *Phys. Rev. Appl.* **13** (2020), 044071. arXiv: 1906.00950 [quant-ph].

[Col+05] J. H. Cole, S. G. Schirmer, A. D. Greentree, C. J. Wellard, D. K. L. Oi, and L. C. L. Hollenberg. "Identifying an Experimental Two-State Hamiltonian to Arbitrary Accuracy". *Phys. Rev. A* **71** (2005), 062312.

[Col+06] J. H. Cole, A. D. Greentree, D. K. L. Oi, S. G. Schirmer, C. J. Wellard, and L. C. L. Hollenberg. "Identifying a Two-State Hamiltonian in the Presence of Decoherence". *Phys. Rev. A* **73** (2006), 062333.

[Col03] B. Collins. "Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability". *Int. Math. Res. Not.* **2003** (2003), 953–982. eprint: math-ph/0205010.

[Com+17] J. Combes, C. Granade, C. Ferrie, and S. T. Flammia. "Logical Randomized Benchmarking". 2017. arXiv: 1702.03688 [quant-ph].

[CP11] E. J. Candes and Y. Plan. "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements". *IEEE Trans. Inf. Th.* **57** (2011), 2342–2359.

[CPR16] S. O. Chan, D. Papailliopoulos, and A. Rubinstein. "On the approximability of sparse PCA". In: *PMLR*. Vol. 49. 2016, 623–646. arXiv: 1507.05950 [stat.ML].

[Cro+16] A. W. Cross, E. Magesan, L. S. Bishop, J. A. Smolin, and J. M. Gambetta. “Scalable randomised benchmarking of non-Clifford gates”. *npj Quant. Inf.* **2** (2016), 16012. arXiv: 1510.02720 [quant-ph].

[CS06] B. Collins and P. Sniady. “Integration with respect to the Haar measure on unitary, orthogonal and symplectic group”. *Commun. Math. Phys.* **264** (2006), 773. arXiv: math-ph/0402073 [math-ph].

[CSV13] E. J. Candès, T. Strohmer, and V. Voroninski. “Phaselift: exact and stable signal recovery from magnitude measurements via convex programming.” *Commun. Pure Appl. Math.* **66** (2013), 1241–1274. arXiv: 1109.4499 [cs.IT].

[CT12] T. M. Cover and J. A. Thomas. *Elements of information theory.* John Wiley & Sons, 2012.

[CT21] J. Carrasquilla and G. Torlai. “Neural networks in quantum many-body physics: a hands-on tutorial”. 2021. arXiv: 2101.11099 [quant-ph].

[CTV17] E. T. Campbell, B. M. Terhal, and C. Vuillot. “Roads towards fault-tolerant universal quantum computation”. *Nature* **549** (2017), 172–179. arXiv: 1612.07330 [quant-ph].

[CW15] T. Chasseur and F. K. Wilhelm. “Complete randomized benchmarking protocol accounting for leakage errors”. *Phys. Rev. A* **92** (2015), 042333. arXiv: 1505.00580 [quant-ph].

[CWE15] A. Carignan-Dugas, J. J. Wallman, and J. Emerson. “Characterizing universal gate sets via dihedral benchmarking”. *Phys. Rev. A* **92** (2015), 060302(R). arXiv: 1508.06312 [quant-ph].

[CWE19] A. Carignan-Dugas, J. J. Wallman, and J. Emerson. “Bounding the average gate fidelity of composite channels using the unitarity”. *New J. Phys.* **21** (2019), 053016. arXiv: 1610.05296 [quant-ph].

[CZ12] J. I. Cirac and P. Zoller. “Goals and opportunities in quantum simulation”. *Nat. Phys.* **8** (2012), 264.

[Cze21] A. Czerwinski. “Hamiltonian Tomography by the Quantum Quench Protocol with Random Noise”. *Phys. Rev. A* **104** (2021), 052431. arXiv: 2107.04033 [quant-ph].

[Dal+20] A. M. Dalzell, A. W. Harrow, D. E. Koh, and R. L. La Placa. “How many qubits are needed for quantum computational supremacy?” *Quantum* **4** (2020), 264. arXiv: 1805.05224 [quant-ph].

[Dan+09] C. Dankert, R. Cleve, J. Emerson, and E. Livine. "Exact and approximate unitary 2-designs and their application to fidelity estimation". *Phys. Rev. A* **80** (2009), 012304. arXiv: quant-ph/0606161 [quant-ph].

[Der+20] E. Derbyshire, J. Yago Malo, A. Daley, E. Kashefi, and P. Wallden. "Randomized Benchmarking in the Analogue Setting". *Quantum Sci. Technol.* **5** (2020), 034001. arXiv: 1909.01295 [quant-ph].

[DGS77] P. Delsarte, J. Goethals, and J. Seidel. "Spherical codes and designs". *Geom. Dedicata* **6** (1977), 363–388.

[DHW19] B. Dirkse, J. Helsen, and S. Wehner. "Efficient unitarity randomized benchmarking of few-qubit Clifford gates". *Phys. Rev. A* **99** (2019), 012315. arXiv: 1808.00850 [quant-ph].

[DLA19] W. Dobrautz, H. Luo, and A. Alavi. "Compact numerical solutions to the two-dimensional repulsive Hubbard model obtained via nonunitary similarity transformations". *Phys. Rev. B* **99** (2019), 075119. arXiv: 1811.03607 [cond-mat.str-el].

[DLT02] D. P. DiVincenzo, D. W. Leung, and B. M. Terhal. "Quantum data hiding". *IEEE Trans. Inf. Th.* **48** (2002), 580–598.

[DPK09] C. Di Franco, M. Paternostro, and M. S. Kim. "Hamiltonian Tomography in an Access-Limited Setting without State Initialization". *Phys. Rev. Lett.* **102** (2009), 187203.

[DR96] E. Dagotto and T. M. Rice. "Surprises on the Way from One- to Two-Dimensional Quantum Magnets: The Ladder Materials". *Science* **271** (1996), 618–623.

[DS13] M. H. Devoret and R. J. Schoelkopf. "Superconducting Circuits for Quantum Information: An Outlook". *Science* **339** (2013), 1169–1174.

[DS93] P. Diaconis and L. Saloff-Coste. "Comparison Theorems for Reversible Markov Chains". *Ann. Appl. Probab.* **3** (1993), 696–730.

[DS94] P. Diaconis and M. Shahshahani. "On the Eigenvalues of Random Matrices". *J. Appl. Probab.* **31** (1994), 49–62.

[EAS98] A. Edelman, T. A. Arias, and S. T. Smith. "The geometry of algorithms with orthogonality constraints". *SIAM J. Matr. Ana. App.* **20** (1998), 303–353. eprint: https://doi.org/10.1137/S0895479895290954.

[EAŻ05a] J. Emerson, R. Alicki, and K. Życzkowski. "Scalable noise estimation with random unitary operators". *J. Opt. B: Quantum Semiclass. Opt.* **7** (2005), 347. arXiv: quant-ph/0503243 [quant-ph].

[EAŻ05b] J. Emerson, R. Alicki, and K. Życzkowski. "Scalable noise estimation with random unitary operators". *J. Opt. B: Quantum Semiclass. Opt.* **7** (2005), S347.

[ECP10] J. Eisert, M. Cramer, and M. B. Plenio. "Colloquium: Area laws for the entanglement entropy". *Rev. Mod. Phys.* **82** (2010), 277–306.

[EFG15] J. Eisert, M. Friesdorf, and C. Gogolin. "Quantum many-body systems out of equilibrium". *Nat. Phys.* **11** (2015).

[Erh+19] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt. "Characterizing large-scale quantum computers via cycle benchmarking". *Nat. Commun.* **10** (2019), 5347. arXiv: 1902.08543 [quant-ph].

[Eva+22] T. J. Evans, W. Huang, J. Yoneda, R. Harper, T. Tanttu, K. W. Chan, F. E. Hudson, K. M. Itoh, A. Saraiva, C. H. Yang, A. S. Dzurak, and S. D. Bartlett. "Fast Bayesian tomography of a two-qubit gate set in silicon". *Phys. Rev. A* **17** (2022), 024068. arXiv: 2107.14473 [quant-ph].

[Fan16] A. Fannjiang. "Compressive spectral estimation with single-snapshot esprit: Stability and resolution". 2016. arXiv: 1607.01827 [cs.IT].

[FBC17] T. Farrelly, F. G. Brandão, and M. Cramer. "Thermalization and Return to Equilibrium on Finite Quantum Lattice Systems". *Phys. Rev. Lett.* **118** (2017), 140601.

[Fen+16] G. Feng, J. J. Wallman, B. Buonacorsi, F. H. Cho, D. K. Park, T. Xin, D. Lu, J. Baugh, and R. Laflamme. "Estimating the coherence of noise in quantum control of a solid-state qubit". *Phys. Rev. Lett.* **117** (2016), 260501. arXiv: 1603.03761 [quant-ph].

[Fey75] P. Feyerabend. *Against method.* New Left Books, 1975.

[Fey86] R. P. Feynman. "Quantum mechanical computers". *Found. Phys.* **16** (1986), 507–531.

[FH16] E. Farhi and A. W. Harrow. "Quantum Supremacy through the Quantum Approximate Optimization Algorithm". 2016. arXiv: 1602.07674 [quant-ph].

[FH18] D. S. França and A. K. Hashagen. "Approximate randomized benchmarking for finite groups". *J. Phys. A* **51** (2018), 395302. arXiv: 1803.03621 [quant-ph].

[FH91] W. Fulton and J. Harris. *Representation theory.* Vol. 129. Springer Science & Business Media, 1991.

[FHB01] M. Fazel, H. Hindi, and S. Boyd. "A rank minimization heuristic with application to minimum order system approximation". In: *Proceedings American Control Conference.* Vol. 6. 2001, 4734–4739.

[FHT10] J. Friedman, T. Hastie, and R. Tibshirani. "A note on the group lasso and a sparse group lasso" (2010). arXiv: 1001.0736 [math.ST].

[FL11] S. T. Flammia and Y.-K. Liu. "Direct fidelity estimation from few Pauli measurements". *Phys. Rev. Lett.* **106** (2011), 230501. arXiv: 1104.4695 [quant-ph].

[Fla+12] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert. "Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators". *New J. Phys.* **14** (2012), 095022. arXiv: 1205.2300 [quant-ph].

[Fla17] S. T. Flammia. *Characterization of quantum devices.* Tutorial session at QIP conference 2017. slides available under: https://www.microsoft.com/en-us/research/wp-content/uploads/2017/09/2017-01-14-Morning-Tutorial-Steve-Flammia-2.pdf (accessed Feb 12, 2020). 2017.

[Flu+20] E. Flurin, L. S. Martin, S. Hacohen-Gourgy, and I. Siddiqi. "Using a Recurrent Neural Network to Reconstruct Quantum Dynamics of a Superconducting Qubit from Physical Observations". *Phys. Rev. X* **10** (2020), 011006.

[FNW92] M. Fannes, B. Nachtergaele, and R. F. Werner. "Finitely correlated states on quantum spin chains". *Commun. Math. Phys.* **144** (1992), 443–490.

[Fou+19] S. Foucart, R. Gribonval, L. Jacques, and H. Rauhut. "Jointly low-rank and bisparse recovery: Questions and partial answers". *arXiv e-prints* (2019). arXiv: 1902.04731 [math.NA].

[Fou11] S. Foucart. "Hard thresholding pursuit: An algorithm for compressive sensing". *SIAM J. Num. An.* **49** (2011), 2543–2563.

[FR13] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing.* Heidelberg: Springer, 2013.

[Fri+15] M. Friesdorf, A. H. Werner, W. Brown, V. B. Scholz, and J. Eisert. "Many-Body Localization Implies that Eigenvectors are Matrix-Product States". *Phys. Rev. Lett.* **114** (2015), 170505.

[FU16] B. Fefferman and C. Umans. "On the Power of Quantum Fourier Sampling". In: *11th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2016)*. Ed. by A. Broadbent. Vol. 61. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016, 1:1–1:19. arXiv: 1507.05592 [cs.CC].

[Fv99] C. A. Fuchs and J. van de Graaf. "Cryptographic Distinguishability Measures for Quantum Mechanical States". *IEEE Trans. Inf. Th.* **45** (1999), 1216–1227. arXiv: quant-ph/9712042 [quant-ph].

[FW20] S. T. Flammia and J. J. Wallman. "Efficient estimation of Pauli channels". *ACM Transactions on Quantum Computing* **1** (2020). arXiv: 1907.12976 [quant-ph].

[Gae+12] J. P. Gaebler, A. M. Meier, T. R. Tan, R. Bowler, Y. Lin, D. Hanneke, J. D. Jost, J. P. Home, E. Knill, D. Leibfried, and D. J. Wineland. "Randomized Benchmarking of Multiqubit Gates". *Phys. Rev. Lett.* **108** (2012), 260503. arXiv: 1203.3733 [quant-ph].

[GAE07] D. Gross, K. Audenaert, and J. Eisert. "Evenly distributed unitaries: on the structure of unitary designs". *J. Math. Phys.* **48** (2007), 052104. arXiv: quant-ph/0611002.

[Gal+10] R. Gallego, N. Brunner, C. Hadley, and A. Acín. "Device-independent tests of classical and quantum dimensions". *Phys. Rev. Lett.* **105** (2010), 230501.

[Gal+18] R. Gallego, H. Wilming, J. Eisert, and C. Gogolin. "What it takes to avoid equilibration". *Phys. Rev. A* **98** (2018), 022135.

[Gam+12] J. M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, M. B. Ketchen, and M. Steffen. "Characterization of Addressability by Simultaneous Randomized Benchmarking". *Phys. Rev. Lett.* **109** (2012), 240504. arXiv: 1204.6308 [quant-ph].

[Gar+10] S. Garnerone, T. R. de Oliveira, S. Haas, and P. Zanardi. "Statistical properties of random matrix product states". *Phys. Rev. A* **82** (2010), 052312.

[GE16] C. Gogolin and J. Eisert. "Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems". *Rep. Prog. Phys.* **79** (2016), 056001.

[GG18] J. R. Garrison and T. Grover. "Does a Single Eigenstate Encode the Full Hamiltonian?" *Phys. Rev. X* **8** (2018), 021026.

[GKK15] D. Gross, F. Krahmer, and R. Kueng. “A partial derandomization of PhaseLift using spherical designs”. *J. Fourier Anal. Appl.* **21** (2015), 229–266. arXiv: 1310.2267 [cs.IT].

[GKK19] A. Gheorghiu, T. Kapourniotis, and E. Kashefi. “Verification of quantum computation: An overview of existing approaches”. *Theory of computing systems* **63** (2019), 715–808. arXiv: 1709.06984 [quant-ph].

[GL89] G. H. Golub and C. F. van Loan. *Matrix computations.* The Johns Hopkins University Press, 1989.

[GNW21] D. Gross, S. Nezami, and M. Walter. “Schur-Weyl Duality for the Clifford Group with Applications”. *Commun. Math. Phys.* **385** (2021), 1325–1393. arXiv: 1712.08628 [quant-ph].

[Gol+10] S. Goldstein, J. L. Lebowitz, C. Mastrodonato, R. Tumulka, and N. Zhangì. “Normal typicality and von Neumann’s quantum ergodic theorem”. *Proc. R. Soc. A* **466** (2010), 3203–3224.

[Got09] D. Gottesman. “An Introduction to Quantum Error Correction and Fault-Tolerant Quantum Computation”. 2009. arXiv: 0904.2557 [quant-ph].

[Gra+12] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory. “Robust online Hamiltonian learning”. *New J. Phys.* **14** (2012), 103013.

[Gre+02] M. Greiner, O. Mandel, T. Esslinger, T. W. HÃ€nsch, and I. Bloch. “Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms”. *Nature* **415** (3, 2002), 39–44.

[Gre15] D. Greenbaum. “Introduction to Quantum Gate Set Tomography”. 2015. arXiv: 1509.02921 [quant-ph].

[Gro+10] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert. “Quantum state tomography via compressed sensing”. *Phys. Rev. Lett.* **105** (2010), 150401. arXiv: 0909.3304 [quant-ph].

[Gro11] D. Gross. “Recovering low-rank matrices from few coefficients in any basis”. *IEEE Trans. Inf. Th.* **57** (2011), 1548–1566. arXiv: 0910.1879 [cs.IT].

[Gu+21] Y. Gu, R. Mishra, B.-G. Englert, and H. Khoon Ng. “Randomized linear gate set tomography”. *PRX Quantum* **2** (2021), 030328. arXiv: 2010.12235 [quant-ph].

[GW00] R. Goodman and N. R. Wallach. *Representations and invariants of the classical groups.* Vol. 68. Cambridge University Press, 2000.

[GWD17] X. Gao, S.-T. Wang, and L.-M. Duan. "Quantum supremacy for simulating a translation-invariant Ising spin model". *Phys. Rev. Lett.* **118** (2017), 040502. arXiv: 1607.04947 [quant-ph].

[HA08] R. Hanson and D. D. Awschalom. "Coherent manipulation of single spins in semiconductors". *Nature* **453** (2008), 1043–1049.

[Haa+17] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu. "Sample-Optimal Tomography of Quantum States". *IEEE Trans. Inf. Theory* **63** (2017), 5628–5641. arXiv: 1508.01797 [quant-ph].

[Haf+20a] J. Haferkamp, D. Hangleiter, A. Bouland, B. Fefferman, J. Eisert, and J. Bermejo-Vega. "Closing gaps of a quantum advantage with short-time Hamiltonian dynamics". *Phys. Rev. Lett.* **125** (2020), 250501. arXiv: 1908.08069 [quant-ph].

[Hal13] B. Hall. *Quantum Theory for Mathematicians.* Vol. 267. Graduate Texts in Mathematics. Springer-Verlag New York, 2013.

[Han+18] D. Hangleiter, J. Bermejo-Vega, M. Schwarz, and J. Eisert. "Anticoncentration theorems for schemes showing a quantum speedup". *Quantum* **2** (2018), 65. arXiv: 1706.03786 [quant-ph].

[Han20] D. Hangleiter. *Sampling and the complexity of nature.* PhD thesis, Freie Universität Berlin. 2020.

[Har+19] R. Harper, I. Hincks, C. Ferrie, S. T. Flammia, and J. J. Wallman. "Statistical analysis of randomized benchmarking". *Phys. Rev. A* **99** (2019), 052350. arXiv: 1901.00535 [quant-ph].

[Has+18] A. K. Hashagen, S. T. Flammia, D. Gross, and J. J. Wallman. "Real Randomized Benchmarking". *Quantum* **2** (2018), 85. arXiv: 1801.06121 [quant-ph].

[Has15] M. B. Hastings. "How quantum are non-negative wavefunctions?" *J. Math. Phys.* **57** (2015), 015210. arXiv: 1506.08883 [quant-ph].

[HC17] M. Howard and E. Campbell. "Application of a resource theory for magic states to fault-tolerant quantum computing". *Phys. Rev. Lett.* **118** (2017), 090501.

[Hel+19a] J. Helsen, J. J. Wallman, S. T. Flammia, and S. Wehner. "Multiqubit randomized benchmarking using few samples". *Phys. Rev. A* **100** (2019), 032304. arXiv: 1701.04299 [quant-ph].

[Hel+19b] J. Helsen, X. Xue, L. M. K. Vandersypen, and S. Wehner. "A new class of efficient randomized benchmarking protocols". *npj Quant. Inf.* **5** (2019), 71. arXiv: 1806.02048 [quant-ph].

[HF17] R. Harper and S. T. Flammia. "Estimating the fidelity of T gates using standard interleaved randomized benchmarking". *Quantum Sci. Technol.* **2** (2017), 015008. arXiv: 1608.02943 [quant-ph].

[HFW20] R. Harper, S. T. Flammia, and J. J. Wallman. "Efficient learning of quantum noise". *Nat. Phys.* (2020). arXiv: 1907.13022 [quant-ph].

[HH19] Y. Huang and A. W. Harrow. "Instability of localization in translation-invariant systems". 2019. arXiv: 1907.13392 [cond-mat.dis-nn].

[HHH99] M. Horodecki, P. Horodecki, and R. Horodecki. "General teleportation channel, singlet fraction, and quasidistillation". *Phys. Rev. A* **60** (1999), 1888–1898.

[Hir+82] J. E. Hirsch, R. L. Sugar, D. J. Scalapino, and R. Blankenbecler. "Monte Carlo simulations of one-dimensional fermion systems". *Phys. Rev. B* **26** (1982), 5033.

[HKP20] H.-Y. Huang, R. Kueng, and J. Preskill. "Predicting Features of Quantum Systems from Very Few Measurements". *Nat. Phys.* **16** (2020), 1050–1057. arXiv: 1908.08909 [quant-ph].

[HL09] A. W. Harrow and R. A. Low. "Random quantum circuits are approximate 2-designs". *Commun. Math. Phys.* **291** (2009), 257–302. arXiv: 0802.1919 [quant-ph].

[HLL17] S.-Y. Hou, H. Li, and G.-L. Long. "Experimental Quantum Hamiltonian Identification from Measurement Time Traces". *Sci. Bull.* **62** (2017), 863–868.

[HMT11] N. Halko, P.-G. Martinsson, and J. A. Tropp. "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions". *SIAM Rev.* **53** (2011), 217–288.

[Hoa61] C. A. R. Hoare. "Algorithm 65: Find". *Commun. ACM* **4** (1961), 321–322.

[Hof76] D. R. Hofstadter. "Energy levels and wave functions of Bloch electrons in rational and irrational magnetic fields". *Phys. Rev. B* **14** (1976), 2239–2249.

[Hol+15] M. Holzäpfel, T. Baumgratz, M. Cramer, and M. B. Plenio. "Scalable reconstruction of unitary processes and Hamiltonians". *Phys. Rev. A* **91** (2015), 042129.

[Hom+09] J. P. Home, D. Hanneke, J. D. Jost, J. M. Amini, D. Leibfried, and D. J. Wineland. "Complete Methods Set for Scalable Ion Trap Quantum Information Processing". *Science* **325** (2009), 1227–1230.

[Hon+16] A. Honecker, S. Wessel, R. Kerkdyk, T. Pruschke, F. Mila, and B. Normand. “Thermodynamic properties of highly frustrated quantum spin ladders: Influence of many-particle bound states”. *Phys. Rev. B* **93** (2016), 054408. arXiv: 1511.01501 [cond-mat.str-el].

[Hra97] Z. Hradil. “Quantum-state estimation”. *Phys. Rev. A* **55** (1997), 1561–1564.

[HS92] N. Hatano and M. Suzuki. “Representation basis in quantum Monte Carlo calculations and the negative-sign problem”. *Phys. Lett. A* **163** (1992), 246–249.

[Hun19] N. Hunter-Jones. “Unitary designs from statistical mechanics in random quantum circuits”. 2019. arXiv: 1905.12053 [quant-ph].

[HWW18] J. Helsen, J. J. Wallman, and S. Wehner. “Representations of the multi-qubit Clifford group”. *J. Math. Phys.* **59** (2018), 072201. arXiv: 1609.08188 [quant-ph].

[HYF20] R. Harper, W. Yu, and S. T. Flammia. “Fast Estimation of Sparse Quantum Noise”. 2020. arXiv: 2007.07901 [quant-ph].

[Ite+20] R. Iten, T. Metger, H. Wilming, L. del Rio, and R. Renner. “Discovering Physical Concepts with Neural Networks”. *Phys. Rev. Lett.* **124** (2020), 010508. arXiv: 1807.10300 [quant-ph].

[Ivo81] I. D. Ivonovic. “Geometrical description of quantal state determination”. *J. Phys. A* **14** (1981), 3241–3245.

[Jak+98] D. Jaksch, C. Bruder, J. I. Cirac, C. W. Gardiner, and P. Zoller. “Cold bosonic atoms in optical lattices”. *Phys. Rev. Lett.* **81** (1998), 3108.

[Jam+01] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White. “Measurement of qubits”. *Phys. Rev. A* **64** (2001), 052312.

[Jam72] A. Jamiolkowski. “Linear transformations which preserve trace and positive semidefiniteness of operators”. *Rep. Math. Phys.* **3** (1972), 275–278.

[Kab+16] M. Kabanava, R. Kueng, H. Rauhut, and U. Terstiege. “Stable low-rank matrix recovery via null space properties”. *Information and Inference: A Journal of the IMA* **5** (2016), 405–441. arXiv: 1507.07184 [cs.IT].

[Kel+18] J. Kelly, P. O’Malley, M. Neeley, H. Neven, and J. M. Martinis. “Physical Qubit Calibration on a Directed Acyclic Graph”. 2018. arXiv: 1803.03226 [quant-ph].

[KG15] R. Kueng and D. Gross. “Qubit stabilizer states are complex projective 3-designs”. 2015. arXiv: 1510.02767 [quant-ph].

[Kim+14] S. Kimmel, M. P. da Silva, C. A. Ryan, B. R. Johnson, and T. Ohki. “Robust extraction of tomographic information via randomized benchmarking”. *Phys. Rev. X* **4** (2014), 011050. arXiv: 1306.2348 [quant-ph].

[KKD15] A. Kalev, R. L. Kosut, and I. H. Deutsch. “Quantum tomography protocols with positivity are compressed sensing protocols”. *npj Quantum Information* **1** (2015), 15018. eprint: arXiv:1502.00536.

[KL17] S. Kimmel and Y. K. Liu. “Phase retrieval using unitary 2-designs”. In: *2017 International Conference on Sampling Theory and Applications (SampTA)*. 2017, 345–349. arXiv: 1510.08887.

[KL18] F. Krahmer and Y. K. Liu. “Phase Retrieval Without Small-Ball Probability Assumptions”. *IEEE Trans. Inf. Th.* **64** (2018), 485–500. eprint: 1604.07281.

[Kla+20] J. Klassen, M. Marvian, S. Piddock, M. Ioannou, I. Hen, and B. Terhal. “Hardness and Ease of Curing the Sign Problem for Two-Local Qubit Hamiltonians”. *SIAM Journal on Computing* **49** (2020), 1332–1362. arXiv: 1906.08800 [quant-ph].

[Kli+19] M. Kliesch, R. Kueng, J. Eisert, and D. Gross. “Guaranteed recovery of quantum processes from few measurements”. *Quantum* **3** (2019), 171. arXiv: 1701.03135 [quant-ph].

[Kli19] M. Kliesch. *Lecture notes: Validation, certification and characterization of quantum systems*. http://www.mkliesch.eu/docs/lecture_QCVV.pdf. [accessed 08-August-2019]. 2019.

[KM15] V. Koltchinskii and S. Mendelson. “Bounding the smallest singular value of a random matrix without concentration”. *International Mathematics Research Notices* **2015** (2015), rnv096. arXiv: 1312.3580 [math.PR].

[Kni+08] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland. “Randomized benchmarking of quantum gates”. *Phys. Rev. A* **77** (2008), 012307. arXiv: 0707.0963 [quant-ph].

[Köh+05] M. Köhl, H. Moritz, T. Stöferle, K. Günter, and T. Esslinger. “Fermionic atoms in a three dimensional optical lattice: Observing Fermi surfaces, dynamics, and interactions”. *Phys. Rev. Lett.* **94** (2005), 080403.

[KR05] A. Klappenecker and M. Roetteler. “Mutually Unbiased Bases are Complex Projective 2-Designs”. In: *Proc. IEEE International Symposium on Information Theory, ISIT, 2005*. IEEE. 2005, 1740–1744. arXiv: quant-ph/0502031 [quant-ph].

[Kra+19] S. Krastanov, S. Zhou, S. T. Flammia, and L. Jiang. “Stochastic estimation of dynamical variables”. *Quantum Sci. Technol.* **4** (2019), 035003. arXiv: 1812.05120 [quant-ph].

[KRT15] R. Kueng, H. Rauhut, and U. Terstiege. “Low rank matrix recovery from rank one measurements”. *Appl. Comp. Harm. Anal.* (2015). arXiv: 1410.6913 [cs.IT].

[KT19] J. Klassen and B. M. Terhal. “Two-local qubit Hamiltonians: when are they stoquastic?” *Quantum* **3** (2019), 139. arXiv: 1806.05405 [quant-ph].

[Kue+16] R. Kueng, D. M. Long, A. C. Doherty, and S. T. Flammia. “Comparing Experiments to the Fault-Tolerance Threshold”. *Phys. Rev. Lett.* **117** (2016), 170502. arXiv: 1510.05653 [quant-ph].

[Kue15] R. Kueng. “Low rank matrix recovery from few orthonormal basis measurements”. In: *Sampling Theory and Applications (SampTA), 2015 International Conference on*. 2015, 402–406.

[KZG16a] R. Kueng, H. Zhu, and D. Gross. “Distinguishing quantum states using Clifford orbits”. 2016. arXiv: 1609.08595 [quant-ph].

[KZG16b] R. Kueng, H. Zhu, and D. Gross. “Low rank matrix recovery from Clifford orbits”. 2016. arXiv: 1610.08070 [cs.IT].

[Lan+09] B. P. Lanyon, M. Barbieri, M. P. Almeida, T. Jennewein, T. C. Ralph, K. J. Resch, G. J. Pryde, J. L. O’Brien, A. Gilchrist, and A. G. White. “Simplifying quantum logic using higher-dimensional Hilbert spaces”. *Nat. Phys.* **5** (2009), 134–140.

[Lap+12] E. H. Lapasar, K. Maruyama, D. Burgarth, T. Takui, Y. Kondo, and M. Nakahara. “Estimation of Coupling Constants of a Three-Spin Chain: Case Study of Hamiltonian Tomography with NMR”. *New J. Phys.* **14** (2012), 013043. arXiv: 1111.1381 [quant-ph].

[LB00] D. P. Landau and K. Binder. *A guide to Monte Carlo simulations in statistical physics*. Cambridge University Press, 2000.

[Lév+07] B. Lévi, C. C. López, J. Emerson, and D. G. Cory. “Efficient error characterization in quantum information processing”. *Phys. Rev. A* **75** (2007), 022314. arXiv: quant-ph/0608246 [quant-ph].

[Li+11] H.-W. Li, Z.-Q. Yin, Y.-C. Wu, X.-B. Zou, S. Wang, W. Chen, G.-C. Guo, and Z.-F. Han. “Semi-device-independent random-number expansion without entanglement”. *Phys. Rev. A* **84** (2011), 034301.

[Li+12] H.-W. Li, M. Pawłowski, Z.-Q. Yin, G.-C. Guo, and Z.-F. Han. “Semi-device-independent randomness certification using $n \to 1$ quantum random access codes”. *Phys. Rev. A* **85** (2012), 052308.

[Lin+09] N. Linden, S. Popescu, A. Short, and A. Winter. “Quantum mechanical evolution towards thermal equilibrium”. *Phys. Rev. E* **79** (2009), 061103.

[Liu11] Y.-K. Liu. “Universal low-rank matrix recovery from Pauli measurements”. *Adv. Neural Inf. Process. Syst.* **24** (2011), 1638–1646. eprint: 1103.2816.

[LJY15] Z.-X. Li, Y.-F. Jiang, and H. Yao. “Solving the fermion sign problem in quantum Monte Carlo simulations by Majorana representation”. *Phys. Rev. B* **91** (2015). arXiv: 1408.2269 [cond-mat.str-el].

[LJY16] Z.-X. Li, Y.-F. Jiang, and H. Yao. “Majorana-Time-Reversal Symmetries: A Fundamental Principle for Sign-Problem-Free Quantum Monte Carlo Simulations”. *Phys. Rev. Lett.* **117** (2016), 267002. arXiv: 1601.05780 [cond-mat.str-el].

[LKF21] I. A. Luchnikov, M. Krechetov, and S. Filippov. “Riemannian geometry and automatic differentiation for optimization problems of quantum physics and quantum technologies”. *New J. Phys.* **23** (2021), 073006. arXiv: 2007.01287 [quant-ph].

[LLF20] W. Li, W. Liao, and A. Fannjiang. “Super-resolution limit of the ESPRIT algorithm”. *IEEE Trans. Inf. Th.* **66** (2020), 4593–4608. arXiv: 1905.03782 [cs.IT].

[LM19] G. Lugosi and S. Mendelson. “Mean estimation and regression under heavy-tailed distributions–a survey”. *Found Comput. Math.* **19** (2019), 1145–1190. arXiv: 1906.04280 [math.ST].

[Lob+08] M. Lobino, D. Korystov, C. Kupchak, E. Figueroa, B. C. Sanders, and A. I. Lvovsky. “Complete Characterization of Quantum-Optical Processes”. *Science* **322** (2008), 563–566.

[Luc+21] I. A. Luchnikov, A. Ryzhov, S. N. Filippov, and H. Ouerdane. “QGOpt: Riemannian optimization for quantum technologies”. *SciPost Phys.* **10** (2021), 79. arXiv: 2011.01894 [quant-ph].

[LVB11] Y.-C. Liang, T. Vértesi, and N. Brunner. “Semi-device-independent bounds on entanglement”. *Phys. Rev. A* **83** (2011), 022108.

[LZH20] Z. Li, L. Zou, and T. H. Hsieh. “Hamiltonian Tomography via Quantum Quench”. *Phys. Rev. Lett.* **124** (2020), 160502.

[MA15] J. R. McClean and A. Aspuru-Guzik. “Clock Quantum Monte Carlo: an imaginary-time method for real-time quantum dynamics”. *Phys. Rev. A* **91** (2015), 012311. arXiv: 1410.1877 [quant-ph].

[Mag+12] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen. “Efficient Measurement of Quantum Gate Error by Interleaved Randomized Benchmarking”. *Phys. Rev. Lett.* **109** (2012), 080505. arXiv: 1203.4550 [quant-ph].

[Mag17] M. Magdon-Ismail. “NP-hardness and inapproximability of sparse PCA”. en. *Inf. Process. Lett.* **126** (2017), 35–38.

[Man+17] N. M. Mangan, J. N. Kutz, S. L. Brunton, and J. L. Proctor. “Model selection for dynamical systems via sparse regression and information criteria”. *Proc. Roy. Soc. A* **473** (2017), 20170009.

[MB17] R. L. Mann and M. J. Bremner. “On the Complexity of Random Quantum Computations and the Jones Polynomial”. 2017. arXiv: 1711.00686 [quant-ph].

[MBE11] E. Magesan, R. Blume-Kohout, and J. Emerson. “Gate fidelity fluctuations and quantum process invariants”. *Phys. Rev. A* **84** (2011), 012309.

[McC+16] W. McCutcheon, A. Pappa, B. A. Bell, A. McMillan, A. Chailloux, T. Lawson, M. Mafu, D. Markham, E. Diamanti, I. Kerenidis, J. G. Rarity, and M. S. Tame. “Experimental verification of multipartite entanglement in quantum networks”. *Nat. Commun.* **7** (2016), 13251–13251.

[McK+19] D. C. McKay, S. Sheldon, J. A. Smolin, J. M. Chow, and J. M. Gambetta. “Three Qubit Randomized Benchmarking”. *Phys. Rev. Lett.* **122** (2019), 200502. arXiv: 1712.06550 [quant-ph].

[Md16] A. Montanaro and R. de Wolf. “A Survey of Quantum Property Testing”. *Theory of Computing*. Graduate Surveys **7** (2016), 1–81. arXiv: 1310.2035 [quant-ph].

[Men15] S. Mendelson. “Learning without concentration”. *J. ACM* **62** (2015), 21:1–21:25. arXiv: 1401.0304 [cs.LG].

[Mer+13] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen. "Self-consistent quantum process tomography". *Phys. Rev. A* **87** (2013), 062119. arXiv: 1211.0322 [quant-ph].

[MGE11] E. Magesan, J. M. Gambetta, and J. Emerson. "Scalable and Robust Randomized Benchmarking of Quantum Processes". *Phys. Rev. Lett.* **106** (2011), 180504. arXiv: 1009.3639 [quant-ph].

[MGE12] E. Magesan, J. M. Gambetta, and J. Emerson. "Characterizing quantum gates via randomized benchmarking". *Phys. Rev. A* **85** (2012), 042311. arXiv: 1109.6887 [quant-ph].

[MK04] H.-J. Mikeska and A. K. Kolezhuk. "One-dimensional magnetism". In: *Quantum Magnetism.* Ed. by U. Schollwöck, J. Richter, D. J. J. Farnell, and R. F. Bishop. Lecture Notes in Physics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, 1–83.

[MK13] C. Monroe and J. Kim. "Scaling the Ion Trap Quantum Processor". *Science* **339** (2013), 1164–1169.

[MLH19] M. Marvian, D. A. Lidar, and I. Hen. "On the computational complexity of curing non-stoquastic Hamiltonians". *Nat. Commun.* **10** (2019), 1571. arXiv: 1802.03408 [quant-ph].

[Mog10] D. Mogilevtsev. "Calibration of single-photon detectors using quantum statistics". *Phys. Rev. A* **82** (2010), 021807.

[Mor17] T. Morimae. "Hardness of classically sampling the one-clean-qubit model with constant total variation distance error". *Phys. Rev. A* **96** (2017), 040302. arXiv: 1704.03640 [quant-ph].

[Mot+14] L. Motka, B. Stoklasa, J. Rehacek, Z. Hradil, V. Karasek, D. Mogilevtsev, G. Harder, C. Silberhorn, and L. L. Sánchez-Soto. "Efficient algorithm for optimizing data-pattern tomography". *Phys. Rev. A* **89** (2014), 054102.

[MPA11] L. Masanes, S. Pironio, and A. Acin. "Secure device-independent quantum key distribution with causally independent measurement devices". *Nat. Commun.* **2** (2011), 1–7.

[MPF21] S. T. Merkel, E. J. Pritchett, and B. H. Fong. "Randomized Benchmarking as Convolution: Fourier Analysis of Gate Dependent Errors". *Quantum* **5** (2021), 581. arXiv: 1804.05951 [quant-ph].

[MRA13] L. Masanes, A. J. Roncaglia, and A. Acín. "Complexity of energy eigenstates as a mechanism for equilibration". *Phys. Rev. E* **87** (2013), 032137.

[MŘH09] D. Mogilevtsev, J. Řeháček, and Z. Hradil. “Relative tomography of an unknown quantum state”. *Phys. Rev. A* **79** (2009), 020101.

[MŘH12] D. Mogilevtsev, J. Řeháček, and Z. Hradil. “Self-calibration for self-consistent tomography”. *New J. Phys.* **14** (2012), 095001.

[MT19] T. Morimae and S. Tamaki. “Fine-grained quantum computational supremacy”. *Quantum Inf. Comput.* **19** (2019), 1089. arXiv: 1901.01637 [quant-ph].

[MT20] T. Morimae and S. Tamaki. “Additive-error fine-grained quantum supremacy”. *Quantum* **4** (2020), 329. arXiv: 1912.06336 [quant-ph].

[MW09] C. B. Mendl and M. M. Wolf. “Unital quantum channels - convex structure and revivals of Birkhoff’s theorem”. *Comm. Math. Phys.* **289** (2009), 1057–1086. arXiv: 0806.2820 [quant-ph].

[Nac96] B. Nachtergaele. “The spectral gap for some spin chains with discrete symmetry breaking”. *Comm. Math. Phys.* **175** (1996), 565–606.

[Nak98] T. Nakamura. “Vanishing of the negative-sign problem of quantum Monte Carlo simulations in one-dimensional frustrated spin systems”. *Phys. Rev. B* **57** (1998), R3197. arXiv: cond-mat/9707019.

[NC10] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information.* Cambridge University Press, 2010.

[Nec+18] I. Nechita, Z. Puchala, L. Pawela, and K. Zyczkowski. “Almost all quantum channels are equidistant”. *J. Math. Phys.* **59** (2018), 052201. arXiv: 1612.00401 [quant-ph].

[Nie02] M. A. Nielsen. “A simple formula for the average gate fidelity of a quantum dynamical operation”. *Phys. Lett. A* **303** (2002), 249–252. eprint: quant-ph/0205035.

[NPT99] Y. Nakamura, Y. A. Pashkin, and J. S. Tsai. “Coherent control of macroscopic quantum states in a single-Cooper-pair box”. *Nature* **398** (1999), 786–788.

[NT08] D. Needell and J. A. Tropp. “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples”. *Appl. Comp. Harm. An.* (2008).

[NVH18] A. Nahum, S. Vijay, and J. Haah. “Operator Spreading in Random Unitary Circuits”. *Phys. Rev. X* **8** (2018), 021014.

[OH14] K. Okunishi and K. Harada. “Symmetry-protected topological order and negative-sign problem for SO(N) bilinear-biquadratic chains”. *Phys. Rev. B* **89** (2014), 134422. arXiv: 1312.2643 [cond-mat.stat-mech].

[OKC21] E. Onorati, T. Kohler, and T. Cubitt. “Fitting quantum noise models to tomography data”. 2021. arXiv: 2103.17243 [quant-ph].

[OS12] D. K. L. Oi and S. G. Schirmer. “Quantum system characterization with limited resources”. *Phil. Trans. R. Soc. A* **370** (2012), 5386–5395.

[OWE19] E. Onorati, A. H. Werner, and J. Eisert. “Randomized Benchmarking for Individual Quantum Gates”. *Phys. Rev. Lett.* **123** (2019), 060501. arXiv: 1811.11775 [quant-ph].

[Oym+15] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi. “Simultaneously Structured Models With Application to Sparse and Low-Rank Matrices”. *IEEE Trans. Inf. Theory* **61** (2015), 2886–2908.

[PB11] M. Pawłowski and N. Brunner. “Semi-device-independent security of one-way quantum key distribution”. *Phys. Rev. A* **84** (2011), 010302.

[Per+07] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac. “Matrix Product State Representations”. *Quantum Inf. Comput.* **7** (2007), 401–430.

[Pir+09] S. Pironio, A. Acin, N. Brunner, N. Gisin, S. Massar, and V. Scarani. “Device-independent quantum key distribution secure against collective attacks”. *New J. Phys.* **11** (2009), 045021.

[PMH08] K. N. Patel, I. L. Markov, and J. P. Hayes. “Efficient Synthesis of Linear Reversible Circuits”. *Quantum Inf. Comput.* **8** (2008), 282–294.

[Pol12] L. Pollet. “Recent developments in Quantum Monte-Carlo simulations with applications for cold gases”. *Rep. Prog. Phys.* **75** (2012), 094501. arXiv: 1206.0781 [cond-mat.quant-gas].

[Pre13] J. Preskill. “Quantum computing and the entanglement frontier”. *Bull. Am. Phys. Soc.* **58** (2013). arXiv: 1203.5813 [quant-ph].

[Pre18] J. Preskill. “Quantum Computing in the NISQ era and beyond”. *Quantum* **2** (2018). arXiv: 1801.00862 [quant-ph].

[Pro+17] T. Proctor, K. Rudinger, K. Young, M. Sarovar, and R. Blume-Kohout. "What Randomized Benchmarking Actually Measures". *Phys. Rev. Lett.* **119** (2017), 130502. arXiv: 1702.01853 [quant-ph].

[Pro+19] T. J. Proctor, A. Carignan-Dugas, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young. "Direct Randomized Benchmarking for Multiqubit Devices". *Phys. Rev. Lett.* **123** (2019), 030503. arXiv: 1807.07975 [quant-ph].

[Pry+03] G. J. Pryde, T. C. Ralph, D. Branning, A. G. White, and J. L. O'Brien. "Demonstration of an all-optical quantum controlled-NOT gate". *Nature* **426** (2003), 264–267.

[QR19] X.-L. Qi and D. Ranard. "Determining a Local Hamiltonian from a Single Eigenstate". *Quantum* **3** (2019), 159.

[Rei08] P. Reimann. "Foundation of Statistical Mechanics under Experimentally Realistic Conditions". *Phys. Rev. Lett.* **101** (2008), 190403.

[Rei12] P. Reimann. "Equilibration of isolated macroscopic quantum systems under experimentally realistic conditions". *Phys. Scr.* **86** (2012), 058512.

[Ren+04] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves. "Symmetric informationally complete quantum measurements". *J. Math. Phys.* **45** (2004), 2171–2180. eprint: quant-ph/0310075.

[Rio+17] C. A. Riofrio, D. Gross, S. T. Flammia, T. Monz, D. Nigg, R. Blatt, and J. Eisert. "Experimental quantum compressed sensing for a seven-qubit system". *Nat. Commun.* **8** (2017), 15305.

[RK12] P. Reimann and M. Kastner. "Equilibration of isolated macroscopic quantum systems". *New J. Phys.* **14** (2012), 043020.

[ŘMH10] J. Řeháček, D. Mogilevtsev, and Z. Hradil. "Operational tomography: Fitting of data patterns". *Phys. Rev. Lett.* **105** (2010), 010402.

[Rot+22] I. Roth, D. Hangleiter, P. Roushan, and J. Eisert. *Robustly learning non-interacting Hamiltonians from dynamical data.* forthcoming. 2022.

[Rou+17] P. Roushan, C. Neill, J. Tangpanitanon, V. M. Bastidas, A. Megrant, R. Barends, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. Fowler, B. Foxen, M. Giustina, E. Jeffrey, J. Kelly, E. Lucero, J. Mutus, M. Neeley, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. White, H. Neven, D. G. Angelakis, and J. Martinis. "Spectroscopic signatures of localization with interacting photons in superconducting qubits". *Science* **358** (2017), 1175–1179.

[RPK86] R. Roy, A. Paulraj, and T. Kailath. “Estimation of signal parameters via rotational invariance techniques-ESPRIT”. In: *MILCOM 1986-IEEE Military Communications Conference: Communications-Computers: Teamed for the 90's.* Vol. 3. IEEE. 1986, 41–6.

[RS07] A. Roy and A. J. Scott. “Weighted complex projective 2-designs from bases: Optimal state determination by orthogonal measurements”. *J. Math. Phys.* **48** (2007), 072110–072110. arXiv: quant-ph/0703025 [quant-ph].

[RW20] A. Rolandi and H. Wilming. “Extensive Rényi entropies in matrix product states”. 2020. arXiv: 2008.11764 [quant-ph].

[Sah+18] M. B. Sahinoglu, S. K. Shukla, F. Bi, and X. Chen. “Matrix product representation of locality preserving unitaries”. *Phys. Rev. B* **98** (2018), 245122. arXiv: 1704.01943 [quant-ph].

[Sam+21] G. O. Samach, A. Greene, J. Borregaard, M. Christandl, D. K. Kim, C. M. McNally, A. Melville, B. M. Niedzielski, Y. Sung, D. Rosenberg, M. E. Schwartz, J. L. Yoder, T. P. Orlando, J. I.-J. Wang, S. Gustavsson, M. Kjaergaard, and W. D. Oliver. “Lindblad tomography of a superconducting quantum processor”. 2021. arXiv: 2105.02338 [quant-ph].

[SB20] I. Supic and J. Bowles. “Self-testing of quantum systems: a review”. *Quantum* **4** (2020), 337. arXiv: 1904.10042 [quant-ph].

[SC17] A. Sone and P. Cappellaro. “Exact Dimension Estimation of Interacting Qubit Systems Assisted by a Single Quantum Probe”. *Phys. Rev. A* **96** (2017), 062334.

[Sch+12] U. Schneider, L. Hackermüller, J. Ronzheimer, S. Will, S. Braun, T. Best, I. Bloch, E. Demler, S. Mandt, D. Rasch, and A. Rosch. “Fermionic transport and out-of-equilibrium dynamics in a homogeneous Hubbard model with ultracold atoms”. *Nat. Phys.* **8** (2012), 213–218.

[Sch+13] P. Schindler, M. Müller, D. Nigg, J. T. Barreiro, E. A. Martinez, M. Hennrich, T. Monz, S. Diehl, P. Zoller, and R. Blatt. “Quantum simulation of dynamical maps with trapped ions”. *Nat. Phys.* **9** (2013), 361–367.

[Sch+15] M. Schreiber, S. S. Hodgman, P. Bordia, H. P. Lüschen, M. H. Fischer, R. Vosk, E. Altman, U. Schneider, and I. Bloch. “Observation of many-body localization of interacting fermions in a quasirandom optical lattice”. *Science* **349** (2015), 842–845.

[Sco06] A. J. Scott. “Tight informationally complete quantum measurements”. *J. Phys. A Math. Gen.* **39** (2006), 13507–13530. eprint: quant-ph/0604049.

[Sco08] A. J. Scott. “Optimizing quantum process tomography with unitary 2-designs”. *J. Phys. A* **41** (2008), 055308.

[SF12] A. J. Short and T. C. Farrelly. “Quantum equilibration in finite time”. *New J. Phys.* **14** (2012), 013063.

[Sha+11] A. Shabani, R. L. Kosut, M. Mohseni, H. Rabitz, M. A. Broome, M. P. Almeida, A. Fedrizzi, and A. G. White. “Efficient measurement of quantum dynamics via compressive sensing”. *Phys. Rev. Lett.* **106** (2011), 100401.

[She+16] S. Sheldon, L. S. Bishop, E. Magesan, S. Filipp, J. M. Chow, and J. M. Gambetta. “Characterizing errors on qubit operations via iterative randomized benchmarking”. *Phys. Rev. A* **93** (2016), 012301. arXiv: 1504.06597 [quant-ph].

[Sho11] A. J. Short. “Equilibration of quantum systems and subsystems”. *New J. Phys.* **13** (2011), 053009.

[Sim+13] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani. “A sparse-group Lasso”. *J. Comp. Graph. Stat.* **22** (2013), 231–245.

[Sim+19] J. Y. Sim, J. Shang, H. K. Ng, and B.-G. Englert. “Proper error bars for self-calibrating quantum tomography”. *Phys. Rev. A* **100** (2019), 022333. arXiv: 1904.11202 [quant-ph].

[Sim96] B. Simon. *Representations of finite and compact groups.* Am. Math. Soc., 1996.

[SK17] A. Sawicki and K. Karnas. “Universality of Single-Qudit Gates”. *Ann. Henri Poincaré* **18** (2017), 3515–3552. arXiv: 1609.05780 [quant-ph].

[SKO04] S. G. Schirmer, A. Kolli, and D. K. L. Oi. “Experimental Hamiltonian Identification for Controlled Two-Level Systems”. *Phys. Rev. A* **69** (2004), 050306.

[SL09] M. Schmidt and H. Lipson. “Distilling free-form natural laws from experimental data”. *Science* **324** (2009), 81–85.

[SLP11] M. P. da Silva, O. Landon-Cardinal, and D. Poulin. “Practical characterization of quantum devices without tomography”. *Phys. Rev. Lett.* **107** (2011), 210404. arXiv: 1104.3835 [quant-ph].

[SO09] S. G. Schirmer and D. K. L. Oi. “Two-Qubit Hamiltonian Tomography by Bayesian Analysis of Noisy Data”. *Phys. Rev. A* **80** (2009), 022333.

[SOD08] S. G. Schirmer, D. K. L. Oi, and S. J. Devitt. “Physics-Based Mathematical Models for Quantum Devices via Experimental System Identification”. *J. Phys.: Conf. Ser.* **107** (2008), 012011.

[SPC11] N. Schuch, D. Perez-Garcia, and I. Cirac. “Classifying quantum phases using matrix product states and projected entangled pair states”. *Phys. Rev. B* **84** (2011), 165139.

[Spr+10] P. Sprechmann, I. Ramirez, G. Sapiro, and Y. Eldar. “Collaborative hierarchical sparse modeling”. In: *2010 44th Annual Conference on Information Sciences and Systems (CISS).* 2010, 1–6.

[Spr+11] P. Sprechmann, I. Ramirez, G. Sapiro, and Y. C. Eldar. “C-HiLasso: A collaborative hierarchical sparse modeling framework”. *IEEE Trans. Sig. Proc.* **59** (2011), 4183–4198.

[SSS21] D. Stilck França, S. Strelchuk, and M. Studziński. “Efficient Classical Simulation and Benchmarking of Quantum Processes in the Weyl Basis”. *Phys. Rev. Lett.* **126** (2021), 210502. arXiv: 2008.12250 [quant-ph].

[Sta12] C. Stark. “Simultaneous Estimation of Dimension, States and Measurements: Computation of representative density matrices and POVMs” (2012). arXiv: 1210.1105 [quant-ph].

[Sta14] C. Stark. “Self-consistent tomography of the state-measurement Gram matrix”. *Phys. Rev. A* **89** (2014), 052109. arXiv: 1209.5737 [quant-ph].

[Ste+06] M. Steffen, M. Ansmann, R. C. Bialczak, N. Katz, E. Lucero, R. McDermott, M. Neeley, E. M. Weig, A. N. Cleland, and J. M. Martinist. “Measurement of the Entanglement of Two Superconducting Qubits via State Tomography”. *Science* **313** (2006), 1423–1425.

[SW19] T. Strohmer and K. Wei. “Painless breakups-efficient demixing of low rank matrices”. *J. Four. Ana. App.* **25** (2019), 1–31.

[SWS16] Y. R. Sanders, J. J. Wallman, and B. C. Sanders. “Bounding quantum gate error rate based on reported average fidelity”. *New J. Phys.* **18** (2016), 012002. arXiv: 1501.04932 [quant-ph].

[Tak96] M. Takano. “Spin ladder compounds”. *Physica C: Superconductivity.* Proceedings of the International Symposium on Frontiers of High - Tc Superconductivity **263** (1996), 468–474.

[Tas98] H. Tasaki. "From quantum dynamics to the canonical distribution: general picture and a rigorous example". *Phys. Rev. Lett.* **80** (1998).

[TD04] B. M. Terhal and D. P. DiVincenzo. "Adaptive Quantum Computation, Constant Depth Quantum Circuits and Arthur-Merlin Games". *Quant. Inf. Comp.* **4** (2004), 134–145. arXiv: quant-ph/0205133.

[Tho+15] R. E. Thomas, Q. Sun, A. Alavi, and G. H. Booth. "Stochastic Multiconfigurational Self-Consistent Field Theory". *J. Chem. Theory Comput.* **11** (2015), 5316–5325. arXiv: 1510.03635 [physics.chem-ph].

[Tro+03] M. Troyer, F. Alet, S. Trebst, and S. Wessel. "Non?local Updates for Quantum Monte Carlo Simulations". *AIP Conf. Proc.* **690** (2003), 156–169. arXiv: physics/0306128.

[Tro+10] S. Trotzky, L. Pollet, F. Gerbier, U. Schnorrberger, I. Bloch, N. Prokof'ev, B. Svistunov, and M. Troyer. "Suppression of the critical temperature for superfluidity near the Mott transition: validating a quantum simulator". *Nat. Phys.* **6** (2010), 998.

[Tro+12] S. Trotzky, Y.-A. Chen, A. Flesch, L. P. McCulloch, U. Schollwöck, J. Eisert, and I. Bloch. "Probing the relaxation towards equilibrium in an isolated strongly correlated one-dimensional Bose gas". *Nat. Phys.* **8** (2012), 325–330.

[Tro12] J. A. Tropp. "User friendly tools for random matrices. An introduction." *Preprint* (2012).

[Tro15] J. A. Tropp. "Convex recovery of a structured signal from independent random linear measurements". In: *Sampling Theory, a Renaissance.* Ed. by E. G. Pfander. Springer, 2015, 67–101. arXiv: 1405.1102.

[Val+19] A. Valenti, E. van Nieuwenburg, S. Huber, and E. Greplova. "Hamiltonian Learning for Quantum Error Correction". *Phys. Rev. Research* **1** (2019), 033092. arXiv: 1907.02540 [quant-ph].

[Van13] B. Vandereycken. "Low-rank matrix completion by Riemannian optimization". *SIAM J. Opt.* **23** (2013), 1214–1236.

[Var13] P. P. Varjú. "Random walks in compact groups". *Doc. Math.* **18** (2013), 1137–1175. arXiv: 1209.1745 [math.GR].

[Vei+14] V. Veitch, S. H. Mousavian, D. Gottesman, and J. Emerson. "The resource theory of stabilizer quantum computation". *New J. Phys.* **16** (2014), 013009. arXiv: 1307.7171 [quant-ph].

[Ver12] R. Vershynin. “Introduction to the non-asymptotic analysis of random matrices”. In: *Compressed Sensing: Theory and Applications.* Cambridge University Press, 2012, 210–268. arXiv: 1011.3027 [math.PR].

[Wal+15] J. J. Wallman, C. Granade, R. Harper, and S. T. Flammia. “Estimating the coherence of noise”. *New J. Phys.* **17** (2015), 113020. arXiv: 1503.07865 [quant-ph].

[Wal15] J. J. Wallman. “Bounding experimental quantum error rates relative to fault-tolerant thresholds”. 2015. arXiv: 1511.00727 [quant-ph].

[Wal18] J. J. Wallman. “Randomized benchmarking with gate-dependent noise”. *Quantum* **2** (2018), 47. arXiv: 1703.09835 [quant-ph].

[Wat11] J. Watrous. *Theory of Quantum Information.* lecture notes. 2011.

[Wat18] J. Watrous. *The Theory of Quantum Information.* Cambridge University Press, 2018.

[WBE15] J. J. Wallman, M. Barnhill, and J. Emerson. “Robust Characterization of Loss Rates”. *Phys. Rev. Lett.* **115** (2015), 060501. arXiv: 1510.01272 [quant-ph].

[WBE16] J. J. Wallman, M. Barnhill, and J. Emerson. “Robust characterization of leakage errors”. *New J. Phys.* **18** (2016), 043021. arXiv: 1412.4126 [quant-ph].

[WDD15] S.-T. Wang, D.-L. Deng, and L.-M. Duan. “Hamiltonian Tomography for Quantum Many-Body Systems with Arbitrary Couplings”. *New J. Phys.* **17** (2015), 093017.

[Web16] Z. Webb. “The Clifford Group Forms a Unitary 3-design”. *Quantum Info. Comput.* **16** (2016), 1379–1400. arXiv: 1510.02769 [quant-ph].

[Wei+16] K. Wei, J.-F. Cai, T. F. Chan, and S. Leung. “Guarantees of Riemannian optimization for low rank matrix recovery”. *SIAM J. Mat. An. App.* **37** (2016), 1198–1222.

[Wei78] D. Weingarten. “Asymptotic behavior of group integrals in the limit of infinite rank”. *J. Math. Phys.* **19** (1978), 999–1001.

[Wes+17] S. Wessel, B. Normand, F. Mila, and A. Honecker. “Efficient Quantum Monte Carlo simulations of highly frustrated magnets: the frustrated spin-1/2 ladder”. *SciPost Physics* **3** (2017), 005.

[WF14] J. J. Wallman and S. T. Flammia. “Randomized benchmarking with confidence”. *New J. Phys.* **16** (2014), 103032. arXiv: 1404.6025 [quant-ph].

[WF89] W. K. Wootters and B. D. Fields. “Optimal state-determination by mutually unbiased measurements”. *Ann. Phys.* **191** (1989), 363–381.

[WFH18] J. Wallman, S. Flammia, and I. Hincks. *Quantum Characterization, Verification, and Validation.* Oxford Research Encyclopedia of Physics, doi: 10.1093/acrefore/9780190871994.013.38 (accessed Feb 02, 2019). 2018.

[WHZ03] C. Wu, J.-P. Hu, and S.-C. Zhang. “Exact SO(5) Symmetry in the Spin- 3 / 2 Fermionic System”. *Phys. Rev. Lett.* **91** (2003). arXiv: cond-mat/0302165.

[Wig59] E. P. Wigner. *Group theory and its application to the quantum mechanics of atomic spectra.* London: Academic Press, 1959.

[WM10] M. Wieśniak and M. Markiewicz. “Finding Traps in Nonlinear Spin Arrays”. *Phys. Rev. A* **81** (2010), 032340.

[Wun+15] G. Wunder, H. Boche, T. Strohmer, and P. Jung. “Sparse signal processing concepts for efficient 5G system design”. *IEEE Access* **3** (2015), 195–208.

[Zau99] G. Zauner. “Quantendesigns: Grundzüge einer nichtkommutativen Designtheorie”. English translation in International Journal of Quantum Information (IJQI) 9 (1) ,445–507, 2011. PhD thesis. University of Vienna, 1999.

[Zha+17] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe. “Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator”. *Nature* **551** (2017), 601–604. arXiv: 1708.01044 [quant-ph].

[Zha+21] D. Zhao, C. Wei, S. Xue, Y. Huang, X. Nie, J. Li, D. Ruan, D. Lu, T. Xin, and G. Long. “Characterizing Quantum Simulations with Quantum Tomography on a Spin Quantum Simulator”. *Phys. Rev. A* **103** (2021), 052403.

[Zho+19] S. Zhou, Z.-C. Yang, A. Hamma, and C. Chamon. “Single T gate in a Clifford circuit drives transition to universal entanglement spectrum statistics”. *SciPost Physics* (2019). arXiv: 1906.01079 [cond-mat.stat-mech].

[Zhu+16] H. Zhu, R. Kueng, M. Grassl, and D. Gross. “The Clifford group fails gracefully to be a unitary 4-design”. 2016. arXiv: 1609.08172 [quant-ph].

[Zhu17] H. Zhu. “Multiqubit Clifford groups are unitary 3-designs”. *Phys. Rev. A* **96** (2017), 062336. arXiv: 1510.02619 [quant-ph].

[ZS14] J. Zhang and M. Sarovar. “Quantum Hamiltonian Identification from Measurement Time Traces”. *Phys. Rev. Lett.* **113** (2014), 080401. arXiv: 1401.5780 [quant-ph].

Selbständigkeitserklärung

Ich, Ingo Roth, erkläre gegenüber der Freien Universität Berlin, dass ich die vorliegende Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe. Die vorliegende Arbeit ist frei von Plagiaten. Alle Ausführungen, die wörtlich oder inhaltlich aus anderen Schriften entnommen sind, habe ich als solche kenntlich gemacht. Diese Dissertation wurde in gleicher oder ähnlicher Form noch in keinem früheren Promotionsverfahren eingereicht. Mit einer Prüfung meiner Arbeit durch ein Plagiatsprüfungsprogramm erkläre ich mich einverstanden.