Quantum information in the real world

Diagnosing and correcting errors in practical quantum devices

Helsen, Jonas

DOI
10.4233/uuid:312b719d-32bc-4219-82bb-8e6febc2abcc

Publication date
2019

QUANTUM COMPUTING IN THE REAL WORLD

DIAGNOSING AND CORRECTING ERRORS IN PRACTICAL QUANTUM DEVICES

Quantum computing in the real world
Diagnosing and correcting errors in practical quantum devices

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op 5 juni 2019 om 12:30 uur

door

Jonas Helsen

Master of science in Physics,
KU Leuven, University of Leuven, Leuven, Belgium
geboren te Geel, België.

Dit proefschrift is goedgekeurd door de

promotor: prof. dr. S. D. C. Wehner

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. S. D. C. Wehner	Technische Universiteit Delft
Onafhankelijke leden:	
Prof. dr. Ir. R. Hanson	Technische Universiteit Delft
Prof. dr. D. Gross	University of Cologne
Prof. dr. A. van Deursen	Technische Universiteit Delft
Dr. M. Walter	Universiteit van Amsterdam
Overige leden:	
Prof. dr. ir. L. M. K. Vandersypen	Technische Universiteit Delft
Prof. dr. B. M. Terhal	Technische Universiteit Delft

Prof. dr. ir. Vandersypen heeft in belangrijke mate bijgedragen aan het totstaankomen van dit proefschrift.

Keywords: quantum computing, randomized benchmarking, representation theory, quantum error correction, quantum medium-scale integration

Printed by: Gildeprint - www.gildeprint.nl

Front & Back: kabinet.studio

ISBN 978-94-6384-042-2

An electronic version of this dissertation is available at
`http://repository.tudelft.nl/`.

Curriculum Vitæ

Jonas Helsen

11-02-1992 Born in Geel, Belgium.

Education

2004-2010 High School
St. Lambertus, Westerlo, Belgium

2010-2013 Undergraduate in Physics (Minor in Mathematics)
KU Leuven University of Leuven, Belgium

2012-2013 Erasmus exchange placement
University of Goteborg, Sweden

2013-2015 Masters in Theoretical Physics
KU Leuven University of Leuven, Belgium

2015-2019 PhD in Quantum Information
Technical University of Delft, Netherlands
Thesis: Quantum computing in the real world
Promotor: Prof. dr. S.D.C. Wehner

List of Publications

11. **Jonas Helsen**, Francesco Battistel, Barbara M. Terhal, "Spectral Quantum Tomography", *arXiv preprint* , arXiv:1904.00177 (2019).

10. Xiao Xue, Tom F. Watson, **Jonas Helsen**, Daniel R. Ward, Donald E. Savage, Max G. Lagally, Susan N. Coppersmith, Mark A. Eriksson, Stephanie Wehner, Lieven M.K. Vandersypen, "Benchmarking Gate Fidelities in a Si/SiGe Two-Qubit Device", *Physical Review X* **9**, 021011 (2019).

9. Le Phuc Thinh, Philippe Faist, **Jonas Helsen**, David Elkouss, Stephanie Wehner, "Practical and reliable error bars for quantum process tomography", *Physical Review A* **99**, 052311 (2019).

8. Bas Dirkse, **Jonas Helsen**, Stephanie Wehner, "Efficient Unitarity Randomized Benchmarking of Few-qubit Clifford Gates", *Physical Review A* **99**, 012315 (2019).

7. **Jonas Helsen**, Xiao Xue, Lieven M.K. Vandersypen, Stephanie Wehner, "A new class of efficient randomized benchmarking protocols", *arXiv preprint* , arXiv:1806.02048 (2018).

6. Axel Dahlberg, **Jonas Helsen**, Stephanie Wehner, "How to transform graph states using single-qubit operations: computational complexity and algorithms", *arXiv preprint* , arXiv:1805.05306 (2018).

5. **Jonas Helsen**, Mark Steudtner, Menno Veldhorst, Stephanie Wehner, "Quantum error correction in crossbar architectures", *Quantum Science and Technology* **3**, 3 (2018).

4. Ruoyu Li, Luca Petit, David P. Franke, Juan Pablo Dehollain, **Jonas Helsen**, Mark Steudtner, Nicole K. Thomas, Zachary R. Yoscovits, Kanwal J. Singh, Stephanie Wehner, Lieven M.K. Vandersypen, James S. Clarke, Menno Veldhorst, "A crossbar network for silicon quantum dot qubits", *Science advances* **4**, 7 eear3960 (2018).

3. **Jonas Helsen**, Joel J. Wallman, Steven T. Flammia, Stephanie Wehner, "Multi-qubit randomized benchmarking using few samples", *arXiv preprint* , arXiv:1701.04299 (2017).

2. **Jonas Helsen**, Joel J. Wallman, Stephanie Wehner, "Representations of the multi-qubit Clifford group", *Journal of Mathematical Physics* **59**, 7 072201 (2018).

1. Jeremy Ribeiro, Le Phuc Thinh, Jedrek Kaniewski, **Jonas Helsen**, Stephanie Wehner, "Device independence for two-party cryptography and position verification with memoryless devices", *Physical Review A* **97**, (6) 062307 (2018).

Summary

Quantum computers promise to be a revolutionary new technology. However, in order to realise this promise many hurdles must first be overcome. In this thesis we investigate two such hurdles: the presence of noise in quantum computers and limitations on the connectivity and control in large scale quantum computing architectures.

In order to combat noise in quantum devices we must first characterize this noise. To do this several diagnostic tools have been developed over the last two decades. The current industry standard for such a diagnostic tool is called randomized benchmarking. Randomized benchmarking doesn't give a full characterization of the noise afflicting the quantum device but rather attempts to give some indication of of the device's average behavior, captured in a quantity called the average fidelity. Because it does not endeavor to characterize every small detail of the noise it can be efficiently applied even to very large quantum devices.

However, with this power also comes increased complexity. Randomized benchmarking has a lot of moving parts, and some fairly strong assumptions must be made in order to guarantee its correctness. In this thesis we attempt to justify these assumptions and if possible remove or weaken them, making randomized benchmarking a more robust and general tool. In chapter 6 of this thesis we investigate the finite statistics of randomized benchmarking. We prove strong bounds on the number of samples needed to perform rigorous randomized benchmarking. To do this we make use tools from representation theory. In particular we use a characterization of certain representations of the Clifford group, which we develop in chapter 5. In chapter 7 we re-use these tools to also bound the number of samples needed to perform rigorous unitarity randomized benchmarking, a newer variant of randomized benchmarking quickly gaining in popularity. These results retroactively justify the use of randomized benchmarking in an experimental setting and also provide guidance on optimal statistical practices cin the context of randomized benchmarking.

In chapter 8 we expand upon the standard randomized benchmarking protocol and formulate a new class of protocols which we call character randomized benchmarking. This new class of protocols removes a critical assumption made in standard randomized benchmarking, making character randomized benchmarking vastly more generally applicable. To show the advantages of character randomized benchmarking we implement it in an experiment characterizing the noise in a Si\SiGe quantum dot device. This experiment is detailed in chapter 9.

Finally we deal with the second main topic of this thesis in chapter 10. Large scale quantum computer will, like classical computers, face limitations in the connectivity between

different parts of the computer. This is due to a fundamental law in computer design called Rent's rule, which states that the number of wires connecting a (quantum) computer chip to the outside world is much smaller than the number of components in that chip. This means the individual components of the chip can not be controlled individually in parallel. Given that parallelism is absolutely critical for the functioning of quantum computers this is a serious problem for the development of large scale quantum computers. Luckily it is possible to organize quantum computing devices in such a way that they can be controlled using a relatively small amount of input wires. One example of such an organization is called a crossbar architecture. Recently a proposal was made for a crossbar architecture quantum computer in quantum dots, and in chapter 10 of this thesis we investigate in detail the advantages and disadvantages of such an architecture. We focus in particular on its effect on standard quantum error correction procedures, a key part of a functioning quantum computer, and one where parallel control of all parts of the quantum device is essential.

Samenvatting

Kwantumcomputers beloven een revolutie teweeg te brengen op technologisch vlak. Voor het zover is moeten echter vele problemen opgelost worden. Twee van deze problemen vormen het onderwerp van deze scriptie. Het eerste probleem is de aanwezigheid van imperfecties in de operaties van de kwantumcomputer, en het tweede probleem wordt gevormd door de scherpe limieten op de simultane controle van alle onderdelen (kwantumbits) in een grote kwantumcomputer.

Voor we imperfecties in kwantumcomputers kunnen verhelpen, moeten we eerst methodes ontwikkelen om deze imperfecties te detecteren. Hiervoor zijn in de laatste jaren verschillende goede methodes ontstaan. De meest prominente methode op dit moment is de zogenaamde stochastische-ijk methode (randomized benchmarking). De stochastische-ijk methode werkt goed omdat ze niet probeert een gedetailleerd beeld te scheppen van de aanwezige imperfecties maar in plaats daarvan een uitspraak doet over de gemiddelde grootte van deze imperfecties. Dit zorgt ervoor dat the stochastische-ijk methode op een efficiente manier kan gebruikt worden in grootschalige kwantumcomputers.

De kracht van de stochastische-ijk methode komt natuurlijk met een prijskaartje. De werking van de methode is complex en er moeten sterke aannames gemaakt worden op het gedrag van de geteste kwantumcomputer om de correctheid van de methode te garanderen. Een belangrijk onderdeel van deze scriptie is het onderzoeken van deze aannames, met het doel ze ofwel te rechtvaardigen, ofwel ze te vervangen door minder sterke aannames. In hoofstukken 6 en 7 onderzoeken we de statistische fluctuaties van de stochastistische-ijk methode en de unitaire stochastische-ijk methode. We gebruiken wiskundige methoden uit de representatie theorie, welke we ontwikkelen in hoofdstuk 5, om sterke garanties te geven over de hoeveelheid data die verzameld moet worden om correcte conclusies te trekken uit de stochastische-ijk en unitaire stochastische-ijk methodes. We formuleren ook een aantal richtlijnen voor het correcte gebruik van statistische methodes in de context van de stochastische-ijk methode.

In hoofstuk 8 formuleren we een nieuwe, meer algemene versie van de stochastische-ijk methode. Deze methode, de karakter-stochastische-ijk methode, laat ons toe een belangrijke aanname in de werking van de stochastische-ijk methode te verwijderen. Dit maakt de karakter-stochastische-ijk methode toepasbaar op een bredere klasse van problemen. In hoofstuk 9 tonen we aan dat deze nieuwe methode ook werkt in de praktijk door de imperfecties van een kwantumcomputer bestaande uit twee Si\SiGe kwantumstippen te analyseren.

In het laatste hoofdstuk van deze scriptie bespreken we het tweede onderwerp van deze thesis: het gebrek aan simultane controle in grote kwantumcomputers. Dit is in zekere zin

een probleem voor de toekomst, maar het is wel een probleem waar we niet onderuit kunnen. Dit is zo omdat het probleem een gevolg is van een fundamentele wet in het ontwerp van computerchips: de zogenaamde regel van Rent. Deze regel stelt dat voor een grote (kwantum) chip het aantal connecties tussen de chip en de buitenwereld steeds veel kleiner is dan het aantal componenten in de chip zelf. Dit betekent dat de componenten in de chip nooit allemaal apart, simultaan kunnen aangestuurd worden. Gelukkig bestaan er methodes om de componenten van een chip collectief aan te sturen. Een standaardmethode in het ontwerp van klassieke computerchips is de zogenaamde 'crossbar'-architectuur. Niet zo lang geleden werd een voorstel voor een versie van deze crossbar-architectuur voor kwantumcomputers, specifiek voor kwantumstippen, gelanceerd. In hoofdstuk 10 bestuderen we de voor-en nadelen van deze architectuur voor kwantumcomputers, waarbij we extra aandacht besteden aan methodes voor kwantum-foutencorrectie. Dit laatste is een centraal puzzelstuk in een werkende kwantumcomputer, en is speciaal gevoelig voor beperkingen in de simultane controle van kwantumbits.

Acknowledgements

It takes a village to raise a child, and I think the same is true for a PhD. Even though this thesis has my name on it, many people have contributed to it in one form or another, and I would like to take a few moments here to acknowledge the people who were vital to this effort. **Stephanie**, someone once told me that the main thing you learn from an advisor is style, and if that is true I could not have had a better teacher. Your mastery of all the aspects of scientific practice, as well as your willingness to teach and your patience with my -occasionally comical, occasionally tragic- mistakes were invaluable to any successes I may claim from my time in Delft. I hope to have been half as good a PhD student as you were an advisor. **Jeremy, Axel, Mark, Glaucia, Valentina, Filip, David, Corsin, Jed, Nelly, Victoria, Bas, Guus, Lennart, Ben, Thinh, Leon, Matt, Tim, Raja, Kaushik, Liangzhong, Stefan, Willem, Marius, Kanvi, Carlo, Constantijn, Sebastian**, thank you for being the best colleagues, students, teachers, coworkers, drinking buddies and friends a young PhD student could possibly wish for. You made working in our little theory corner of QuTech a joy beyond words. You are all brilliant scientists and people and I hope you live lives of happiness and success.

James, Ramiro, Adriaan, Christian, Ben, Michiel, Joel, Christophe, Willemijn, Anne-Marije, Niels, Youri, Suzanne, Romy, Will, Nandini, Rafal, Franscesco, Florian, Udit, Xiao, Xavi, Tom, Norbert, Peter, Alex, Marianne, Julia, Lingling, Nathan, Alei, Stephan, and many others, QuTech is a very special place and it is so because of the hard work and incredible kindness from all of you. I felt truly at home here and all of you participated in making it that home. **Mark, Jeremy, Thinh, Axel, Xiao, Roy, Menno, Lieven, Bas, Franscesco**, thanks for being part of my projects or letting me be a part of yours, my publication list would look a lot more meagre without your help. **Adriaan, James, Suzanne, Anne-Marije, Guan and Tim**, thank you for starting the QuTech blog with me and for your continuing enthusiasm and commitment to this project. Thanks also to **Christian** for being such a driving force behind many of our best posts.

Many thanks to my committee: **Prof. Gross, Prof. Hanson, Prof. van Deursen, Dr. Walter, Prof. Vandersypen, and Prof. Terhal** for reading this, rather too long, thesis.

Wo, Jaime, Enrico, Emma, Ruby, Karthik, Jan, Fred, Loic, Jasper, Michele, Irene, Benedetta, Julien, Rajeev, Balint, Alvaro, Henk, Kamila, Karolina, Radu, Daniela, Joreen, , and many others, improv, and the extended family that came with it, was such a big part of keeping me sane through the trials and tribulations of PhD life. DIG and DILF are safe spaces for people's -in particular mine- quirky sides. They are hidden treasures built out of the hearts and souls and fantasies of all of you and I hope they never go away. Keep on being weird peeps, I love you all. **Christian**, your Wednesday jam nights were the linchpin of my social life here in Delft, and I can only admire your generosity and mu-

sical spirit. Also thanks to **Holger, Deniz, Jochen, Anna, Albert, Rob, James, Güzin, Steven, Ramiro, Ege, Tolga, Raj, Vladimir, Almira, Matteo, Joel, Jasmin, Nandini, Rafal, Aynur**, and many others for showing up, making music and having a good time. **Will**, the jam is yours now, I know you will keep the fire burning and I'm proud of you for it. **Joel, Ben, Kenneth, Christophe, Pierre**, thanks for rolling dice and pretending to be heroes with me. We never did quite finish that adventure, but it was a lot of fun.

Bram, Alex, Ruben, Arnaud, Kobe, An, Tom, Duncan, Natasa, Anna, Freya, Elisabeth, you are all better friends than I deserve and I don't visit you half as much as you deserve.

Gilles, Pieter-Jan, jullie oudste kortste broer is een doctor nu, ik ben blij dat ik jullie als mijn kleinere grotere broers heb. **Oma, Opa, Moeke, Pappie**, ik had dit nooit gekund zonder jullie liefde en aanmoediging. Jullie hebben me geleerd, in woorden en in daden, wat het betekent om een goed mens te zijn. Ik hoop dat ik naar jullie voorbeeld kan leven. **Mams**, je liefste vriend wordt groot. Bedankt voor je luisterend oor en bemoedigende woorden, je enthousiasme en levenskracht. Je bent alles wat ik van een moeder kan wensen. **Paps**, bedankt voor je kennis en wijsheden, voor de rust die je brengt in mijn gejaagd bestaan en voor je eindeloze interesse in alles wat ik doe en leer. Ik kan me geen betere vader voorstellen. En **Lisa**, dit was het, we hebben het overleefd. Ik ben dolblij dat jij de mijne bent en nog blijer dat ik de jouwe ben. Dit leven is een avontuur, en ik zou er met niemand anders willen induiken. Bedankt voor alles, ik zie je graag.

Contents

1

Quantum computing in the real world

A spectre is haunting quantum computing – The spectre of decoherism.

Karl Quarks

In this general introductory chapter we talk about the challenges one faces when trying to build quantum computers in real physical systems, and also about the techniques developed to meet these challenges. The first of these challenges, the presence of noise in real devices, has been a topic of study for more than two decades, and plenty is known about diagnosing and mitigating the errors that arise from this noise. The second challenge, dealing with connectivity and classical control limitations in quantum devices, has only recently been garnering more attention.

1

This thesis is about quantum computing. In particular it is about the challenges faced when building and operating quantum computers. There are many ways to approach these challenges. One could focus on one particular platform and try to work out al the details of operation for this platform, or one could take a wider view and try to develop tools and ideas that can be applied across platforms and experiments. In this thesis we mostly focus on this latter category. We will in particular deal with two generic problems one faces when trying to build and operate quantum computers: how to diagnose noise in quantum devices and how to deal with limitations in connectivity and control in quantum devices.

In this short introductory chapter we aim to present a somewhat more personal view on the 'state-of-the-field'. In particular we would like to outline some thoughts on how to diagnose noise in quantum computers and in particular formulate some criteria for what we consider to be a good diagnostic tool for quantum computers. This we will do in section 1.1. We would also like to discuss the problems of limited connectivity and limited classical control in quantum computers. This we will do in section 1.2. Both of these problems will become much more of an issue as quantum computers scale up and little thought has been given in particular to the latter one. Finally we give an overview of the chapters in this thesis in section 1.3.

1.1. Testing quantum computers

A key part of building quantum computers is dealing with unwanted behavior. This unwanted behavior might be due to uncontrolled interactions with an environment (stray photons, background magnetic field fluctuations, people spilling coffee,...) but might also be due to experimental deficiencies such as mis-calibrated signals or the limitations of finite precision arithmetic. This unwanted behavior inside the quantum computer can result in erroneous outcomes of computations, an outcome we would like to avoid. There are generally two ways of mitigating such errors. The first one is to spend a lot of time and energy tracking down all possible sources of unwanted behavior and engineering them away. The second one is to accept that unwanted behavior will happen and to use error correction techniques to limit the sensitivity of computations to this behavior. It is generally accepted that both of these approaches will be necessary for building functional quantum computers. Quantum systems are highly sensitive to outside interference and building completely isolated, yet perfectly controlled quantum systems is likely in the realm of fantasy. On the other hand, quantum error correction techniques only work if the rate of error inside a device is sufficiently low. This means we must perform the hard work of engineering away errors at least up to the point where error correction can be trusted to take care of any leftovers.

When optimizing quantum devices to minimize unwanted behavior, a basic necessity is the ability to tell whether anything is actually wrong. Moreover, one would like to do this in a somewhat structured manner, so as to spend a minimal amount of time on guesswork. This calls for the development of diagnostic procedures that can extract information about a quantum device in a principled manner. Over the past decades, many such procedures

have been devised. Later in this thesis we will investigate two currently popular protocols, randomized benchmarking and unitarity randomized benchmarking, in great detail. But first we would like to discuss the criteria a good diagnostics protocol should adhere to. The items on this list are neither necessary nor sufficient conditions for a good diagnostic protocol, but should rather be read as broad guidelines, informed by our personal experience, for the future design of such protocols

- **Efficiency:** We would like our diagnostic protocols to be efficient. This can mean efficiency in the complexity-theoretic sense that only a polynomial number of operations (relative to the number of qubits in the device) must be performed, but also efficiency in the absolute sense (the protocol shouldn't take weeks to run).
- **Generality:** We want diagnostic protocols to be as general as possible. There are many different proposals for quantum computing platforms and a prospective diagnostic protocol should work for as many platforms as possible and moreover yield results that are comparable across platforms.
- **Interpretability:** We want the outcome of the protocol (this is typically one or more real numbers) to have a clear interpretation. This could either be in terms of some clear physical effect (e.g. the magnitude of electric field fluctuations) or in terms of some metric with a clear operational interpretation (e.g. the mean number of consecutive error free operations).
- **Robustness:** We want the protocol to be subject to as few assumptions the working of the device as possible, and we would ideally like to be able to detect whether these assumptions are violated.
- **Transparency:** We want the protocol to be 'easy' to understand. This means it should be easy to use and hard to abuse, even by someone who doesn't know all details that make the protocol tick.

We will later see that guaranteeing these attributes in quantum diagnostic protocols can take a lot of work and often making progress in one category will mean ceding ground in another. However, the goal is not to design the best possible diagnostic protocol, but rather to help the people that build quantum computers, build quantum computers. To this end we will, in this thesis, be mostly concerned with the exploration of a diagnostic protocol called randomized benchmarking. This protocol is considered to be the gold standard of diagnostic protocols currently used in experimental practice. It is however quite difficult to rigorously justify that randomized benchmarking fulfills some of the above criteria, and we will spend a large portion of this thesis developing the necessary guarantees.

1.2. Overcoming limited connectivity and control

There is a curious disconnect between the theory and practice of quantum computing. On the theory side quantum algorithms are consistently designed having in mind some idealized version of the quantum computer the algorithm is going to run on. This idealized quantum computer has many properties that are seldom satisfied in real devices. The main assumption made is that operations on the quantum computer are completely error

free. As we have seen in the previous section, this is never satisfied. However with good diagnostic tools and error correction techniques we can probably get arbitrarily close to this ideal at some point in the future. On the other hand other assumptions are almost always made (often without explicitly stating them) where the solution is less clear-cut. We single out two important such assumptions:

- **All-to-all connectivity:** We assume any qubit can exchange information with any other qubit, even when they are physically far removed from each other.
- **Unlimited classical control:** We assume that we can perform different operations on all qubits in a quantum device at the same time.

Neither of these assumptions is satisfied in real devices. Moreover, looking at classical computers (which is in some sense like looking into the future of quantum computers), it is unlikely that they will ever be satisfied. Classical transistors are confided to living on a 2 dimensional plane and can usually only connect to a limited number of other connectors. Moreover, the number of wires coming into a microchip (and thus the number of instructions that can be exchanged) is typically orders of magnitude smaller than the number of transistors in a chip. Overcoming these limitations in classical devices takes up a large portion of the intellectual energy in the microchip design world and it is a small miracle that these issues are almost never directly relevant for an end user. Quantum computers will likely follow the same path, with current proposals for large scale quantum devices by-and-large sticking to qubits on a plane with only nearest neighbor connectivity. Moreover these devices restrict how much control can be exerted on all qubits in parallel. One might for instance restrict the control of the qubits to a crossbar system, where the qubits are arranged on a grid and control signals are sent to whole rows and columns of qubits. This means a qubit can be uniquely addressed at the intersection of a row and a column. However not all qubits can be individually addressed at the same time. This is similar to the solutions to large scale control found in classical devices. However quantum computers have a critical need for parallelism that is not shared by classical devices. This is so because quantum error correction is inherently a parallel process. If the time spent doing error correction grows too quickly with the number of qubits involved in the error correction procedure the error correction will become counterproductive. This means applying quantum algorithms in general, and quantum error correction in particular in a large scale quantum device with limited classical control is highly non-trivial. We will spend the latter part of this thesis (chapter 10) investigating these issues in the context of a concrete proposal for a large scale quantum computing device in quantum dots.

1.3. Chapter overview

This thesis has eleven chapters. Of these eleven chapters the first four (including this one) can be seen as introductory, setting notation and reviewing relevant concepts. The fifth to ninth chapters deal with various aspects of noise diagnosis, mostly in the context of a popular class of diagnostic protocols called randomized benchmarking. The tenth chapter on the other hand deals with the second main topic of this thesis: dealing with limited control and connectivity.

Chapter 2 is a preliminary chapter where we discuss some basic quantum mechanical and mathematical notions that will be used throughout the thesis. This chapter is also where a lot of the notation will be set.

In **chapter 3** we introduce representations of finite groups and discuss some basic notions of representations theory. We also introduce Schur's lemma and the character projection formula, two powerful representations theoretic tools that will be of great use in later chapters.

In **chapter 4** we discuss two prominent diagnostic protocols for quantum computers: randomized benchmarking and unitarity randomized benchmarking. We will discuss the what-and-why of these protocols and outline their advantages and disadvantages.

In **chapter 5** we derive explicit expressions for all irreducible subrepresentations of the 'two-copy' representation of the Clifford group. The results of this chapter will form the foundation on which the results of chapter 6 and chapter 7 are built.

In **chapter 6** we provide an improved version of the randomized benchmarking protocol and leverage the results of chapter 5 to analyze the finite-statistics behavior of this protocol. In particular we prove strong bounds on the number of random sequences needed to perform rigorous randomized benchmarking. We also raise issues with the current approach to the curve fitting part of the randomized benchmarking protocol and propose an improved method that remedies these issues.

In **chapter 7** we prove, similarly to the results in chapter 4, strong bounds on the number of random sequences needed to perform rigorous unitarity randomized benchmarking. We also provide an in depth discussion of the unitarity randomized benchmarking protocol, clarifying a number of issues involving the statistics and scalability of the protocol.

In **chapter 8** we introduce a new class of randomized benchmarking protocols which we call character randomized benchmarking. These protocols extend standard randomized benchmarking and allows one to reliable extract the average fidelity from a much broader class of groups of quantum gates. We prove that the protocol works as advertised, discuss its finite statistics properties and give examples of scenarios where character randomized benchmarking might be of use.

In **chapter 9** we describe the outcome of an experiment in Si\SiGe quantum dots. In this experiments we use an instance of the character randomized benchmarking protocol described in chapter 8 to extract the fidelity of a CPHASE gate.

In **chapter 10** we deal with the problem of limited connectivity and control in quantum devices. In particular we present schemes for quantum error correction in a recent proposal for a large scale quantum dot processor. This processor arranges qubits in a grid and addresses them at the intersection of row and column lines, a so called crossbar architecture. This saves on control architecture but creates issues with the parallel operation

of the device qubits. This lack of parallelism is especially problematic for quantum error correction. We present a detailed study of the planar surface code and the 4.8.8. and 6.6.6. planar color codes on this device, explicitly describing each operation necessary for error correction in terms of the native operations of the device. We then give efficient schemes for performing error correction and analyze their performance. Finally we outline some algorithms for more general computation in crossbar architectures.

Finally in **chapter 11** we present the conclusions of this thesis and discuss some future research avenues.

2

The basics of quantum computing

Wie het kleine niet eert, is het grote niet weert.

Old Belgian proverb, origin unknown

This chapter recalls basic notions of quantum mechanics and quantum computing that will be of use throughout this thesis. We will also set most of the notation used in this thesis. We will discuss quantum states and operations, groups of quantum operations, ways to represent noise in quantum devices and also some quality measures for quantum operations; in particular the average fidelity and the unitarity. These two measures will be the topic of later chapters.

In this preliminary chapter we will introduce a variety of basic quantum mechanical concepts that are needed for the rest of this thesis. This will mostly be an exercise in notation setting. For a much more expansive and didactic introduction to the basics of quantum computing we refer the reader to Nielsen & Chuang's excellent textbook [1] or the more advanced lecture notes by John Watrous [2]. We will assume a fair amount of familiarity with the basic concepts of linear algebra. For a review of linear algebra we recommend Horn & Johnson [3] or Bhatia [4]. In section 2.1 we will introduce the fundamental building blocks of quantum mechanics such as quantum states, measurements and unitary operations. In section 2.2 we will introduce the mathematical concept of a group, in the context of operations in quantum computers. In section 2.3 we will introduce methods to describe the dynamics of noisy quantum computers, and also discuss some common types of noise. Finally in section 2.4 we will discuss some common methods to quantify noise in quantum devices. This last section is far from expansive, merely covering the quantities we will discuss in greater detail later in this thesis.

2.1. Quantum states and quantum operations

In this section we outline the very basics of quantum mechanics through the lens of quantum computation. We will deal with qubits, measurements, operations on qubits and the density matrix formalism for noisy quantum states.

2.1.1. Quantum states

The fundamental building block of a quantum computer, and arguably quantum mechanics itself, is the qubit. Qubits come in many different physical forms but for the purpose of this thesis we shall abstract them away to their very core. We define a qubit to be a two-dimensional complex Hilbert space $\mathcal{H}_2$. The state of a qubit, denoted using Dirac notation as $|\psi\rangle$ (this is called a 'ket') is then a normalized vector in this Hilbert space. We will generally describe this state $|\psi\rangle$ in terms of a privileged orthonormal basis of $\mathcal{H}_2$ which we call the *computational basis* and denote $\{|0\rangle, |1\rangle\}$. These two states are often called the 'zero-state' and the 'one-state'. Thus we can in general write $|\psi\rangle = a|0\rangle + b|1\rangle$ with $a, b \in \mathbb{C}$ and $|a|^2 + |b|^2 = 1$. It is also useful to have a notation for the dual vector of $|\psi\rangle$. We will denote this dual vector as $\langle\psi|$ (this is called a 'bra'). This allows us to write the inner product between states as $\langle\phi, \psi\rangle = \langle\psi|\phi\rangle$ (a bra-ket).

We can combine multiple qubits together to create states in larger Hilbert spaces. This is done using the tensor product, denoted $\otimes$. So the state $|\psi\rangle \otimes |\phi\rangle$ for $|\psi\rangle, |\phi\rangle \in \mathcal{H}_2$ is a 2-qubit state and thus an element of the 4-dimensional complex Hilbert space $\mathcal{H}_4$. We can write down a computational basis for $\mathcal{H}_4$ as well given by $\{|i\rangle \otimes |j\rangle \quad \| \quad i, j \in \{0, 1\}\}$. In the future we will skip the tensor product where it is not needed and write e.g. $|ij\rangle = |i\rangle \otimes |j\rangle$. We can repeat this construction to describe the space of q-qubit Hilbert spaces $\mathcal{H}_{2^q}$ (of dimension 2^q) which has computational basis $\{|x\rangle \quad \| \quad x \in \{0, 1\}^n\}$. It is of course also possible to consider states in Hilbert spaces which don't have a power of two dimension, and we will often write $\mathcal{H}_d$ to indicate a complex Hilbert space of dimension d, leaving it up to context whether or not $d = 2^q$ for some q.

2.1.2. Noisy quantum states

Up until now we have assumed that our qubit is always in a definite state. However it is often useful to consider probabilistic mixtures of states. This is what the density matrix formalism is for. In this formalism we think of states as linear transformations rather than vectors. In particular we will identify the state $|\psi\rangle$ with the rank one orthogonal projector onto the vector $|\psi\rangle$. We denote this projector as $|\psi\rangle\langle\psi|$. A qubit (or multiple qubits) in a probabilistic mixture of orthogonal states can then be seen as a non-negative operator ρ in the Hilbert space of linear transformations of $\mathcal{H}_d$ which we denote as $\mathcal{M}_d$. Moreover we demand that $\mathrm{Tr}(\rho) = 1$ (this follows from the fact that probabilities must add up to 1). Such operators are called density operators (or density matrices). Conversely (by the spectral theorem) we can consider every non-negative operator with trace one as a probabilistic mixture of rank one orthogonal projectors (and thus quantum states). Abusing nomenclature somewhat we will also refer to such operators as states and reserve the term pure state for a rank one projector and will use the term mixed state when specifically referring to a non-trivial probabilistic mixture of states.

2.1.3. Measurements

We can also 'measure' qubits. Measurement is a difficult topic but will abstract most of it away and simply define a measurement to be associated to an orthonormal basis $\{|x\rangle\}_x$ in $\mathcal{H}_2$. When measuring a qubit state $|\psi\rangle$ we will observe a 'measurement outcome' x associated to the basis state $|x\rangle$. This will happen with probability $|\langle x, \psi\rangle|^2$ (this is called the 'Born rule'). Moreover, after the measurement and the observation of the outcome x the qubit will have 'collapsed' to the state $|x\rangle$. Measurement thus changes the state of the qubit. The concept of measurement extends straightforwardly to multiple qubits. It is also possible to perform more general types of measurements (these can be seen as the measurement analogues of density matrices). A general measurement on a quantum state (described by a density matrix) is described by a Positive Operator Valued Measure (POVM). This is a set of positive operators $\{Q_i\}_{i\in I}$, where I is the set of possible outcomes, such that $\sum_{i\in} Q_i = \mathbb{1}$. Given a state ρ and a POVM $\{Q_i\}_{i\in I}$ the probability of obtaining an outcome $i \in I$ is given by $p_i = \mathrm{Tr}(\rho Q_i)$. Note that our previous notion of measurement can be seen as the POVM $\{|x\rangle\langle x|\}_x$.

2.1.4. Quantum operations

In order to do computation with quantum states we need some way to transform one state into another. This is done using unitary operations. Unitary operations (often called 'unitary gates' or simply 'gates') are linear transformations of states that preserve the inner product between states. We thus require that $\langle\phi| U^\dagger U |\psi\rangle = \langle\phi|\psi\rangle$ for all $|\phi\rangle, |\psi\rangle \in \mathcal{H}_d$ where $U^\dagger$ is the Hermitian conjugate of the transformation U, or equivalently that $U^\dagger U = \mathbb{1}$. We will denote the set of unitary matrices as $U(d)$. Later we will see that this set naturally has a group structure. In the next section we will delve deeper into the behavior of particular unitaries. But first we most deal with the idea of noisy quantum states.

We can also perform unitary operations on density matrices in the obvious way; U acts on a state ρ by conjugation, i.e. $U\rho U^\dagger$. Note that the unitary operation in some sense preserves the probability distribution associated to ρ. In particular unitary operations

will send pure states to pure states. Later in this chapter we will deal with more general operations on quantum states that do not have this property, i.e. they will map pure states to mixed states or the other way around. But first we will spend some time discussing the group structure inherent to unitary quantum operations.

2.2. Groups of quantum operations

So far we have considered unitary operations merely as a subset of the linear operators $\mathcal{M}_d$. However it turns out they have a lot more structure. Recall that the unitary operators preserve the inner product between all states. This is in fact equivalent to stating that $U^\dagger U = \mathbb{1}$ where $\mathbb{1}$ is the identity transformation. Moreover it is clear that $U^\dagger$ is unitary if and only if U is and moreover if U and V are unitary, then so is their composition UV. This means that the set of unitaries has a notion of inverse and a notion of closure under composition (here given by matrix multiplication). A set together with some composition rule that has such properties is called a group. Groups show up everywhere in physics and mathematics and they are central enough to this thesis (and unfamiliar enough to quantum computing researchers) to merit our first definition.

Definition 1 (Groups)**.** Let G be a set and let $* : \mathsf{G} \times \mathsf{G} \to \mathsf{G}$ be a function such that the following statements hold:

1. There is a $e \in \mathsf{G}$ such that $g * e = e * g = g$ for all $g \in \mathsf{G}$. (identity element)
2. For all $g \in \mathsf{G}$ there is an $h \in \mathsf{G}$ such that $g * h = h * g = e$. (inverse)
3. For all $g_1, g_2, g_3 \in \mathsf{G}$ we have $g_1 * g_2 * g_3 = (g_1 * g_2) * g_3 = g_1 * (g_2 * g_3)$. (associativity)

Then $\mathsf{G}, *$ is a group

For a broad introduction to the theory of group one can consult [5]. We will almost always drop the composition law $*$ when talking about a group and just refer to G as a group. Likewise we will almost always suppress it in computation, writing $g_1 * g_2 = g_1 g_2$ when the composition is obvious from the context. We have already seen a first example of a group, namely the group of unitary operators on q qubits $U(2^q)$, where the composition law is given by matrix multiplication. Note that this group has an (uncountably) infinite number of elements. There are two more group-related definitions we must get out of the way before we can really start talking about groups in the context of quantum computing. The first is the notion of a subgroup.

Definition 2 (Subgroups)**.** Let G be a group consider a strict subset H of G such that for all $h_1, h_2 \in \mathsf{H}$ we have $h_1 h_2 \in \mathsf{H}$. Then we call H a subgroup of G.

Often we will specify subgroups of the unitary group implicitly, and for this the concept of group generators is very handy. We have the following definition.

Definition 3 (Generators)**.** Let G be a group and let A be a subset of G. We call A a generating set of G, denoted as $< A >= \mathsf{G}$ if and only if for all elements $g \in \mathsf{G}$ there exists an integer k and ordered sequence $g_1, g_2, \ldots, g_k$ with $g_i \in A$ (possibly with repeats) such that $g_1 g_2 \cdots g_k = g$. We call the elements of a generating set A generators.

Note here that for any $g \in \mathsf{G}$ there might be multiple ordered sequences of generators that reach g. Now we are ready to define what are probably the two most important (in the context of quantum computing) subgroups of the unitary group, namely the Pauli and Clifford groups.

2.2.1. The Pauli group

The q-qubit Pauli group is in a sense the most basic 'really quantum' subgroup of the unitary group. It is generated by the basic bit-flip and phase-flip operators and pops up all over quantum computing. It has also, under a variety of names, been a key structure in quantum mechanics since the very beginning. The Pauli group is formally defined as follows:

Definition 4 (Pauli group)**.** Let $\{|0\rangle, |1\rangle\}$ be an orthonormal basis of $\mathcal{H}_2$ and in this basis define the following linear operators by their action on the basis

$$X|l\rangle = |l+1\rangle\,,\ Z|l\rangle = (-1)^l|l\rangle\,,\ Y|l\rangle = iZX|l\rangle = i(-1)^{l+1}|l+1\rangle\,,$$

for $l \in \{0,1\}$ and addition over indices is taken modulo 2. Note that $X, Y, Z \in U(2)$. The q-qubit Pauli group P_q is now defined as the subgroup of the unitary group $U(2^q)$ consisting of all q-fold tensor products of q elements of $\mathsf{P}_1 := \langle X, Z, i\mathbb{1}_2\rangle$.

We will make frequent use of a subset of P_q defined as all q-fold tensor products of q elements of $\{X, Y, Z\}$. We will call this subset (which contains only Hermitian elements of the Pauli group) P_q^*.

The Pauli group P_q has a number of notable properties which we will use throughout the thesis. Its first useful property is a specific set of commutation relations. Let's briefly recall the definition of an (anti-) commutator of two operators in $\mathcal{M}_d$.

$$[A,B] = AB - BA \quad \forall A, B \in \mathcal{M}_d \qquad \text{(commutator)}$$
$$\{A,B\} = AB + BA \quad \forall A, B \in \mathcal{M}_d \qquad \text{(anti-commutator)}$$

Elements P, P' of the Pauli group have the property that they either *commute* or *anti-commute*, that is

$$[P,P'] := PP' - P'P = 0 \qquad \text{or} \qquad \{P,P'\} := PP' + P'P = 0. \tag{2.1}$$

Also notable here is that every non-identity Pauli operator commutes with exactly half of the elements of the Pauli group, and anti-commutes with the other half. This moreover stays true if we restrict ourselves to elements of the set P_q^*. This fact is easy to verify but extremely useful. We will use it to simplify expressions and prove theorems throughout the thesis.

2.2.2. The Clifford group

The next important subgroup of the unitary group is the so-called Clifford group. In order to define the Clifford group we must first recall a basic concept from group theory called a 'normalizer'.

Definition 5. Let G be a group and let S be a subset of G. The normalizer $N(S)$ (with respect to G) of S is then defined as

$$N(S) = \{g \in \mathsf{G} \parallel \forall s \in S : \ gsg^{-1} \in S\} \tag{2.2}$$

We can think of the normalizer of a set S as all elements of the group G that maps elements of S to elements of S. It is easy to see that $N(S)$ is in fact a subgroup of G for any set S. The Clifford group is now defined as the normalizer of the Pauli group in the unitary group. A small technicality crops up here, namely that the normalizer of the Pauli group in the unitary group is not actually a finite group. This is so because if a unitary U is in the normalizer of P_q then so is $e^{i\theta}U$ for an arbitrary angle θ. Hence we will define the Clifford group as the normalizer of the Pauli group up to this angle. We have concretely

Definition 6 (Clifford group)**.** The q-qubit Clifford group C_q is the normalizer (up to complex phases) of P_q in $\mathrm{U}(2^q)$, that is,

$$\mathsf{C}_q := \{U \in \mathrm{U}(2^q) \parallel U\mathsf{P}_q U^\dagger \subseteq \mathsf{P}_q\}/\mathrm{U}(1).$$

This definition illuminates the main feature of the Clifford group, namely that it maps elements of the Pauli group to elements of the Pauli group under conjugation. However, if one wants to implement elements of the Clifford group as gates in a quantum computer this definition does not give much insight. There is however an equivalent definition of the Clifford group in terms of generators which is much more useful in that regard

Definition 7 (Clifford group (equivalent))**.** The q-qubit Clifford group C_q is the subgroup of $\mathrm{U}(2^q)$ generated as

$$\mathsf{C}_q = \langle i\mathbb{1}, H_i, S_j, \mathrm{CNOT}_{lk} \ \ \| i,j,k,l \in [1:q], \ \ l \neq k \rangle \tag{2.3}$$

where $[1:q] = \{1, \ldots, q\}$, and $H_i = \mathbb{1} \otimes \mathbb{1} \ldots \otimes H \otimes \ldots \mathbb{1}$ and similarly for S_j, CNOT_{lk} with H, S and CNOT given as

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \ \ \mathrm{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{2.4}$$

The Clifford group acts on the Pauli group by conjugation, and this action has some notable properties. The first is that the action of the Clifford group preserves the commutation relations of the Pauli operators. This can be seen quite easily by the definition of the commutator

$$[CPC^\dagger, CP'C^\dagger] = C[P, P']C^\dagger, \quad P, P' \in \mathsf{P}_q, \ C \in \mathsf{C}_q \tag{2.5}$$

which is zero whenever $[P, P'] = 0$. The same calculation goes for the anti-commutator. Less trivial, and more interesting is the fact that this is the only restriction on the action of the Clifford group. By this we mean that if $[P, P'] = 0$ and $[\hat{P}, \hat{P}'] = 0$ then there exists a Clifford element C such that $CPC^\dagger = \hat{P}$ and $CP'C^\dagger = \hat{P}'$, with the same statement holding for the anti-commutator. See e.g. [6] for a proof of this fact.

2.3. NOISY QUANTUM OPERATIONS

The ideal quantum computer is a device that implements unitary operations on pure quantum states. However, as we already discussed in chapter 2 this is not how things work in the real world. In the real world the qubits inside the quantum computer will invariably experience uncontrolled interaction with the outside world. These interactions typically manifest in a stochastic manner, turning pure states into probabilistic mixtures of states. As we have already seen, it is not possible to describe such a transition using unitary operations. To describe such interactions we must use a more general formalism. This is the formalism of quantum channels. Quantum channels are linear operators that map elements of $\mathcal{M}_d$ to $\mathcal{M}_d$ (one can also define more general quantum channels that map $\mathcal{M}_d$ to $\mathcal{M}_{d'}$ but we will not deal with this here). This means they are 'superoperators' (operators acting on operators). We will denote the Hilbert space (for it is also a Hilbert space) of superoperators as $\mathcal{S}_d$. Moreover we require that our quantum channels map quantum states to quantum states. This means they must preserve the trace of an operator in $\mathcal{M}_d$ and moreover they must preserve the positivity of an operator in $\mathcal{M}_d$. In fact it is useful to make this last requirement slightly stricter to also take into account that quantum channels sometimes only act on certain tensor factors of a larger quantum states (that might be entangled). This brings us to the following definition of a quantum channel

Definition 8 (Quantum channel)**.** Let $\mathcal{E}$ be a superoperator in $\mathcal{S}_d$. If $\mathcal{E}$ satisfies the following properties

$$\begin{aligned} &\mathrm{Tr}(\mathcal{E}(X)) = \mathrm{Tr}(X) && \forall X \in \mathcal{M}_d && \text{(trace preservation)} \\ &\mathrm{id} \otimes \mathcal{E}(\rho) \geq 0 && \forall \rho \in \mathcal{M}_{d^2}, \;\; \rho \geq 0 && \text{(complete positivity)} \end{aligned}$$

then $\mathcal{E}$ is a quantum channel.

Quantum channels are also sometimes called CPTP maps (Completely Positive and Trace Preserving). We will use this nomenclature occasionally as well. We will generally use calligraphic script to refer to quantum channels $\mathcal{E}, \mathcal{D}$ etc. It is useful to consider a few examples of quantum channels that are physically relevant and also show up throughout this thesis.

Pauli channel: Pauli channels (denoted $\mathcal{P}$) are quantum channels that stochastically apply a Pauli operator $P \in \mathsf{P}_q^*$ with probability p_P. They are defined by a probability distribution $\{p_P\}_{P \in \mathsf{P}_q^*}$ that has 2^{2q} elements. Formally we have

$$\mathcal{P}(\rho) = \sum_{P \in \mathsf{P}_q^*} p_P P \rho P^\dagger. \tag{2.6}$$

Note that the identity is part of the set P_q^*. Note also that eq. (2.6) doesn't quite cover the full range of quantum channels of the functional form $\mathcal{P}(\rho) = \sum_{P \in \mathsf{P}_q^*} \lambda_P P \rho P^\dagger$. This is so because we restrict the set $\{\lambda_P\}_{P \in \mathsf{P}_q^*}$ to be a probability distribution. The most general set of values for which eq. (2.6) would still satisfy the conditions in definition 8 is given by $\{\lambda_P \;\; \| \;\; 1/(d^2-1) \leq \lambda_P \leq 1, \;\; \sum_{P \in \mathsf{P}_q^*} \lambda_P = 1\}$. One could also call this more general set of channels the Pauli channels, but common convention is to restrict the definition to

probability distributions. We will sometimes use this more general definition, most notably in chapter 6.

Depolarizing channel: The depolarizing channel (denoted $\mathcal{D}$) is a special case of the Pauli channels where any $P \in \mathsf{P}_q^* \backslash \{\mathbb{1}\}$ happens with equal probability. It turns out this is equivalent to writing

$$\mathcal{D}_p(\rho) = (1-p)\rho + p\frac{\mathbb{1}}{d}. \tag{2.7}$$

with $0 \leq p \leq 1$. We again note that it is possible to adopt a more general definition of the depolarizing channel which still satisfies definition 8, where $-1/(d^2-1) \leq p \leq 1$. However the first definition is more common in the literature.

Unitary channel: Unitary channels are defined by the conjugation action of a unitary operator U on a quantum state ρ. We will denote them by calligraphic letters, that is

$$\mathcal{U}(\rho) = U\rho U^\dagger. \tag{2.8}$$

Amplitude damping channel: The amplitude damping channel (denoted $\mathcal{A}$) models the physical process of a system relaxing into a lowest energy state. It is parametrized by relaxation parameter λ. We will only give the explicit description of a special case, namely amplitude damping towards the $|0\rangle$ state in a qubit system:

$$\mathcal{A}_\lambda(\rho) = A_1\rho A_1^\dagger + A_2\rho A_2^\dagger, \tag{2.9}$$

with

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\lambda} \end{pmatrix} \qquad A_2 = \begin{pmatrix} 0 & \sqrt{\lambda} \\ 0 & 0 \end{pmatrix}. \tag{2.10}$$

The general case of a multi-qubit system relaxing to some general state $|\psi\rangle$ is easily inferred from here.

The amplitude damping channel is notably different from all other mentioned channels in that it does not preserve the maximally mixed state $\rho = \mathbb{1}/d$, that is $\mathcal{A}(\mathbb{1}/d) \neq \mathbb{1}/d$. Channels that do have this property are called unital channels. We will be dealing with unital channels extensively in chapter 7 so we will write down the definition here explicitly

Definition 9 (Unital quantum channels)**.** Let $\mathcal{E}$ be a quantum channel and let $\rho = \mathbb{1}/d$ be the maximally mixed state. If $\mathcal{E}(\mathbb{1}/d) = \mathbb{1}/d$ then we call $\mathcal{E}$ a unital quantum channel.

2.3.1. Liouville representation of quantum channels

Thus far we have been defining quantum channels by explicitly describing their action on a quantum state ρ. However it is often useful to have an explicit matrix description of a quantum channel. There are several ways to obtain such a description. The one we will make use of is variously known as the Liouville [7], affine [8] or Pauli Transfer Matrix representation.

This representation corresponds to fixing an orthonormal basis for $\mathcal{M}_d$ according to the Hilbert-Schmidt or trace-inner product and then expressing elements of $\mathcal{M}_d$ as vectors in $\mathcal{H}_{d^2}$. The Hilbert-Schmidt inner product is defined as

$$\langle A, B\rangle := \mathrm{Tr}(A^\dagger B), \qquad \forall A, B \in \mathcal{M}_d. \tag{2.11}$$

Now let $\{B_j\}_j$ be an orthonormal basis for $\mathcal{M}_d$ with respect to the Hilbert-Schmidt inner product. We can construct a map $|.\rangle\!\rangle : \mathcal{M}_d \to \mathcal{H}_{d^2}$ by setting $|B_j\rangle\!\rangle = e_j$ where e_j is the jth canonical basis vector for $\mathcal{H}_{d^2}$. Linearly extending the map $|\cdot\rangle\!\rangle$ to all elements $M \in \mathcal{M}_d$ we get

$$|M\rangle\!\rangle = \sum_j \mathrm{Tr}(B_j^\dagger M)|B_j\rangle\!\rangle. \tag{2.12}$$

Defining $\langle\!\langle M| = |M\rangle\!\rangle^\dagger$, we then have

$$\langle\!\langle M|N\rangle\!\rangle = \langle M, N\rangle = \mathrm{Tr}(M^\dagger N), \tag{2.13}$$

so that the Hilbert-Schmidt inner product is equivalent to the standard vector inner product.

We will generally construct the Liouville representation using the basis spanned by the *normalized* (with respect to the Hilbert-Schmidt inner product) Pauli matrices $\{\sigma_0\} \cup \boldsymbol{\sigma}_q$ where $\sigma_0 := \mathbb{1}/\sqrt{d}$ is the normalized identity matrix and

$$\boldsymbol{\sigma}_q := \frac{1}{\sqrt{d}}\{\mathbb{1}, X, Y, Z\}^{\otimes q} \backslash \{\sigma_0\}, \tag{2.14}$$

is the set of normalized Hermitian Pauli matrices excluding the identity. This is the origin of the term 'Pauli transfer matrix'.

As any quantum channel $\mathcal{E}$ is a linear map from $\mathcal{M}_d$ to $\mathcal{M}_d$ we have

$$|\mathcal{E}(\rho)\rangle\!\rangle = \sum_{\sigma \in \boldsymbol{\sigma}_q \cup \sigma_0} |\mathcal{E}(\sigma)\rangle\!\rangle\langle\!\langle\sigma|\rho\rangle\!\rangle, \tag{2.15}$$

so that we can represent $\mathcal{E}$ by the matrix

$$\mathcal{E} = \sum_{\sigma \in \boldsymbol{\sigma}_q \cup \sigma_0} |\mathcal{E}(\sigma)\rangle\!\rangle\langle\!\langle\sigma|, \tag{2.16}$$

where we abuse notation by using the same symbol to refer to an abstract channel and its matrix representation. The action of a channel $\mathcal{E}$ on a density matrix ρ now corresponds to the standard matrix action on the vector $|\rho\rangle\!\rangle$, hence for a density matrix ρ and a POVM element Q in $\mathcal{M}_d$ we have

$$\mathcal{E}|\rho\rangle\!\rangle = |\mathcal{E}(\rho)\rangle\!\rangle, \tag{2.17}$$

$$\mathrm{Tr}(Q\mathcal{E}(\rho)) = \langle\!\langle Q|\mathcal{E}|\rho\rangle\!\rangle. \tag{2.18}$$

The Liouville representation has the nice properties (as can be easily checked) that the composition of quantum channels is equivalent to matrix multiplication of their Liouville matrices and that tensor products of channels correspond to tensor products of the corresponding Liouville matrices, that is, for all channels $\mathcal{E}_1$ and $\mathcal{E}_2$ and all $A \in \mathcal{M}_d$,

$$\begin{aligned} |\mathcal{E}_1 \circ \mathcal{E}_2(A)\rangle\rangle &= \mathcal{E}_1\mathcal{E}_2|A\rangle\rangle \\ |\mathcal{E}_1 \otimes \mathcal{E}_2(A^{\otimes 2})\rangle\rangle &= \mathcal{E}_1 \otimes \mathcal{E}_2|A^{\otimes 2}\rangle\rangle. \end{aligned} \tag{2.19}$$

Note that some properties of quantum channels are more obvious than others in the Liouville representation. In general any superoperator $\mathcal{E}$, in the Pauli basis, can be written as

$$\mathcal{E} = \begin{pmatrix} \alpha & v \\ w & \mathcal{E}_u \end{pmatrix} \tag{2.20}$$

where $\alpha = \mathrm{Tr}(\sigma_0\mathcal{E}(\sigma_0)) \in \mathbb{C}$, $v := [\mathrm{Tr}(\sigma\mathcal{E}(\sigma_0))]_\sigma \in \mathbb{C}^{d^2-1}$, $w = [\mathrm{Tr}(\sigma_0\mathcal{E}(\sigma))]_\sigma \in \mathbb{C}^{d^2-1}$ and $\mathcal{E}_u = [\mathrm{Tr}(\sigma\mathcal{E}(\sigma')]_{\sigma,\sigma'} \in \mathbb{C}^{d^2-1\times d^2-1}$. For a TP map $\mathcal{E}$ it is immediately clear that $\alpha = 1$ and $v = 0$, and moreover that this is also a sufficient condition for trace preservation. Moreover $\mathcal{E}$ is unital if and only if $w = 0$. Hence we will often refer to $\mathcal{E}_u$ as the unital block of $\mathcal{E}$. The CP condition on the other hand is very hard to verify in this representation. This is the most important downside when working with the Liouville representation, and we will occasionally have to work hard to overcome it.

2.4. Quality measures of operations

We have introduced density matrices and POVMs to deal with imperfect quantum states and measurements and have introduced quantum channels to deal with imperfect operations. However, when implementing a quantum operation, we would also like to know precisely how well we have implemented this operation. For this a number of different measures of quality have been thought up over the years, each with their advantages and disadvantages, and each measuring different aspects of the quality of a quantum operation. We will make no attempt at listing all such measures here, merely showing the ones that we think are most relevant to this thesis.

2.4.1. Average fidelity

Typically we will be interested in how closely a quantum channel $\mathcal{E}$ approximates some unitary channel $\mathcal{U}$. In this scenario a popular measure of quality is the so called average fidelity. This measure captures how much the output of $\mathcal{E}$ deviates from that of $\mathcal{U}$ in Hilbert-Schmidt inner product when a random pure state is given as input. Therefore it is a measure of the average behavior of the channel $\mathcal{E}$.

Definition 10 (Average fidelity)**.** Let $\mathcal{E}$ be a quantum channel and $\mathcal{U}$ a unitary quantum channel. Then the average fidelity of $\mathcal{E}$ (w.r.t. $\mathcal{U}$) is defined as

$$F(\mathcal{E},\mathcal{U}) := \int_{\mathrm{Haar}} d\psi\, \mathrm{Tr}(\mathcal{U}(|\psi\rangle\langle\psi|)^\dagger \mathcal{E}(|\psi\rangle\langle\psi|)), \tag{2.21}$$

where the integral is taken over the uniform (or Haar) measure on pure quantum states.

Note that, since unitary channels are invertible, we can always write $F(\mathcal{E},\mathcal{U}) = F(\mathcal{U}^\dagger\mathcal{E},\mathcal{I})$, which is the average fidelity of the channel $\mathcal{U}\mathcal{E}$ w.r.t. the identity channel. We will often drop the $\mathcal{I}$ qualifier from this notation and use the shorthand $F(\mathcal{E})$. When speaking about 'the' average fidelity of a quantum channel we will always mean its average fidelity w.r.t. the identity channel.

The average fidelity has a few rather noteworthy features. The first is that it is linear in it's argument $\mathcal{E}$. This makes it easy to manipulate, but it also means it is insensitive to certain types of behaviors one might like to detect. For instance, if one considers two quantum channels $\mathcal{E}_1, \mathcal{E}_2$ that have the same fidelity, then any convex combination of such channels is also a quantum channel with the same fidelity. The second feature is that it is invariant under unitary conjugation, that is we have

$$F(\mathcal{E}) = F(\mathcal{U}^\dagger\mathcal{E}\mathcal{U}) \tag{2.22}$$

for all unitary quantum channels $\mathcal{U}$. This is a straightforward consequence of the fact that the Haar measure is invariant under unitary action. This invariance, together with the insensitivity to convex combinations of channels means that the average fidelity only provides very crude information about how a quantum channel $\mathcal{E}$ deviates from the identity (or another unitary). However, as we shall see later, these two properties are also key to the protocols that efficiently estimate the average fidelity in real devices.

Finally we would like to note that one can also define the average fidelity of a quantum channel $\mathcal{E}$ in terms of its Liouville representation.

Definition 11 (Average fidelity (equivalent))**.** Let $\mathcal{E}$ be a quantum channel. Then the average fidelity of $\mathcal{E}$ is given by

$$F(\mathcal{E}) = \frac{1}{d+1}\left(\frac{1}{d}\operatorname{Tr}(\mathcal{E}) + 1\right) \tag{2.23}$$

where the trace is taken over the Liouville matrix representation of $\mathcal{E}$.

For a proof that these two definitions are equivalent see e.g. [9].

2.4.2. Unitarity

As mentioned above, the average fidelity only gives crude indications of the behavior of a quantum channel, and sometimes we would like to know more about a specific implementation. One particular thing we would like to learn is whether the noise on some quantum operation is unitary or not. That is, when implementing some unitary $\mathcal{U}$ we would like to see the difference between imperfectly implementing the right unitary (maybe some depolarizing noise crept in) or perfectly implementing the wrong unitary. The reason we would like to know the difference between these two types of errors is because they must be fixed in different manners in actual devices. To distinguish between these two scenarios, recently an interesting measure was proposed [10] which is called the unitarity.

Definition 12 (Unitarity)**.** Let $\mathcal{E}$ be a quantum channel let $\mathcal{E}'$ be the superoperator created by subtracting off the identity component of $\mathcal{E}$, that is, for all ρ we have $\mathcal{E}'(\rho) = \mathcal{E}(\rho) - \mathrm{Tr}(\rho)\mathcal{E}(\mathbb{1}/d)$. The unitarity $u(\mathcal{E})$ of $\mathcal{E}$ is defined as

$$u(\mathcal{E}) = \frac{d}{d-1}\int_{\mathrm{Haar}} d\psi \,\mathrm{Tr}\left(\mathcal{E}'(|\psi\rangle\langle\psi|)^\dagger \mathcal{E}'(|\psi\rangle\langle\psi|)\right), \tag{2.24}$$

where the integral is again taken over the Haar measure on pure quantum states.

There are a few noteworthy things about this definition. Firstly, the inclusion of the subtracted term $\mathrm{Tr}(\rho)\mathcal{E}(\mathbb{1}/d)$ might strike one as odd. The motivation for defining the unitarity this way is that we would like (as we shall later see) that the unitarity is maximal (equal to 1) only for unitary channels. If we had chosen to define the unitarity in terms of $\mathcal{E}$ instead of $\mathcal{E}'$ then there are non-unitary channels that have unitarity equal to one. An example of this is the amplitude damping channel discussed before, for $\lambda = 1$ this channel would have unitarity equal to one (under the naive definition) but is decidedly not a unitary channel. However if one uses $\mathcal{E}'$ instead we have the following theorem.

Theorem 2.1. Let $\mathcal{E}$ be a quantum channel. Then $u(\mathcal{E}) \leq 1$ and moreover $u(\mathcal{E}) = 1$ if and only if $\mathcal{E}$ is a unitary quantum channel.

A proof of this theorem can be found in [10].

The definition of unitarity in definition 12 is thus justified. It is however quite difficult to work with. Luckily there is an equivalent definition involving the Liouville representation of the quantum channel $\mathcal{E}$ that is a lot easier to work with.

Definition 13 (Unitarity (equivalent))**.** Let $\mathcal{E}$ be a quantum channel and let $\mathcal{E}_u$ be the unital block of its Liouville representation. Then the unitarity of $\mathcal{E}$ is defined as

$$u(\mathcal{E}) = \frac{1}{d^2-1}\,\mathrm{Tr}(\mathcal{E}_u^\dagger \mathcal{E}_u), \tag{2.25}$$

where the trace is taken over the matrix $\mathcal{E}_u^\dagger \mathcal{E}_u$.

A proof that definition 12 and definition 13 are equivalent can be found in [10].

Lastly we would like to point out that the unitarity is not by itself strictly speaking a quality measure of a quantum channel $\mathcal{E}$. By this we mean that an implementation $\mathcal{E}$ of some unitary channel $\mathcal{U}$ could have high unitarity (even equal to one) and still be arbitrarily far away from its intended implementation. The unitarity is more intended as an additional diagnostic tool to be used in conjunction with the average fidelity.

3

Representation theory

Not to be confused with group presentations.

Wikipedia page on representation theory

This chapter deals with the representation theory of finite groups. All of the material covered here is standard but we have made an effort to be didactic. We will introduce representations, characters of representations and two powerful tools called Schur's lemma and the character projection formula. We will also discuss some examples of representations that show up in a quantum computing setting as well as some useful lemma's involving tensor products of representations.

In chapter 2 we introduced the basic concept of a finite group. In this chapter we will deal with representation theory, which describes how abstract groups can be represented in terms linear operations. In fact we have defined the Clifford and Pauli groups in terms of linear operations. However such a representation is not necessarily unique and it is worth understanding what the possibilities are. Representation theory will give us powerful tools to understand and analyze protocols like randomized benchmarking, which we will do later in this thesis. In this chapter we will discuss the basics of representation theory. We will occasionally state theorems and lemmas, but since these are all standard results we will not give proofs but rather refer to the textbooks of Fulton & Harris [1] and Goodman & Wallach [2] which cover this material in great detail. In section 3.1 we will introduce the notion of a representation and discuss its structure. We also provide some examples of where representations appear in quantum computing. In section 3.2 we introduce the notion of a character of a representation. Finally in section 3.3 we deal with two extremely useful representation theoretic tools, namely Schur's lemma and the character projection formula. These two tools will feature prominently in later chapters.

3.1. Representations

We will begin with defining what we mean by a representation of a group. A representation of a group is a map between the abstract group and a set of matrices (linear operators) such that the multiplication of those matrices maps back correctly to the multiplication rule inside the abstract group. Formally this means:

Definition 14 (group representation)**.** Let G be a finite group and $\mathcal{M}_V$ be the space of linear transformations of a complex vector space V . Let φ be a map

$$\varphi : \mathsf{G} \to \mathcal{M}_d : g \mapsto \varphi(g). \tag{3.1}$$

If the following property holds:

$$\varphi(g)\varphi(h) = \varphi(gh), \qquad \forall g, h \in \mathsf{G}. \tag{3.2}$$

then we call φ a representation of G.

We will in general assume that the matrices $\varphi(g)$ are unitary. For finite groups this does not result in a loss of generality. Note that the sets of matrices we used to define the Pauli and Clifford groups thus form a representation of these groups (basically by definition). Moreover these representations have the property that every element in G corresponds to a unique matrix in $\mathcal{M}_d$. Representations that have this property are called *faithful representations*. However not all representations have this property. For instance, every finite group G has a particularly simple representation called the *trivial representation* φ_{tr} which is defined as

$$\varphi_{\mathrm{tr}} : \mathsf{G} \to \mathbb{C} : g \mapsto 1. \tag{3.3}$$

We see that this map satisfies definition 14 and is thus a representation (for any group G). It is thus also clear that groups will generally have many possible representations. Luckily it is possible to discover quite a lot of regularity among them.

A first step towards deepening our understanding of representations is given by the notion of a subrepresentation.

Definition 15 (subrepresentation)**.** Let G be a finite group and φ a representation of G on a vector space V. If there exists a non-trivial vector space W such that $\varphi(g)W \subset W$ for all $g \in \mathsf{G}$ then the map $\varphi' = P_W \varphi P_W$ where P_W is the projection onto the subspace W, is also a representation of G and is called a subrepresentation of φ.

One can think of subrepresentations as being able the jointly block-diagonalize all matrices $\varphi(g)$ into a block acting only on W and a block acting only on $V \backslash W$. Note that this implies that $P_{V\backslash W} \varphi P_{V\backslash W}$ is also a subrepresentation of φ.

Given a group G an important subclass of its representations are given by what are called *irreducible representations*. These are precisely the representations that have no subrepresentations. They are in a sense the building blocks of all representations of a group.

Definition 16 (irreducible representations)**.** Let G be a finite group and let φ be a representation of G on a vector space V. If there exists no non-trivial subspace W of V such that $\varphi(g)W \subset W$ for all $g \in \mathsf{G}$ then we call φ an irreducible representation (irrep) of G.

Finally we would like to note that two representations φ, φ' that act on the same space V aren't necessarily the same representation. They might be genuinely different. We call two representations 'equivalent' if there is an invertible linear map that connects them. Formally:

Definition 17 (Equivalent representations)**.** Let G be a finite group and let φ, φ' be representations on a space V. We call φ, φ' *equivalent* if there exists an invertible linear map $T \in \mathcal{M}_V$ such that

$$T\varphi(g) = \varphi'(g)T, \qquad \forall g \in \mathsf{G}. \tag{3.4}$$

Now we see that there is a fair amount of structure in representations. Representations are either irreducible or reducible and if they are reducible then they have subrepresentations (that may be further subdivided into representations). It is also possible that a representation contain multiple equivalent copies of some subrepresentation. Now we can make formal our previous statement that irreducible representations are really the building blocks of all other representations of a group. This result is called Maschke's lemma

Lemma 3.1 (Maschke's Lemma)**.** Let G be a group and let φ be a representation G. Then there are irreducible representations φ_λ of G (for λ in some index set R_G) such that, up to basis transformations, the following holds for all $g \in \mathsf{G}$:

$$\varphi(g) \simeq \bigoplus_{\lambda \in R_\mathsf{G}} \varphi_\lambda(g)^{\oplus n_\lambda} \tag{3.5}$$

where n_λ indicates the number of equivalent copies of the representation φ_λ present in φ.

A proof of this lemma can be found in [1][proposition 1.8].
This means all representations of a group can be built up from (copies of) irreducible representation of that group. We will sometimes (see chapter 8) find it handy to look at a restricted class of representations which we call multiplicity-free representations.

Definition 18 (multiplicity-free representation)**.** Let G be a finite group and φ a representation of G. We call φ a multiplicity-free representation if there exist mutually inequivalent irreps φ_λ ($\lambda \in R_\mathsf{G}$) such that

$$\varphi(g) \simeq \bigoplus_{\lambda \in R_\mathsf{G}} \varphi_\lambda(g) \tag{3.6}$$

for all $g \in \mathsf{G}$.

3.1.1. Examples of representations

Having gone through a fair amount of theory it is useful to see some examples of representations that are relevant to quantum computation. We begin with some representations of the d dimensional unitary group $U(d)$ and the Pauli and Clifford groups.

Standard representation: The line between the unitary group and its standard representation is thin. Usually they are considered the same thing. The unitary group of dimension d is the set of $d \times d$ unitary matrices with matrix multiplication as a composition. This group has a trivial mapping into the linear transformations of a vector space of dimension d. We call this mapping the standard representation. This representation is irreducible (this is easy to see) and also faithful.

Note that the Clifford and Pauli groups also have a standard representation (as they are subgroups of the unitary group). These representations are also irreducible.

Liouville representation: It was already implicit in the naming, but the Liouville representation is a representation of the unitary group $U(d)$ (in the sense of definition 14) on the vector space $\mathcal{M}_d$. This representation is not irreducible, as one can see by noting that $\mathcal{U}(\mathbb{1}) = U\mathbb{1}U^\dagger = \mathbb{1}$ for all $U \in U(d)$. Hence the Liouville representation of the unitary group decomposes into the trivial representation (which is irreducible) and another representation. This other representation is called the adjoint representation, and it turns out that this representation is in fact irreducible [2]. Note that this representation coincides with the restriction of the Liouville representation to its unital component, as discussed in the previous chapter.

The Liouville representation again also is a representation for the Pauli and Clifford groups, and it also decomposes into trivial and adjoint representations (as they are both subgroups of the unitary group). For the Clifford group the adjoint representation is also irreducible [3]. However for the Pauli group this is not the case, and it decomposes further. We discuss this decomposition in chapter 8.

3.1.2. TENSOR REPRESENTATIONS

In the following chapters we will often deal with tensor products of representations. Note that for any pair of representations φ, φ' the tensor product $\varphi \otimes \varphi'$ is also a representation. This representation might however not be irreducible, even if φ, φ' are. However, $\varphi \otimes \varphi'$ does inherit some of the structure of φ, φ'. In this section we will explore this a bit more. In particular we can formulate the following rather powerful lemma which connects φ and φ' to the trivial subrepresentations of $\varphi \otimes \varphi'$ if φ and φ' are irreducible. The proof is somewhat technical, relying on the canonical isomorphism between the Hilbert space $\mathcal{H} \otimes \mathcal{H}^*$ and the space of linear transformations of $\mathcal{H}$ (here $\mathcal{H}^*$ is the dual of $\mathcal{H}$) so we will not discuss it here. A nice proof can be found in the appendix of [4] but also in many standard textbooks on representation theory.

Lemma 3.2. Let φ and φ' be unitary, irreducible finite-dimensional representations of a finite-dimensional group G and let $\{v_i\}, \{w_i\}$ be an orthonormal basis for the spaces V, W carrying φ, φ' respectively. If φ, φ' are equivalent representations (and the basis vectors are labeled such that the intertwining map θ between V and W maps $v_i \mapsto w_i$), then the representation $\varphi \otimes \varphi'^*$ has exactly one trivial subrepresentation spanned by the vector

$$v_{\text{triv}} = \sum_i v_i \otimes w_i^\dagger \tag{3.7}$$

If φ and φ' are not equivalent then $\varphi \otimes \varphi'$ contains no trivial subrepresentation.

Note that this lemma of course also applies to the tensor product of φ with the dual of itself. Moreover, if φ is a real representation it will be equal to its dual, a situation we will see often. We can lift lemma 3.2 from irreducible representations to general representations by invoking lemma 3.1 (Maschke's lemma). This gives the following corollary.

Corollary 3.1. Let φ be a real representation of a finite group G. From Maschke's lemma we know that $\varphi = \oplus_{\lambda \in R_\mathsf{G}} \varphi_\lambda^{\oplus n_\lambda}$ where φ_λ are irreducible representations with multiplicity n_λ. Denote V_{λ_s} the s-th copy of the space V_λ ($s \in \{1, ..., n_\lambda\}$) spanning the s'th copy of φ_λ, and denote $\{v_j^{(\lambda_s)} : j = 1, ..., |V_\lambda|\}$ an orthonormal basis of V_{λ_s} that respect the isomorphisms between equivalent spaces (meaning that $v_j^{(\lambda_s)} \mapsto v_j^{(\lambda_{s'})}$ under the intertwining isomorphism between V_{λ_s} and $V_{\lambda_{s'}}$). Now the trivial subrepresentations of $\varphi \otimes \varphi$ have support exclusively and fully on the space

$$(V \otimes V)^\mathsf{G} = \text{Span}\left\{ \sum_{j=1}^{|V_\lambda|} v_j^{(\lambda_s)} \otimes v_j^{(\lambda_{s'})} \,\middle|\, \forall s, s' = 1, ..., n_\lambda, \, \forall \lambda \in R_\mathsf{G} \right\}. \tag{3.8}$$

We will use this powerful corollary to derive the fitting model of unitarity randomized benchmarking in chapter 4 and use it to even greater effect in chapter 7. In the next section we introduce the notion of the character of a representation, a powerful tool for reasoning about irreducible representations of groups.

3.2. CHARACTERS OF REPRESENTATIONS

The character of a representation is a complex valued function of a group G that is defined in terms of a particular representation:

Definition 19 (Character of a representation)**.** Let G be a finite group and φ a representation of G. The character χ_φ of the representation φ is defined as

$$\chi_\varphi : \mathsf{G} \to \mathbb{C} : g \mapsto \mathrm{Tr}(\varphi(g)). \tag{3.9}$$

It can be easily verified that the character function 'plays nice' with composition of representations. For representations φ, φ' we have the relations

$$\chi_{\varphi\otimes\varphi'} = \chi_\varphi\, \chi_{\varphi'}, \tag{3.10}$$
$$\chi_{\varphi\oplus\varphi'} = \chi_\varphi + \chi_{\varphi'}, \tag{3.11}$$

with suitable generalizations to multiple direct sums and tensor products. Note also that if two representations are equivalent if and only if the they have identical characters (this can be seen from definition 17 and the cyclicity of the trace).

From eq. (3.10) one can see that character functions form a vector space, that is, linear combinations of character functions are again character functions (of some representation). We can define an inner product on this vector space:

Definition 20 (Character inner product)**.** Let G be a finite group and let φ, φ' be representations of G with associated characters $\chi_\varphi, \chi_{\varphi'}$. We can define the character inner product of $\chi_\varphi, \chi_{\varphi'}$ as

$$\langle \chi_\varphi, \chi_{\varphi'} \rangle = \frac{1}{|\mathsf{G}|} \sum_{g\in\mathsf{G}} \overline{\chi_\varphi(g)} \chi_{\varphi'}(g). \tag{3.12}$$

For unitary irreducible representations φ, φ' this inner product has the following striking property

$$\langle \chi_\varphi, \chi_{\varphi'} \rangle = \begin{cases} 0 & \text{if } \varphi \text{ not equivalent to } \varphi', \\ 1 & \text{if } \varphi \text{ equivalent to } \varphi'. \end{cases} \tag{3.13}$$

We can combine this with Maschke's lemma to arrive at the following statement about the inner product of a character χ_φ with itself (we will refer to this as the norm of the character).

Lemma 3.3 (character norm)**.** Let G be a finite group and let φ be a representation of G with a decomposition into irreps $\varphi \simeq \bigoplus_{\lambda\in R_\mathsf{G}} \varphi_\lambda^{\oplus n_\lambda}$. Then the norm of the character χ_φ of φ is given by

$$\langle \chi_\varphi, \chi_\varphi \rangle = \sum_{\lambda\in R_\mathsf{G}} n_\lambda^2. \tag{3.14}$$

This seemingly simple lemma allows us to compute the number of irreducible subrepresentations of a given representation given only information about its character (which is often easier to obtain). Lemma 3.3 will be one of the key ideas behind the results presented in chapter 5.

3.3. SCHUR'S LEMMA AND THE PROJECTION FORMULA

Thus far we haven't really talked about why representations are useful in quantum computation. In this section we will review the interaction between a representation φ acting on a vector space V and the linear transformations on V. In particular we will discuss a seemingly simple but very powerful result called Schur's lemma and also highlight a number of corollaries of Schur's lemma that will form the key structural ideas behind the results in several chapters (chapters 4 and 6 to 8).

3.3.1. SCHUR'S LEMMA AND CONSEQUENCES

We begin by stating Schur's lemma, which is a statement about what kind of operators commute with irreducible representations.

Theorem 3.1 (Schur's lemma)**.** Let G be a finite group and let φ, φ' be irreducible representations of G on spaces V, V'. Let A be a linear map from V to V' such that

$$\varphi(g)'A = \varphi(g)A, \tag{3.15}$$

for all $g \in \mathsf{G}$. Then we have that

$$A = \begin{cases} 0 & \text{if } \varphi \text{ not equivalent to } \varphi', \\ \alpha\mathbb{1} & \text{if } \varphi. \text{ equivalent to } \varphi' \end{cases} \tag{3.16}$$

for some $\alpha \in \mathbb{C}$.

We would like to note here that if φ and φ' are equivalent we can take the trace of both sides of eq. (3.16) to obtain the equality

$$\alpha = \frac{1}{d}\operatorname{Tr}(A) \tag{3.17}$$

where d is the dimension of V. This might seem like a trivial statement but it is extremely useful in cases where A is given implicitly and we want to evaluate α. In particular we will use the above when evaluating 'twirls' of operators. Twirls are a very useful tool to 'force' any operator to conform to eq. (3.15). They are defined as follows:

Definition 21 (Twirl)**.** Let G be a finite group and φ a representation of G on a space V. Let $A \in \mathcal{M}_V$ be a linear transformation of V. Then the twirl of A is defined as

$$\mathcal{T}_\varphi(A) := \frac{1}{|\mathsf{G}|}\sum_{g\in\mathsf{G}} \varphi(g)^\dagger A \varphi(g). \tag{3.18}$$

Note that we have $\varphi(h)\mathcal{T}_\varphi(A) = \mathcal{T}_\varphi(A)\varphi(h)$ for all $h \in \mathsf{G}$. This is easy to verify from the definition of a representation and fact that $h\mathsf{G} = \mathsf{G}$ for all $h \in \mathsf{G}$. This means the twirl of any operator A satisfies eq. (3.15) with respect to the representation φ. If φ is a multiplicity-free representation of G then the twirl of A has a particularly nice form.

Lemma 3.4 (multiplicity-free twirls)**.** Let G be a finite group and let φ be a multiplicity-free representation of G with decomposition

$$\varphi(g) \simeq \bigoplus_{\lambda \in R_{\mathsf{G}}} \varphi_\lambda(g), \qquad \forall g \in \mathsf{G} \tag{3.19}$$

into inequivalent irreducible subrepresentations φ_λ. Then for any linear map transformation A the twirl of A takes the form

$$\mathcal{T}_\varphi(A) = \sum_{\lambda \in R_{\mathsf{G}}} \frac{\mathrm{Tr}(AP_\lambda)}{\mathrm{Tr}(P_\lambda)} P_\lambda, \tag{3.20}$$

where P_λ is the projector onto the support of the representation φ_λ.

This is perhaps the single most useful statement in this chapter. We will use it many times in chapters 4, 6 and 7. Note that this lemma can be easily generalized to deal with the case of non multiplicity-free representations, but we will not need this more general statement.

3.3.2. The projection formula

Finally we would like to point out another useful structural lemma regarding representations. This lemma involves averages over the representation of a group, weighted by a character function. It turns out such averages have a particularly neat structure.

Lemma 3.5 (Projection formula)**.** Let G be a group and let $\varphi \simeq \bigoplus_{\lambda \in R_{\mathsf{G}}} \varphi_\lambda^{\oplus n_\lambda}$ be a representation of G. Choose a particular irreducible subrepresentation $\varphi_{\lambda'}$ of φ with associated character function $\chi_{\varphi_{\lambda'}}$. Then the following formula holds

$$\frac{|\varphi_{\lambda'}|}{|\mathsf{G}|} \sum_{g \in \mathsf{G}} \overline{\chi_{\varphi_{\lambda'}}(g)} \varphi(g) = \sum_{s_{\lambda'} \in \{1, \ldots n_{\lambda'}\}} P_{s'_\lambda} \tag{3.21}$$

where $P_{s_{\lambda'}}$ is the projector onto $V_{s_{\lambda'}}$, the support of the $s_{\lambda'}$'th copy of $\varphi_{\lambda'}$ in φ and $|\varphi_{\lambda'}|$ is the dimension of the representation $\varphi_{\lambda'}$.

Note that in the presence of subrepresentations φ equivalent to $\varphi_{\lambda'}$, the projector on the RHS of eq. (3.21) projects onto all subrepresentations that are equivalent to φ' rather than just $\varphi_{\lambda'}$. This lemma will be one of the key structural elements of chapter 8.

A notable special case of lemma 3.5 occurs when $\varphi_{\lambda'}$ is the trivial representation. In that case the character $\chi_{\varphi_{\lambda'}}$ is just a constant function and eq. (3.21) reduces to

$$\frac{1}{|\mathsf{G}|} \sum_{g \in \mathsf{G}} \varphi(g) = P_{\mathrm{triv}}, \tag{3.22}$$

where P_{triv} is the projector onto the subspace of vectors that are left invariant by $\varphi(g)$ for all g. This special case of lemma 3.5 combined with corollary 3.1 will be of great use in chapter 7.

4

Randomized benchmarking

I look forward to reading your thesis:
everything you never wanted to know about randomized benchmarking.

Christian Dickel

In this chapter we will discuss two protocols to test quantum computers: randomized benchmarking and unitarity randomized benchmarking. Randomized benchmarking is a standard protocol for estimating the average fidelity of a set of quantum gates while unitarity randomized benchmarking is a variant of randomized benchmarking that estimates the unitarity. We aim to give an overview of what makes these protocols tick, with an explicit focus on the underlying representation theoretic notions.

As discussed in chapter 1, one of the goals of this thesis is to advance understanding of how errors can be diagnosed in quantum devices. There we also outlined a list of criteria that a quantum diagnostic should aspire to meet. In this chapter we will discuss two protocols that, wholly or in part, meet these criteria. These protocols are randomized benchmarking [1–4], and its younger variant: unitarity randomized benchmarking [5]. Randomized benchmarking is probably the industry standard for diagnosing errors in quantum computers at this point [6–10], and unitarity randomized benchmarking, while newer, is beginning to find its way to experimental practice [11]. In section 4.1 we will discuss the what-and-how of randomized benchmarking, describing how to perform randomized benchmarking in practice, working out the representation theoretic basis of its functioning and finally discussing its performance with respect to the criteria listed in chapter 1. This raises a number of issues with the protocol that we will attempt to resolve later in this thesis. In section 4.2 we will do the same for unitarity randomized benchmarking.

4.1. THE RANDOMIZED BENCHMARKING PROTOCOL

Randomized benchmarking is a protocol for assessing the quality of a set of quantum operations, often called a gateset. This quality can be quantified by the average fidelity (see chapter 2). Almost always the gateset will be a group, so we will use the notation G for gatesets as well. The most prominent features of randomized benchmarking are its efficiency, meaning it can be used to asses the quality of devices with many qubits; and its resistance to state preparation and measurement (SPAM) errors. This last feature is especially important because state preparation and readout are often the most error prone operations of a quantum device [12], setting a noise floor that limits how accurately the gate quality can be assessed through other protocols such as process tomography [13]. These two qualities have made it a standard tool in recent quantum computing experiments. However this efficiency and robustness come at a price. The protocol has a lot of moving parts and a fair amount of assumptions must be made on the workings of quantum devices that must be satisfied in order for the protocol to produce correct results. We will discuss this more after describing how randomized benchmarking works.

4.1.1. THE PROTOCOL

Randomized benchmarking, following [1], works roughly in the following manner. We begin by preparing our quantum device in some state ρ and then sampling a sequence of gates $\vec{G} = G_1, \ldots, G_m$ at random from a gateset and applying these to ρ. We then compute the inversion gate $(G_1 \cdots G_m)^\dagger$ and also apply the resultant operation (ideally bringing the state of the system back to ρ). Measuring the expectation value the resulting state with respect to some POVM element Q (and repeating this enough times to obtain good statistics) we obtain some expectation value $p_m(\vec{G})$. We then repeat this process for many random sequences $\vec{G}$ and compute the average expectation value $p_m = \mathbb{E}_{\vec{G}}(p_m(\vec{G}))$. Finally we perform this process for different sequence lengths m obtaining a dataset $\{p_m\}_m$. This data can then be fitted to an exponential model

$$p_m =_{\text{fit}} Af^m + B \tag{4.1}$$

obtaining fitting parameters A, B, f. We will later see that the exponential decay factor f only depends on the quality of the gates in the gateset G while all dependence on the state ρ and the POVM element Q is absorbed into the parameters A, B. The intuition behind this is that the total error on the implemented gates grows with the sequence length m while the SPAM dependence stays constant (as the amount of state preparation and measurement is the same for each sequence length m). The RB protocol is outlined in more detail in fig. 4.1. Note that we have made no comment on what states ρ to prepare and observables Q to measure. We will discuss good choices for ρ and Q in chapter 6, where we also present an adapted version of the RB protocol that has better statistical properties.

1. Choose a random sequence $\vec{G} = (G_1, \ldots, G_m)$ of m. gates independently and uniformly at random from the gateset G and compute the gate $G_{m+1} = (G_m \ldots G_1)^\dagger$.
2. Prepare q qubits in a state ρ.
3. For $t = 1, \ldots, m+1$, apply the gate G_t to ρ.
4. Measure the expectation value $p_m(\vec{G})$ of some observable Q to a suitable precision (By repeating 1-3 for the same sequence L times).
5. Repeat steps 1–5 a total of N times to estimate
$$p_m = |\mathsf{G}|^{-m} \sum_{\vec{G} \in \mathsf{G}^m} p_m(\vec{G})$$
to a suitable precision (implicitly regarding the $p_m(\vec{G})$ as realizations of a random variable P_m). We call the empirical average over the N sampled Clifford sequences $p_{m,N}$.
6. Repeat steps 1–6 for multiple values of m and fit to the decay model
$$p_m =_{\text{fit}} Af^m + B$$
extracting fit parameters A, B and f.

Figure 4.1: **The Randomized Benchmarking Protocol.**

4.1.2. DERIVATION OF THE FIT MODEL

From the description of the randomized benchmarking protocol in fig. 4.1 it is not clear that the data $\{p_m\}_m$ follows the exponential decay model given in eq. (4.1). We will now justify this model under some assumptions. The first assumption is the **assumption of gate-independent noise**. Formally this means:

Definition 22 (Gate-independent noise)**.** Let G be a gateset. We say that an implemen-

tation of G is subject to gate-independent noise if there exists some quantum channel $\mathcal{E}$ such that for all $G \in \mathsf{G}$ the implementation of G is given by $\mathcal{E}\mathcal{G}$.

Secondly we must make some assumption on the representation theoretic structure of the gateset G (which we assume to be a group). In particular we must assume that the Liouville representation of the group G is the same as the Liouville representation of the full unitary group. This is called the 2-design property (for groups)

Definition 23 (2-design)**.** Let G be a group with Liouville representation $\mathcal{G}$. We call G a 2-design if $\mathcal{G}$ decomposes into two irreducible representations; the trivial representation carried by the space $\mathrm{Span}\{\sigma_0\}$ and the adjoint representation carried by the space $\mathrm{Span}\{\sigma \;\|\; \sigma \in \boldsymbol{\sigma}_q\}$.

We note that the Clifford group is in fact a 2-design and is thus a suitable gateset [14]. Under these two assumptions we can derive the fit model eq. (4.1). In fact we will derive a slightly more general statement that holds for groups G that have a multiplicity-free Liouville representation. Note that any group 2-design falls under this category. Original versions of this proof are due to [1, 15]. We will give a version given in [16, 17] which emphasizes generality and representation theory. We have:

Lemma 4.1 (Fitting model)**.** Let G be a group gateset such that the Liouville representation $\mathcal{G} = \oplus_{\lambda \in R_{\mathsf{G}}} \phi_\lambda(G)$ is multiplicity-free. Let $\tilde{\mathcal{G}} = \mathcal{E}\mathcal{G}$ be some implementation of the operation $G \in \mathsf{G}$ with $\mathcal{E}$ a quantum channel. Consider the average survival probability p_m of a RB experiment involving G of sequence length m with an input state ρ and an output two-component POVM $\{Q, \mathbb{1} - Q\}$,

$$p_m = |\mathsf{G}|^{-m} \sum_{G_1,\ldots,G_m \in \mathsf{G}} \langle\langle Q|\widetilde{\mathcal{G}}_{\mathrm{inv}}\widetilde{\mathcal{G}}_m \cdots \widetilde{\mathcal{G}}_1|\rho\rangle\rangle. \tag{4.2}$$

We now have that

$$p_m = |\mathsf{G}|^{-m} \langle\langle Q| \left(\sum_{G \in \mathsf{G}} \mathcal{G}^\dagger \mathcal{E}\mathcal{G} \right)^m |\rho\rangle\rangle = \sum_{\lambda \in R_{\mathsf{G}}} \langle\langle Q|\mathcal{P}_\lambda|\rho\rangle\rangle f_\lambda^m. \tag{4.3}$$

Proof. We begin by noting that $\tilde{\mathcal{G}}_{\mathrm{inv}} = \mathcal{E}\mathcal{G}_1^\dagger \cdots \mathcal{G}_m^\dagger$ Using this and the fact that $\tilde{\mathcal{G}} = \mathcal{E}\mathcal{G}$ for all $G \in \mathsf{G}$ we can write

$$p_m = |\mathsf{G}|^{-m} \sum_{G_1,\ldots G_m} \langle\langle Q|\widetilde{\mathcal{G}}_{\mathrm{inv}}\widetilde{\mathcal{G}}_m \cdots \widetilde{\mathcal{G}}_1|\rho\rangle\rangle, \tag{4.4}$$

$$= |\mathsf{G}|^{-m} \langle\langle Q| \sum_{G_1,\ldots,G_m \in \mathsf{G}} \mathcal{E}\mathcal{G}_1^\dagger \cdots \mathcal{G}_m^\dagger \mathcal{E}\mathcal{G}_m \mathcal{E}\mathcal{G}_{m-1} \cdots \mathcal{E}\mathcal{G}_1|\rho\rangle\rangle. \tag{4.5}$$

Noting that the operator $\sum_{G_m \in \mathsf{G}} \mathcal{G}_m^\dagger \mathcal{E}\mathcal{G}_m$ is an unnormalized twirl over G and thus com-

mutes with $\mathcal{G}$ for all $G \in \mathsf{G}$ we can write

$$p_m = |\mathsf{G}|^{-m} \langle\langle Q| \sum_{G_1,\ldots,G_{m-1}\in\mathsf{G}} \mathcal{E}\mathcal{G}_1^\dagger \cdots \mathcal{G}_{m-1} \left(\sum_{G_m\in\mathsf{G}} \mathcal{G}_m^\dagger \mathcal{E}\mathcal{G}_m \right) \mathcal{E}\mathcal{G}_{m-1} \cdots \mathcal{E}\mathcal{G}_1 |\rho\rangle\rangle, \tag{4.6}$$

$$= |\mathsf{G}|^{-m} \langle\langle Q| \sum_{G_1,\ldots,G_{m-2}\in\mathsf{G}} \mathcal{E}\mathcal{G}_1^\dagger \cdots \mathcal{G}_{m-2} \left(\sum_{G_m\in\mathsf{G}} \mathcal{G}_m^\dagger \mathcal{E}\mathcal{G}_m \right) \left(\sum_{G_{m-1}\in\mathsf{G}} \mathcal{G}_{m-1}^\dagger \mathcal{E}\mathcal{G}_{m-1} \right) \mathcal{E}\mathcal{G}_{m-2} \cdots \mathcal{E}\mathcal{G}_1 |\rho\rangle\rangle. \tag{4.7}$$

Repeating this procedure for G_{m-1} and then G_{m-2} and so forth we obtain

$$p_m = \langle\langle Q| \left(|\mathsf{G}|^{-1} \sum_{G\in\mathsf{G}} \mathcal{G}^\dagger \mathcal{E}\mathcal{G} \right)^m |\rho\rangle\rangle, \tag{4.8}$$

where we have set $Q \to \mathcal{E}^\dagger(Q)$. Now we use Schur's lemma (lemma 3.4) and the fact that $\mathcal{G} = \bigoplus_{\lambda\in R_\mathsf{G}} \phi_\lambda(G)$ to obtain

$$p_m = \langle\langle Q| \left(\sum_{\lambda\in R_\mathsf{G}} f_\lambda \mathcal{P}_\lambda \right)^m |\rho\rangle\rangle, \tag{4.9}$$

$$= \sum_{\lambda\in R_\mathsf{G}} f_\lambda^m \langle\langle Q|\mathcal{P}_\lambda|\rho\rangle\rangle, \tag{4.10}$$

where $\mathcal{P}_\lambda$ is the projector onto the support of ϕ_λ and we have set $f_\lambda := \operatorname{Tr}(\mathcal{P}_\lambda\mathcal{E})/\operatorname{Tr}(\mathcal{P}_\lambda)$. In the last equality we have also used that the $\mathcal{P}_\lambda$ are mutually orthogonal projectors and hence $\mathcal{P}_\lambda\mathcal{P}_{\lambda'} = \mathcal{P}_\lambda\delta_{\lambda\lambda'}$. ■

From this and the 2-design assumption we see that randomized benchmarking data $\{p_m\}_m$ can be fitted to

$$p_m = s^m \langle\langle Q|\sigma_0\rangle\rangle\langle\langle\sigma_0|\rho\rangle\rangle + f^m \frac{1}{|\boldsymbol{\sigma}_q|} \sum_{\sigma\in\boldsymbol{\sigma}_q} \langle\langle Q|\sigma\rangle\rangle\langle\langle\sigma|\rho\rangle\rangle, \tag{4.11}$$

where

$$s = \langle\langle\sigma_0|\mathcal{E}|\sigma_0\rangle\rangle, \tag{4.12}$$

$$f = \frac{1}{|\boldsymbol{\sigma}_q|} \sum_{\sigma\in\boldsymbol{\sigma}_q} \langle\langle\sigma|\mathcal{E}|\sigma\rangle\rangle. \tag{4.13}$$

Noting that $s = 1$ if $\mathcal{E}$ is trace preserving we obtain eq. (4.1). From definition 11 we also see that f is related to the average fidelity of $\mathcal{E}$ (assuming $\mathcal{E}$ is TP). We have

$$F(\mathcal{E}) = 1 - \frac{d-1}{d}(1-f). \tag{4.14}$$

4.1.3. Discussion

In this section we will spend some time critically discussing the randomized benchmarking protocol. We will do this using the criteria we developed in chapter 1. We will not go over all the criteria in detail but focus on the ones where we think there are challenges to be addressed or conversely where we think randomized benchmarking passes a test with flying colors. In particular we will discuss the efficiency of randomized benchmarking and also its robustness to the violation of its underlying assumptions.

Efficiency: When thinking about efficiency we must consider two types of resources: the amount of classical pre-and post-processing that must be done to perform the protocol and also the amount of times a quantum operation such as applying a gate, preparing a state or making a measurement must be done. When considering classical processing, randomized benchmarking is on the face of it quite scalable. There are three types of classical operations that must be performed. The first is that we must be able to generate gates from the gateset uniformly at random. This if the gateset is the Clifford group (which is the most popular choice) this can be done in classical polynomial time using for instance the algorithm given in [18] which is even quite efficient in practice. Secondly we must, given a sequence of gates $\vec{G}$ be able to compute its inversion gate G_{inv}. If we choose the gateset G to be the Clifford group then this can be done in classical polynomial time as per the Gottesman-Knill theorem [19, 20]. The last form of classical computation is to, given the data points p_m, fit these to a single exponential. For this problem a variety of fast standard methods such as least-squares optimization are available. Note that if we relax the requirement that G be a 2-design this last conclusion becomes less straightforward as the RB data can no longer be described in terms of a single exponential decay, making the fitting process harder in practice. We will address this in great detail in chapter 8.

Less clear is the amount of quantum resources needed. When computing p_m we implicitly assume that we can perform an average over all $p_m(\vec{G})$. Note however that the number of sequences $\vec{G}$ is equal to $|\mathsf{G}|^m$. Given that $|\mathsf{G}| = O(2^{q^2})$ if G is the Clifford group [19], this means that the number of sequences $\vec{G}$ grows exponentially in both sequence length m and number of qubits q. We can find a way out of this unsatisfactory situation by realizing that we must not know p_m exactly but rather can compute some empirical estimate $p_{m,N}$ by sampling N sequences uniformly at random and averaging only over those. Using the law of large numbers in a quantitative form, such as the Hoeffding inequality [21], and recognizing that p_m is the mean of a bounded random variable (for all m and q) we see that we can approximate p_m with a resource cost independent of q and m, making the protocol efficient in theory. However the actual resource costs stemming from this argument are impractically high, and we will expend a lot of energy in chapter 6 to get this resource cost down to a level that is actually feasible in experiments.

So in summary, the randomized benchmarking protocol is efficient both in the complexity theoretic sense and in the practical sense, provided some extra statistical work is done and provided the group being benchmarked is efficiently tractable (the Clifford group being the prime example of such a group).

Robustness: The randomized benchmarking protocol is subject to a number of assumptions. We have already made explicit two of these, namely the gate-independent noise assumption and the 2-design assumption. One can see that these two assumptions exhibit different levels of 'danger'. The 2-design assumption is not an assumption made on the quantum device being benchmarked but rather is baked into the design of the protocol itself. If one has chosen the correct gateset, and can sample from this gateset in a correct manner (both of which are requirements that do not depend on the quantum device being tested) then the 2-design assumption will be satisfied. It is therefore very hard to break this assumption without being aware of doing so.

The gate-independent noise assumption on the other hand is much more problematic. It is critical to the derivation in lemma 4.1, and is moreover unlikely to be satisfied in an experimental setting because (a) different types of gates will typically suffer from different types of noise: in many platforms for instance the 2-qubit gates such as the CNOT gates will have a much higher noise level than the single qubit gates , and (b) general Clifford gates will not be implemented directly but rather composed out of generator operations (see definition 6). This means that different Clifford gates will have different associated noise maps even if the generators all have equal noise.

To make matters worse, the randomized benchmarking protocol is not robust to violations of the assumption of gate independent noise. To see this consider the simple example where every gate $G \in \mathsf{G}$ has an implementation $\mathcal{E}_G\mathcal{G}$ with $\mathcal{E}_G$ being precisely the inverse of $\mathcal{G}$; that is $\mathcal{E}_G = \mathcal{G}^\dagger$. This means that every every gate will be implemented as effectively the identity. As one can easily infer from the RB protocol description in this case we would have p_m a constant for all sequence lengths m. The naive conclusion drawn would be that the depolarizing parameter f and hence the average gate fidelity is equal to 1. However the implemented gates are clearly not perfect! It is also possible to find less nefarious examples of gate dependent noise where randomized benchmarking gives mistaken conclusions [22]. The above example is part of a class of noise models called 'Markovian noise'. Here the assumption is made that associated to every gate $G \in \mathsf{G}$ there is a unique CPTP map $\mathcal{E}_G$ such that G is implemented as $\mathcal{E}_G\mathcal{G}$. The lack of robustness of RB against violations of the gate-independent noise assumption was already recognized in the initial proposals for RB [3] where an updated model taking into account gate dependent noise was proposed. This model was critiqued and subsequently improved in [22, 23] leading to a much more robust analysis of randomized benchmarking. In this analysis randomized benchmarking data can always be fitted to a single exponential decay provided the noise is Markovian and the gate dependence not too strong (in a well specified sense), and the decay constant can be interpreted as the average fidelity of the average noise between two random gates. This analysis is too sophisticated for this introductory chapter but we will revisit it in chapter 8 where we will generalize it to a broader class of RB style protocols.

Lastly we note that Markovian noise is not the most general noise model possible, we could for instance let the map $\mathcal{E}_G$ depend on the time at which G is applied, or let it depend on which gates have been applied before G. Such more general noise model are collectively called 'non-Markovian'. This type of noise is something that actually occurs

in experiments [24], and it is unclear how randomized benchmarking (or any other type of diagnostic tool) can really be made robust against non-Markovianity. Some research in this direction has been done [25] and we suspect interesting things can be said by considering non-Markovian but restricted noise models that are sourced from experimental practice.

There are some other, less explicitly stated assumptions underlying the workings of randomized benchmarking. The first is the assumption that the measurement POVM $\{Q, \mathbb{1} - Q\}$ and the input state ρ do not depend on the drawn sequence $\vec{G}$. Violation of this assumption can be considered a form of non-Markovianity. However this assumption is not always respected in real devices, where the quality of a measurement can change as the sequence length increases, due to e.g. heating effects [26, 27]. The second is the assumption of CPTP-ness of the implementations of the gates G. In particular the trace preserving property of the implementation is often not respected. This phenomenon is known as leakage in the experimental literature and it occurs when the qubit (or qubits) being benchmarked has extra, non-computational, degrees of freedom. This could for instance be a nearby energy level (as in superconducting qubits [28]) , or an extra spin degree of freedom (as in NV centers [29]). This leakage can be characterized separately [30] and taken into account when performing randomized benchmarking. However, some subtleties arise when the measurement POVM also has access to these hidden degrees of freedom. This is for instance the case in superconducting qubits and there is no rigorous framework for dealing with this scenario (but see for instance [31] for some work in this direction).

This concludes our initial discussion of randomized benchmarking. Next we will discuss a related protocol called unitarity randomized benchmarking.

4.2. The unitarity randomized benchmarking protocol

The unitarity randomized benchmarking protocol (introduced first in [5]) is a protocol derived from standard randomized benchmarking, using many of the same ideas. The goal of URB is to measure the unitarity of the gates in a gateset G. It shares the strengths of RB, in that it delivers this estimate of the unitarity in a manner that is independent of SPAM errors. It also shares some of the weaknesses of RB, in that we must assume that each gate in the gateset is implemented with approximately the same noise in other to reliable extract a number we might feasibly call the unitarity. Moreover, it is less clear that the URB protocol is actually scalable in the number of qubits. It is possible to write down a scalable implementation of the URB protocol but this requires the use of ancilla qubits and moreover some strong assumptions on the structure of the noise affecting the gates. There are also non-scalable versions of the protocol that do not need these extra assumptions. We will discuss the differences between these implementations in more detail in chapter 7.

4.2.1. The protocol

In this section we discuss a bare bones version of the URB protocol. We will skip over some of the more thorny aspects of its practical implementation, leaving these for a more

detailed discussion in chapter 7 where we propose an adapted version of the URB protocol and perform a detailed analysis of this adapted protocol.

The URB protocol works similar to the standard RB protocol with some key modifications. We begin by sampling a sequence of random gates $\vec{G}$ of length m from the gateset G. We then apply these to a state ρ and measure, obtaining some probability $p_m(\vec{G})$. Note however that we have not inverted the sequence of gates back to the identity, as we did in standard RB. We repeat this procedure for many sequences $\vec{G}$ and finally compute the average *of the squares* $q_m = \mathbb{E}_{\vec{G}}(p_m^2(\vec{G}))$. This squaring of the probabilities is a second key difference between URB and standard RB. Finally we repeat this procedure for many different sequence lengths m and fit the resulting data to the model

$$q_m =_{\text{fit}} Au^{m-1} + B \tag{4.15}$$

We shall see in the next section that the fitting parameter u, under some assumptions, can be seen as the unitarity of the gates in the gateset G. We give a more detailed outline of the protocol in fig. 4.2.

1. Choose a random sequence $\vec{G} = (G_1, \ldots, G_m)$ of m. gates independently and uniformly at random from the gateset G.
2. Prepare q qubits in a state ρ.
3. For $t = 1, \ldots, m$, apply the gate G_t to ρ.
4. Measure the expectation value $p_m(\vec{G})$ of some POVM element Q to a suitable precision (By repeating 1-3 for the same sequence L times).
5. Repeat steps 1–5 a total of N times to estimate
$$q_m = |\mathsf{G}|^{-m} \sum_{\vec{G} \in \mathsf{G}^m} p_m^2(\vec{G})$$
to a suitable precision. We call the empirical average over the N sampled Clifford sequences $q_{m,N}$.
6. Repeat steps 1–6 for multiple values of m and fit to the decay model
$$q_m =_{\text{fit}} Au^{m-1} + B$$
extracting fit parameters A, B and u.

Figure 4.2: **The Unitarity Randomized Benchmarking Protocol.**

Throughout the rest of this thesis we shall refer to the number $q_m(\vec{G}) = p_m^2(\vec{G})$ as the sequence purity (of the sequence $\vec{G}$) and to q_m as the average sequence purity.

4.2.2. DERIVATION OF THE FIT MODEL

In this section we will justify the fitting model given in eq. (4.15). We will make the same assumptions we made in the standard RB protocol, namely the assumption of gate-independent noise (definition 22) and the assumption that G is a 2-design (definition 23). With these two assumptions we can derive the following lemma:

Lemma 4.2 (URB fitting model)**.** Let G be a finite group that is also a 2-design, and let $\mathcal{E}$ be a quantum channel such that for every gate $G \in \mathsf{G}$ there is an implementation $\widetilde{\mathcal{G}} = \mathcal{E}\mathcal{G}$ of G. For some input state ρ and two-component output POVM $\{Q, \mathbb{1} - Q\}$ we now have that the average sequence purity

$$q_m = |\mathsf{G}|^{-m} \sum_{\vec{G} \in \mathsf{G}^m} p_m^2(\vec{G}),$$

with $p_m(\vec{G}) = \langle\!\langle Q | \widetilde{\mathcal{G}}_m \dots \widetilde{\mathcal{G}}_1 | \rho \rangle\!\rangle$ is of the form

$$q_m = A u^{m-1} + B, \tag{4.16}$$

where $u = u(\mathcal{E})$ is the unitarity of the map $\mathcal{E}$ and the parameters A, B depend only on SPAM (ρ and Q).

Proof. We begin by explicitly writing out the average sequence purity q_m:

$$q_m = |\mathsf{G}|^{-m} \sum_{\vec{G} \in \mathsf{G}^m} \langle\!\langle Q | \widetilde{\mathcal{G}}_m \dots \widetilde{\mathcal{G}}_1 | \rho \rangle\!\rangle^2, \tag{4.17}$$

$$= |\mathsf{G}|^{-m} \sum_{\vec{G} \in \mathsf{G}^m} \langle\!\langle Q^{\otimes 2} | \mathcal{E}^{\otimes 2} \mathcal{G}^{\otimes 2} \cdots \mathcal{E}^{\otimes 2} \mathcal{G}_1^{\otimes 2} | \rho \rangle\!\rangle, \tag{4.18}$$

where we have used the gate-independent noise assumption as well as the basic fact that $\operatorname{Tr}(A)\operatorname{Tr}(B) = \operatorname{Tr}(A \otimes B)$ for all linear operators A, B. Now we use the linearity of the trace to rewrite this as

$$q_m = \langle\!\langle Q | \left[\mathcal{E}^{\otimes 2} \frac{1}{|\mathsf{G}|} \sum_{G \in \mathsf{G}} \mathcal{G}^{\otimes 2} \right]^m | \rho \rangle\!\rangle. \tag{4.19}$$

Next, we recognize that since $\mathcal{G}$ is a representation of G, so is $\mathcal{G}^{\otimes 2}$. This means we can apply the techniques from chapter 3 to analyze the above equation. In particular we will use the corollary of the projection formula (lemma 6.2) given in eq. (3.22), which states that

$$\frac{1}{|\mathsf{G}|} \sum_{G \in \mathsf{G}} \mathcal{G}^{\otimes 2} = \mathcal{P}_{\text{triv}}, \tag{4.20}$$

where $\mathcal{P}_{\text{triv}}$ is the projection onto the subspace carrying the trivial subrepresentations of $\mathcal{G}^{\otimes 2}$. Using corollary 3.1 and the assumption that G is a 2-design we can characterize this projector as

$$\mathcal{P}_{\text{triv}} = |\sigma_0^{\otimes 2}\rangle\!\rangle\langle\!\langle \sigma_0^{\otimes 2}| + \left| \frac{1}{\sqrt{|\boldsymbol{\sigma}_q|}} \sum_{\sigma \in \boldsymbol{\sigma}_q} \sigma^{\otimes 2} \right\rangle\!\!\right\rangle \left\langle\!\!\left\langle \frac{1}{\sqrt{|\boldsymbol{\sigma}_q|}} \sum_{\sigma \in \boldsymbol{\sigma}_q} \sigma^{\otimes 2} \right| . \tag{4.21}$$

This means we can further rewrite the expression for q_m as

$$q_m = \langle\langle Q^{\otimes 2} | (\mathcal{P}_{\text{triv}} \mathcal{E}^{\otimes 2} \mathcal{P}_{\text{triv}})^{m-1} | \rho^{\otimes 2} \rangle\rangle = \langle\langle Q^{\otimes 2} | \mathcal{M}^{m-1} | \rho^{\otimes 2} \rangle\rangle, \tag{4.22}$$

where we absorbed the $\mathcal{E}^\dagger$ factor into Q and where $\mathcal{M} = \mathcal{P}_{\text{triv}} \mathcal{E}^{\otimes 2} \mathcal{P}_{\text{triv}}$ is a rank two matrix. Using the definition of $\mathcal{P}_{\text{triv}}$ we can find the non-zero components of $\mathcal{M}$ as

$$\mathcal{M}_{00} = \langle\langle \sigma_0^{\otimes 2} | \mathcal{E}^{\otimes 2} | \sigma_0^{\otimes 2} \rangle\rangle = 1, \tag{4.23}$$

$$\mathcal{M}_{01} = \langle\langle \sigma_0^{\otimes 2} | \mathcal{E}^{\otimes 2} | \frac{1}{\sqrt{|\boldsymbol{\sigma}_q|}} \sum_{\sigma \in \boldsymbol{\sigma}_q} \sigma^{\otimes 2} \rangle\rangle = 0, \tag{4.24}$$

$$\mathcal{M}_{10} = \langle\langle \frac{1}{\sqrt{|\boldsymbol{\sigma}_q|}} \sum_{\sigma \in \boldsymbol{\sigma}_q} \sigma^{\otimes 2} | \mathcal{E}^{\otimes 2} | \sigma_0^{\otimes 2} \rangle\rangle = \frac{1}{\sqrt{|\boldsymbol{\sigma}_q|}} \sum_{\sigma \in \boldsymbol{\sigma}_q} \text{Tr}(\sigma \mathcal{E}(\sigma_0))^2, \tag{4.25}$$

$$\mathcal{M}_{11} = \langle\langle \frac{1}{\sqrt{|\boldsymbol{\sigma}_q|}} \sum_{\sigma \in \boldsymbol{\sigma}_q} \sigma^{\otimes 2} | \mathcal{E}^{\otimes 2} | \frac{1}{\sqrt{|\boldsymbol{\sigma}_q|}} \sum_{\sigma \in \boldsymbol{\sigma}_q} \sigma^{\otimes 2} \rangle\rangle = \frac{1}{|\boldsymbol{\sigma}_q|} \text{Tr}(\mathcal{E}_u \mathcal{E}_u^\dagger) = u(\mathcal{E}), \tag{4.26}$$

where we used the fact that $\mathcal{E}$ is trace preserving and the definition of the unitarity. Note that $\mathcal{M}$ is a lower-triangular matrix with eigenvalues $u, 1$. This means $\mathcal{M}^{m-1}$ is also a lower triangular matrix with eigenvalues $1, u^{m-1}$. Feeding this back into the expression for q_m we obtain the stated fitting relation. ■

4.2.3. DISCUSSION

Similarly to randomized benchmarking, URB has advantages and weaknesses that can be assessed using the criteria proposed in chapter 1. Since URB is structurally quite similar to RB we will not repeat the much of the discussion of section 4.1 but rather point out where URB differs substantially from RB. We will again discuss only the criteria we consider the most relevant, namely the efficiency of the protocol and its robustness.

Efficiency There are again two things to consider: classical computational resources and quantum (sampling) costs. The classical resources needed do for URB do not significantly differ from those needed for standard RB (provided the gateset used is the Clifford group), so we will not repeat this discussion. With regards to the quantum resources needed, in particular the amount of times the device must be measured to obtain enough data for statistical inference, there is a subtlety that makes the URB protocol as introduced in fig. 4.2 non-scalable in the number of qubits q. To see this we must dive into the proof of lemma 4.2, investigating the $\mathcal{M}_{11}$ component of the matrix $\mathcal{M}$ given in eq. (4.26). This is the matrix component that contributes the parameter $u(\mathcal{E})$, i.e. the unitarity, to the fitting relation eq. (4.15). However, as one can surmise from the structure of the matrix $\mathcal{M}$, the prefactor A in eq. (4.15) is given by

$$A = \frac{1}{|\boldsymbol{\sigma}_q|} \langle\langle Q | \sum_{\sigma \in \boldsymbol{\sigma}_q} \sigma^{\otimes 2} \rangle\rangle \langle\langle \sum_{\sigma \in \boldsymbol{\sigma}_q} \sigma^{\otimes 2} | \rho^{\otimes 2} \rangle\rangle \tag{4.27}$$

for any input state ρ and POVM element Q. We will not prove it here but it can be seen (see for instance proposition 7.8 in chapter 7) that A scales as $O(2^{-q/2})$ for any state ρ and POVM element Q . This means that the prefactor A will become small exponentially

quickly as we increase the number of qubits. Hence the signal from which we want to extract an exponential decay becomes small as well, meaning we need to know each data point exponentially (in q) precise. This means the amount of times we must query the quantum system (in order to reconstruct the probabilities $p_m(\vec{G})$) grows rapidly with q. Thus the URB protocol is, as stated, not scalable in the number of qubits q. However it will be in practice efficient enough to use for systems of not too many qubits (say 10). It is possible to write down a version of the URB protocol that avoids this issue and is explicitly scalable with respect to q but it requires two copies of the device under investigation, the ability to create entanglement between these devices and moreover that we make extra assumptions on the noise affecting the quantum devices. We will discuss this scalable protocol in more detail in chapter 7.

Robustness The URB protocol, as it is presented above, also relies on the assumption of gate-independent noise for its proper functioning. However what happens when this assumption is broken has been studied far less than in the case of standard RB. Therefore it is not clear how robust unitarity randomized benchmarking is in the presence of gate-dependent noise. One way to make progress in this regard is by noting that URB can be seen as a special case of the extended randomized benchmarking protocol discussed in chapter 8. Hence the robustness results we derive in chapter 8, which deal with the case of Markovian noise, also hold for URB. However some interpretative issues remain here.

5

Representation theory of the Clifford group

The q-qubit Clifford group, that is, the normalizer of the q-qubit Pauli group in $U(2^q)$, is a fundamental structure in quantum information with a wide variety of applications. We characterize all irreducible subrepresentations of the two-copy representation $\varphi^{\otimes 2}$ of the Clifford group on the two-fold tensor product of the space of linear operators $\mathcal{M}_{2^q}^{\otimes 2}$. In chapter 6 and chapter 7 we will use this decomposition to analyze the statistics of randomized benchmarking and unitarity randomized benchmarking.

This chapter has been published, with minor changes, in J. Helsen, J.J. Wallman & S. Wehner, *Representation theory of the mulitqubit Clifford group*, Journal of Mathematical Physics, 59 072201 (2018)

5.1. Introduction

Symmetric structures, encoded as groups, play a fundamental role in the study of quantum information theory and quantum mechanics in general. The Pauli group and its normalizer, the Clifford group, are particularly important in quantum information, with applications such as quantum error-correcting codes [1], quantum tomographic methods [2], and quantum data hiding [3]. Furthermore, operations within the Clifford group can be efficiently simulated [4] and the Clifford group is a unitary 2-design [3], that is, averages over the fundamental representation of the Clifford group reproduce the first two moments of the Haar average over the full unitary group [5]. These properties make the Clifford group useful for characterization protocols for quantum systems such as randomized benchmarking [6].

A subgroup of the unitary group (in our case the Clifford group) is a unitary t-design if the irreducible subrepresentations of t tensor copies of its standard representation are in one-to-one correspondence with the irreducible subrepresentations of the same construction involving the the full unitary group [7]. This equivalent definition is useful because the tensor representations of the unitary group are well understood via Schur-Weyl duality [8].

Recently it has been shown that the q-qubit Clifford group is also a unitary 3-design [9, 10]. However, simultaneously it was shown that the multi-qubit Clifford group is not a unitary 4-design. Consequently, the representation of 4 tensor copies of the standard representation of the Clifford group differs from the same construction using the unitary group. In this chapter we will analyze a closely related representation of the Clifford group which we call the two-copy representation. This representation is the tensor product of two tensor copies of the standard representation and two tensor copies of the dual of the standard representation. The structure of the two-copy representation of the single-qubit Clifford group was analyzed in [11] and used to analyze the statistical performance of randomized benchmarking.

In this chapter we provide a complete analysis of the two-copy representation of the multi-qubit Clifford group for any number of qubits. In chapter 6 and chapter 7 we use these results to analyze multi-qubit randomized benchmarking and unitarity randomized benchmarking, leading to a substantial reduction in the amount of data required to obtain rigorous and precise estimates using these procedures.

5.2. The two-copy representation of the multi-qubit Clifford group

We begin by defining what we mean by the two-copy representation of the Clifford group. In order to do this we must first set some notation regarding the Pauli matrices (definition 4). Recall from chapter 2 that the set of normalized Pauli matrices $\boldsymbol{\sigma}_q$ (see eq. (2.14), together with the normalized identity σ_0 forms an orthonormal basis (with respect to the Hilbert-Schmidt inner product) for the Hilbert space $\mathcal{M}_d$ (for $d = 2^q$). For convenience we will define $\hat{\boldsymbol{\sigma}}_q := \boldsymbol{\sigma}_q \cup \{\sigma_0\}$. We will denote the elements of the set $\boldsymbol{\sigma}_q$ by Greek letters

$(\sigma, \tau, \nu, ...)$. For the case of a single qubit we denote the normalized X, Y, Z Pauli matrices by $\sigma_X, \sigma_Y, \sigma_Z$. We also, for later convenience, introduce the *normalized* matrix product of two normalized Pauli matrices as

$$\sigma \cdot \tau := \sqrt{d}\sigma\tau \qquad \sigma, \tau \in \hat{\boldsymbol{\sigma}}_q. \tag{5.1}$$

Note that $\sigma \cdot \tau \in \pm\hat{\boldsymbol{\sigma}}_q$ if $[\sigma, \tau] = 0$ and $i\sigma \cdot \tau \in \pm\boldsymbol{\sigma}_q$ if $\{\sigma, \tau\} = 0$. Lastly we define the following parametrized subsets of $\boldsymbol{\sigma}_q$ and $\hat{\boldsymbol{\sigma}}_q$. For all $\tau \in \boldsymbol{\sigma}_q$ we define

$$\boldsymbol{N}_\tau := \{\sigma \in \boldsymbol{\sigma}_q \mid \{\sigma, \tau\} = 0\}, \tag{5.2}$$
$$\boldsymbol{C}_\tau := \{\sigma \in \boldsymbol{\sigma}_q \backslash \{\tau\} \mid [\sigma, \tau] = 0\}, \tag{5.3}$$
$$\hat{\boldsymbol{C}}_\tau := \{\sigma \in \hat{\boldsymbol{\sigma}}_q \mid [\sigma, \tau] = 0\}. \tag{5.4}$$

Note that we have $|\hat{\boldsymbol{C}}_\tau| = |\boldsymbol{N}_\tau| = \frac{d^2}{2}$ and $\hat{\boldsymbol{C}}_\tau$ and $\boldsymbol{N}_\tau$ are disjoint for all $\tau \in \boldsymbol{\sigma}_q$. With regard to these sets we can also prove the following lemma which will be useful in later derivations:

Lemma 5.1. Let $\tau, \tau' \in \boldsymbol{\sigma}_q$ and $\tau \neq \tau'$. The following equalities hold

$$|\boldsymbol{N}_\tau \cap \hat{\boldsymbol{C}}_{\tau'}| = |\hat{\boldsymbol{C}}_\tau \cap \hat{\boldsymbol{C}}_{\tau'}| = |\hat{\boldsymbol{C}}_\tau \cap \boldsymbol{N}_{\tau'}| = |\boldsymbol{N}_\tau \cap \boldsymbol{N}_{\tau'}| = \frac{d^2}{4}. \tag{5.5}$$

Also for all $\tau \in \boldsymbol{\sigma}_q$ we have

$$|\boldsymbol{N}_{\sigma_0} \cap \hat{\boldsymbol{C}}_\tau| = |\boldsymbol{N}_{\sigma_0} \cap \boldsymbol{N}_\tau| = 0, \tag{5.6}$$
$$|\hat{\boldsymbol{C}}_{\sigma_0} \cap \hat{\boldsymbol{C}}_\tau| = |\hat{\boldsymbol{C}}_{\sigma_0} \cap \boldsymbol{N}_\tau| = \frac{d^2}{2}. \tag{5.7}$$

Proof. Let $\tau, \tau' \in \boldsymbol{\sigma}_q$ and $\tau \neq \tau'$. We begin by noting that $\boldsymbol{N}_\tau$ is the complement of $\hat{\boldsymbol{C}}_\tau$ in $\hat{\boldsymbol{\sigma}}_q$ and that $|\hat{\boldsymbol{C}}_\tau| = |\boldsymbol{N}_\tau| = \frac{d^2}{2}$ for all $\tau \in \boldsymbol{\sigma}_q$. This allows us to make the following statements

$$|\hat{\boldsymbol{C}}_\tau \cap \hat{\boldsymbol{C}}_{\tau'}| + |\boldsymbol{N}_\tau \cap \hat{\boldsymbol{C}}_{\tau'}| = \frac{d^2}{2}, \qquad |\hat{\boldsymbol{C}}_\tau \cap \hat{\boldsymbol{C}}_{\tau'}| + |\hat{\boldsymbol{C}}_\tau \cap \boldsymbol{N}_{\tau'}| = \frac{d^2}{2}, \tag{5.8}$$
$$|\boldsymbol{N}_\tau \cap \hat{\boldsymbol{C}}_{\tau'}| + |\boldsymbol{N}_\tau \cap \boldsymbol{N}_{\tau'}| = \frac{d^2}{2}, \qquad |\hat{\boldsymbol{C}}_\tau \cap \boldsymbol{N}_{\tau'}| + |\boldsymbol{N}_\tau \cap \boldsymbol{N}_{\tau'}| = \frac{d^2}{2}. \tag{5.9}$$

We can solve this system of equations to obtain

$$|\hat{\boldsymbol{C}}_\tau \cap \hat{\boldsymbol{C}}_{\tau'}| = |\boldsymbol{N}_\tau \cap \boldsymbol{N}_{\tau'}|, \tag{5.10}$$
$$|\boldsymbol{N}_\tau \cap \hat{\boldsymbol{C}}_{\tau'}| = \frac{d^2}{2} - |\boldsymbol{N}_\tau \cap \boldsymbol{N}_{\tau'}|, \tag{5.11}$$
$$|\hat{\boldsymbol{C}}_\tau \cap \boldsymbol{N}_{\tau'}| = \frac{d^2}{2} - |\boldsymbol{N}_\tau \cap \boldsymbol{N}_{\tau'}|. \tag{5.12}$$

The rest of the argument will proceed by induction on the number of qubits q (recall that $d = 2^q$). For $q = 1$ we have that

$$|\boldsymbol{N}_\tau \cap \boldsymbol{N}_{\tau'}| = |\{\tau', i\tau \cdot \tau'\} \cap \{\tau, i\tau \cdot \tau'\}| = |\{i\tau \cdot \tau'\}| = 1 = \frac{2^2}{4}. \tag{5.13}$$

From eqs. (5.10) to (5.12) we then have that

$$|\boldsymbol{N}_\tau \cap \hat{\boldsymbol{C}}_{\tau'}| = |\hat{\boldsymbol{C}}_\tau \cap \hat{\boldsymbol{C}}_{\tau'}| = |\boldsymbol{N}_\tau \cap \hat{\boldsymbol{C}}_{\tau'}| = |\boldsymbol{N}_\tau \cap \boldsymbol{N}_{\tau'}| = 1. \tag{5.14}$$

Now assume eq. (5.5) to hold up to $q-1$. For $\tau \in \boldsymbol{\sigma}_q$ we can write

$$\boldsymbol{N}_\tau = \Big(\boldsymbol{N}_{\tau_1} \otimes \hat{\boldsymbol{C}}_{\tau_{q-1}}\Big) \cup \Big(\hat{\boldsymbol{C}}_{\tau_1} \otimes \boldsymbol{N}_{\tau_{q-1}}\Big), \quad \tau_1 \in \hat{\boldsymbol{\sigma}}_1,\ \tau_{q-1} \in \boldsymbol{\sigma}_{q-1},\ \text{s.t. } \tau_1 \otimes \tau_{q-1} = \tau. \tag{5.15}$$

Where by $\boldsymbol{A} \otimes \boldsymbol{B}$ we mean $\boldsymbol{A} \otimes \boldsymbol{B} := \{a \otimes b \parallel a \in \boldsymbol{A},\ b \in \boldsymbol{B}\}$. Now we can write

$$\begin{aligned}
|\boldsymbol{N}_\tau \cap \boldsymbol{N}_{\tau'}| &= \left|\left[\Big(\boldsymbol{N}_{\tau_1} \otimes \hat{\boldsymbol{C}}_{\tau_{q-1}}\Big) \cup \Big(\hat{\boldsymbol{C}}_{\tau_1} \otimes \boldsymbol{N}_{\tau_{q-1}}\Big)\right] \cap \left[\Big(\boldsymbol{N}_{\tau_1'} \otimes \hat{\boldsymbol{C}}_{\tau_{q-1}'}\Big) \cup \Big(\hat{\boldsymbol{C}}_{\tau_1'} \otimes \boldsymbol{N}_{\tau_{q-1}'}\Big)\right]\right| \\
&= \Bigg| \Big(\{\sigma_0\} \otimes (\boldsymbol{N}_{\tau_{q-1}} \cap \boldsymbol{N}_{\tau_{q-1}'})\Big) \cup \Big(\{i\tau_1 \cdot \tau_1'\} \otimes (\hat{\boldsymbol{C}}_{\tau_{q-1}} \cap \hat{\boldsymbol{C}}_{\tau_{q-1}'})\Big) \\
&\qquad \cup \Big(\{\tau_1\} \otimes (\boldsymbol{N}_{\tau_{q-1}} \cap \hat{\boldsymbol{C}}_{\tau_{q-1}'})\Big) \cup \Big(\{\tau_1'\} \otimes (\hat{\boldsymbol{C}}_{\tau_{q-1}} \cap \boldsymbol{N}_{\tau_{q-1}'})\Big) \Bigg| \\
&= \frac{d^2}{4},
\end{aligned} \tag{5.16}$$

where the last line holds by the induction hypothesis and the fact that all sets in the equation are disjoint. This proves the first half of the lemma. Now take $\tau \in \boldsymbol{\sigma}_q$ and consider the sets $\boldsymbol{N}_{\sigma_0}, \hat{\boldsymbol{C}}_{\sigma_0}$. It is trivial to see that $\boldsymbol{N}_{\sigma_0} = \emptyset$ and $\hat{\boldsymbol{C}}_{\sigma_0} = \hat{\boldsymbol{\sigma}}_q$. Since $|\boldsymbol{N}_\tau| = |\hat{\boldsymbol{C}}_\tau| = \frac{d^2}{2}$ the second half of the lemma also follows. ∎

As mentioned above, $\hat{\boldsymbol{\sigma}}_q$ forms an orthonormal basis for $\mathcal{M}_d$. We can define a representation φ of the Clifford group by its action by conjugation on this basis, we have

$$\varphi : \mathsf{C}_q \to \mathcal{M}_d : C \mapsto \varphi(C)\sigma = C\sigma C^\dagger, \quad \sigma \in \hat{\boldsymbol{\sigma}}_q, \tag{5.17}$$

This representation is the Liouville representation of the Clifford group, as discussed in chapter 3. We will refer to this representation as the 'one-copy' representation, for reasons that will immediately become obvious.

Note that we have for all $C \in \mathsf{C}_q$ and $\sigma \in \hat{\boldsymbol{\sigma}}_q$ that $\varphi(C)\sigma = \pm\tau$ for some $\tau \in \hat{\boldsymbol{\sigma}}_q$. This means that in the basis $\hat{\boldsymbol{\sigma}}_q$ the Clifford group is represented by signed permutation matrices. Note also that the action of the Clifford group through φ is transitive on $\boldsymbol{\sigma}_q$ [12].

Now we define analogously the two-copy representation of the q-qubit Clifford group C_q on the Hilbert space $\mathcal{M}_d \otimes \mathcal{M}_d = \mathcal{M}_d^{\otimes 2}$ (hence the name two-copy representation). We define $\varphi^{\otimes 2}$ with respect to its action on the basis

$$\boldsymbol{\mathcal{B}} = \{\sigma_0 \otimes \sigma_0, \sigma_0 \otimes \sigma, \sigma \otimes \sigma_0, \sigma \otimes \tau \ \parallel\ \sigma, \tau \in \boldsymbol{\sigma}_q\}, \tag{5.18}$$

of $\mathcal{M}_d^{\otimes 2}$. We define the action of $\varphi^{\otimes 2}$ on $\boldsymbol{\mathcal{B}}$ as

$$\varphi^{\otimes 2}(C)\sigma \otimes \tau = \big(C\sigma C^\dagger\big) \otimes \big(C\tau C^\dagger\big), \qquad \sigma \otimes \tau \in \boldsymbol{\mathcal{B}}. \tag{5.19}$$

Note that this representation is equivalent to the representation $C \otimes C^* \otimes C \otimes C^*$. For brevity we will often forget about the tensor product symbol and write $\sigma \otimes \tau$ as $\sigma\tau$ when it is clear from the context. Note that in the basis $\boldsymbol{\mathcal{B}}$ the action of a Clifford element $C \in \mathsf{C}_q$ again takes the form of a signed permutation matrix. The rest of this chapter will be concerned with identifying the irreducible subrepresentations of $\varphi^{\otimes 2}$.

5.3. FINDING THE IRREDUCIBLE REPRESENTATIONS

The characterization of the two-copy representation of the Clifford group for multiple qubits is more complicated than the single-qubit case considered previously [11] because non-trivial elements of the multi-qubit Pauli group can commute, while others can anti-commute and these relations must be preserved under the action of the Clifford group [12]. This section will be composed of several lemmas, ultimately culminating in theorem 5.1. In these lemmas we will introduce a variety of subspaces of $\mathcal{M}_d^{\otimes 2}$ and prove that they all carry subrepresentations of $\varphi^{\otimes 2}$. In theorem 5.1 we will then exactly characterize which of the subspaces carry irreducible subrepresentations. We begin by calculating how many subrepresentations we require for each q. The following lemma, proven in [9], characterizes the inner product with itself of the character $\chi_{\varphi^{\otimes 2}}$ of the two-copy representation of the Clifford group.*

Lemma 5.2. Let C_q be the q-qubit Clifford group and $\varphi^{\otimes 2}$ its two-copy representation with character $\chi_{\varphi^{\otimes 2}}$. The character inner of this representation with itself is

$$\langle \chi_{\varphi^{\otimes 2}}, \chi_{\varphi^{\otimes 2}} \rangle = \begin{cases} 15 & q = 1 \\ 29 & q = 2 \\ 30 & q \geq 3. \end{cases} \tag{5.20}$$

By lemma 3.3, this number provides an upper limit to how many (in)equivalent irreducible subrepresentations the representation $\varphi^{\otimes 2}$ can contain. We will now, over the course of several lemmas (lemmas 5.4 to 5.6, 5.8 and 5.9), divide the space $\mathcal{M}_d^{\otimes 2}$ into subspaces carrying subrepresentations of $\varphi^{\otimes 2}$. This will eventually culminate in theorem 5.1 where we prove that all the subrepresentations derived in lemmas 5.4 to 5.6, 5.8 and 5.9 are in fact irreducible.

We continue by defining subspaces of the space $\mathcal{M}_d^{\otimes 2}$ (spanned by $\boldsymbol{\mathcal{B}}$) that carry subrepresentations of C_q. Not all of these spaces will carry irreducible representations, these will be divided further in lemmas 5.4 to 5.6, 5.8 and 5.9.

Definition 24. Let $\boldsymbol{\mathcal{B}}$ be the basis for $\mathcal{M}_d^{\otimes 2}$ as in eq. (5.18) and define the vectors

$$A_{\sigma,\tau} := \frac{1}{\sqrt{2}}(\sigma\tau - \tau\sigma), \tag{5.21}$$

$$S_{\sigma,\tau} := \frac{1}{\sqrt{2}}(\sigma\tau + \tau\sigma) \tag{5.22}$$

*Technically the character inner product of the representation $C^{\otimes 4}$ rather than $C \otimes C^* \otimes C \otimes C^*$ is calculated in [9], but it can be easily seen that the character inner product is invariant under complex conjugation of some or all tensor factors of the representation.

for $\sigma, \tau \in \boldsymbol{\sigma}_q$ and $\sigma \neq \tau$. We define the following subspaces of $\mathcal{M}_d^{\otimes 2}$:

$$
\begin{aligned}
V_{\mathrm{id}} &:= \mathrm{Span}\{\sigma_0\sigma_0\}, && \text{(trivial)}\\
V_{\mathrm{r}} &:= \mathrm{Span}\{\sigma_0\tau \;\|\; \tau \in \boldsymbol{\sigma}_q\}, && \text{(right adjoint)}\\
V_{\mathrm{l}} &:= \mathrm{Span}\{\tau\sigma_0 \;\|\; \tau \in \boldsymbol{\sigma}_q\}, && \text{(left adjoint)}\\
V_{\mathrm{d}} &:= \mathrm{Span}\{\tau\tau \;\|\; \tau \in \boldsymbol{\sigma}_q\}, && \text{(diagonal sector)}\\
V_{[\mathrm{S}]} &:= \mathrm{Span}\Big\{S_{\sigma,\tau} \;\|\; \sigma \in \boldsymbol{C}_\tau,\ \tau \in \boldsymbol{\sigma}_q\Big\}, && \text{(symmetric commuting sector)}\\
V_{\{\mathrm{S}\}} &:= \mathrm{Span}\Big\{S_{\sigma,\tau} \;\|\; \sigma \in \boldsymbol{N}_\tau,\ \tau \in \boldsymbol{\sigma}_q\Big\}, && \text{(symmetric anti-commuting sector)}\\
V_{[\mathrm{A}]} &:= \mathrm{Span}\Big\{A_{\sigma,\tau} \;\|\; \sigma \in \boldsymbol{C}_\tau,\ \tau \in \boldsymbol{\sigma}_q\Big\}, && \text{(antisymmetric commuting sector)}\\
V_{\{\mathrm{A}\}} &:= \mathrm{Span}\Big\{A_{\sigma,\tau} \;\|\; \sigma \in \boldsymbol{N}_\tau,\ \tau \in \boldsymbol{\sigma}_q\Big\}. && \text{(antisymmetric anti-commuting sector)}
\end{aligned}
$$

5

These spaces do not all carry irreducible subrepresentations of $\varphi^{\otimes 2}$ but they do all carry subrepresentations. This is proven in the following lemma:

Lemma 5.3. All spaces W defined in definition 24 carry a subrepresentation of the representation $\varphi^{\otimes 2}$ of the Clifford group C_q, that is

$$\varphi^{\otimes 2}(C)W \subset W \qquad \forall C \in \mathsf{C}_q. \tag{5.23}$$

Note that W may be empty for $q = 1$, in which case the statement holds trivially.

Proof. First note that $C\sigma_0 C^\dagger = \sigma_0$ for all $C \in \mathsf{C}_q$ and that for any $C \in \mathsf{C}_q$ and $\sigma \in \boldsymbol{\sigma}_q$ there exists a $\tau \in \boldsymbol{\sigma}_q$ such that $C\sigma C^\dagger = \pm\tau$. This means the spaces $V_{\mathrm{id}}, V_{\mathrm{r}}, V_{\mathrm{l}}$ and V_{d} carry a subrepresentation of $\varphi^{\otimes 2}$. Note also that we have

$$\varphi^{\otimes 2}(C)S_{\sigma,\tau} = S_{C\sigma C^\dagger, C\tau C^\dagger} \qquad C \in \mathsf{C}_q, \tag{5.24}$$

$$\varphi^{\otimes 2}(C)A_{\sigma,\tau} = A_{C\sigma C^\dagger, C\tau C^\dagger} \qquad C \in \mathsf{C}_q, \tag{5.25}$$

for all $\sigma, \tau \in \boldsymbol{\sigma}_q$ and $\sigma \neq \tau$ and also

$$\{C\sigma C^\dagger, C\tau C^\dagger\} = 0 \iff \{\sigma, \tau\} = 0 \qquad C \in \mathsf{C}_q, \tag{5.26}$$

$$[C\sigma C^\dagger, C\tau C^\dagger] = 0 \iff [\sigma, \tau] = 0 \qquad C \in \mathsf{C}_q \tag{5.27}$$

for all $\sigma, \tau \in \hat{\boldsymbol{\sigma}}_q$. From these equations it is easy to see that $V_{[\mathrm{S}]}, V_{\{\mathrm{S}\}}, V_{[\mathrm{A}]}$ and $V_{\{\mathrm{A}\}}$ carry subrepresentations of $\varphi^{\otimes 2}$ as well. ∎

Note that since V_{id} is a trivial representation it is automatically irreducible. Over the next few lemmas we will further characterize the other spaces defined in definition 24, beginning with the diagonal sector, i.e. the space V_{d}.

Lemma 5.4 (Diagonal sector)**.** Take the space V_{d} as defined in definition 24 and define the following 3 subspaces

$$V_0 := \mathrm{Span}\left\{ w \in V_{\mathrm{d}} \;\|\; w = \frac{1}{\sqrt{d^2-1}} \sum_{\sigma \in \boldsymbol{\sigma}_q} \sigma\sigma \right\} \tag{5.28}$$

$$V_1 := \mathrm{Span}\left\{ v \in V_{\mathrm{d}} \;\|\; v = \sum_{\sigma \in \boldsymbol{\sigma}_q} \lambda_\sigma \sigma\sigma,\ \sum_{\sigma \in \boldsymbol{\sigma}_q} \lambda_\sigma = 0,\ \sum_{\sigma \in \boldsymbol{N}_\tau} \lambda_\sigma = -\frac{d}{2}\lambda_\tau,\ \ \forall \tau \in \boldsymbol{\sigma}_q \right\} \tag{5.29}$$

$$V_2 := \mathrm{Span}\left\{ v \in V_{\mathrm{d}} \;\|\; v = \sum_{\sigma \in \boldsymbol{\sigma}_q} \lambda_\sigma \sigma\sigma,\ \sum_{\sigma \in \boldsymbol{\sigma}_q} \lambda_\sigma = 0,\ \sum_{\sigma \in \boldsymbol{N}_\tau} \lambda_\sigma = \frac{d}{2}\lambda_\tau,\ \ \forall \tau \in \boldsymbol{\sigma}_q \right\} \tag{5.30}$$

with $|V_1| = \frac{d(d+1)}{2} - 1$ and $|V_2| = \frac{d(d-1)}{2} - 1$. We have the following statements

- For $q = 1$ the spaces V_0 and V_1 carry irreducible subrepresentations of $\varphi^{\otimes 2}$ and $V_2 = \emptyset$.
- For $q \geq 2$ the spaces V_0, V_1 and V_2 carry irreducible subrepresentations of $\varphi^{\otimes 2}$.

Proof. The special case of $q = 1$ was treated in [11]. We will treat the case $q \geq 2$. We begin by establishing that the space $V_{\mathrm{d}} = \mathrm{Span}\{\sigma\sigma \;\|\; \sigma \in \boldsymbol{\sigma}_q\}$ has exactly three subspaces carrying inequivalent subrepresentations of $\varphi^{\otimes 2}$. One can see this by considering the character χ_d of $\varphi^{\otimes 2}$ restricted to V_{d}. It is easy to see by direct calculation that for all $C \in \mathsf{C}_q$ we have $\chi_{\mathrm{d}}(C) = F(C)$ where $F(C)$ is the number of non-identity Pauli matrices fixed under conjugation by C up to a sign. This means the character inner product $\langle \chi_{\mathrm{d}}, \chi_{\mathrm{d}} \rangle$ is given by

$$\langle \chi_{\mathrm{d}}, \chi_{\mathrm{d}} \rangle = \frac{1}{|C|} \sum_{C \in \mathsf{C}_q} F(C)^2. \tag{5.31}$$

By a generalized version of Burnside's Lemma (see [9]) we can relate this to the number of orbits of the Clifford group (up to signs) on the set $\boldsymbol{\sigma}_q \times \boldsymbol{\sigma}_q$. These orbits were characterized in [9] which yielded $\langle \chi_{\mathrm{d}}, \chi_{\mathrm{d}} \rangle = 3$ for $q \geq 2$. This means, by lemma 3.3, that V_{d} must contain exactly three inequivalent irreducible subrepresentations (all with multiplicity one). It is easy to see that V_0 carries a trivial subrepresentation by noting that $\varphi^{\otimes 2}$ acts as a permutation on the basis of V_{d}. Hence we can write $V_{\mathrm{d}} = V_0 \oplus V_{\mathrm{orth}}$ where

$$V_{\mathrm{orth}} := \mathrm{Span}\left\{ v \in V_{\mathrm{d}} \;\|\; v = \sum_{\sigma \in \boldsymbol{\sigma}_q} \lambda_\sigma \sigma\sigma,\ \sum_{\sigma \in \boldsymbol{\sigma}_q} \lambda_\sigma = 0 \right\}. \tag{5.32}$$

Because of the character argument given above we know this space must decompose into exactly two orthogonal subspaces V_1, V_2 which carry irreducible inequivalent subrepresentations of $\varphi^{\otimes 2}$. We now characterize these subrepresentations. We define the linear

map $\boldsymbol{\mathcal{T}} : V_{\mathrm{d}} \to V_{\mathrm{d}}$ by its action on the basis of V_{d}. For all $\tau \in \boldsymbol{\sigma}_q$ we have

$$\boldsymbol{\mathcal{T}}(\tau\tau) := \sum_{\sigma \in \boldsymbol{N}_\tau} \sigma\sigma. \tag{5.33}$$

It is easy to see that this map commutes with the action of $\varphi^{\otimes 2}$ on V_{d}. Hence, by the character argument above and Schur's lemma theorem 3.1, it must be of the form

$$\boldsymbol{\mathcal{T}} = a_0 P_0 + a_1 P_1 + a_2 P_2, \tag{5.34}$$

where P_0 is the projector onto the space V_0 and P_1, P_2 are projectors onto the eigenspaces of $\boldsymbol{\mathcal{T}}$ with eigenvalues a_1, a_2 respectively. We will label these eigenspaces V_1 and V_2. Note that $a_1, a_2 \in \mathbb{R}$ since $\boldsymbol{\mathcal{T}}$ is symmetric. We will also assume that $a_1 \leq a_2$. This can always be achieved by relabeling. By direct calculation we see that $a_0 = \frac{d^2}{2}$. We find can a_1, a_2 by considering the squared operator $\boldsymbol{\mathcal{T}}^2$. We can compute its matrix elements in the given basis of V_{d} as

$$\left[\boldsymbol{\mathcal{T}}^2\right]_{\tau\tau'} = \langle \tau\tau, \mathcal{T}^2(\tau'\tau')\rangle = \sum_{\sigma \in \boldsymbol{N}_{\tau'}} \sum_{\hat{\sigma} \in \boldsymbol{N}_\sigma} \langle \tau, \hat{\sigma}\rangle^2 \tag{5.35}$$

$$= |\boldsymbol{N}_\tau \cap \boldsymbol{N}_{\tau'}| \tag{5.36}$$

$$= \frac{d^2}{4} + \frac{d^2}{4}\delta_{\tau,\tau'}, \tag{5.37}$$

where the last equality follows from lemma 5.1 and $|\boldsymbol{N}_\tau| = \frac{d^2}{2}$ for all $\tau \in \boldsymbol{\sigma}_q$. From this characterization we can find the action of $\boldsymbol{\mathcal{T}}^2$ on $v \in V_{\mathrm{orth}}$:

$$\boldsymbol{\mathcal{T}}^2(v) = \sum_{\sigma \in \boldsymbol{\sigma}_q} \lambda_\sigma \boldsymbol{\mathcal{T}}^2(\sigma\sigma) \tag{5.38}$$

$$= \sum_{\sigma \in \boldsymbol{\sigma}_q} \left(\lambda_\sigma \frac{d^2}{2} + \sum_{\hat{\sigma} \in \boldsymbol{\sigma}_q \setminus \{\sigma\}} \lambda_{\hat{\sigma}} \frac{d^2}{4} \right) \sigma\sigma \tag{5.39}$$

$$= \sum_{\sigma \in \boldsymbol{\sigma}_q} \left(\lambda_\sigma \frac{d^2}{2} - \lambda_\sigma \frac{d^2}{4} \right) \sigma\sigma \tag{5.40}$$

$$= \frac{d^2}{4} v, \tag{5.41}$$

where we used the definition of $v \in V_{\mathrm{orth}}$. This means that we must have $a_1^2 = a_2^2 = \frac{d^2}{4}$. There are hence two options: either $a_1 = a_2$ or $a_1 = -a_2$. We can exclude the first option by noting that the operator $\boldsymbol{\mathcal{T}}$ is traceless. Hence we must have

$$\mathrm{Tr}(\boldsymbol{\mathcal{T}}) = 0 = a_0 + a_1|V_1| + a_2|V_2| = \frac{d^2}{2} + a_1|V_1| + a_2|V_2|, \tag{5.42}$$

where $|V_i|$ is the dimension of the space V_i. By noting that $|V_1| + |V_2| = d^2 - 2$ and that $V_1, V_2 \neq \emptyset$ (this is a consequence of the character argument above) we find the only

possible solution to be

$$|V_1| = \frac{d(d+1)}{2} - 1, \qquad a_1 = -\frac{d}{2}, \tag{5.43}$$

$$|V_2| = \frac{d(d-1)}{2} - 1, \qquad a_2 = \frac{d}{2}. \tag{5.44}$$

We can now diagonalize the operator $\boldsymbol{\mathcal{T}}$ to find the description for the spaces V_1, V_2 given in the lemma statement.

■

Next we establish an equivalence between the representations carried by V_{r} and V_{l} and two subspaces in the symmetric and antisymmetric sectors. All four of these representations will be equivalent to the adjoint representation of the Clifford group.

Lemma 5.5 (Adjoint representations)**.** Take the vector spaces $V_{\mathrm{r}}, V_{\mathrm{l}}$ as defined in definition 24. Also define the vector spaces

$$V_{[\mathrm{adj}]} := \mathrm{Span}\Big\{ v_\tau^{[\mathrm{adj}]} \in V_{[\mathrm{S}]} \;\Big\|\; v_\tau^{[\mathrm{adj}]} = \frac{1}{\sqrt{2|\boldsymbol{C}_\tau|}} \sum_{\sigma \in \boldsymbol{C}_\tau} S_{\sigma,\sigma\cdot\tau}, \;\; \tau \in \boldsymbol{\sigma}_q \Big\}$$

(symmetric adjoint)

$$V_{\{\mathrm{adj}\}} := \mathrm{Span}\Big\{ v_\tau^{\{\mathrm{adj}\}} \in V_{\{\mathrm{A}\}} \;\Big\|\; v_\tau^{\{\mathrm{adj}\}} = \frac{1}{\sqrt{2|\boldsymbol{N}_\tau|}} \sum_{\sigma \in \boldsymbol{N}_\tau} A_{\sigma,i\sigma\cdot\tau}, \;\; \tau \in \boldsymbol{\sigma}_q \Big\}$$

(antisym adjoint)

located in the symmetric commuting and antisymmetric anti-commuting sectors. The spaces $V_{\mathrm{r}}, V_{\mathrm{l}}, V_{\{\mathrm{adj}\}}$ and $V_{[\mathrm{adj}]}$ carry equivalent irreducible representations.

Proof. Note that the representations carried by the spaces $V_{\mathrm{r}}, V_{\mathrm{l}}$ are trivially equivalent to the adjoint representation of the Clifford group, which is irreducible [3]. This leaves us with the spaces $V_{\{\mathrm{adj}\}}$ and $V_{[\mathrm{adj}]}$. We begin by noting that the spaces $V_{[\mathrm{adj}]}, V_{\{\mathrm{adj}\}}$ carry subrepresentations. This is easily seen by by taking $v_\tau^{[\mathrm{adj}]} \in V_{[\mathrm{adj}]}$ as defined in the lemma statement and writing

$$\varphi^{\otimes 2}(C) v_\tau^{[\mathrm{adj}]} = \frac{1}{\sqrt{2|\boldsymbol{C}_\tau|}} \sum_{\sigma \in \boldsymbol{C}_\tau} S_{C\sigma C^\dagger,(C\sigma C^\dagger)\cdot(C\tau C^\dagger)} \tag{5.45}$$

$$= \frac{1}{\sqrt{2|\boldsymbol{C}_\tau|}} \sum_{C^\dagger \sigma C \in \boldsymbol{C}_\tau} S_{\sigma,\sigma\cdot C\tau C^\dagger} \tag{5.46}$$

$$= \frac{1}{\sqrt{2|\boldsymbol{C}_\tau|}} \sum_{\sigma \in \boldsymbol{C}_{C\tau C^\dagger}} S_{\sigma,\sigma\cdot C\tau C^\dagger} \tag{5.47}$$

$$= v_{C\tau C^\dagger}^{[\mathrm{adj}]} \in V_{[\mathrm{adj}]}, \tag{5.48}$$

where we used the fact that the action of the Clifford group preserves commutativity of elements of the Pauli group and acts transitively on $\boldsymbol{\sigma}_q$. We have a similar argument for

$V_{\{\mathrm{adj}\}}$. Note also that the vectors spanning $V_{[\mathrm{adj}]}$ as given in the lemma statement form an orthonormal basis for $V_{[\mathrm{adj}]}$. For $\tau, \tau' \in \boldsymbol{\sigma}_q$ we have

$$\langle v_\tau^{[\mathrm{adj}]}, v_{\tau'}^{[\mathrm{adj}]} \rangle = \frac{1}{2|\boldsymbol{C}_\tau|} \sum_{\sigma \in \boldsymbol{C}_\tau} \sum_{\hat{\sigma} \in \boldsymbol{C}_{\tau'}} \langle S_{\sigma, \sigma\cdot\tau}, S_{\hat{\sigma}, \tau'\cdot\hat{\sigma}} \rangle \tag{5.49}$$

$$= \frac{1}{2|\boldsymbol{C}_\tau|} \sum_{\sigma \in \boldsymbol{C}_\tau} \sum_{\hat{\sigma} \in \boldsymbol{C}_{\tau'}} \langle \sigma, \hat{\sigma} \rangle \langle \sigma \cdot \tau, \hat{\sigma}\tau' \rangle + \frac{1}{2|\boldsymbol{C}_\tau|} \sum_{\sigma \in \boldsymbol{C}_\tau} \sum_{\hat{\sigma} \in \boldsymbol{C}_{\tau'}} \langle \sigma \cdot \tau, \hat{\sigma} \rangle \langle \sigma, \tau' \cdot \hat{\sigma} \rangle \tag{5.50}$$

$$= \frac{1}{2|\boldsymbol{C}_\tau|} \sum_{\sigma \in \boldsymbol{C}_\tau} \sum_{\hat{\sigma} \in \boldsymbol{C}_{\tau'}} \delta_{\tau,\tau'}\delta_{\sigma,\hat{\sigma}} + \delta_{\tau,\tau'} \frac{1}{2|\boldsymbol{C}_\tau|} \sum_{\sigma \in \boldsymbol{C}_\tau} \sum_{\hat{\sigma} \in \boldsymbol{C}_\tau \cap \boldsymbol{C}_\sigma} \langle \sigma \cdot \hat{\sigma}, \tau \rangle^2 \tag{5.51}$$

$$= \frac{1}{2}\delta_{\tau,\tau'} + \delta_{\tau,\tau'} \frac{1}{|\boldsymbol{C}_\tau|} \sum_{\sigma \in \boldsymbol{C}_\tau} \sum_{\hat{\sigma} \in \boldsymbol{C}_\tau} \delta_{\sigma,\hat{\sigma}} \tag{5.52}$$

$$= \delta_{\tau,\tau'}, \tag{5.53}$$

where we obtained the second to last equality by using lemma 5.1 and noting that $\langle \sigma \cdot \sigma_0, \tau \rangle = 0$ if $\sigma \in C_\tau$. We can make a similar argument for the vectors spanning $V_{\{\mathrm{adj}\}}$. Now since $|V_{\mathrm{r}}| = |V_{[\mathrm{adj}]}|$ we can construct the isomorphism

$$\theta : V_{\mathrm{r}} \to V_{[\mathrm{adj}]} : \sigma_0\tau \mapsto v_\tau^{[\mathrm{adj}]}. \tag{5.54}$$

We can check that this isomorphism commutes with the action of $\varphi^{\otimes 2}$. We have for all $\tau \in \boldsymbol{\sigma}_q$

$$\theta(\varphi^{\otimes 2}(C)\sigma_0\tau) = \frac{1}{\sqrt{2|\boldsymbol{C}_\tau|}} \sum_{\sigma \in \boldsymbol{C}_{C\tau C^\dagger}} S_{\sigma, \sigma\cdot C\tau C^\dagger} \tag{5.55}$$

$$= \frac{1}{\sqrt{2|\boldsymbol{C}_\tau|}} \sum_{C^\dagger \sigma C \in \boldsymbol{C}_\tau} S_{\sigma, \sigma\cdot C\tau C^\dagger} \tag{5.56}$$

$$= \frac{1}{\sqrt{2|\boldsymbol{C}_\tau|}} \sum_{\sigma \in \boldsymbol{C}_\tau} S_{C\sigma C^\dagger, C\sigma C^\dagger \cdot C\tau C^\dagger} \tag{5.57}$$

$$= \varphi^{\otimes 2}(C)\theta(\sigma_0\tau), \tag{5.58}$$

for all $C \in \mathsf{C}_q$. This means that the spaces V_{r} and $V_{[\mathrm{adj}]}$ carry equivalent subrepresentations of $\varphi^{\otimes 2}$. We can make the same argument for $V_{\{\mathrm{adj}\}}$ and hence $V_{[\mathrm{adj}]}, V_{\{\mathrm{adj}\}}, V_{\mathrm{r}}, V_{\mathrm{l}}$ carry equivalent irreducible representations. ■

Now we turn our attention to the antisymmetric sector, i.e. the spaces $V_{[\mathrm{A}]}, V_{\{\mathrm{A}\}}$ as defined in definition 24, where we can formulate the following lemma.

Lemma 5.6 (Antisymmetric sector)**.** Take the space $V_{\{\mathrm{A}\}}$ as defined in definition 24 and note that it contains the space $V_{\{\mathrm{adj}\}}$ (defined in lemma 5.5). Denote the orthogonal complement of $V_{\{\mathrm{adj}\}}$ in $V_{\{\mathrm{A}\}}$ as $V_{\{\mathrm{adj}\}}^\perp$. We have that the subrepresentations of $\varphi^{\otimes 2}$ carried by $V_{[\mathrm{A}]}$ and $V_{\{\mathrm{adj}\}}^\perp$ are equivalent.

Proof. Note that $V_{\{\mathrm{adj}\}}^{\perp}$ carries a subrepresentation of $\varphi^{\otimes 2}$ by Maschke's lemma [13] since $V_{\{\mathrm{adj}\}}$ and $V_{\{\mathrm{A}\}}$ carry subrepresentations. We will prove that the representations carried by $V_{\{\mathrm{adj}\}}^{\perp}$ and $V_{[\mathrm{A}]}$ are equivalent by constructing an isomorphism between them that commutes with the action of $\varphi^{\otimes 2}$. Note first that we can write down the following orthogonal basis for $V_{[\mathrm{A}]}$ as

$$V_{[\mathrm{A}]} = \mathrm{Span}\{A_{\sigma,\sigma\cdot\tau} \;\|\; \sigma \in \boldsymbol{C}_\tau, \;\; \tau \in \boldsymbol{\sigma}_q\}. \tag{5.59}$$

Now consider the following linear map (defined as the linear extension of its action on the basis defined above) between $V_{[\mathrm{A}]}$ and $V_{\{\mathrm{A}\}}$.

$$\Theta : V_{[\mathrm{A}]} \to V_{\{\mathrm{A}\}} : A_{\sigma,\sigma\cdot\tau} \mapsto \sum_{\hat{\sigma}\in \boldsymbol{N}_\tau \cap \boldsymbol{C}_\sigma} A_{\hat{\sigma},i\hat{\sigma}\cdot\tau} - \sum_{\hat{\sigma}\in \boldsymbol{N}_\tau \cap \boldsymbol{N}_\sigma} A_{\hat{\sigma},i\hat{\sigma}\cdot\tau} \tag{5.60}$$

for all $\sigma \in \boldsymbol{C}_\tau, \;\; \tau \in \boldsymbol{\sigma}_q$. We now argue that the image of Θ is orthogonal to the space $V_{\{\mathrm{adj}\}}$. We do this by direct calculation. For all $\nu \in \boldsymbol{\sigma}_q$ and all $\sigma \in \boldsymbol{C}_\tau, \;\; \tau \in \boldsymbol{\sigma}_q$ we can calculate

$$\begin{aligned}
\sqrt{2|\boldsymbol{N}_\tau|}\langle v_\nu^{\{\mathrm{adj}\}}, \Theta\big(A_{\sigma,\sigma\cdot\tau}\big)\rangle &= \sum_{\sigma'\in \boldsymbol{N}_\nu} \sum_{\hat{\sigma}\in \boldsymbol{N}_\tau\cap\boldsymbol{C}_\sigma} \langle A_{\sigma',i\sigma'\cdot\nu}, A_{\hat{\sigma},i\hat{\sigma}\cdot\tau}\rangle && (5.61)\\
&\qquad - \sum_{\sigma'\in \boldsymbol{N}_\nu} \sum_{\hat{\sigma}\in \boldsymbol{N}_\tau\cap\boldsymbol{N}_\sigma} \langle A_{\sigma',i\sigma'\cdot\nu}, A_{\hat{\sigma},i\hat{\sigma}\cdot\tau}\rangle && (5.62)\\
&= \sum_{\sigma'\in \boldsymbol{N}_\nu} \sum_{\hat{\sigma}\in \boldsymbol{N}_\tau\cap\boldsymbol{C}_\sigma} (\delta_{\sigma',\hat{\sigma}} + \delta_{\sigma',i\hat{\sigma}\cdot\tau})\delta_{\nu,\tau}\\
&\qquad - \sum_{\sigma'\in \boldsymbol{N}_\nu} \sum_{\hat{\sigma}\in \boldsymbol{N}_\tau\cap\boldsymbol{N}_\sigma} (\delta_{\sigma',\hat{\sigma}} + \delta_{\sigma',i\hat{\sigma}\cdot\tau})\delta_{\nu,\tau} && (5.63)\\
&= 2\left(|\boldsymbol{N}_\nu \cap \boldsymbol{N}_\tau \cap \boldsymbol{C}_\sigma| - |\boldsymbol{N}_\nu \cap \boldsymbol{N}_\tau \cap \boldsymbol{N}_\sigma|\right)\delta_{\tau,\nu} && (5.64)\\
&= 2\left[|\boldsymbol{N}_\tau \cap \boldsymbol{C}_\sigma| - |\boldsymbol{N}_\tau \cap \boldsymbol{N}_\sigma|\right]\delta_{\tau,\nu} && (5.65)\\
&= 0 && (5.66)
\end{aligned}$$

where in the last line we used lemma 5.1 and $|\boldsymbol{N}_\tau \cap \boldsymbol{C}_\sigma| = |\boldsymbol{N}_\tau \cap \hat{\boldsymbol{C}}_\sigma|$ if $\sigma \in \boldsymbol{C}_\tau$. This means that $\mathrm{Im}(\Theta) \subset V_{\{\mathrm{adj}\}}^{\perp}$. We now argue that $\mathrm{Im}(\Theta) = V_{\{\mathrm{adj}\}}^{\perp}$. We first note that $|V_{\{\mathrm{adj}\}}^{\perp}| = |V_{[\mathrm{A}]}|$. Furthermore we can show that Θ preserves orthogonality under the Hilbert-Schmidt inner product and that $\mathrm{Ker}(\Theta) = \emptyset$. By direct calculation we have for all

$\tau, \tau' \in \boldsymbol{\sigma}_q$ and $\sigma \in \boldsymbol{C}_\tau, \sigma' \in \boldsymbol{C}_{\tau'}$

$$\langle \Theta\big(A_{\sigma,\sigma\cdot\tau}\big), \Theta\big(A_{\sigma',\sigma'\cdot\tau'}\big)\rangle$$

$$\begin{aligned}
&= \sum_{\substack{\hat\sigma\in \boldsymbol{N}_\tau\cap\boldsymbol{C}_\sigma\\ \hat\sigma'\in\boldsymbol{N}_{\tau'}\cap\boldsymbol{C}_{\sigma'}}} \langle A_{\hat\sigma,i\hat\sigma\cdot\tau}, A_{\hat\sigma',i\hat\sigma'\cdot\tau'}\rangle - \sum_{\substack{\hat\sigma\in \boldsymbol{N}_\tau\cap\boldsymbol{N}_\sigma\\ \hat\sigma'\in\boldsymbol{N}_{\tau'}\cap\boldsymbol{C}_{\sigma'}}} \langle A_{\hat\sigma,i\hat\sigma\cdot\tau}, A_{\hat\sigma',i\hat\sigma'\cdot\tau'}\rangle \\
&\qquad - \sum_{\substack{\hat\sigma\in \boldsymbol{N}_\tau\cap\boldsymbol{C}_\sigma\\ \hat\sigma'\in\boldsymbol{N}_{\tau'}\cap\boldsymbol{N}_{\sigma'}}} \langle A_{\hat\sigma,i\hat\sigma\cdot\tau}, A_{\hat\sigma',i\hat\sigma'\cdot\tau'}\rangle + \sum_{\substack{\hat\sigma\in \boldsymbol{N}_\tau\cap\boldsymbol{N}_\sigma\\ \hat\sigma'\in\boldsymbol{N}_{\tau'}\cap\boldsymbol{N}_{\sigma'}}} \langle A_{\hat\sigma,i\hat\sigma\cdot\tau}, A_{\hat\sigma',i\hat\sigma'\cdot\tau'}\rangle
\end{aligned} \tag{5.67}$$

$$\begin{aligned}
&= \sum_{\substack{\hat\sigma\in \boldsymbol{N}_\tau\cap\boldsymbol{C}_\sigma\\ \hat\sigma'\in\boldsymbol{N}_{\tau'}\cap\boldsymbol{C}_{\sigma'}}} (\delta_{\hat\sigma',\hat\sigma} + \delta_{\hat\sigma',i\hat\sigma\cdot\tau})\delta_{\tau,\tau'} - \sum_{\substack{\hat\sigma\in \boldsymbol{N}_\tau\cap\boldsymbol{N}_\sigma\\ \hat\sigma'\in\boldsymbol{N}_{\tau'}\cap\boldsymbol{C}_{\sigma'}}} (\delta_{\hat\sigma',\hat\sigma} + \delta_{\hat\sigma',i\hat\sigma\cdot\tau})\delta_{\tau,\tau'} \\
&\qquad - \sum_{\substack{\hat\sigma\in \boldsymbol{N}_\tau\cap\boldsymbol{C}_\sigma\\ \hat\sigma'\in\boldsymbol{N}_{\tau'}\cap\boldsymbol{N}_{\sigma'}}} (\delta_{\hat\sigma',\hat\sigma} + \delta_{\hat\sigma',i\hat\sigma\cdot\tau})\delta_{\tau,\tau'} + \sum_{\substack{\hat\sigma\in \boldsymbol{N}_\tau\cap\boldsymbol{N}_\sigma\\ \hat\sigma'\in\boldsymbol{N}_{\tau'}\cap\boldsymbol{N}_{\sigma'}}} (\delta_{\hat\sigma',\hat\sigma} + \delta_{\hat\sigma',i\hat\sigma\cdot\tau})\delta_{\tau,\tau'}
\end{aligned} \tag{5.68}$$

$$\begin{aligned}
&= \Big(|\boldsymbol{N}_\tau \cap \boldsymbol{C}_\sigma \cap \boldsymbol{C}_{\sigma'}| - |\boldsymbol{N}_\tau \cap \boldsymbol{C}_\sigma \cap \boldsymbol{N}_{\sigma'}| \\
&\qquad\qquad - |\boldsymbol{N}_\tau \cap \boldsymbol{N}_\sigma \cap \boldsymbol{C}_{\sigma'}| + |\boldsymbol{N}_\tau \cap \boldsymbol{N}_\sigma \cap \boldsymbol{N}_{\sigma'}| \Big)\delta_{\tau,\tau'}
\end{aligned} \tag{5.69}$$

To further evaluate this expression we use the following fact. Let $\nu \in \hat{\boldsymbol{\sigma}}_q$ such that $\sigma\cdot\sigma' \propto \nu$ (note that this implies that $\nu \in \hat{\boldsymbol{C}}_\tau$) . We then have

$$\forall \mu \in \boldsymbol{\sigma}_q: \quad \mu \in \boldsymbol{C}_\nu \iff \mu \in (\boldsymbol{C}_\sigma \cap \boldsymbol{C}_{\sigma'}) \cup (\boldsymbol{N}_\sigma \cap \boldsymbol{N}_{\sigma'}) \tag{5.70}$$

$$\forall \mu \in \boldsymbol{\sigma}_q: \quad \mu \in \boldsymbol{N}_\nu \iff \mu \in (\boldsymbol{C}_\sigma \cap \boldsymbol{N}_{\sigma'}) \cup (\boldsymbol{C}_\sigma \cap \boldsymbol{N}_{\sigma'}). \tag{5.71}$$

We use this together with the fact that $\boldsymbol{C}_\tau \cap \boldsymbol{N}_\tau = \emptyset$ for all $\tau \in \boldsymbol{\sigma}_q$ to reduce eq. (5.69) to

$$\left(|\boldsymbol{N}_\tau \cap \boldsymbol{C}_\nu| - |\boldsymbol{N}_\tau \cap \boldsymbol{N}_\nu|\right)\delta_{\tau,\tau'} = \left(\frac{d^2}{2} - 1\right)\delta_{\tau,\tau'}(\delta_{\sigma,\sigma'} + \delta_{\sigma,i\sigma'\cdot\tau}) \tag{5.72}$$

where in the last equality we used lemma 5.1 together with $\sigma\cdot\sigma' \propto \nu$ and that $\boldsymbol{N}_\tau \cap \hat{\boldsymbol{C}}_\nu = \boldsymbol{N}_\tau \cap \boldsymbol{C}_\nu$ if $\nu \in \boldsymbol{C}_\tau$ and that $\boldsymbol{C}_\nu = \boldsymbol{\sigma}_q$ if $\nu = \sigma_0$ which occurs if and only if $\sigma = \sigma'$. Since $\langle A_{\sigma,i\sigma\cdot\tau}, A_{\sigma',\sigma'\cdot\tau'}\rangle = \delta_{\tau,\tau'}(\delta_{\sigma,\sigma'} + \delta_{\sigma,i\sigma'\cdot\tau})$ this means that Θ preserves orthogonality and that $\mathrm{Ker}(\Theta) = \emptyset$. Together with the fact that $|V^\perp_{\{\mathrm{adj}\}}| = |V_{[\mathrm{A}]}|$ this implies that $\mathrm{Im}(\Theta) = V^\perp_{\{\mathrm{adj}\}}$. This means we can restrict Θ to an isomorphism from $V_{[\mathrm{A}]}$ to $V^\perp_{\{\mathrm{adj}\}}$. We will abuse notation and refer to this isomorphism as Θ as well.

To prove that the representations carried by $V_{[\mathrm{A}]}$ and $V^\perp_{\{\mathrm{adj}\}}$ are equivalent we now still have to argue that Θ commutes with $\varphi^{\otimes 2}$. We can do this by direct calculation. For all

$\tau \in \boldsymbol{\sigma}_q$ and $\sigma \in \boldsymbol{C}_\tau$ and $C \in \mathsf{C}_q$ we have

$$\Theta(\varphi^{\otimes 2}(C)(A_{\sigma,\sigma\cdot\tau})) = \Theta(A_{C\sigma C^\dagger, C\sigma C^\dagger \cdot C\tau C^\dagger}) \tag{5.73}$$

$$= \sum_{\hat{\sigma}\in \boldsymbol{N}_{C\tau C^\dagger}\cap \boldsymbol{C}_{C\sigma C^\dagger}} A_{\hat{\sigma}, i\hat{\sigma}\cdot C\tau C^\dagger} - \sum_{\hat{\sigma}\in \boldsymbol{N}_{C\tau C^\dagger}\cap \boldsymbol{N}_{C\sigma C^\dagger}} A_{\hat{\sigma}, i\hat{\sigma}\cdot C\tau C^\dagger} \tag{5.74}$$

$$= \sum_{C^\dagger\hat{\sigma}C\in \boldsymbol{N}_\tau\cap \boldsymbol{C}_\sigma} A_{\hat{\sigma}, i\hat{\sigma}\cdot C\tau C^\dagger} - \sum_{C^\dagger\hat{\sigma}C\in \boldsymbol{N}_\tau\cap \boldsymbol{N}_\sigma} A_{\hat{\sigma}, i\hat{\sigma}\cdot C\tau C^\dagger} \tag{5.75}$$

$$= \sum_{\hat{\sigma}\in \boldsymbol{N}_\tau\cap \boldsymbol{C}_\sigma} A_{C\hat{\sigma}C^\dagger, iC\hat{\sigma}C^\dagger \cdot C\tau C^\dagger} - \sum_{\hat{\sigma}\in \boldsymbol{N}_\tau\cap \boldsymbol{N}_\sigma} A_{C\hat{\sigma}C^\dagger, iC\hat{\sigma}C^\dagger \cdot C\tau C^\dagger} \tag{5.76}$$

$$= \varphi^{\otimes 2}(C)\left(\sum_{\hat{\sigma}\in \boldsymbol{N}_\tau\cap \boldsymbol{C}_\sigma} A_{\hat{\sigma}, i\hat{\sigma}\cdot\tau} - \sum_{\hat{\sigma}\in \boldsymbol{N}_\tau\cap \boldsymbol{N}_\sigma} A_{\hat{\sigma}, i\hat{\sigma}\cdot\tau}\right) \tag{5.77}$$

$$= \varphi^{\otimes 2}(C)\left(\Theta(A_{\sigma,\sigma\cdot\tau})\right). \tag{5.78}$$

This proves the equivalence of the subrepresentations carried by $V_{[\mathrm{A}]}$ and $V^\perp_{\{\mathrm{adj}\}}$. ■

Note that we have not proven that the subrepresentations carried by $V_{[\mathrm{A}]}$ and $V^\perp_{\{\mathrm{adj}\}}$ are irreducible. We will get this irreducibility for free when proving theorem 5.1.

Next up are the symmetric sectors. In order to facilitate the analysis of these spaces we begin by proving the following technical lemma. This technical lemma allows us to draw conclusions about the subrepresentations of $\varphi^{\otimes 2}$ carried by subspaces of $V_{[\mathrm{S}]}$ and $V_{\{\mathrm{S}\}}$ by considering the action of a strict subgroup of the Clifford group C_q on particular subspaces of $V_{[\mathrm{S}]}$ and $V_{\{\mathrm{S}\}}$.

Lemma 5.7 (space reduction)**.** For every $\tau \in \boldsymbol{\sigma}_q$ define a subgroup C_q^τ of C_q as

$$\mathsf{C}_q^\tau := \{C \in \mathsf{C}_q \;\|\; C\tau C^\dagger = \pm\tau\}. \tag{5.79}$$

Also define subspace $\mathrm{V} = \mathrm{Span}\{\, \sigma\hat{\sigma} \;\|\; \sigma, \hat{\sigma} \in \boldsymbol{\sigma}_q,\ \ \sigma \neq \hat{\sigma}\} \subset \mathcal{M}_d^{\otimes 2}$ and for all $\tau \in \boldsymbol{\sigma}_q$ define the subspace

$$V^\tau := \mathrm{Span}\{\, \sigma_\tau\hat{\sigma}_\tau \;\|\; \sigma_\tau, \hat{\sigma}_\tau \in \boldsymbol{\sigma}_q,\ \ \sigma_\tau \cdot \hat{\sigma}_\tau \propto \tau\}. \tag{5.80}$$

The first claim of the lemma is:

- The space V decomposes with respect to V^τ, that is

$$V = \bigoplus_{\tau'\in\boldsymbol{\sigma}_q} V^{\tau'}. \tag{5.81}$$

5

Now assume that for some $\tau \in \boldsymbol{\sigma}_q$ there exists a subspace W^τ of V^τ such that

$$\varphi^{\otimes 2}(\hat{C})W^\tau \subset W^\tau, \quad \forall \hat{C} \in \mathsf{C}_q^\tau. \tag{5.82}$$

The second claim of the lemma is:

- For all $\tau' \in \boldsymbol{\sigma}_q$ there exist $W^{\tau'} \subset V^{\tau'}$ such that W^τ and $W^{\tau'}$ are isomorphic and that
$$\varphi^{\otimes 2}(C)W \subset W, \quad \forall C \in \mathsf{C}_q, \tag{5.83}$$
with
$$W := \bigoplus_{\tau' \in \boldsymbol{\sigma}_q} W^{\tau'}. \tag{5.84}$$

Proof. Note first that $\cup_{\tau' \in \boldsymbol{\sigma}_q} V^{\tau'} = V$ and also for $\tau, \tau' \in \boldsymbol{\sigma}_q$ we have for all $\sigma_\tau \hat{\sigma}_\tau \in V^\tau$, $\sigma_{\tau'} \hat{\sigma}_{\tau'} \in V^{\tau'}$ that

$$\langle \sigma_\tau \hat{\sigma}_\tau, \sigma_{\tau'} \hat{\sigma}_{\tau'} \rangle = \delta_{\sigma_\tau, \sigma_{\tau'}} \delta_{\hat{\sigma}_\tau, \hat{\sigma}_{\tau'}} = \delta_{\sigma_\tau, \sigma_{\tau'}} \delta_{\tau, \tau'}, \tag{5.85}$$

since if $\sigma_\tau = \sigma_{\tau'}$ we must have $(\hat{\sigma}_\tau = \hat{\sigma}_{\tau'} \iff \tau = \tau')$. This immediately implies

$$V = \bigoplus_{\tau' \in \boldsymbol{\sigma}_q} V^{\tau'}. \tag{5.86}$$

This proves the first claim of the lemma.

Now assume that there exists a $\tau \in \boldsymbol{\sigma}_q$ such that there is a subspace $W^\tau \subset V^\tau$ such that for all $\hat{C} \in \mathsf{C}_q^\tau$ we have $\varphi^{\otimes 2}(\hat{C})W^\tau \subset W^\tau$. For all $\tau' \in \boldsymbol{\sigma}_q$ we can define the following subset $S_{\tau'}$ of C_q:

$$S_{\tau'} := \{C \in \mathsf{C}_q \parallel C\tau C^\dagger = \pm\tau'\}. \tag{5.87}$$

Because the C_q acts transitively on $\boldsymbol{\sigma}_q$ this set is never empty. Now for every $C \in S_{\tau'}$ we can define the subspace W^C as

$$W^{C,\tau'} := \{\varphi^{\otimes 2}(C)v \parallel v \in W^\tau\}. \tag{5.88}$$

Note that for every $C \in S_{\tau'}$ we have $W^{C,\tau'} \subset V^{\tau'}$. We also have for $C_1, C_2 \in S_{\tau'}$ that

$$C_1^\dagger C_2 \in \mathsf{C}_q^\tau, \tag{5.89}$$
$$C_2^\dagger C_1 \in \mathsf{C}_q^\tau. \tag{5.90}$$

The first equation implies that

$$\varphi^{\otimes 2}(C_1^\dagger)\varphi^{\otimes 2}(C_2)W^\tau \subset W^\tau \implies \varphi^{\otimes 2}(C_1^\dagger)W^{C_2,\tau'} \subset W^\tau, \tag{5.91}$$

which we can left-multiply by $\varphi^{\otimes 2}(C_1)$ to get

$$W^{C_2,\tau'} \subset W^{C_1,\tau'}. \tag{5.92}$$

We can repeat this reasoning with C_2, C_1 interchanged to obtain $W^{C_2,\tau'} \subset W^{C_1,\tau'}$ and thus

$$W^{C_1,\tau'} = W^{C_2,\tau'}, \quad \forall C_1, C_2 \in S_{\tau'}, \tag{5.93}$$

for all τ'. Let us label this single subspace by $W^{\tau'}$. Note that since $W^{\tau} \subset V^{\tau}$ for all $\tau \in \boldsymbol{\sigma}_q$ the spaces $W^{\tau}, W^{\tau'}$ are orthogonal for $\tau \neq \tau'$. Hence we can consider the space

$$W = \bigoplus_{\tau \in \boldsymbol{\sigma}_q} W^{\tau}. \tag{5.94}$$

Now take $w \in W$. We can write

$$w = \sum_{\tau \in \boldsymbol{\sigma}_q} v^{\tau}, \quad v^{\tau} \in W^{\tau}. \tag{5.95}$$

Now for all $C \in \boldsymbol{C}_q$ and $\tau \in \boldsymbol{\sigma}_q$ there exist unique vectors $u^{\tau'} \in V^{\tau'}$ with $\tau' = \pm C\tau C^{\dagger}$ such that

$$\varphi^{\otimes 2}(C)w = \sum_{\tau \in \boldsymbol{\sigma}_q} \varphi^{\otimes 2}(C)v^{\tau} = \sum_{\tau' \in \boldsymbol{\sigma}_q} u^{\tau'} \in W, \tag{5.96}$$

which proves the lemma.

■

Next we turn our attention to the symmetric commuting sector i.e., the space $V_{[\mathrm{S}]}$. We will decompose this space by using a curious connection between the representation $\varphi^{\otimes 2}$ of C_q on $V_{[\mathrm{S}]}$ and the representation $\varphi^{\otimes 2}$ of C_{q-1} (the Clifford group on $q-1$ qubits) on its diagonal sector V_{d}^{q-1}. We have the following lemma.

Lemma 5.8 (Symmetric commuting sector)**.** Take the space $V_{[\mathrm{S}]}$ as defined in definition 24, the space $V_{[\mathrm{adj}]}$ as defined in lemma 5.5 and define the spaces

$$V_{[1]} := \bigoplus_{\tau \in \boldsymbol{\sigma}_q} V_{[1]}^{\tau}, \qquad V_{[2]} := \bigoplus_{\tau \in \boldsymbol{\sigma}_q} V_{[2]}^{\tau}, \tag{5.97}$$

where for all $\tau \in \boldsymbol{\sigma}_q$

$$V_{[1]}^{\tau} := \mathrm{Span}\left\{ v^{\tau} \in V_{[\mathrm{S}]} \;\|\; v^{\tau} = \sum_{\sigma \in \boldsymbol{N}_{\tau}} \lambda_{\sigma} S_{\sigma, i\sigma\cdot\tau}, \; \sum_{\sigma \in \boldsymbol{C}_{\tau} \cap \boldsymbol{N}_{\nu}} \lambda_{\sigma} = -\frac{d}{4}\lambda_{\nu}, \quad \forall \nu \in \boldsymbol{C}_{\tau} \right\}, \tag{5.98}$$

$$V_{[2]}^{\tau} := \mathrm{Span}\left\{ v^{\tau} \in V_{[\mathrm{S}]} \;\|\; v^{\tau} = \sum_{\sigma \in \boldsymbol{N}_{\tau}} \lambda_{\sigma} S_{\sigma, i\sigma\cdot\tau}, \; \sum_{\sigma \in \boldsymbol{C}_{\tau} \cap \boldsymbol{N}_{\nu}} \lambda_{\sigma} = \frac{d}{4}\lambda_{\nu}, \quad \forall \nu \in \boldsymbol{C}_{\tau} \right\}. \tag{5.99}$$

Note that $V_{[1]}, V_{[2]}$ and $V_{[\mathrm{adj}]}$ are subspaces of $V_{[\mathrm{S}]}$. We have the following:

- For $q = 1$ we have $V_{[\mathrm{S}]} = \emptyset$ and hence $V_{[1]} = V_{[2]} = V_{[\mathrm{adj}]} = \emptyset$.

5

- For $q = 2$ we have $V_{[\mathrm{S}]} = V_{[\mathrm{adj}]} \oplus V_{[1]}$ and $V_{[2]} = \emptyset$. The spaces $V_{[\mathrm{adj}]}$ and $V_{[1]}$ carry irreducible subrepresentations of $\varphi^{\otimes 2}$.
- For $q \geq 3$ have $V_{[\mathrm{S}]} = V_{[\mathrm{adj}]} \oplus V_{[1]} \oplus V_{[2]}$. The spaces $V_{[\mathrm{adj}]}, V_{[1]}$ and $V_{[2]}$ carry irreducible subrepresentations of $\varphi^{\otimes 2}$.

Proof. We begin the proof by noting that the space $V_{[\mathrm{S}]}$ can be block decomposed in the following way

$$V_{[\mathrm{S}]} = \bigoplus_{\tau \in \boldsymbol{\sigma}_q} V_{[\mathrm{S}]}^{\tau}, \tag{5.100}$$

with

$$V_{[\mathrm{S}]}^{\tau} := \mathrm{Span}\left\{ S_{\sigma,\sigma\cdot\tau} \;\middle\|\; \sigma \in \boldsymbol{C}_{\tau},\ \ \tau \in \boldsymbol{\sigma}_q \right\}. \tag{5.101}$$

Using lemma 5.7 we can, to find subspaces of $V_{[\mathrm{S}]}$ carrying subrepresentations of $\varphi^{\otimes 2}$, restrict ourselves to finding, for some $\tau \in \boldsymbol{\sigma}_q$, subspaces of $V_{[\mathrm{S}]}^{\tau}$ that are invariant under the representation $\varphi^{\otimes 2}$ restricted the subgroup $\mathsf{C}_q^{\tau} \subset \mathsf{C}_q$ where C_q^{τ} is defined as in lemma 5.7. For the purposes of this proof we choose $\tau = \sigma_Z\sigma_0$. This means that we can write any $\hat{\sigma} \in \boldsymbol{C}_{\tau}$ as

$$\hat{\sigma} = \sigma_Z\sigma \ \text{ or } \ \sigma_0\sigma, \quad \sigma \in \boldsymbol{\sigma}_{q-1}, \tag{5.102}$$

with $\boldsymbol{\sigma}_{q-1}$ the normalized, hermitian, non-identity Pauli elements on $q-1$ qubits. We also recall the definition of the diagonal sector on $q-1$ qubits:

$$V_{\mathrm{d}}^{q-1} := \mathrm{Span}\{\ \sigma\sigma \ \|\ \ \sigma \in \boldsymbol{\sigma}_{q-1}\}. \tag{5.103}$$

Since we have that

$$S_{\sigma_0\sigma,\sigma_Z\sigma} = S_{\sigma_Z\sigma,\sigma_0\sigma} \tag{5.104}$$

for all $\sigma \in \boldsymbol{\sigma}_{q-1}$ there is an isomorphism θ between the vector spaces V_{d}^{q-1} and $V_{[\mathrm{S}]}^{\tau}$ of the form

$$\theta : V_{\mathrm{d}}^{q-1} \to V_{[\mathrm{S}]}^{\tau} : \sigma\sigma \mapsto S_{\sigma_0\sigma,\sigma_Z\sigma}. \tag{5.105}$$

Now consider the Clifford group on $q-1$ qubits, C_{q-1}. It can be seen as a subgroup of the group C_q^{τ} through the embedding

$$\hat{\theta} : \mathsf{C}_{q-1} \to \mathsf{C}_q^{\tau} : C \mapsto \mathbb{1} \otimes C. \tag{5.106}$$

Now note that C_q^{τ} preserves the commutation relations of the set $\boldsymbol{\sigma}_{q-1}$, that is, for all $\sigma, \hat{\sigma} \in \boldsymbol{\sigma}_{q-1}$ and $\sigma_1, \hat{\sigma}_1 \in \{\sigma_0, \sigma_Z\}$ we have

$$\left[C(\sigma_1\sigma)C^{\dagger}, C(\hat{\sigma}_1\hat{\sigma})C^{\dagger}\right] = 0 \iff [\sigma_1\sigma, \hat{\sigma}_1\hat{\sigma}] = 0 \iff [\sigma, \hat{\sigma}] = 0 \tag{5.107}$$

for all $C \in \mathsf{C}_q^{\tau}$ with the same conclusion holding for the anti-commutator. Now from this and eq. (5.104) one can see that for all $C \in \mathsf{C}_q^{\tau}$ there exists a $\hat{C} \in \mathsf{C}_{q-1}$ such that $\varphi^{\otimes 2}(C)v = \varphi^{\otimes 2}(\hat{\theta}(\hat{C}))v$ for all $v \in V_{[\mathrm{S}]}^{\tau}$. This means that for any subspace W of $V_{[\mathrm{S}]}^{\tau}$ we have

$$\varphi^{\otimes 2}(C)W \subset W, \ \ \forall C \in \mathsf{C}_q^{\tau} \iff \varphi^{\otimes 2}(\hat{\theta}(\hat{C}))W \subset W, \ \ \forall \hat{C} \in \mathsf{C}_{q-1} \tag{5.108}$$

Now let us consider the representation $\varphi^{\otimes 2}$ of C_{q-1} on $q-1$ qubits. Let's label the restriction of this representation to V_{d}^{q-1} as φ_{d}. From lemma 5.4 we see that V_{d} decomposes into three spaces carrying irreducible subrepresentations of φ_{d}. We shall label these V_0^{q-1}, V_1^{q-1} and V_2^{q-1}. Now note that we have for all $\hat{C} \in \mathsf{C}_{q-1}$ and all $\sigma \in \boldsymbol{\sigma}_{q-1}$ that

$$\varphi^{\otimes 2}(\hat{\theta}(\hat{C}))\theta(\sigma\sigma) = S_{\sigma_0 \hat{C}\sigma\hat{C}^\dagger, \sigma_Z \hat{C}\sigma\hat{C}^\dagger} = \theta(\varphi_{\mathrm{d}}(\hat{C})\sigma\sigma) \tag{5.109}$$

which implies that the representations φ_{d} and the subrepresentation of $\varphi^{\otimes 2}$ carried by $V_{[\mathrm{S}]}^{\tau}$ restricted to the image of $\hat{\theta}$ are equivalent with the equivalence given by the map θ. This means that the subspace $V_{[\mathrm{S}]}^{\tau}$ (with $\tau = \sigma_Z\sigma_0$) decomposes into three subspaces carrying irreducible subrepresentations of $\varphi^{\otimes 2}$ restricted to C_q^{τ}. We label these three spaces as

$$V_{[0]}^{\tau} := \theta(V_0^{q-1}), \qquad V_{[1]}^{\tau} := \theta(V_1^{q-1}), \qquad V_{[2]}^{\tau} := \theta(V_2^{q-1}), \tag{5.110}$$

with $\tau = \sigma_Z\sigma_0$. From lemma 5.7 and identifying the spaces $\oplus_{\tau' \in \boldsymbol{\sigma}_q} V_{[0]}^{\tau'}$ and $V_{[\mathrm{adj}]}$ we now arrive at the lemma statement. ∎

Finally we analyze the symmetric anti-commuting sector, i.e the space $V_{\{\mathrm{S}\}}$. This space carries an irreducible subrepresentation for $q = 1$ and falls apart into two subspaces carrying irreducible subrepresentations for $q \geq 2$. We have the following lemma.

Lemma 5.9 (Symmetric anti-commuting sector)**.** Take the space $V_{\{\mathrm{S}\}}$ as defined in definition 24 and define the subspaces

$$V_{\{1\}} := \bigoplus_{\tau \in \boldsymbol{\sigma}_q} V_{\{1\}}^{\tau}, \qquad V_{\{2\}} := \bigoplus_{\tau \in \boldsymbol{\sigma}_q} V_{\{2\}}^{\tau}, \tag{5.111}$$

where for all $\tau \in \boldsymbol{\sigma}_q$:

$$V_{\{1\}}^{\tau} := \mathrm{Span}\left\{ v^{\tau} \in V_{\{\mathrm{S}\}} \parallel v^{\tau} = \sum_{\sigma \in \boldsymbol{N}_\tau} \lambda_\sigma S_{\sigma, i\sigma\cdot\tau}, \sum_{\sigma \in \boldsymbol{N}_\tau \cap \boldsymbol{C}_\nu} \lambda_\sigma - \sum_{\sigma \in \boldsymbol{N}_\tau \cap \boldsymbol{N}_\nu} \lambda_\sigma = \frac{d}{2}\lambda_\nu, \forall \nu \in \boldsymbol{N}_\tau \right\},$$

$$V_{\{2\}}^{\tau} := \mathrm{Span}\left\{ v^{\tau} \in V_{\{\mathrm{S}\}} \parallel v^{\tau} = \sum_{\sigma \in \boldsymbol{N}_\tau} \lambda_\sigma S_{\sigma, i\sigma\cdot\tau}, \sum_{\sigma \in \boldsymbol{N}_\tau \cap \boldsymbol{C}_\nu} \lambda_\sigma - \sum_{\sigma \in \boldsymbol{N}_\tau \cap \boldsymbol{N}_\nu} \lambda_\sigma = -\frac{d}{2}2\lambda_\nu, \forall \nu \in \boldsymbol{N}_\tau \right\}.$$

We have the following statements:

- For $q = 1$ the space $V_{\{\mathrm{S}\}}$ carries an irreducible subrepresentation of $\varphi^{\otimes 2}$
- For $q \geq 2$ we have $V_{\{\mathrm{S}\}} = V_{\{1\}} \oplus V_{\{2\}}$ and the spaces $V_{\{1\}}$ and $V_{\{2\}}$ carry subrepresentations of $\varphi^{\otimes 2}$

Proof. The $q = 1$ case was dealt with in [11], we will deal with the case of $q \geq 2$. The argument goes by a combination of the arguments in lemma 5.4 and lemma 5.8. First note that we can write $V_{\{\mathrm{S}\}}$ as

$$V_{\{\mathrm{S}\}} = \bigoplus_{\tau \in \boldsymbol{\sigma}_q} V_{\{\mathrm{S}\}}^{\tau}, \qquad V_{\{\mathrm{S}\}}^{\tau} = \mathrm{Span}\{S_{\sigma, i\sigma\cdot\tau} \parallel \sigma \in \boldsymbol{N}_\tau\}. \tag{5.112}$$

We can again use lemma 5.7 to look for subspaces of $V_{\{S\}}$ carrying subrepresentations of $\varphi^{\otimes 2}$ by considering the action of the strict subgroup C_q^τ of C_q on the space $V_{\{S\}}^\tau$ (where C_q^τ is defined as in lemma 5.7). As in lemma 5.8 we choose $\tau = \sigma_Z\sigma_0$. The elements of $\boldsymbol{\sigma}_q$ that anti-commute with τ can now be seen to be

$$\boldsymbol{N}_\tau = \{\ \sigma_X\sigma, \sigma_Y\sigma\ \ \|\ \ \sigma \in \hat{\boldsymbol{\sigma}}_{q-1}\}. \tag{5.113}$$

Note that the set $\boldsymbol{N}_\tau$ leads to an ambiguous definition of a basis for $V_{\{S\}}^\tau$ as we have that

$$S_{\sigma_X\sigma, i(\sigma_X\sigma)\cdot(\sigma_Z\sigma_0)} = -S_{\sigma_Y\sigma, i(\sigma_Y\sigma)\cdot(\sigma_Z\sigma_0)}. \tag{5.114}$$

for all $\sigma \in \hat{\boldsymbol{\sigma}}_{q-1}$ (recall that $\hat{\boldsymbol{\sigma}}_{q-1} = \boldsymbol{\sigma}_{q-1} \cup \{\sigma_0\}$). We resolve this ambiguity by choosing the set $\{\sigma_X\sigma\ \ |\ \ \sigma \in \hat{\boldsymbol{\sigma}}_{q-1}\}$ to generate a basis of $V_{\{S\}}^\tau$. This makes that

$$V_{\{S\}}^\tau = \mathrm{Span}\{S_{\sigma_X\sigma,\sigma_Y\sigma}\ \ \|\ \ \sigma \in \hat{\boldsymbol{\sigma}}_{q-1}\}. \tag{5.115}$$

In the spirit of lemma 5.4 we define the following linear map $\boldsymbol{\mathcal{A}}$ as a linear extension of the action on the basis of $V_{\{S\}}^\tau$ as

$$\boldsymbol{\mathcal{A}}\left(S_{\sigma_X\sigma, i(\sigma_X\sigma)\cdot\tau}\right) = \sum_{\hat{\sigma}\in \boldsymbol{C}_{\sigma_X\sigma}\cap\boldsymbol{N}_\tau} S_{\hat{\sigma}, i\hat{\sigma}\cdot\tau} - \sum_{\hat{\sigma}\in \boldsymbol{N}_{\sigma_X\sigma}\cap\boldsymbol{N}_\tau} S_{\hat{\sigma}, i\hat{\sigma}\cdot\tau}. \tag{5.116}$$

We can argue that this map commutes with the action of $\varphi^{\otimes 2}$ restricted to C_q^τ (Where C_q^τ is defined as in lemma 5.7 with $\tau = \sigma_Z\sigma_0$) by direct calculation. We have for $C \in \mathsf{C}_q^\tau$ and $\sigma \in \hat{\boldsymbol{\sigma}}_{q-1}$:

$$\begin{aligned}
\boldsymbol{\mathcal{A}}\left[\varphi^{\otimes 2}(C)\left(S_{\sigma_X\sigma, i(\sigma_X\sigma)\cdot\tau}\right)\right] &= \boldsymbol{\mathcal{A}}\left[S_{C(\sigma_X\sigma)C^\dagger, iC(\sigma_X\sigma)\cdot(\tau)C^\dagger}\right] && (5.117)\\
&= \sum_{\hat{\sigma}\in \boldsymbol{C}_{C(\sigma_X\sigma)C^\dagger}\cap\boldsymbol{N}_\tau} S_{\hat{\sigma}, i\hat{\sigma}\cdot C\tau C^\dagger} - \sum_{\hat{\sigma}\in \boldsymbol{N}_{C(\sigma_X\sigma)C^\dagger}\cap\boldsymbol{N}_\tau} S_{\hat{\sigma}, i\hat{\sigma}\cdot C\tau C^\dagger} && (5.118)\\
&= \sum_{C^\dagger\hat{\sigma}C\in \boldsymbol{C}_{\sigma_X\sigma}\cap\boldsymbol{N}_\tau} S_{\hat{\sigma}, i\hat{\sigma}\cdot C\tau C^\dagger} - \sum_{C^\dagger\hat{\sigma}C\in \boldsymbol{N}_{\sigma_X\sigma}\cap\boldsymbol{N}_\tau} S_{\hat{\sigma}, i\hat{\sigma}\cdot C\tau C^\dagger} && (5.119)\\
&= \sum_{\hat{\sigma}\in \boldsymbol{C}_{\sigma_X\sigma}\cap\boldsymbol{N}_\tau} S_{C\hat{\sigma}C^\dagger, iC\hat{\sigma}\cdot\tau C^\dagger} - \sum_{\hat{\sigma}\in \boldsymbol{N}_{\sigma_X\sigma}\cap\boldsymbol{N}_\tau} S_{C\hat{\sigma}C^\dagger, iC\hat{\sigma}\cdot\tau C^\dagger} && (5.120)\\
&= \varphi^{\otimes 2}(C)\left(\boldsymbol{\mathcal{A}}\left[S_{\sigma_X\sigma, i(\sigma_X\sigma)\cdot\tau}\right]\right). && (5.121)
\end{aligned}$$

This means that, through Schur's lemma the map $\boldsymbol{\mathcal{A}}$ tells us something about the subrepresentations of $\varphi^{\otimes 2}$ restricted to C_q^τ carried by $V_{\{S\}}^\tau$. Because $\tau = \sigma_Z\sigma_0$ we can write $\boldsymbol{\mathcal{A}}$

in a slightly better form by noting

$$\boldsymbol{\mathcal{A}}\left(S_{\sigma_X\sigma,i(\sigma_X\sigma)\cdot\tau}\right) = \sum_{\hat{\sigma}\in\boldsymbol{C}_{\sigma_X\sigma}\cap\boldsymbol{N}_\tau} S_{\hat{\sigma},i\hat{\sigma}\cdot\tau} - \sum_{\hat{\sigma}\in\boldsymbol{N}_{\sigma_X\sigma}\cap\boldsymbol{N}_\tau} S_{\hat{\sigma},i\hat{\sigma}\cdot\tau} \tag{5.122}$$

$$= \left[\sum_{\sigma'\in\hat{\boldsymbol{C}}_\sigma} S_{\sigma_X\sigma',\sigma_Y\sigma'} + \sum_{\sigma'\in\boldsymbol{N}_\sigma} S_{\sigma_Y\sigma',-\sigma_X\sigma'}\right] - \left[\sum_{\sigma'\in\boldsymbol{N}_\sigma} S_{\sigma_X\sigma',\sigma_Y\sigma'} + \sum_{\sigma'\in\hat{\boldsymbol{C}}_\sigma} S_{\sigma_Y\sigma',-\sigma_X\sigma'}\right] \tag{5.123}$$

$$= 2\left[\sum_{\sigma'\in\hat{\boldsymbol{C}}_\sigma} S_{\sigma_X\sigma',\sigma_Y\sigma'} - \sum_{\sigma'\in\boldsymbol{N}_\sigma} S_{\sigma_X\sigma',\sigma_Y\sigma'}\right], \tag{5.124}$$

where we recall $\hat{\boldsymbol{C}}_\sigma$ to be $\hat{\boldsymbol{C}}_\sigma = \boldsymbol{C}_\sigma \cup \{\sigma_0, \sigma\}$. We now analyze the properties of the map $\boldsymbol{\mathcal{A}}$ by calculating $\mathrm{Tr}(\boldsymbol{\mathcal{A}})$ and $\boldsymbol{\mathcal{A}}^2$. We have

$$\frac{1}{2}\mathrm{Tr}(\boldsymbol{\mathcal{A}}) = \frac{1}{2}\sum_{\sigma\in\hat{\boldsymbol{\sigma}}_{q-1}} \langle S_{\sigma_X\sigma,\sigma_Y\sigma}, \boldsymbol{\mathcal{A}}\left(S_{\sigma_X\sigma,\sigma_Y\sigma}\right)\rangle \tag{5.125}$$

$$= \sum_{\sigma\in\hat{\boldsymbol{\sigma}}_{q-1}}\left[\sum_{\hat{\sigma}\in\hat{\boldsymbol{C}}_\sigma}\delta_{\sigma,\hat{\sigma}} - \sum_{\hat{\sigma}\in\boldsymbol{N}_\sigma}\delta_{\sigma,\hat{\sigma}}\right] \tag{5.126}$$

$$= |\hat{\boldsymbol{\sigma}}_{q-1}| = \left(\frac{d}{2}\right)^2. \tag{5.127}$$

We can calculate $\boldsymbol{\mathcal{A}}^2$ entry-wise. We abuse notation a little bit by denoting the entries of $\boldsymbol{\mathcal{A}}^2$ as $[\boldsymbol{\mathcal{A}}^2]_{\sigma,\hat{\sigma}}$ with $\sigma, \hat{\sigma} \in \hat{\boldsymbol{\sigma}}_{q-1}$ (this set has a one-to-one correspondence with the basis of $V_{\{\mathrm{S}\}}$ given in eq. (5.114)). We calculate:

$$\frac{1}{4}[\boldsymbol{\mathcal{A}}^2]_{\sigma,\hat{\sigma}} = \frac{1}{4}\langle S_{\sigma_X\sigma,\sigma_Y\sigma}, \boldsymbol{\mathcal{A}}^2\left[S_{\sigma_X\hat{\sigma},\sigma_Y\hat{\sigma}}\right]\rangle \tag{5.128}$$

$$= \sum_{\substack{\sigma''\in\hat{\boldsymbol{C}}_{\sigma'}\\ \sigma'\in\hat{\boldsymbol{C}}_\sigma}}\delta_{\sigma'',\hat{\sigma}} - \sum_{\substack{\sigma''\in\boldsymbol{N}_{\sigma'}\\ \sigma'\in\hat{\boldsymbol{C}}_\sigma}}\delta_{\sigma'',\hat{\sigma}} - \sum_{\substack{\sigma''\in\hat{\boldsymbol{C}}_{\sigma'}\\ \sigma'\in\boldsymbol{N}_\sigma}}\delta_{\sigma'',\hat{\sigma}} + \sum_{\substack{\sigma''\in\boldsymbol{N}_{\sigma'}\\ \sigma'\in\boldsymbol{N}_\sigma}}\delta_{\sigma'',\hat{\sigma}} \tag{5.129}$$

$$= |\hat{\boldsymbol{C}}_\sigma \cap \hat{\boldsymbol{C}}_{\hat{\sigma}}| - |\hat{\boldsymbol{C}}_\sigma \cap \boldsymbol{N}_{\hat{\sigma}}| - |\boldsymbol{N}_\sigma \cap \hat{\boldsymbol{C}}_{\hat{\sigma}}| + |\boldsymbol{N}_\sigma \cap \boldsymbol{N}_{\hat{\sigma}}| \tag{5.130}$$

$$= \delta_{\sigma,\hat{\sigma}}|\hat{\boldsymbol{\sigma}}_{q-1}| = \delta_{\sigma,\hat{\sigma}}\left(\frac{d}{2}\right)^2, \tag{5.131}$$

where the last equality follows directly from lemma 5.1. We see that $\boldsymbol{\mathcal{A}}^2$ is proportional to the identity. This means that the eigenvalues of $\boldsymbol{\mathcal{A}}$ must be $\pm d$. Since $\boldsymbol{\mathcal{A}}$ is not proportional to the identity this means that both eigenvalues must be associated with non-trivial eigenspaces. Schur's lemma thus implies that $V^\tau_{\{\mathrm{S}\}}$ carries a reducible subrepresentation of

$\varphi^{\otimes 2}$ restricted to C_q^τ and moreover that the eigenspaces of $\boldsymbol{\mathcal{A}}$ must be subrepresentations. We will call the spaces carrying these these subrepresentations $V_{\{1\}}^\tau$ and $V_{\{2\}}^\tau$ where we identify $V_{\{1\}}^\tau$ with the d eigenvalue of $\boldsymbol{\mathcal{A}}$ and $V_{\{2\}}^\tau$ with the $-d$ eigenvalue of $\boldsymbol{\mathcal{A}}$. We can find out the dimensions of these spaces by noting that

$$\mathrm{Tr}(\boldsymbol{\mathcal{A}}) = d|V_{\{1\}}^\tau| - d|V_{\{2\}}^\tau| = \frac{d^2}{2}, \tag{5.132}$$

$$|V_{\{1\}}^\tau| + |V_{\{2\}}^\tau| = \left(\frac{d}{2}\right)^2. \tag{5.133}$$

Solving these equations yields

$$|V_{\{1\}}^\tau| = \frac{d}{4}\left(\frac{d}{2}+1\right), \qquad |V_{\{2\}}^\tau| = \frac{d}{4}\left(\frac{d}{2}-1\right). \tag{5.134}$$

Diagonalizing $\boldsymbol{\mathcal{A}}$ then yields the equations given in the lemma statement for $V_{\{1\}}^\tau$ and $V_{\{2\}}^\tau$ and by lemma 5.7 we also get that $V_{\{1\}}$ and $V_{\{2\}}$ as defined in the lemma statement carry subrepresentations of the subrepresentation carried by $V_{\{\mathrm{S}\}}$. ■

5

Note that we have not argued that the spaces $V_{\{1\}}, V_{\{2\}}$ carry irreducible subrepresentations. We will get the irreducibility for free in the full decomposition theorem, which we will deal with now. Using lemmas 5.4 to 5.6, 5.8 and 5.9 we can prove the main result of this chapter: a decomposition of the two-copy representation $\varphi^{\otimes 2}$ of the Clifford group C_q valid for any number of qubits q. We have:

Theorem 5.1 (Decomposition of the two-copy representation)**.** The decomposition of the vector space $\mathcal{M}_d^{\otimes 2} = \mathrm{Span}\{\boldsymbol{\mathcal{B}}\}$ into subspaces carrying irreducible subrepresentations of C_q in $\varphi^{\otimes 2}$ for different values of q is:

$$V_{\mathrm{id}} \oplus V_{\mathrm{r}} \oplus V_{\mathrm{l}} \oplus V_0 \oplus V_1 \oplus V_{\{\mathrm{S}\}} \oplus V_{\{\mathrm{A}\}}, \qquad (q=1)$$

$$V_{\mathrm{id}} \oplus V_{\mathrm{r}} \oplus V_{\mathrm{l}} \oplus V_0 \oplus V_1 \oplus V_2 \oplus V_{[\mathrm{adj}]} \oplus V_{[1]} \oplus V_{\{\mathrm{adj}\}} \oplus V_{\{1\}} \oplus V_{\{2\}} \oplus V_{[\mathrm{A}]} \oplus V_{\{\mathrm{adj}\}}^\perp, \qquad (q=2)$$

$$V_{\mathrm{id}} \oplus V_{\mathrm{r}} \oplus V_{\mathrm{l}} \oplus V_0 \oplus V_1 \oplus V_2 \oplus V_{[\mathrm{adj}]} \oplus V_{[1]} \oplus V_{[2]} \oplus V_{\{\mathrm{adj}\}} \oplus V_{\{1\}} \oplus V_{\{2\}} \oplus V_{[\mathrm{A}]} \oplus V_{\{\mathrm{adj}\}}^\perp, \qquad (q \geq 3)$$

where all spaces are as defined in definition 24 and lemmas 5.4 to 5.6, 5.8 and 5.9 and are gathered in table 5.1 in the appendix.

Proof. The $q = 1$ case is dealt with in [11]. We will now deal with the cases $q = 2$ and $q \geq 3$. Beginning with $q \geq 3$ note that we have already argued (in lemmas 5.4 to 5.6, 5.8 and 5.9) that all spaces given in theorem 5.1 are non-trivial and carry subrepresentations of $\varphi^{\otimes 2}$. It remains to argue that these subrepresentations are all irreducible. We will do this using the Schur orthogonality relations (lemma 3.3) and lemma 5.2. Begin by noting that the representations carried by the spaces $V_{\mathrm{r}}, V_{\mathrm{l}}, V_{[\mathrm{adj}]}$ and $V_{\{\mathrm{adj}\}}$ are equivalent (lemma 5.5), the representations carried by the spaces $V_{[\mathrm{A}]}$ and $V_{\{\mathrm{adj}\}}^\perp$ are equivalent (lemma 5.6) and the representations carried by V_{id} and V_0 are equivalent (Because they are both the trivial representation). Denote the character of the representations spanned by the direct sum of

these representations by χ_{sum}. By the Schur orthogonality relations we have the following relation

$$\langle \chi_{\text{sum}}, \chi_{\text{sum}} \rangle \geq 16 + 4 + 4 = 24, \tag{5.135}$$

with equality if and only if all these spaces carry irreducible subrepresentations. Noting that we have yet to include the spaces $V_1, V_2, V_{\{1\}}, V_{\{2\}}, V_{[1]}$ and $V_{[2]}$ we can lower bound the character of $\varphi^{\otimes 2}$ as

$$\langle \chi_{\varphi^{\otimes 2}}, \chi_{\varphi^{\otimes 2}} \rangle \geq \langle \chi_{\text{sum}}+\chi_1+\chi_2+\chi_{\{1\}}+\chi_{\{2\}}+\chi_{[1]}+\chi_{[2]}, \chi_{\text{sum}}+\chi_1+\chi_2+\chi_{\{1\}}+\chi_{\{2\}}+\chi_{[1]}+\chi_{[2]} \rangle,$$

where χ_i is the character associated with the subrepresentation carried by the space V_i and the inequality accounts for the fact that some of the subrepresentations might a priori be equivalent and/or reducible. From lemma 3.3 now we have

$$\langle \chi_{\varphi^{\otimes 2}}, \chi_{\varphi^{\otimes 2}} \rangle \geq 30. \tag{5.136}$$

From lemma 5.2 we note that $\langle \chi_{\varphi^{\otimes 2}}, \chi_{\varphi^{\otimes 2}} \rangle = 30$ for $q \geq 30$. This means that all spaces mentioned must carry irreducible subrepresentations of $\varphi^{\otimes 2}$ and that the spaces $V_0, V_{\text{r}}, V_{[\text{A}]}, V_1, V_2, V_{\{1\}}, V_{\{2\}}, V_{[1]}$ and $V_{[2]}$ must carry mutually inequivalent irreducible representations. We can make the same argument for $q = 2$ noting that the space $V_{[2]} = \emptyset$ (and hence does not contribute to the character inner product) and that for $q = 2$ we have $\langle \chi_{\varphi^{\otimes 2}}, \chi_{\varphi^{\otimes 2}} \rangle = 29$. This completes the classification of the irreducible representations of the two-copy representation $\varphi^{\otimes 2}$ of the q-qubit Clifford group C_q. ■

5.4. CONCLUSION

We characterized the two-copy representation of the multi-qubit Clifford group and identified three distinct cases, namely, the single-qubit (analyzed in [11]), two-qubit, and many-qubit cases, which contain 7, 13, and 14 irreducible representations respectively.

As the Clifford group plays a central role in quantum information, we expect the present analysis to have many applications such as state & channel tomography, analysis of fault-tolerance thresholds, large-deviation bounds [14] and state distinguishability (as analyzed in [15–17]). As a concrete example, we will use the results derived in this chapter in chapters 6 and 7 to provide a much sharper analysis of the statistical performance of randomized benchmarking [6, 18] and unitarity randomized benchmarking [19]. While this result advances understanding of the representation theory of the Clifford groups, there remain several open questions about general representation theory of multi-qubit Clifford groups. First and foremost, the character table of the Clifford group is unknown. Working out this table would greatly assist future studies. In this chapter we have identified several distinct irreducible representations, which should assist in the construction of the character table. Finally, these results hold for qubits and generalizing them to higher-dimensional systems remains an open problem. For completeness we present in table 5.1 all vector spaces mentioned in this chapter, as well as their relation to each other and the two-copy representation of the multi-qubit Clifford group.

space	definition	irreducible	dimension
V_{id}	$\mathrm{Span}\{\sigma_0\sigma_0\}$	$q \geq 1$	1
V_{r}	$\mathrm{Span}\{\sigma_0\tau \;\|\; \tau \in \boldsymbol{\sigma}_q\}$	$q \geq 1$	d^2-1
V_{l}	$\mathrm{Span}\{\tau\sigma_0 \;\|\; \tau \in \boldsymbol{\sigma}_q\}$	$q \geq 1$	d^2-1
V_{d}	$\mathrm{Span}\{\tau\tau \;\|\; \tau \in \boldsymbol{\sigma}_q\}$	no	d^2-1
$V_{[\mathrm{S}]}$	$\mathrm{Span}\Big\{S_{\sigma,\tau} \;\|\; \sigma \in \boldsymbol{C}_\tau,\ \tau \in \boldsymbol{\sigma}_q\Big\}$	no	$\frac{d^2-1}{2}\left(\frac{d^2}{2}-2\right)$
$V_{\{\mathrm{S}\}}$	$\mathrm{Span}\Big\{S_{\sigma,\tau} \;\|\; \sigma \in \boldsymbol{N}_\tau,\ \tau \in \boldsymbol{\sigma}_q\Big\}$	$q=1$	$\frac{d^2-1}{2}\left(\frac{d^2}{2}\right)$
$V_{[\mathrm{A}]}$	$\mathrm{Span}\Big\{A_{\sigma,\tau} \;\|\; \sigma \in \boldsymbol{C}_\tau,\ \tau \in \boldsymbol{\sigma}_q\Big\}$	$q \geq 2$	$\frac{d^2-1}{2}\left(\frac{d^2}{2}-2\right)$
$V_{\{\mathrm{A}\}}$	$\mathrm{Span}\Big\{A_{\sigma,\tau} \;\|\; \sigma \in \boldsymbol{N}_\tau,\ \tau \in \boldsymbol{\sigma}_q\Big\}$	$q=1$	$\frac{d^2-1}{2}\left(\frac{d^2}{2}\right)$
V_0	$\mathrm{Span}\left\{w \in V_{\mathrm{d}} \;\|\; w = \frac{1}{\sqrt{d^2-1}}\sum_{\sigma\in\boldsymbol{\sigma}_q}\sigma\sigma\right\}$	$q \geq 1$	1
V_1	$\mathrm{Span}\left\{v \in V_d \;\|\; v = \sum_{\sigma\in\boldsymbol{\sigma}_q}\lambda_\sigma\sigma\sigma,\ \sum_{\sigma\in\boldsymbol{\sigma}_q}\lambda_\sigma = 0,\ \sum_{\sigma\in\boldsymbol{N}_\nu}\lambda_\sigma = -\frac{d}{2}\lambda_\tau,\ \tau\in\boldsymbol{\sigma}_q\right\}$	$q \geq 1$	$\frac{d(d+1)}{2}-1$
V_2	$\mathrm{Span}\left\{v \in V_d \;\|\; v = \sum_{\sigma\in\boldsymbol{\sigma}_q}\lambda_\sigma\sigma\sigma,\ \sum_{\sigma\in\boldsymbol{\sigma}_q}\lambda_\sigma = 0,\ \sum_{\sigma\in\boldsymbol{N}_\nu}\lambda_\sigma = \frac{d}{2}\lambda_\tau,\ \tau\in\boldsymbol{\sigma}_q\right\}$	$q \geq 2$	$\frac{d(d-1)}{2}-1$
$V_{[\mathrm{adj}]}$	$\mathrm{Span}\Big\{v_\tau^{[\mathrm{adj}]} \in V_{[\mathrm{S}]} \;\|\; v_\tau^{[\mathrm{adj}]} = \frac{1}{\sqrt{2\lvert\boldsymbol{C}_\tau\rvert}}\sum_{\sigma\in\boldsymbol{C}_\tau}S_{\sigma,\sigma\cdot\tau},\ \tau\in\boldsymbol{\sigma}_q\Big\}$	$q \geq 2$	(d^2-1)
$V_{\{\mathrm{adj}\}}$	$\mathrm{Span}\Big\{v_\tau^{\{\mathrm{adj}\}} \in V_{\{\mathrm{A}\}} \;\|\; v_\tau^{\{\mathrm{adj}\}} = \frac{1}{\sqrt{2\lvert\boldsymbol{N}_\tau\rvert}}\sum_{\sigma\in\boldsymbol{N}_\tau}A_{\sigma,i\sigma\cdot\tau},\ \tau\in\boldsymbol{\sigma}_q\Big\}$	$q \geq 2$	(d^2-1)
$V_{\{\mathrm{adj}\}}^\perp$	$\mathrm{Span}\Big\{v^{\{A\}} \in V_{\{\mathrm{A}\}} \;\|\; \langle v^{\{A\}}, v^{\{\mathrm{adj}\}}\rangle = 0,\ \forall\, v^{\{\mathrm{adj}\}} \in V_{\{\mathrm{adj}\}}\Big\}$	$q \geq 2$	$(d^2-1)\left(\frac{d^2}{2}-2\right)$
$V_{[1]}$	$\mathrm{Span}\left\{v^\tau \in V_{[\mathrm{S}]} \;\|\; v^\tau = \sum_{\sigma\in\boldsymbol{N}_\tau}\lambda_\sigma\sigma\sigma,\ \sum_{\sigma\in\boldsymbol{C}_\tau\cap\boldsymbol{N}_\nu}\lambda_\sigma = -d\lambda_\nu,\ \nu\in\boldsymbol{C}_\tau,\ \tau\in\boldsymbol{\sigma}_q\right\}$	$q \geq 2$	$(d^2-1)\left[\frac{\frac{d}{2}(\frac{d}{2}+1)}{2}-1\right]$
$V_{[2]}$	$\mathrm{Span}\left\{v^\tau \in V_{[\mathrm{S}]} \;\|\; v^\tau = \sum_{\sigma\in\boldsymbol{N}_\tau}\lambda_\sigma\sigma\sigma,\ \sum_{\sigma\in\boldsymbol{C}_\tau\cap\boldsymbol{N}_\nu}\lambda_\sigma = d\lambda_\nu,\ \nu\in\boldsymbol{C}_\tau,\ \tau\in\boldsymbol{\sigma}_q\right\}$	$q \geq 3$	$(d^2-1)\left[\frac{\frac{d}{2}(\frac{d}{2}-1)}{2}-1\right]$
$V_{\{1\}}$	$\mathrm{Span}\left\{v^\tau \in V_{\{\mathrm{S}\}} \;\|\; v^\tau = \sum_{\sigma\in\boldsymbol{N}_\tau}\lambda_\sigma S_{\sigma,i\sigma\cdot\tau},\ \sum_{\sigma\in\boldsymbol{N}_\tau\cap\boldsymbol{C}_\nu}\lambda_\sigma - \sum_{\sigma\in\boldsymbol{N}_\tau\cap\boldsymbol{N}_\nu}\lambda_\sigma = \frac{d}{2}\lambda_\nu,\ \nu\in\boldsymbol{N}_\tau,\ \tau\in\boldsymbol{\sigma}_q\right\}$	$q=1$	$(d^2-1)\frac{\frac{d}{2}(\frac{d}{2}+1)}{2}$
$V_{\{2\}}$	$\mathrm{Span}\left\{v^\tau \in V_{\{\mathrm{S}\}} \;\|\; v^\tau = \sum_{\sigma\in\boldsymbol{N}_\tau}\lambda_\sigma S_{\sigma,i\sigma\cdot\tau},\ \sum_{\sigma\in\boldsymbol{N}_\tau\cap\boldsymbol{C}_\nu}\lambda_\sigma - \sum_{\sigma\in\boldsymbol{N}_\tau\cap\boldsymbol{N}_\nu}\lambda_\sigma = -\frac{d}{2}\lambda_\nu,\ \nu\in\boldsymbol{N}_\tau,\ \tau\in\boldsymbol{\sigma}_q\right\}$	$q \geq 2$	$(d^2-1)\frac{\frac{d}{2}(\frac{d}{2}-1)}{2}$

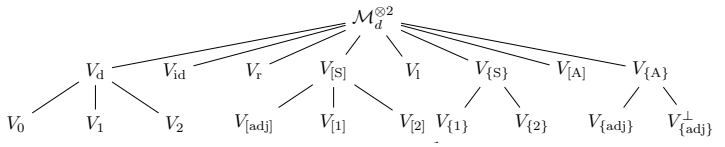


Table 5.1: Table with all subspaces of $\mathcal{M}_d$ carrying subrepresentations of $\varphi^{\otimes 2}$. Given are the name in the text, the definition, for which values (if any) of $q \in \mathbb{N}$ they carry irreducible subrepresentations of $\varphi^{\otimes 2}$ and their dimension as a function of $d = 2^q$. Also given is a tree diagram showing subspace inclusions where every child node is a subspace of its parent nodes.

6

The statistics of randomized benchmarking

Randomized benchmarking (RB) is an efficient and robust method to characterize gate errors in quantum circuits. Averaging over random sequences of gates leads to estimates of gate errors in terms of the average fidelity. These estimates are isolated from the state preparation and measurement errors that plague other methods like channel tomography and direct fidelity estimation. A decisive factor in the feasibility of randomized benchmarking is the number of sampled sequences required to obtain rigorous confidence intervals. Previous bounds were either prohibitively loose or required the number of sampled sequences to scale exponentially with the number of qubits in order to obtain a fixed confidence interval at a fixed error rate.

Here we show that, with a small adaptation to the randomized benchmarking procedure, the number of sampled sequences required for a fixed confidence interval is dramatically smaller than could previously be justified. In particular, we show that the number of sampled sequences required is essentially independent of the number of qubits and scales favorably with the average error rate of the system under investigation. We also investigate the fitting procedure inherent to randomized benchmarking in the light of our results and find that standard methods such as ordinary least squares optimization can give misleading results. We therefore recommend moving to more sophisticated fitting methods such as iteratively reweighted least squares optimization. Our results bring rigorous randomized benchmarking on systems with many qubits into the realm of experimental feasibility.

This chapter has been published, with minor changes, in J. Helsen, J.J. Wallman, S.T. Flammia & S. Wehner, *Multi-qubit Randomized Benchmarking Using Few Samples*, PRXpleaseacceptme.

6.1. Introduction

In chapter 4 we introduced the randomized benchmarking (RB) protocol. There we noted that an important practical problem when using RB is choosing a number of random gate sequences that is sufficiently small to be practical experimentally, and yet gives a good estimate of the gate fidelity. This problem becomes increasingly relevant as error rates improve since estimating small errors accurately ordinarily requires more samples. Early treatments of this problem demanded numbers of sequences that were orders of magnitude larger than were feasible in experiment [1]. A more specialized analysis allowed rigorous confidence intervals to be derived for a number of random sequences comparable to the number used in experiments [2]. However, this analysis only provided reasonable bounds on the number of sequences for short sequence lengths and for single qubit experiments while more general multi-qubit bounds had an unfavorable exponential scaling with the number of qubits being benchmarked. The restriction to short sequence lengths is also problematic because long sequences generally lead to better experimental fits [3, 4].

In this chapter, we propose an adapted version of the standard RB protocol on the set of Clifford gates that requires little experimental overhead. For this protocol we provide a bound on the number of random sequences required to obtain rigorous confidence intervals that is several orders of magnitude sharper than previous multi-qubit bounds. Our result makes rigorous and efficient characterization of multi-qubit systems possible using a reasonable amount of experimental resources. In particular, our bounds are approximately independent of the number of qubits being benchmarked As a special case, we also obtain bounds for the single-qubit version of RB that are valid for all sequence lengths and improve on the bounds of Ref. [2] for long sequence lengths. The key to the analysis of the statistical performance is a novel understanding of the representations of the Clifford group, which we developed in chapter 5. Similar representation-theoretic questions have also been studied independently by Zhu *et al.* [5]. We also prove a precise sense in which the derived bounds are optimal.

In section 6.2 we present an overview of the new contributions of this chapter (equations of note here are eqs. (6.4) and (6.6)) and explain their context. In section 6.3, we discuss the implications of the new bound for experiments, and investigate it in various limits. Finally, in section 6.5 we discuss the derivation of the new bounds and how to apply them in practice. We also prove that our results are optimal in some well specified sense. We focus on intuition and displace most of the technical proofs to the technical statements (section 6.6) section in the back of the chapter.

The randomized benchmarking protocol

In fig. 6.1 we lay out our version of the randomized benchmarking protocol as it was analyzed in [1, 2, 6]. We will perform randomized benchmarking over the Clifford group on q qubits C_q. We refer to chapter 2 for the definition of the Clifford group.

We make two essential changes to the standard randomized benchmarking protocol, both of which lead to better guarantees on the precision of randomized benchmarking.

- A first modification is to perform each randomized benchmarking sequence twice, but with different input states $\rho, \hat{\rho}$ and then subtracting the result. This is equivalent

1. Choose a random sequence $\vec{G} = (G_1, \ldots, G_m)$ of m gates independently and uniformly at random from the q-qubit Clifford group C_q and compute the gate $G_{m+1} = (G_m \ldots G_1)^\dagger$.

2. Prepare q qubits in a state ρ that maximizes $\mathrm{Tr}(\rho\mathbf{P})$ [e.g., $\rho \approx 2^{-q}(I + \mathbf{P})$].

3. For $t = 1, \ldots, m+1$, apply the gate G_t to ρ.

4. Measure the expectation value $p_m(\vec{G})(\rho)$ of POVM $\{Q, \mathbb{1} - Q\}$ with $Q \approx \frac{1}{2}(\mathbb{1} + \mathbb{P})$ to a suitable precision (By repeating 1-3 for the same sequence L times)

5. Repeat these steps for the same string $\vec{G}$ but for a different state $\hat{\rho}$ [ideally, $\rho \approx 2^{-q}(I - \mathbf{P})$]. and compute $k_m(\vec{G}) = \frac{1}{2}(p_m(\vec{G})(\rho) - p_m(\vec{G})(\hat{\rho}))$.

6. Repeat steps 1–5 a total of N times to estimate

$$\mathbb{E}_{\vec{G}}(K_m) = |\mathsf{C}_q|^{-m} \sum_{\vec{G} \in \mathsf{C}_q^m} k_m(\vec{G})$$

   to a suitable precision (implicitly regarding the $k_m(\vec{G})$ as realizations of a random variable K_m). We call the empirical average over the N sampled Clifford sequences $k_{m,N}$

7. Repeat steps 1–6 for multiple values of m and fit to the decay model

$$\mathbb{E}_{\vec{G}}(K_m) = Af^m,$$

   where $f = (dF_{\mathrm{avg}}(\mathcal{E}, \mathcal{I}) - 1)/d - 1$ is the depolarizing parameter as defined in chapter 4 [1] (and $d = 2^q$).

Figure 6.1: **The Randomized Benchmarking Protocol.** We perform randomized benchmarking using the Clifford group C_q, i.e. all gates that can be constructed by successive application of CNOT gates, Hadamard gates and $\pi/4$ phase gates. We assume the input states ρ, $(\hat{\rho})$ to be noisy implementations of the states $2^{-q}(I + \mathbf{P})$, $(2^{-q}(I - \mathbf{P}))$, and Q a noisy implementation of the observable $\mathbf{P}$ where $\mathbf{P}$ is a Pauli operator. We denote the length of an RB sequence by m, the amount of random sequences for a given m by N and the amount of times a single sequence is repeated by L. The goal of this chapter is to provide confidence intervals around the empirical average $k_{m,N}$ assuming that individual realizations $k_m(\vec{G})$ are estimated to very high precision (corresponding to the case $L \to \infty$). In experimental implementations, running the same sequence many times (L) is typically easy, but running many different sequences (N) is hard [3], meaning that the quantity that we want to minimize is N. See section 6.5 for a detailed discussion of the construction of confidence regions around the empirical average $k_{m,N}$

to performing standard randomized benchmarking with the "input operator" $\nu = \frac{1}{2}(\rho - \hat{\rho})$. A similar idea was suggested in [4, 6–8]. The factor $(1/2)$ is not strictly necessary but it allows for a fairer comparison between the original benchmarking protocol and our proposal *.

*In particular this factor of two ensures that "signal ranges" of the two protocols are equal, that is, standard RB starts at 1 and decays to $1/2$ for large m and the new protocol starts at $1/2$ and decays to 0.

- Secondly, we do not assume the ideal measurement operator to be the projector on the $|0\cdots 0\rangle$ state. Instead we perform some stabilizer measurement related to a pre-chosen Pauli matrix $\mathbf{P}$. An experimentally good choice would be for instance $\mathbf{P} = Z^{\otimes q}$ but our results hold for any choice of Pauli operator. Correspondingly we pick the input states to be some (impure) states $\rho, \hat{\rho}$ with support on the positive, resp. negative, eigenspaces of the Pauli operator $\mathbf{P}$. Ideally we would like to to prepare the impure states $\rho = \frac{I+\mathbf{P}}{2d}, \hat{\rho} = \frac{I-\mathbf{P}}{2d}$, but as we explain in section 6.5 this is not a necessity for rigorous randomized benchmarking.

Both of these adjustments are done with the purpose of lowering the experimental requirements for rigorous randomized benchmarking. Our first change to the RB protocol; performing randomized benchmarking with a state difference has two beneficial effects. (1) It changes the regression problem inherent to randomized benchmarking from an exponential fit with a non-zero off-set to an exponential fit (see eq. (6.2)). This eliminates a fitting parameter, lowering experimental requirements. (2) It lowers the statistical fluctuations of randomized benchmarking regardless of what input states are actually used. This improvement is mostly noticeable in the limit of long sequence lengths. We discuss this in more detail in section 6.5.9.
A much stronger improvement to the statistical fluctuations inherent to randomized benchmarking stems from our second change to the RB protocol; preparing states and performing measurements proportional to a Pauli operator $\mathbf{P}$. This change allows us to prove a radically sharper bound on the statistical fluctuations induced by finite sampling relative to preparing other input states. In section 6.5.9 we argue that this behavior is not an artifact of our proof techniques but rather inherent to the statistical behavior of randomized benchmarking. Note that for a single qubit the state $(I \pm \mathbf{P})/2$ is in fact a pure state for any choice of $\mathbf{P}$ (in particular $(I + Z)/2 = |0\rangle\langle 0|$).

As seen in fig. 6.1 the RB protocol starts by, for a given sequence of Clifford operations $\vec{G}$ of length m, computing the expectation value $p_m(\vec{G})(\rho)$ of an observable Q for two different input states ρ and $\hat{\rho}$. We subtract these two numbers to obtain a number $k_m(\vec{G}) := \frac{1}{2}(p_m(\vec{G})(\rho) - p_m(\vec{G})(\hat{\rho}))$. Next we obtain an average of this quantity over all possible sequences $\vec{G}$.

$$\mathbb{E}_{\vec{G}}(K_m) = |\mathsf{C}_q|^{-m} \sum_{\vec{G}\in\mathsf{C}_q^m} k_m(\vec{G}) \tag{6.1}$$

This average over all possible Clifford strings of length m can be fitted for various values m to the exponential decay curve

$$\mathbb{E}_{\vec{G}}(K_m) =_{\text{fit}} Af^m, \tag{6.2}$$

with two fitting parameter A and f. In the case where all gates performed in the experiment suffer from the same noise, that is $\hat{\mathcal{G}} = \mathcal{E}\circ\mathcal{G}$ for all Clifford operations $\mathcal{G}$ the number f can be interpreted as the depolarizing parameter of the channel $\mathcal{E}$ giving an estimate of the average fidelity of the noisy operation $\hat{\mathcal{G}}$ w.r.t. its ideal version $\mathcal{G}$.

As already indicated in chapter 4, the number of possible sequences for a given m is too large to average over completely. Instead one averages over a randomly sampled subset of sequences, which generates an empirical estimate $k_{m,N}$ the validity of which we can interpret using *confidence regions*. A confidence region, for some set confidence level $1-\delta$ and size ϵ, is an interval $[k_{m,N} - \epsilon, k_{m,N} + \epsilon]$ around the estimate $k_{m,N}$ such that the probability that the (unknown) parameter $\mathbb{E}_{\vec{G}}(k_m)$ lies in this interval with probability greater than $1 - \delta$, i.e.,

$$\text{Prob}\big[\mathbb{E}_{\vec{G}}(K_m) \in [k_{m,N} - \epsilon, k_{m,N} + \epsilon]\big] \geq 1 - \delta.$$

These confidence intervals, obtained for various values of sequence length during the experiment can then be used in the fitting procedure eq. (6.2) to generate a confidence interval around the empirical estimate $\hat{F}$ for the true channel average fidelity $F_{avg}(\mathcal{E}, \mathcal{I})$. This can be done using standard statistical procedures (see e.g. [9]). The number of random sequences N used to obtain $k_{m,N}$ will depend on ϵ and δ which are set before the beginning of the experiment, and in general also on some prior estimate of the infidelity r and unitarity u. The rest of the chapter will be mostly concerned with making this N as small as possible given δ and ϵ and (if possible) an a priori bound on the average infidelity r.

6.2. RESULTS

In this section we state the main contributions of this chapter. We present practical bounds on the number of sequences required to obtain rigorous confidence intervals for randomized benchmarking using the Clifford group under the assumption that the expectation value difference $k_m(\vec{G})$ for a given Clifford sequence $\vec{G}$ is estimated easily to a very high precision. This means we assume that the contribution of uncertainty on the numbers $k_m(\vec{G})$ to the uncertainty on k_m is negligible compared to the uncertainty due the random sampling of sequences $\vec{G}$. [2, 3]. In order to construct a $1 - \delta$ confidence interval of size ϵ around a randomized benchmarking sequence average $k_{m,N}$ with sequence length m, system dimension d and a prior estimate of the channel infidelity r and unitarity u one needs to average over N random sequences where N is given by [10]:

$$N(\delta, \epsilon, m, r, \chi, d) = -\log(2/\delta) \left[\log\left(\frac{1}{1-\epsilon}\right) \frac{1-\epsilon}{\mathbb{V}^2 + 1} + \log\left(\frac{\mathbb{V}^2}{\mathbb{V}^2 + \epsilon}\right) \frac{\mathbb{V}^2 + \epsilon}{\mathbb{V}^2 + 1} \right]^{-1}, \tag{6.3}$$

where $\mathbb{V}^2$ is the variance of the distribution of the samples $k_m(\vec{G})$ from a uniform distribution over the Clifford sequences $\vec{G}$. This variance is given below.

THE VARIANCE OF RANDOMIZED BENCHMARKING

The most important contribution of this chapter is a bound on the number of sequences N needed for multi-qubit randomized benchmarking. Previous bounds for multi-qubit RB [1, 2] are either prohibitively loose or scale exponentially with the number of qubits. Our new bounds, which are derived in detail in theorem 6.1, resolve both these issues using techniques from representation theory, enabling multi-qubit RB with practical numbers of random sequences.

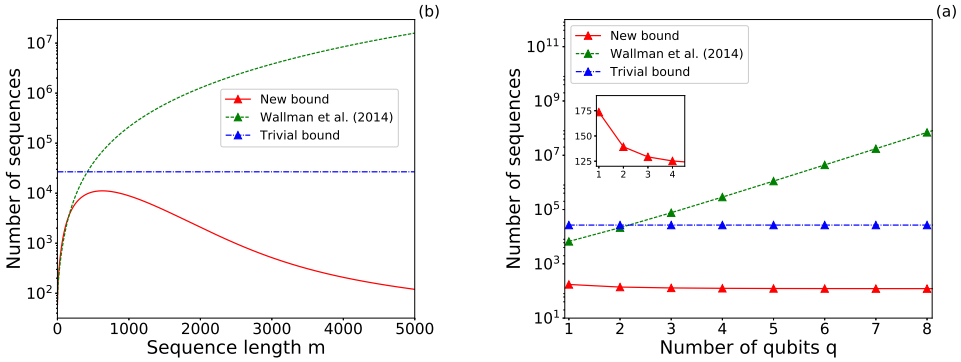


Figure 6.2: **Improvements in dimensional and sequence length scaling** The number of sequences needed (on a log scale) to obtain a 99% confidence interval around $k_{m,N}$ with $\epsilon = 10^{-2}$ for a prior infidelity $r = 10^{-3}$ as a function of **(a)** the sequence length m for a single qubit ($q = 1$) from eq. (6.4) (full line red) compared to the bounds from [2] (dashed green) and a trivial bounds that arises from noting that the distribution sampled from is bounded on the interval $[0, 1]$ and hence has a variance at most $1/4$ (dot-dashed blue) and **(b)** the number of qubits from eq. (6.6) (full line) for sequence length $m = 100$. In both cases, our bounds are asymptotically constant while the bounds from [2] diverge. Our bounds are also substantially smaller than the trivial bound. For multiple qubits, we set the SPAM contribution to $\eta = 0.05$ while for a single qubit we set the SPAM contribution to $\eta = 0$ in both bounds. We also assumed the unitarity to be $u = (1 + f^2)/2$ where f is the depolarizing parameter, corresponding to somewhat, but not fully coherent noise.

6

VARIANCE BOUND FOR SPAM-FREE MULTI-QUBIT RB

For states and measurements that are (very close to) ideal, section 6.5.5 yields a bound on the variance in terms of the sequence length m, the infidelity r, the unitarity u and the system size d. It is given by

$$\mathbb{V}_0^2 \leq \frac{d^2-2}{4(d-1)^2} r^2 m f^{m-1} + \frac{d^2}{(d-1)^2} r^2 u^{m-2} \frac{(m-1)\left(\frac{f^2}{u}\right)^m - m\left(\frac{f^2}{u}\right)^{m-1} + 1}{(1-f^2/u)^2}. \quad (6.4)$$

This bound is asymptotically independent of system size d. A notable upper bound on eq. (6.4) is

$$\mathbb{V}_0^2 \leq f^{m-1} \frac{(d^2-2)m}{4(d-1)^2} r^2 + u^{m-2} \frac{d^2 m(m-1)}{2(d-1)^2} r^2, \quad (6.5)$$

which can be made independent of the unitarity by setting $u = 1$. The bounds will however then only by useful in the regime of small sequence lengths.

To illustrate the improvements due to our bound, consider a single qubit ($d = 2$) RB experiment with sequences of length $m = 100$ and average infidelity $r \leq 10^{-4}$. To obtain a rigorous 99% confidence interval of size $\epsilon = 10^{-2}$ around $k_{m,N}$, Ref. [2] reported that $N = 145$ random sequences were needed (In the case of perfect state preparation and measurement). While our bounds imply that $N = 173$ random sequences are sufficient. However, the new bound has substantially better scaling with m. For instance, with $m = 5000$, $\epsilon = 0.05$ and other parameters as above, our bound only requires $N = 470$ compared to the $N = 1631$ required by the single qubit bound of Ref. [2]. We illustrate the

difference in scaling of the number of sequences needed for a given confidence interval with respect to sequence length m in fig. 6.2.

Variance bound including SPAM

The above variance bound is sensitive to SPAM errors, which introduce terms into the variance which scale linearly in the infidelity r. In theorem 6.1, we prove that in the presence of SPAM errors the variance is bounded by

$$\mathbb{V}^2_{\mathrm{SPAM}} \leq \frac{d^2-2}{4(d-1)^2} r^2 m f^{m-1} + \frac{d^2(1+4\eta)r^2}{(d-1)^2} \frac{(m-1)\left(\frac{f^2}{u}\right)^m - m\left(\frac{f^2}{u}\right)^{m-1} + 1}{(1-f^2/u)^2} u^{m-2} + \frac{2\eta d m r}{d-1} f^{m-1}. \quad (6.6)$$

The correction factor η only depends on SPAM. As we show in section 6.5 this SPAM dependence is impossible to avoid if one wants to retain the preferred quadratic scaling in infidelity r. This bound is also asymptotically independent of the number of qubits. This means we can perform rigorous randomized benchmarking even in the limit of very many qubits. We illustrate the difference in scaling with respect to system size in fig. 6.2.
To illustrate the improvements our methods yield we can again compare to [2]. Consider a system with 4 qubits, that is, $d = 16$, with sequence length $m = 100$, an a priori estimate of $r \leq 10^{-4}$, and $\eta = 0.05$. For a 99% confidence region of size $\epsilon = 10^{-2}$ the previous best known bound for multiple qubits [2] would require $N = 3 \times 10^5$ random sequences, while our dimension independent bound from eq. (6.6) only requires $N = 249$.

Optimality of results

We also prove (see section 6.5) that for *arbitrary* SPAM a bound on the variance which is linear in the infidelity r is in fact optimal. This means the result stated above is in some sense the best possible bound on the variance of a randomized benchmarking sequence. It is important to note that this optimality result also holds when RB is performed using a different set of gates than the Clifford group and also when one considers the standard protocol [6, 11] as opposed to the protocol involving differences of quantum states which we presented in this chapter.

Both the SPAM and SPAM-free variance bound also approach a constant independent of the infidelity r in the limit of large sequence length m when the unitarity is one, that is when the noise in the system is purely coherent. In section 6.5.8 we argue that this behavior is not an artifact of the proof techniques used but is in fact a generic feature of a randomized benchmarking procedure with a unitary noise process. More generally the rate of decay of the variance of randomized benchmarking in the limit of long sequences could be used to yield an estimate of the unitarity of the noise process under study.

Fitting procedure

In section 6.5.3 we discuss the consequences of eqs. (6.4) and (6.6) on the fitting procedure used to fit the data $\{k_{m,N}\}$ generated by fig. 6.1 to the RB fitting relation eq. (6.2).

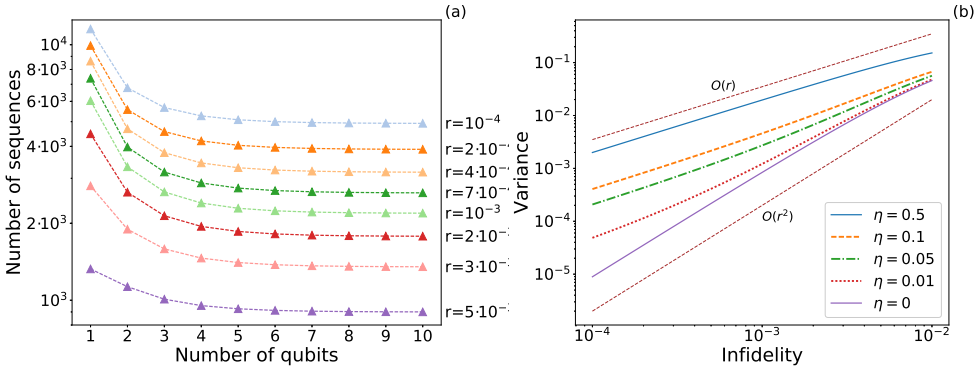


Figure 6.3: **(a)** Number of sequences needed for a 99% confidence interval of size $\epsilon = 5r$ for various infidelities r, number of qubits $q \in [1, 10]$ and sequence length $m = 100$ using eq. (6.5) under the assumption of negligible SPAM. (similar plots can be made without this assumption). The number of sequences needed increases with decreasing infidelity, reflecting the generic statistical rule that higher precision requires more samples. Note that even in the case of infidelity $r = 2 \times 10^{-4}$ the number of sequences required is within experimental limits. **(b)** Variance, as given by eq. (6.6) versus infidelity r (taking $d = 16$ and $m = 100$ for illustration) for various levels of SPAM $\eta \in \{0, 0.01, 0.05, 0.1, 0.5\}$. Note that the size of the SPAM term has a strong influence on the variance and hence the number of sequences required, especially in the small r limit. As indicated by the visual aids this is due to the transition from a variance scaling quadratically in infidelity r (small η) to a variance scaling linearly in the infidelity r (large η).

6

Our results show that the variance of randomized benchmarking data is strongly heterogeneous with respect to the sequence length m. This invalidates the key assumption of homogeneity of variance (homoskedasticity) [12] that is necessary for the correct functioning of Ordinary Least Squares (OLS), the standard method used for fitting RB data [3]. Because of this inferences drawn from can give misleading results when applied to RB data. We recommend switching from OLS to the more sophisticated method of Iteratively Reweighted Least Squares, which can deal with non-homoskedastic data.

6.3. Discussion

In this section we will discuss the behavior of the variance bound eqs. (6.4) and (6.6) in various regimes. Of interest are its scaling with respect to the number of qubits in the system, the presence of state preparation and measurement noise and varying amounts of coherence in the noise process.

6.3.1. Scaling with number of qubits.

We begin by discussing the effect of the number of qubits in the system on the variance and the number of necessary sequences.

As illustrated in fig. 6.2 (red full) and can be seen from eq. (6.4), the derived bound is almost independent of the number of qubits q (where $d = 2^q$). In fact, the bound on the variance decreases asymptotically to a constant in the limit of many qubits despite the number of possible sequences (that is, $|\mathsf{C}_q|^m$) increasing exponentially with the number of qubits. This constitutes a notable improvement over previous multi-qubit variance bounds

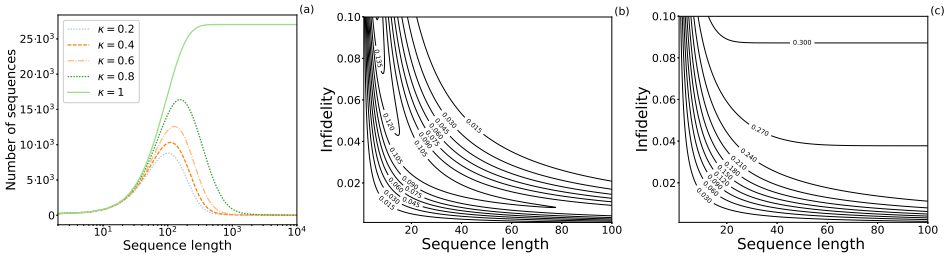


Figure 6.4: **(a)** Number of sequences needed for a 99% confidence interval of size $\epsilon = 0.01$ around $k_{m,N}$ for various values of the unitarity (given by a linear interpolation between f^2 and 1 where $\kappa = 1$ corresponds to $u = 1$ (unitary noise) and $\kappa = 0$ corresponds to $u = f^2$ (depolarizing noise)) for fixed infidelity $r = 0.01$ and sequence length in the interval $m \in [1, 10000]$ (log scale) using the variance eq. (6.4). We also assume $d = 16$ (four qubits) and ideal SPAM (η =0). Note that the number of sequences differs radically for $u = 1$ (unitary noise). In the case of $u < 1$ the number of sequences needed rises with increasing sequence length m, peaks and then decays to zero but for $u = 1$ the number of sequences keeps rising with increasing sequence length m until it converges to a non-zero constant (which will be independent of r). In section 6.5.8 we argue that this is expected behavior for randomized benchmarking with unitary noise. **(b),(c)** Contour plot of the variance bound with infidelity on the y-axis ($r \in [0.01, 0.1]$) and sequence length m on the y-axis ($m \in [1, 100]$). For **(b)** we have set the unitarity to $u = (1+f^2)/2$ corresponding to relatively incoherent noise and for **(c)** we have set the unitarity $u = 1$ corresponding to coherent noise. Note again the radical difference in behavior. For $u = 1$ the variance rises monotonically in the sequence length m to a constant independent of the infidelity r . Moreover the variance is monotonically increasing in infidelity r. However for incoherent noise the variance will peak strongly around $mr \approx 1$ and then decay to zero with increasing sequence length m. This means that both an upper and lower bound on the infidelity is required to make full use of the bound in eq. (6.4). The looser bound of eq. (6.5) does not share this property and can be used with only an upper bound on the infidelity r.

with an explicit dependence on the infidelity (dashed green in fig. 6.2), given in [2] which had a linear scaling with infidelity but scaled exponentially with the number of qubits. The qualitative behavior of the variance bound in terms of dimension matches a trivial bound on the number of sequences, which can be made by noting that the numbers k_m are sampled from a distribution bounded on an interval of unit size (and hence has variance at most $1/4$ (dashed blue in fig. 6.2)) but is much sharper in absolute terms due to it quadratic dependence on the infidelity r.

To further illustrate the behavior of the bound, fig. 6.3(a) shows the number of sequences needed for a 99% confidence interval around $k_{m,N}$ of size $5r$ versus the number of qubits in the system for various values of r ranging from 10^{-2} to 10^{-4} and sequence length $m = 100$. The size of ϵ was chosen to reflect that for fixed sequence length a smaller infidelity will lead to the need for greater precision around $k_{m.N}$ for a successful fit to the exponential eq. (6.2) [3]. This plot was made using the unitarity independent bound in eq. (6.5) for ideal SPAM, but similar plots can be made for non-negligible SPAM errors using eq. (6.6). Note also that greater numbers of sequences are needed when the infidelity is small even though the variance in eq. (6.4) decreases with infidelity. This is due to our setting of the size of the confidence interval and reflects the statistical truism that more samples are in general needed to detect small differences.

6.3.2. Effects of SPAM terms

In practice it will always be the case that the input state difference ν and the output measurement POVM element Q are not ideal. This means that in general we must take into account the contributions from non-ideal SPAM when calculating the number of required sequences. These contributions scale linearly in the infidelity r (see eq. (6.6)) rather than quadratically and so will increase the amount of required sequences. The degree to which ν and Q deviate from the ideal situation is captured by the prefactor η (see section 6.5 for more on this factor). To illustrate the effect of the SPAM terms on the variance we plot in fig. 6.3(b) the variance versus the infidelity r using eq. (6.6) taking the sequence length $m = 100$ and the dimension of the system $d = 16$ (four qubits) for SPAM of size $\eta \in \{0, 0.01, 0.05, 0.1, 0.5\}$. From this plot we note that for non-zero η the variance, and hence the amount of sequences needed increases rapidly, especially in the regime of small r. This is due to the fact that increasing the SPAM contribution interpolates the variance between a regime where the terms quadratic in infidelity r are dominant and a regime where the terms linear in infidelity r are dominant. This means that, especially when dealing with systems with very small r it is advantageous to try to suppress SPAM errors. In section 6.5.7 we show that this type of quadratic-to-linear interpolation behavior is in fact optimal for the variance of randomized benchmarking.

6.3.3. Scaling with sequence length

Of more immediate relevance is the scaling of the bound with the sequence length. It is easy to see that the variance bound eq. (6.4) scales quadratically in the sequence length m for any noise process when the sequence length is small (see also eq. (6.5)) but when the sequence length is very long the precise nature of the noise under consideration heavily impacts the variance. If the noise is purely coherent, i.e. the unitarity $u = 1$, we see that the scaling of the second term in eq. (6.4) is set by the factor

$$\frac{(m-1)f^{2m} - mf^{2(m-1)} + 1}{(1-f^2)^2}. \tag{6.7}$$

In the limit of m going to infinity this factor goes to

$$\frac{1}{(1-f^2)^2} \approx O(1/r^2) \tag{6.8}$$

which means the variance eq. (6.4) converges to a constant independent of the infidelity r. This behavior for unitary noise is strikingly different from the behavior for incoherent noise, that is $u < 1$. Here we see that the variance in the limit of long sequences is dominated by the exponential terms u^{m-2} and $f^{2(m-1)}$. Since f and u are strictly less than one by the assumption of incoherence, the variance will decay to zero in the limit of long sequences. As $u \geq f^2$ for all possible noise processes [13] the decay rate will be dominated by the size of the unitarity. This is also evident in fig. 6.4(a). In this figure we see the number of sequences needed (as given by eq. (6.4)) versus sequence length m for fixed infidelity $r = 0.1$ and dimension $d = 16$, and a fixed confidence interval $\delta = 0.99, \epsilon = 0.01$ but for different values of the unitarity u. Here we have chosen $u = (\kappa + (1-\kappa)f^2)$ for $\kappa \in \{0.2, 0.4, 0.6, 0.8, 1\}$ corresponding to the situations where

the noise is relatively incoherent going all the way up to a situation where the unitarity is one. We see that for $u < 1$ the number of sequences needed first rises quadratically, tops out and subsequently decays to zero whereas in the case of $u = 1$ the number of sequences needed keeps rising with sequence length m until it tops out at some asymptotic value. In section 6.5.8 we argue that this behavior is not a feature of the variance bound but rather a feature of the variance of randomized benchmarking itself. Therefore, in the case of highly unitary noise, we recommend performing more experiments at shorter sequence lengths rather than trying to map out the entire decay curve.

Another noteworthy feature of the variance bound eq. (6.4) is the fact that, for non-unitary noise (that is $u < 1$) it is in general not monotonically increasing in infidelity r. Rather, for a fixed sequence length, the variance increases at first with increasing infidelity but then peaks and decays towards zero. This behavior is illustrated in fig. 6.4(c). Here we plot a contour plot of the variance with infidelity on the y-axis ($r \in [0.01, 0.1]$) and sequence length m on the y-axis ($m \in [1, 100]$) and have set the unitarity to $u = (f^2 + 1)/2$ corresponding to relatively incoherent noise. The take-away from this plot is that it is not enough to have an upper bound on the infidelity to get an upper bound on the variance, rather one must have both and upper and a lower bound on the variance to make full use of the bound eq. (6.4). Note that the looser upper bound eq. (6.5) does not share this behavior and always yields an upper bound on the variance given an upper bound on the infidelity r.

On the other hand, when the underlying noise process is unitary, that is $u = 1$ the variance does increase monotonically with increasing r. This strikingly different behavior is illustrated in fig. 6.4 (b). Here we plot a contour plot of the variance with infidelity on the y-axis ($r \in [0.01, 0.1]$) and sequence length m on the y-axis ($m \in [1, 100]$) and have set the unitarity to $u = 1$ corresponding to fully coherent noise.

6.4. FUTURE WORK

An important caveat when applying the confidence bounds is the assumption of gate and time independent noise (this can be relaxed to Markovian, gate independent noise [2]). This is an assumption that many analyses of RB suffer from to various degrees, hence a major open problem would be to generalize the current bounds to encompass more general noise models.

Our work can be straightforwardly extended to interleaved RB [14]. However the dominant source of error in the interleaved RB protocol is usually systematic rather than stochastic (due to the fact that the protocol does not yield an estimate of the interleaved gate fidelity but rather provides upper and lower bounds).

Also, successful and rigorous randomized benchmarking not only depends on the number of random sequences needed per sequence length but also on the fitting procedure used to fit the points generated by randomized benchmarking of various lengths to a decay curve in order to extract an estimate of the average gate fidelity. Finding the optimal way to perform this fitting procedure is still an open problem [3]. Finally, a major theoretical open problem is the extension of the present bounds to non-qubit systems, different varieties of randomized benchmarking [15, 16], and to different 2-designs [15, 17, 18] or even orthogonal 2-designs [19, 20]. If these 2-designs are assumed to be groups, similar techniques from representation theory might be used [21] but how this would be done is

currently unknown.

6.5. METHODS

In this section, we will discuss the new contributions in detail, and explain how to apply them in an experimental setting. We will give a high level overview of the proof of the bound on the variance of a randomized benchmarking sequence; full details can be found in theorem 6.1 in section 6.6. We will also discuss the behavior of noise terms in the case of non-ideal SPAM and prove that the bounds we obtain are in some sense optimal. Finally we briefly comment on how the variance changes when performing regular randomized benchmarking (using an input state ρ rather than an input state difference $\nu = \frac{1}{2}(\rho - \hat{\rho})$).

6.5.1. ESTIMATION THEORY

In this section, we review confidence intervals and relate the bounding of confidence intervals to the bounding of the variance of a distribution. A first thing of note is that all the variance bounds stated in section 6.2 are dependent on the infidelity r. The appearance of r in the bound might strike one as odd since this is precisely the quantity one tries to estimate through RB. It is however a general feature of estimation theory that one needs some knowledge of the quantity one tries to estimate in order to use nontrivial estimation methods [9]. Note also that while our results are stated in frequentist language, they should also be translatable to Bayesian language, that is, as credible regions on the infidelity given prior beliefs as in Ref. [4] for example. Bayesian methods are more natural because our bounds depend on prior information about the infidelity, however, a full Bayesian treatment would involve the fitting process, obscuring our primary technical result, i.e. the variance bounds.

Let us now discuss how to use the variance bounds to construct confidence intervals around numbers $k_{m,N}$. We can in general define a $1-\delta$ confidence interval of size ϵ to be

$$\text{Prob}\left[|k_{m,N} - \mathbb{E}_{\vec{G}}(K_m)| \leq \epsilon\right] \geq 1 - \delta. \tag{6.9}$$

Once we have an upper bound on the variance $\mathbb{V}_m^2$ of an RB distribution we can relate this to an upper bound on number of required sequences through the use of concentration inequalities.

Note that for the case of randomized benchmarking there are two sets of confidence parameters. (δ_N, ϵ_N) is associated with estimating the average over all possible Clifford sequences, where the relevant parameter is the number of performed sequences N and (δ_L, ϵ_L) is associated with getting an estimate for the survival probability $p_m(\vec{G})$ for a given fixed sequence. Here the relevant parameter is L, the number of times a single sequence is performed. Since in practice $L < \infty$ there will be some finite (δ_L, ϵ_L) confidence region around the survival probability $p_m(\vec{G})$ for a given sequence $\vec{G}$. So in general, when looking at a ϵ, δ confidence region for an RB procedure of a given length one should look at $(\epsilon_N + \epsilon_L, \delta_N + \delta_L)$ confidence regions. In what follows we will assume that L is high enough such that ϵ_L, δ_L are negligible relative to (δ_N, ϵ_N). This approach is motivated by experimental realities where it is usually much easier to perform a single string of Clif-

fords many times quickly than it is to generate, store and implement a large number of random sequences.

For a given variance $\mathbb{V}^2$ we can relate the number of sequences N needed to obtain $1-\delta$ confidence intervals of size ϵ using the following concentration inequality due to Hoeffding [10]:

$$\begin{aligned}\text{Prob}\left[|k_{m,N} - \mathbb{E}_{\vec{G}}(K_m)| \leq \epsilon\right] &\geq 1-\delta \\ &\geq 1 - 2H(\mathbb{V}^2,\epsilon)^N,\end{aligned}$$

with

$$H(\mathbb{V}^2,\epsilon) = \left(\frac{1}{1-\epsilon}\right)^{\frac{1-\epsilon}{\mathbb{V}^2+1}} \left(\frac{\mathbb{V}^2}{\mathbb{V}^2+\epsilon}\right)^{\frac{\mathbb{V}^2+\epsilon}{\mathbb{V}^2+1}}. \tag{6.10}$$

We can invert this statement to express the number of necessary sequences N as a function of δ, r, ϵ as

$$N = -\frac{\log(2/\delta)}{\log(H(\mathbb{V}^2,\epsilon))}. \tag{6.11}$$

Note that this expression can also be inverted to yield a bound on δ, ϵ in terms of a given number of samples N. This identity heavily depends on the size of the variance $\mathbb{V}_m^2$.

6.5.2. STATE PREPARATION AND MEASUREMENT COSTS

We have argued that our adapted RB protocol allows for a reduction in the number of needed sequences to make rigorous estimated. However implicit in this cost reduction argument is the assumption that estimating the number $k_m(\vec{G})$ for a fixed sequence $\vec{G}$ is not more costly than estimating the number $p_m(\vec{G})$. Here we justify this assumption for the two changes we made to the randomized benchmarking protocol; using a state difference as input and using an impure input state proportional to a single Pauli matrix. In the following we forgo rigor in favor of intuition. We are however only applying standard statistical techniques that can easily be made rigorous.

State difference

At first glance one might think that estimating the same sequence twice for difference input states as we propose yields a two-fold overhead in the number of samples per sequence. To see that this is not the case consider the variance $\mathbb{V}_\rho^2$ associated with estimating the expectation value for a single sequence for a single state ρ. From the standard rules of error addition we now have, for the state difference $\nu = (\rho - \hat{\rho})/2$ that

$$\mathbb{V}_\nu^2 = \mathbb{V}_{(\rho-\hat{\rho})/2}^2 \leq \frac{1}{2^2}(\mathbb{V}_\rho^2 + \mathbb{V}_{\hat{\rho}}^2) \tag{6.12}$$

since the random variables associated to ρ and $\hat{\rho}$ are independently distributed (making the covariance zero). Now assuming that ρ incurs the largest variance, we get

$$\mathbb{V}_\nu^2 \leq \frac{1}{2}\mathbb{V}_\rho^2 \tag{6.13}$$

which means that estimating the expectation value of a single sequence for a difference of states is statistically not harder than estimating it for a single state.

Preparing the optimal input state and measurement
In our adapted RB procedure we call for preparing the input states $\rho = \frac{\mathbb{1}+\mathbf{P}}{2}, \rho = \frac{\mathbb{1}-\mathbf{P}}{2}$ for some Pauli matrix $\mathbf{P}$ and measuring the output POVM $\{Q, \mathbb{1} - Q\}$ with $Q = \frac{1}{2}(1 + \mathbf{P})$. This is different from standard RB where one is asked to prepare and project onto the all zero state $|0 \dots 0\rangle\langle 0 \dots 0|$. We argue that performing RB this way is not more costly than using the standard approach. For concreteness we shall set $\mathbf{P} = Z^{\otimes q}$. Measuring the expectation value of the operator $Z^{\otimes q}$ is trivial; one simply measures all qubits in the standard basis (as one would do in standard RB) and then computes the parity of the outcome. Since standard basis states with even parity precisely span the positive eigenspace of $Z^{\otimes q}$ this amounts to measuring the expectation value of $Z^{\otimes q}$. Preparing the states $\rho = \frac{\mathbb{1}+Z^{\otimes q}}{2}, \rho = \frac{\mathbb{1}-Z^{\otimes q}}{2}$ is a little more involved. The state ρ is a probabilistic mixture of all computational basis states $|x\rangle$ of even parity. By the linearity of expectation one could compute (for a fixed Clifford sequence $\vec{G}$) the survival probability $p_m(\vec{G}, |x\rangle)$ and then compute $p_m(\vec{G}, \rho) = 2^{-q/2} \sum_x p_m(\vec{G}, |x\rangle)$. This requires measuring $2^{2/q}$ expectation values $p_m(\vec{G}, |x\rangle)$, making this approach not scalable. We can remedy this by realizing that we are only interested in a good estimate of the mean $p_m(\vec{G}, \rho)$. Considering $p_m(\vec{G}, |x\rangle)$ to be the mean of a Bernoulli random variable with outcomes 0 and 1, and thus $p_m(\vec{G}, \rho)$ to be the mean of a normalized binomial distribution we can estimate this mean efficiently by sampling $|x\rangle$ at random (with even parity), estimating $p_m(\vec{G}, |x\rangle)$ and then computing the empirical mean. Moreover, since we do not need to know the means $p_m(\vec{G}, |x\rangle)$ very well to get a good estimate of $p_m(\vec{G}, \rho)$ the about of single data points (clicks) gathered to estimate $p_m(\vec{G}, \rho)$ is not higher than it would be to accurately estimate $p_m(\vec{G}, |\psi\rangle)$ for $|\psi\rangle$ some pure state.

6.5.3. The fitting procedure

In the previous section we outlined how to use the bound eq. (6.4) to construct confidence intervals around $k_{m,N}$. However, we have not yet discussed how to integrate the variance bound eq. (6.4) into the fitting procedure required by eq. (6.2). A fitting procedure is any method that takes in the set of data points $\{k_{m,N}\}_m$ with $m \in \mathbb{M}$ where $\mathbb{M}$ is some set of integers and outputs a tuple (A^*, f^*) such that $A^* f^{*m}$ is a 'good' description of the data $\{k_{m,N}\}_m$. There are many ways to approach this problem, we refer to [12] for a good overview, and finding an optimal procedure is outside the scope of this thesis. However we would like to discuss the most commonly used fitting procedure: Ordinary Least Squares (OLS) in the light of the bounds eqs. (6.4) and (6.6).

Ordinary least squares
Given data $\{k_{m,N}\}_m$ and the function $F(A, f) = Af^m$ the OLS procedure returns estimates $(\hat{A}, \hat{f})$. Through a linearization procedure, as outlined for RB in [3], confidence intervals can then be constructed around these estimates. However, for this procedure to yield correct results each data point $k_{m,N}$ must be distributed around $\mathbb{E}_{\vec{G}}(K_m)$ *with the same variance* [12, Chapter 2.8]. This assumption, called homoskedasticity in the statis-

tics literature, is not universally valid for randomized benchmarking data $\{k_{m,N}\}_m$. This shows in the functional form of the upper bound eq. (6.4), which strongly depends on the sequence length and from eq. (6.30) one can see that this is not an artifact of bounding techniques but rather an innate feature of RB data. Moreover OLS assumes that the variance of $k_{m,N}$ is independent of the fitting parameters A, f, an assumption which is also explicitly violated in RB data. The violation of these two assumptions (homoskedasticity and independence of fitting parameters) creates problems when performing OLS on the RB data $\{k_{m,N}\}_m$. In particular OLS no longer provides an unbiased estimate of the standard error on the fitting parameters (f, A) [12, Chapter 3.3], which can lead to mis-estimation of confidence intervals around the fitting parameters. Therefore we recommend using a more sophisticated approach.

Iteratively re-weighted least squares

Heteroskedasticity (violation of homoskedasticity) and functional dependence of the data distribution on the fitting parameters are well studied problems, and many robust solutions are available. Here we will focus on one particular solution called Iteratively Reweighted Least Squares (IRLS). For the purposes of this construction we will assume that the data $\{k_{m,N}\}_m$ is drawn from a random variable with mean $\mathbb{E}_{\vec{G}}(K_m)$ and variance $\mathbb{V}_m^2(m,r)/N$. IRLS constructs estimates for the parameters (A, f) by minimizing the function

$$\min_{A,f} \sum_{m\in\mathbb{M}} w_m (k_{m,N} - A f^m)^2 \tag{6.14}$$

where the weights w_m can depend on the functional variables. Under the assumption that eq. (6.4) is the actual variance $\mathbb{V}_m^2$ up to a constant factor we can set the weights [12, Section 2.8.8] to be $w_m = w(f,u,m) = 1/\sigma(f,u,m)$ where σ is the RHS of eq. (6.4) (if one suspects that $\eta \neq 0$ the eq. (6.6) can be used instead). We note that this procedure is fairly robust against misspecification of the weights, and moreover that σ captures the behavior of $\mathbb{V}_m^2$ with respect to the sequence length very well (see section 6.5.5). IRLS now proceeds in the following manner:

Algorithm 1 Iteratively Reweighted Least Squares

Input: Initial estimates f_0, u_0, A_0 and a dataset $k_{m,N}$
Output: Final estimates $\hat{f}, \hat{A}$

Set $f_{-1} = 0$
Set $i = 0$
// Optimization loop (here ϵ is some preset sensitivity)

while $|f_{i-1} - f_i| \geq \epsilon$ **do**
    Set $w_m = w(f_i, u_i, m) = \sigma(f_i, u_i, m)^{-1}$
    Optimize eq. (6.14) with weights w_m to get A_{i+1}, f_{i+1}
    Estimate u_{i+1} by fitting $\sigma(f_{i+1}, u_i, m)/N$ to the empirical variance of $k_{m,N}$
    Set $i = i + 1$
end while
Set $\hat{A} = A_i, \hat{f} = f_i$
return $\hat{A}, \hat{f}$

It as been shown [22, Page 45] (under some mild regularity conditions) that this algorithm converges to estimates $\hat{A}, \hat{f}$. If the weights w_m are exactly proportional to the variance $\mathbb{V}_m^2$ then these estimates are asymptotically consistent. In section 6.6.4 we provide a detailed estimate of how close the estimate $\hat{f}$ is to the real depolarizing parameter f in terms of the number of data points in $\{k_{m,N}\}_m$ and the number of sequences N sampled per data point.

Finally we would like to note that we have in this procedure kept the number of sequences N constant for varying N. It is however possible to let N depend on the sequence length m. One choice would be to vary N proportionally to $\mathbb{V}_0^2$ (assuming a good estimate of f is available). In this scenario, since $k_{m,N}$ is drawn from a distribution with variance $\mathbb{V}_m^2/N$ this would remedy the issue with heteroskedasticity and OLS could be used to provide reliable fitting.

6.5.4. GATE DEPENDENT NOISE AND GAUGE INVARIANCE

In recent work [23, 24] it has been noted that the relation between the parameter estimated by randomized benchmarking and the average fidelity is less than straightforward when the noise channel is allowed to depend on the gate being implemented, that is $\tilde{\mathcal{G}} = \mathcal{E}_G\mathcal{G}$. At the heart of the issue is that the only quantities measurable in the lab, probabilities of the form $\mathrm{Tr}(Q\tilde{\mathcal{G}}(\rho))$ for a state ρ and an observable Q are *gauge invariant*. That is, for any invertible superoperator $\mathcal{S}$ we have that

$$\mathrm{Tr}(Q\tilde{\mathcal{G}}(\rho)) = \mathrm{Tr}(\mathcal{S}^{-1}(Q)\mathcal{S}\tilde{\mathcal{G}}\mathcal{S}^{-1}(\mathcal{S}(\rho))). \tag{6.15}$$

This difficulty can be remedied by considering a more general noise model. Instead of choosing $\tilde{\mathcal{G}} = \mathcal{E}\mathcal{G}$ one chooses $\tilde{\mathcal{G}} = \mathcal{L}_G\mathcal{G}\mathcal{R}_G$ for superoperators $\mathcal{R}_G, \mathcal{L}_G$ [24]. The individual operators $\mathcal{R}_G, \mathcal{L}_G$ are not gauge invariant but the combined operator $\mathcal{R}_G\mathcal{L}_G$ is. Since here we deal exclusively with gate-independent noise we can choose the gauge such that $\mathcal{L} = \mathcal{I}$ and $\mathcal{R} = \mathcal{E}$ but our results also hold for the more general choice of gauge with the express caveat that our bounds then work in terms of the infidelity r and unitarity u of

the noise in between gates $\mathcal{RL}$. That is we have $r = r(\mathcal{RL})$ and $u = u(\mathcal{RL})$. It is possible to see this explicitly by making the substitution $\mathcal{E} \to \mathcal{RL}$ in all steps of the derivation of the variance bound in section 6.5.5 (and theorem 6.1 in section 6.6).

6.5.5. VARIANCE BOUND

In this section we present a derivation of the multi-qubit variance bound in eq. (6.4) under the assumption of ideal input difference operator $\nu = \frac{1}{2}(\rho - \hat{\rho})$ and output POVM element Q, i.e.

$$\nu = \frac{\mathbf{P}}{2d} \tag{6.16}$$

$$Q = \frac{1}{2}(\mathbb{1} + \mathbf{P}) \tag{6.17}$$

where $\mathbf{P}$ is some pre-specified target Pauli matrix (fig. 6.1). Under these ideal conditions we can guarantee that the variance scales quadratically in the infidelity r. We will focus on intuition and relegate most technical work to section 6.6. For the remainder of the text we will choose a basis for the space of linear operators $\mathcal{M}_d$. This means we can think of density matrices and POVM elements as column and row vectors which we denote with a Dirac-like notation, i.e. $\nu \to |\nu\rangle\rangle$ and $Q \to \langle\langle Q|$. Quantum channels can then be though of as matrices acting on vectors (which represent density matrices). Moreover, in this picture, composition of channels corresponds to matrix multiplication. When measuring the state $\mathcal{E}(\rho)$ using a two component POVM $\{Q, \mathbb{1} - Q\}$ for some quantum channel $\mathcal{E}$ and state ρ and positive operator Q we can write the expectation value $\mathrm{Tr}(Q\mathcal{E}(\rho)$ as a vector inner product

$$\mathrm{Tr}(Q\mathcal{E}(\rho)) = \langle\langle Q|\mathcal{E}(\rho)\rangle\rangle = \langle\langle Q|\mathcal{E}|\rho\rangle\rangle \tag{6.18}$$

where we abuse notation by referring to the matrix representation of the quantum channel $\mathcal{E}$ as $\mathcal{E}$ as well. This is variously called the affine or Liouville representation [2, 25].

We assume that every experimental implementation of a Clifford gate $\tilde{\mathcal{G}}$ can be written as $\tilde{\mathcal{G}} = \mathcal{E}\mathcal{G}$ for some fixed CPTP map $\mathcal{E}$ where G is the ideal Clifford gate. That is, we assume the noise is Markovian, constant and independent of the target gate. These assumptions can be relaxed partially [2, 3, 24, 26].

The key to randomized benchmarking is that randomly applying elements of the Clifford group C_q and then inverting produces, on average, the depolarizing channel [27]

$$\mathcal{D}_f(\rho) = f\rho + \frac{1-f}{d}\mathbb{1}_d, \tag{6.19}$$

that is, we have

$$\sum_{G \in \mathsf{C}_q} \mathcal{G}^\dagger \mathcal{E} \mathcal{G} = \mathcal{D}_f \tag{6.20}$$

with the depolarizing parameter f related to the fidelity by [28]

$$F_{\mathrm{avg}}(\mathcal{E}, \mathcal{I}) = \frac{(d-1)f + 1}{d}\,. \tag{6.21}$$

Therefore applying a sequence of independently-random gates and then inverting produces $\mathcal{D}_{f^m}$ on average. Hence the expectation value of any operator decays as f^m on average.
The value of $k_m(\vec{G})$ for a fixed sequence of Clifford gates $\vec{G}$ (as defined in fig. 6.1), and the variance over $\vec{G} \in \mathsf{C}_q$ are

$$k_m(\vec{G}) = \langle\langle Q|\mathcal{G}_m^\dagger \mathcal{E}\mathcal{G}_m \cdots \mathcal{G}_1^\dagger \mathcal{E}\mathcal{G}_1|\nu\rangle\rangle \tag{6.22}$$

$$\mathbb{V}_m^2 = \mathbb{E}_{\vec{G}}[k_m(\vec{G})^2] - [\mathbb{E}_{\vec{G}}(k_m(\vec{G}))]^2 \tag{6.23}$$

respectively. We can use the identity $a^2 = a \otimes a$ for $a \in \mathbb{C}$, the distributivity and associativity of the tensor product, and the linearity of quantum channels to write this as [2, 29]

$$\mathbb{V}_m^2 = \langle\langle Q^{\otimes 2}\big|T_{\mathsf{C}_q}(\mathcal{E}^{\otimes 2})^m - \big[T_{\mathsf{C}_q}(\mathcal{E})^m\big]^{\otimes 2}\big|\nu^{\otimes 2}\rangle\rangle \tag{6.24}$$

where

$$T_{\mathsf{C}_q}(\mathcal{E}) = \frac{1}{|\mathsf{C}_q|}\sum_{G\in\mathsf{C}_q} \mathcal{G}^\dagger \mathcal{E}\mathcal{G} = \mathcal{D}_f, \tag{6.25}$$

$$T_{\mathsf{C}_q}(\mathcal{E}^{\otimes 2}) = \frac{1}{|\mathsf{C}_q|}\sum_{G\in\mathsf{C}_q} \mathcal{G}^{\dagger\otimes 2} \mathcal{E}^{\otimes 2} \mathcal{G}^{\otimes 2}. \tag{6.26}$$

The superoperator $T_{\mathsf{C}_q}(\mathcal{E})$ is often referred to as the *twirl* of the quantum channel $\mathcal{E}$.
At this point, our analysis diverges from that of Ref. [2]. First, note that for our modified scheme, $\nu^{\otimes 2}$ is traceless and symmetric under the interchange of the tensor factors (we will refer to such a matrix as a traceless symmetric matrix) so

$$\big[T_{\mathsf{C}_q}(\mathcal{E})^m\big]^{\otimes 2}\,\big|\nu^{\otimes 2}\rangle\rangle = f^{2m}\big|\nu^{\otimes 2}\rangle\rangle. \tag{6.27}$$

Furthermore, $T_{\mathsf{C}_q}(\mathcal{E}^{\otimes 2})$ preserves the trace and symmetry under interchange of tensor factors. Therefore we can define $T_{\mathrm{TS}}(\mathcal{E}^{\otimes 2})$ to be the restriction of $T_{\mathsf{C}_q}(\mathcal{E}^{\otimes 2})$ to the space of traceless symmetric matrices. As we prove in lemma 6.1 using the results from chapter 5, the representation $\mathcal{G}^{\otimes 2}$ of the Clifford group restricted to the traceless symmetric subspace decomposes into inequivalent irreducible representations. Therefore by Schur's lemma (see chapter 3 for an explanation of Schur's lemma),

$$T_{\mathrm{TS}}(\mathcal{E}^{\otimes 2}) = \sum_{i\in\mathcal{Z}} \chi_i \mathcal{P}_i \tag{6.28}$$

where $\mathcal{Z}$ indexes the irreducible subrepresentations of $\mathcal{G}^{\otimes 2}$ on the space of traceless symmetric matrices, $\mathcal{P}_i$ are projectors associated to each representation and $\chi_i = \chi_i(\mathcal{E}) \in \mathbb{R}$ are numbers that depend on the quantum channel $\mathcal{E}$. [†] As the $\mathcal{P}_i$ are orthogonal projectors that span the space of traceless symmetric matrices, we can write the variance as

$$\mathbb{V}_m^2 = \sum_{i\in\mathcal{Z}} \langle\langle Q^{\otimes 2}|\mathcal{P}_i|\nu^{\otimes 2}\rangle\rangle(\chi_i^m - f^{2m}). \tag{6.29}$$

[†] When ν is not traceless, as is the case in regular randomized benchmarking, we can not restrict the two-copy twirl to a twirl over the traceless-symmetric subspace. However the derivation below will still hold, up to the addition of extra terms stemming from equivalent irreducible subrepresentations present in eq. (6.28). This extra term is discussed in section 6.5.9 and also [2]

Now we use a telescoping series trick (lemma 6.6 and in particular corollary 6.1) on the last factor to write this as

$$\mathbb{V}_m^2 = \sum_{i\in\mathcal{Z}} \langle\!\langle Q^{\otimes 2}|\mathcal{P}_i|\nu^{\otimes 2}\rangle\!\rangle \Big[m f^{2(m-1)}(\chi_i - f^2) + (\chi_i - f^2)^2 \sum_{j=1}^{m}(j-1)\chi_i^{m-j} f^{2(j-2)}\Big]. \tag{6.30}$$

Here we see that getting a sharp bound on the variance will depend on getting sharp bounds on the difference between the χ_i prefactors and the square of the depolarizing parameter f^2. Up to this point the derivation has been valid for any input state difference ν and output positive operator Q. However now we will restrict to the case of ideal Q and ν. For the general case of non-ideal Q and ν see theorem 6.1 in section 6.6. In the case of ideal Q and ν we can use lemmas 6.2 and 6.3 to upper bound

$$\sum_{i\in\mathcal{Z}} \langle\!\langle Q^{\otimes 2}|\mathcal{P}_i|\nu^{\otimes 2}\rangle\!\rangle (\chi_i - f^2) \leq \frac{1}{4}\frac{d^2-1}{(d-1)^2} r^2 \tag{6.31}$$

where $r = 1 - F_{\text{avg}}(\mathcal{E}, I)$ is the infidelity of the quantum channel. Note that the assumption of ideal Q and ν is necessary in order to apply lemma 6.2 and obtain an upper bound that scales quadratically in the infidelity r. In general, for $r \leq \frac{1}{3}$, we can say that (lemma 6.5)

$$|\chi_i - f^2| \leq \frac{2dr}{d-1}. \tag{6.32}$$

Hence we can say

$$\mathbb{V}_m^2 \leq m f^{2(m-1)} \frac{d^2-2}{4(d-1)^2} r^2 + \sum_{i\in\mathcal{Z}} \frac{4d^2 r^2 \langle\!\langle Q^{\otimes 2}|\mathcal{P}_i|\nu^{\otimes 2}\rangle\!\rangle}{(d-1)^2} \sum_{j=1}^{m}(j-1)\chi_i^{m-j} f^{2(j-2)} \tag{6.33}$$

for ideal Q and ν. Now we only need to deal with the χ_i factors in the sum. To do this we will use the fact that every χ_i term is upper bounded by the unitarity u of the quantum channel $\mathcal{E}$. This is derived in lemma 6.4 in section 6.6. Inserting this we get

$$\mathbb{V}_m^2 \leq m f^{2(m-1)} \frac{d^2-2}{4(d-1)^2} r^2 + \sum_{i\in\mathcal{Z}} \frac{4d^2 r^2 \langle\!\langle Q^{\otimes 2}|\mathcal{P}_i|\nu^{\otimes 2}\rangle\!\rangle}{(d-1)^2} \sum_{j=1}^{m}(j-1)u^{m-j} f^{2(j-2)}. \tag{6.34}$$

Now we factor u^{m-2} out of the sum over j and use the fact that this sum has a closed form. Using this and lemma 6.2 to bound the projector inner products we obtain a final bound on the variance

$$\mathbb{V}_m^2 \leq m f^{2(m-1)} \frac{d^2-2}{4(d-1)^2} r^2 + \frac{d^2}{(d-1)^2} r^2 u^{m-2} \frac{(m-1)(\frac{f^2}{u})^m - m(\frac{f^2}{u})^{m-1} + 1}{(1-(\frac{f^2}{u}))^2}, \tag{6.35}$$

which is the bound we set out to find. To obtain from this the bound given in eq. (6.5) we note that $u \geq f^2$ and moreover that the fractional term in eq. (6.35) is monotonically decreasing in u (for fixed f^2) and reaches a limiting value of $m(m-1)/2$ in the limit of $u \to f^2$ (This can be seen by using l'Hôpital's rule).

6.5.6. State preparation and measurement

When Q, ν do not satisfy eq. (6.16), (which will always happen in practice) the above derivation will not hold exactly and the deviation of Q, ν from their ideal forms will introduce terms of order ηr i.e., terms which scale linearly and not quadratically in the infidelity r. Deriving an expression of the variance taking into account these these contributions is a little tedious so we will relegate the details to section 6.6 and instead discuss the form of the prefactor η. Let ν be some non-ideal input state difference and let Q be some non-ideal observable. Note from eq. (6.16) that the ideal input state difference ν and output POVM Q are related to a pre-chosen "target Pauli matrix" $\mathbf{P}$. We hence have

$$Q_{\mathrm{id}} = \frac{1}{2}(\mathbb{1} + \mathbf{P}) \tag{6.36}$$

$$\nu_{\mathrm{id}} = \frac{\mathbf{P}}{2d} \tag{6.37}$$

the ideal Q and ν. Suppressing some prefactors (the exact expression can be found in eq. (6.91) in section 6.6) we get the following approximate expression for the SPAM factor η:

$$\begin{aligned} \eta \approx & \left\| Q - Q_{\mathrm{id}} \right\|_2 \left\| \nu - \nu_{\mathrm{id}} \right\|_2 \\ & + \left\| Q - Q_{\mathrm{id}} \right\|_2^2 + \left\| \nu - \nu_{\mathrm{id}} \right\|_2^2 \end{aligned} \tag{6.38}$$

where $\|\cdot\|_2$ is the Schatten-2 norm [25] and Q, ν are the non-ideal operators that are actually implemented. There are several important things to notice here:

- η goes to zero in the limit of ideal Q, ν. This justifies our choice of the ideal Q and ν as being proportionate to a Pauli matrix rather than preparing and measuring in the $|0\rangle$ state as was the case in the original randomized benchmarking proposal [1]
- η scales quadratically in the deviation from the ideal of Q and ν. This means that for small deviations η is likely to be small.
- η is non-zero for non-ideal Q even when ν is ideal and vice versa. This is unfortunate as it means that both state preparation and measurement must be good to ensure small variance. However, as we argue in section 6.5.7, this is actually optimal.

To get a feel for how the parameter η behaves we discuss a particular error model for state preparation and measurement errors, inspired by recent research in superconducting qubits [30]. Here we see that the dominant error source when preparing states in the computational basis is given by decay to the ground state when in the excited ($|1\rangle$) state and residual excitations when preparing the ground ($|0\rangle$) state. The dominant contribution to measurement errors when measuring in the computational basis are here discrimination errors (mistaking 0 for 1 and vice versa) as well as errors due to finite sampling. When performing our version of RB, and choosing $\mathbf{P} = Z$, we see that $\nu_{\mathrm{id}} = (|0\rangle\langle 0| - |1\rangle\langle 1|)/2$ and hence we want to ideally prepare the states $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$. Following [30] we assume 0.5% residual excitations when preparing the $|0\rangle\langle 0|$ state, 0.8% decay to the ground when preparing $|1\rangle\langle 1|$ and a 1% discrimination error (modeled by a symmetric bit-flip channel) (Here we use the discrimination fidelity given in [31]). Plugging these numbers

into the assumed error models and calculating η using eq. (6.91) in the Supplementary Material we see that in this case $\eta = 0.001$. Hence we can say that under realistic scenarios η will be quite small.

6.5.7. OPTIMALITY OF MAXIMAL VARIANCE

In this section we will argue that the bounds on the variance in the case of non-ideal SPAM are optimal in the sense that it is impossible for the variance to scale better that linearly in the infidelity r for arbitrary noise maps when the input POVM element Q is non-ideal even when the input state difference ν is ideal. The same reasoning will also hold for non-ideal ν even when Q is ideal. (More generally the reasoning below will also work when randomized benchmarking is performed using a state rather than a state difference but we will not show this explicitly here).

Consider the variance as in eq. (6.24) for a randomized benchmarking experiment with a quantum channel $\mathcal{E}$ with infidelity r and for simplicity set the sequence length $m = 1$ (the argument will work for general m). Then we have an expression for the variance

$$\mathbb{V}^2 = \langle\langle Q^{\otimes 2} \big| T_{\mathsf{C}_q}(\mathcal{E}^{\otimes 2}) - T_{\mathsf{C}_q}(\mathcal{E})^{\otimes 2} \big| \nu^{\otimes 2} \rangle\rangle \tag{6.39}$$

with the $T_{\mathsf{C}_q}(\mathcal{E}^{\otimes 2}), T_{\mathsf{C}_q}(\mathcal{E})^{\otimes 2}$ defined in eq. (6.25). Now consider setting $\nu = \nu_{\mathrm{id}}$ and maximizing over the POVM element Q. That is consider

$$\mathbb{V}^2 = \max_{0 \leq Q \leq \mathbb{1}} \langle\langle Q^{\otimes 2} \big| T_{\mathsf{C}_q}(\mathcal{E}^{\otimes 2}) - T_{\mathsf{C}_q}(\mathcal{E})^{\otimes 2} \big| \nu_{\mathrm{id}}^{\otimes 2} \rangle\rangle.$$

Now note that for any unitary U the operator $\mathcal{U}(Q) = UQU^\dagger$ is also a POVM element. This means we can write

$$\begin{aligned}
\mathbb{V}^2 &= \max_{0 \leq Q \leq \mathbb{1}} \langle\langle Q^{\otimes 2} \big| T_{\mathsf{C}_q}(\mathcal{E}^{\otimes 2}) - T_{\mathsf{C}_q}(\mathcal{E})^{\otimes 2} \big| \nu_{\mathrm{id}}^{\otimes 2} \rangle\rangle \\
&= \max_{0 \leq Q \leq \mathbb{1}} \langle\langle (\mathcal{U}(Q))^{\otimes 2} \big| T_{\mathsf{C}_q}(\mathcal{E}^{\otimes 2}) - T_{\mathsf{C}_q}(\mathcal{E})^{\otimes 2} \big| \nu_{\mathrm{id}}^{\otimes 2} \rangle\rangle \\
&\geq \max_{0 \leq Q \leq \mathbb{1}} \langle\langle \int dU (\mathcal{U}(Q))^{\otimes 2} \big| T_{\mathsf{C}_q}(\mathcal{E}^{\otimes 2}) - T_{\mathsf{C}_q}(\mathcal{E})^{\otimes 2} \big| \nu_{\mathrm{id}}^{\otimes 2} \rangle\rangle,
\end{aligned}$$

where we used the linearity of the inner product and the definition of maximum and the integral is taken over the uniform or Haar measure of the unitary group. Now we use a well known fact from the representation theory of the unitary group which states that the integrated operator $\int dU(\mathcal{U}(Q))$ is precisely proportional to one of the projectors defined in eq. (6.28). [13]. In particular it is proportional to the rank one projector $\mathcal{P}_{\mathrm{tr}} = |\Delta\rangle\rangle\langle\langle\Delta|$ where $\Delta \in \mathcal{M}_d$ is some matrix operator (see lemma 6.1 in the appendix) and tr is an element of the set $\mathcal{Z}$ which indexed the irreducible representations of the Clifford group in eq. (6.28). This means we can we can write using eq. (6.29)

$$\begin{aligned}
\mathbb{V}^2 &\geq \max_{0 \leq Q \leq \mathbb{1}} \sum_{i \in \mathcal{Z}} \alpha(Q) \langle\langle \Delta | \mathcal{P}_i | \nu \rangle\rangle (\chi_i - f^2) \\
&= \max_{0 \leq Q \leq \mathbb{1}} \alpha(Q) \langle\langle \Delta | \mathcal{P}_{\mathrm{tr}} | \nu \rangle\rangle (\chi_{\mathrm{tr}} - f^2)
\end{aligned} \tag{6.40}$$

where $\alpha(Q)$ is some positive prefactor function of Q. From lemma 6.4 and [13] it can be seen that χ_{tr} is precisely the unitarity u of the quantum channel $\mathcal{E}$. If we now consider $\mathcal{E}$ to be a unitary channel (that is $u = 1$), we get (ignoring the prefactors, which can be proven to be strictly positive)

$$\mathbb{V}^2 \approx 1 - f^2 = \frac{dr}{d-1}\left(2 - \frac{dr}{d-1}\right) \tag{6.41}$$

which is linear in infidelity r. Hence when the POVM element Q is allowed to vary freely a linear scaling of the variance with the infidelity r can not be avoided even when the input state difference ν is ideal. One can perform a similar thought experiment maximizing over ν while setting $Q = Q_{\mathrm{id}}$ and get the same result. Hence the expression for η we discussed in the above section is essentially optimal.

6.5.8. Asymptotic behavior of the variance

When looking at the bound on the variance eq. (6.4) the difference between unitary and non-unitary noise is striking. When the noise is non-unitary, and thus $u < 1$ the upper bound on the variance (and hence the variance itself) decays exponentially to zero in the sequence length m but when the noise process is unitary the variance keeps increasing and eventually saturates on a constant that is independent of the infidelity of the noise process. Here we argue that this is not an artifact of the bounding techniques but rather a fundamental feature of performing randomized benchmarking over unitary noise. Moreover this effect is independent of whether RB is performed using a state difference input ν or a state input ρ (as in standard RB).

Consider a unitary noise process $\mathcal{U} = U \cdot U^\dagger$ with infidelity $r > 0$ (That is $\mathcal{U}$ is not the identity). Now consider a randomized benchmarking experiment of sequence length m. That is, for a random sequence of Clifford unitaries $G_1, \ldots G_m$ we perform the unitary

$$V_m = U(G_m \cdots G_1)^\dagger U G_m U \cdots U G_1 \tag{6.42}$$

Following the reasoning of [1] we can write V_m as

$$V_m = U {G'}_m^\dagger U G'_m \cdots {G'}_1^\dagger U G'_1 \tag{6.43}$$

where the unitaries $G'_m, \ldots G'_1$ are sampled uniformly at random from the Clifford group. We can equally well think of the unitary $U^\dagger V_m$ as being the product of m uniformly random samples from the set

$$\mathbb{G}_U = \{G^\dagger U G \;\|\; G \in \mathsf{C}_q\}. \tag{6.44}$$

Note that this set depends on the unitary U. In [32] it was shown that the distribution of the product of m unitaries sampled uniformly at random from a set of unitaries converges to the Haar measure (uniform measure) on the unitary group in the limit of large m as long as this set contains a *universal* set of gates. Note that this convergence phenomenon is independent of the initial set [‡].

[‡]This is similar to how the limiting distribution of a random walk is independent of the initial step-size

Note now that as long as the unitary U is not a Clifford gate the set $\mathbb{G}_U$ will contain a universal gateset [33]. This means that the distribution from which V_m is sampled will converge to the Haar measure in limit of long sequence length (the extra $U^\dagger$ factor gets absorbed into the Haar measure). This will happen independently of the unitary U (as long as U is not Clifford). From this we can conclude that the variance of randomized benchmarking with unitary noise must, in the limit of long sequences, converge to the variance of the randomized benchmarking expectation value over the Haar measure independently of what the original unitary noise process is. Note again that the above argument is independent of whether RB is performed using a state difference input or a state input.

6.5.9. RELATION TO REGULAR RANDOMIZED BENCHMARKING

When performing regular randomized benchmarking, that is using an input state $\rho = \frac{1}{2}(\mathbb{1} + \mathbf{P})$ rather than an input state difference $\nu = \frac{\mathbf{P}}{2}$ the upper bounds on the variance given in eqs. (6.4) and (6.6) still hold provided an extra additive term is added to them. This term will stem from the addition of an extra superoperator (that is not a projector) in the sum in eq. (6.28) which stem from the appearance of two equivalent trivial subrepresentations of the two-copy representation $\mathcal{G}^{\otimes 2}$ of the Clifford group. This term is of the form

$$\begin{aligned} T &= \frac{1}{4}\left\|\mathcal{E}(\mathbb{1}/d) - \mathbb{1}/d\right\|_2^2 \frac{1-u^m}{1-u} \\ &\leq \frac{(d+1)^2}{2d^2} r^2 \frac{1-u^m}{1-u} \end{aligned} \tag{6.45}$$

where $\mathcal{E}$ is the noise process under investigation, with infidelity r and unitarity u and system dimension d. Here $\left\|\mathcal{E}(\mathbb{1}/d) - \mathbb{1}/d\right\|_2^2$ is a measure of how 'non-unital' the quantum channel $\mathcal{E}$, that is how far its output deviates from the identity when the identity is the input. This measure can be upper bounded using [34, Theorem 3] and is already implicitly analyzed in [2]. We will not prove the above explicitly but it can be derived straightforwardly by following the derivation in theorem 6.1 using ρ as input state. Note however that the upper bound on T does not decay to zero exponentially but rather converges to a non-zero constant even for non-unitary channels. This is not a feature of the upper bound itself but rather of the long sequence behavior of standard randomized benchmarking. It was proven in [2, Theorem 17] that the upper bound T is actually saturated for almost al non-unitary channels. Moreover, for physically relevant noise models such as amplitude damping T can be quite substantial. This very different behavior in the limit of long sequence lengths further motivates the use the state difference ν for rigorous randomized benchmarking.

6.6. TECHNICAL STATEMENTS

In this section we will give more rigorous versions of the statements made in the previous sections. This section will not provide insights beyond those already provided in the previous sections and can be skipped by readers uninterested in the mathematical details of our results. We refer to chapter 2 and chapter 3 for introductions to basic notions of quantum mechanics and representation theory that will be needed to follow the proofs and which we will not repeat here.

6.6.1. Traceless-Symmetric representation

In the rest of the chapter we will often work with quantum channels which have a tensor product structure. That is we will often be dealing with channels of the form

$$\mathcal{W} := \sum_i \lambda_i \mathcal{E}_i^{\otimes 2} \tag{6.46}$$

where $\mathcal{E}_i$ is a CPTP map for all i and $\lambda_i \in \mathbb{C}$ is some abstract parameter. Note that $\mathcal{W}$ is now a linear map from $\mathcal{M}_d^{\otimes 2}$ to $\mathcal{M}_d^{\otimes 2}$. Maps of these form have a number of useful properties which we will now consider. We begin by defining the *traceless-symmetric* subspace V_{TS} which is a subspace of $\mathcal{M}_d^{\otimes 2}$ of the form

$$V_{\mathrm{TS}} := \mathrm{Span}\left\{ S_{\sigma,\tau} := \frac{1}{\sqrt{2}}(|\sigma\tau\rangle\rangle + |\tau\sigma\rangle\rangle) \;\; \| \;\; \sigma,\tau \in \boldsymbol{\sigma}_q \right\}. \tag{6.47}$$

where we have suppressed the tensor product (that is $\sigma\tau := \sigma \otimes \tau$). The traceless-symmetric subspace has several desirable properties which we note here. First let $\rho, \hat{\rho} \in \mathcal{M}_d$ be density matrices and call their difference $\nu := \rho - \hat{\rho}$, then we have that

$$|\nu^{\otimes 2}\rangle\rangle = |(\rho - \hat{\rho})^{\otimes 2}\rangle\rangle \in V_{\mathrm{TS}} \tag{6.48}$$

Moreover, for any quantum channel $\mathcal{W}$ of the form defined in eq. (6.46) we have that

$$\mathcal{W}|v\rangle\rangle \in V_{\mathrm{TS}}, \qquad \forall |v\rangle\rangle \in V_{\mathrm{TS}}, \tag{6.49}$$

or equivalently we have that

$$\mathcal{P}_{\mathrm{TS}}\mathcal{W} = \mathcal{W}\mathcal{P}_{\mathrm{TS}} \tag{6.50}$$

where $\mathcal{P}_{\mathrm{TS}}$ is the projector onto the space V_{TS} (note that $\mathcal{P}_{\mathrm{TS}}$ is a linear map from $\mathcal{M}_d^{\otimes 2}$ to $\mathcal{M}_d^{\otimes 2}$). This observation follows from the fact than $\mathcal{W}$ is a linear combination of two-fold tensor products of quantum channels (which preserve the trace and map operators that are symmetric under interchange of the two copies of $\mathcal{M}_d^{\otimes 2}$ to operators that are symmetric under interchange of the two copies of $\mathcal{M}_d^{\otimes 2}$).

We will in particular be interested in how a representation of of the Clifford group C_q behaves on the traceless symmetric subspace. Define the two-fold tensor product representation of the Clifford group on $\mathcal{M}_d^{\otimes 2}$ as

$$\phi_2 : G \longrightarrow \mathcal{G}^{\otimes 2} \tag{6.51}$$

for all where $\mathcal{G}$ is the Liouville representation of G for all $G \in \mathsf{C}_q$. This representation has a natural restriction to the subspace V_{TS} since $\mathcal{G}^{\otimes 2}$ is of the form described in eq. (6.46). We can define the subrepresentation ϕ_{TS} of ϕ_2 as

$$\phi_{\mathrm{TS}} : G \longrightarrow \mathcal{P}_{\mathrm{TS}}\mathcal{G}^{\otimes 2}\mathcal{P}_{\mathrm{TS}} \tag{6.52}$$

for all $G \in \mathsf{C}_q$. This representation is in general not irreducible but decomposes further into a collection of irreducible subrepresentations. In chapter 5 we derived these irreducible subrepresentations of ϕ_{TS} and studied their properties. In the following lemma we will quote several results from chapter 5 which will be useful for our purposes here.

Lemma 6.1. Let C_q be the Clifford group and let ϕ_{TS} be the traceless symmetric representation. This representation is a direct sum of three subrepresentations ϕ_{d} (diagonal), $\phi_{[\mathrm{S}]}$ (symmetric commuting) and $\phi_{\{\mathrm{S}\}}$ (symmetric anti-commuting) acting on the spaces

$$\begin{aligned} V_{\mathrm{d}} &:= \mathrm{Span}\left\{|\sigma\sigma\rangle\rangle \;\;\|\;\; \sigma \in \sigma_{\mathbf{q}}\right\} && \text{(diagonal)} \\ V_{[\mathrm{S}]} &:= \mathrm{Span}\{S_{\nu,\nu\cdot\tau} \;\;\|\;\; \tau \in \boldsymbol{\sigma}_q,\;\; \nu \in \mathbf{C}_\tau\} && \text{(symmetric commuting)} \\ V_{\{\mathrm{S}\}} &:= \mathrm{Span}\{S_{\nu,i\nu\cdot\tau} \;\;\|\;\; \tau \in \boldsymbol{\sigma}_q,\;\; \nu \in \mathbf{N}_\tau\} && \text{(symmetric anti-commuting)} \end{aligned}$$

The diagonal subrepresentation ϕ_{d} decomposes into three subrepresentations denoted by $\phi_{\mathrm{tr}}, \phi_1, \phi_2$ with ϕ_{tr} the trivial representation spanned by

$$V_{\mathrm{tr}} = \left\{ \frac{1}{\sqrt{d^2-1}} \sum_{\tau\in\boldsymbol{\sigma}_q} |\tau\tau\rangle\rangle \right\}. \qquad \text{(trivial)}$$

We will index these representations by the set $\mathcal{Z}_{\mathrm{d}} := \{\mathrm{tr}, 1, 2\}$.
The symmetric commuting representation $\phi_{[S]}$ decomposes into 3 irreducible subrepresentations denoted as $\phi_{[\mathrm{adj}]}, \phi_{[1]}, \phi_{[2]}$. We will index these representations by the set $\mathcal{Z}_{[S]} := \{[\mathrm{adj}], [1], [2]\}$. The spaces carrying these representations can be written as a direct sum of subspaces in the following way

$$V_i = \bigoplus_{\tau\in\boldsymbol{\sigma}_q} V_i^\tau \tag{6.53}$$

where $V_i^\tau \subset V^{[\tau]}$ with

$$V^{[\tau]} := \mathrm{Span}\{S_{\nu,\nu\cdot\tau} \;\;\|\;\; \nu \in \mathbf{C}_\tau\}. \tag{6.54}$$

The symmetric anti-commuting representation $\phi_{\{S\}}$ decomposes into 2 irreducible subrepresentations denoted as $\phi_{\{1\}}, \phi_{\{2\}}$. We will index these representations by the set $\mathcal{Z}_{\{S\}} := \{\{1\}, \{2\}\}$. The spaces carrying these representations can be written as a direct sum of subspaces in the following way

$$V_i = \bigoplus_{\tau\in\boldsymbol{\sigma}_q} V_i^\tau \tag{6.55}$$

where $V_i^\tau \subset V^{\{\tau\}}$ with

$$V^{\{\tau\}} := \mathrm{Span}\{S_{\nu,i\nu\cdot\tau} \;\;\|\;\; \nu \in \mathbf{N}_\tau\}. \tag{6.56}$$

Finally we denote the set indexing all irreducible subrepresentations of ϕ_{TS} as $\mathcal{Z} = \mathcal{Z}_{\mathrm{d}} \cup \mathcal{Z}_{[S]} \cup \mathcal{Z}_{\{S\}}$ and we note that all irreducible representations indexed by $\mathcal{Z}$ are mutually inequivalent.

Note that we have only given an explicit basis for the space on which the representation ϕ_{tr} acts. It is possible to write down explicit bases for all relevant vector spaces but we will not need to do this here.

6.6.2. Variance bound

In this section we prove the main theorem of this chapter. Concretely we prove the following.

Theorem 6.1. Let Q be an observable and $\rho, \hat{\rho}$ density matrices and set $\nu = \frac{1}{2}(\rho - \hat{\rho})$. Consider a randomized benchmarking experiment using the Clifford group C_q with noisy implementation $\tilde{\mathcal{G}} = \mathcal{E}\mathcal{G}$ for all $G \in \mathsf{C}_q$. Then the variance $\mathbb{V}_m^2$ of this experiment is upper bounded by

$$\begin{aligned}\mathbb{V}_m^2 \leq mf^{m-1}\frac{d^2-2}{(d+1)^2}r^2 + \frac{d^2}{(d-1)^2}r^2u^{m-2}\frac{(m-1)\left(f^2/u\right)^m - m\left(f^2/u\right)^{m-1}+1}{(1-(f^2/u))^2} \\ +\eta(Q,\nu)mf^{m-1}r + \eta(Q,\nu)r^2u^{m-2}\frac{(m-1)\left(f^2/u\right)^m - m\left(f^2/u\right)^{m-1}+1}{(1-(f^2/u))^2}\end{aligned} \tag{6.57}$$

where $u = u(\mathcal{E})$ is the unitarity, $r = r(\mathcal{E})$ is the infidelity, d is the system dimension, m is the sequence length, $f = 1 - \frac{dr}{d-1}$ is the depolarizing parameter and η is a function capturing the deviation from the ideal Q and ν. This bound is valid for $r \leq \frac{1}{3}$.

Proof. We begin from an exact expression of the variance expressed in the Liouville representation eq. (6.24):

$$\mathbb{V}_m^2 = \langle\langle Q^{\otimes 2}|\mathcal{T}_{\phi_2}(\mathcal{E}^{\otimes 2})^m|\nu^{\otimes 2}\rangle\rangle - \langle\langle Q^{\otimes 2}|\left(\mathcal{T}_{\phi}(\mathcal{E})^{\otimes 2}\right)^m|\nu^{\otimes 2}\rangle\rangle \tag{6.58}$$

where $\mathcal{T}_{\phi_2}$ is the twirl over the two-copy representation of the Clifford group as defined in eq. (6.51) and $\mathcal{T}_\phi$ is the twirl over the (single copy) Liouville representation. Note now that $|\nu^{\otimes 2}\rangle\rangle \in V_{\mathrm{TS}}$ and that both $\mathcal{T}_{\phi_2}(\mathcal{E}^{\otimes 2})$ and $\mathcal{T}_\phi(\mathcal{E})^{\otimes 2}$ are CPTP maps of the form described in eq. (6.46). This means we can restrict both twirls to the traceless symmetric subspace. In this subspace we have from lemma 3.4 and lemma 6.1 that $\mathcal{T}_{\phi_2}(\mathcal{E}^{\otimes 2})$ and $\mathcal{T}_\phi(\mathcal{E})^{\otimes 2}$ are of the form

$$\mathcal{T}_{\phi_2}(\mathcal{E}^{\otimes 2}) = \sum_{i\in\mathcal{Z}}\chi_i\mathcal{P}_i \tag{6.59}$$

$$\mathcal{T}_{\phi}(\mathcal{E})^{\otimes 2} = \sum_{i\in\mathcal{Z}}f^2\mathcal{P}_i \tag{6.60}$$

where $\mathcal{Z}$ (as defined in lemma 6.1) indexes the irreducible subrepresentations of the traceless symmetric representation of the Clifford group and $\chi_i = \mathrm{Tr}(\mathcal{P}_i\mathcal{E}^{\otimes 2})/\mathrm{Tr}(\mathcal{P}_i)$ are the prefactors associated to the different subrepresentations. We also used that $\mathcal{T}_\phi(\mathcal{E})$ is a depolarizing channel with depolarizing parameter f [2]. Using that $\mathcal{P}_i^2 = \mathcal{P}_i$ and $\mathcal{P}_i\mathcal{P}_j = 0$ for $i, j \in \mathcal{Z}, i \neq j$ we can rewrite the variance as

$$\mathbb{V}_m^2 = \langle\langle Q^{\otimes 2}|\sum_{i\in\mathcal{Z}}\chi_i^m\mathcal{P}_i|\nu^{\otimes 2}\rangle\rangle - \langle\langle Q^{\otimes 2}|\sum_{i\in\mathcal{Z}}f^{2m}\mathcal{P}_i|\nu^{\otimes 2}\rangle\rangle \tag{6.61}$$

$$= \sum_{i\in\mathcal{Z}}\langle\langle Q^{\otimes 2}|\mathcal{P}_i|\nu^{\otimes 2}\rangle\rangle(\chi_i^m - f^{2m}). \tag{6.62}$$

We now apply a telescoping series identity, which is proven in corollary 6.1 of lemma 6.6, to the factor $\chi_i^m - f^{2m}$ in the above equation (for all $i \in \mathcal{Z}$). This gives

$$\mathbb{V}_m^2 = mf^{2(m-1)} \sum_{i\in\mathcal{Z}} \langle\!\langle Q^{\otimes 2}|\mathcal{P}_i|\nu^{\otimes 2}\rangle\!\rangle(\chi_i - f^2) \tag{6.63a}$$

$$+ \sum_{i\in\mathcal{Z}} \langle\!\langle Q^{\otimes 2}|\mathcal{P}_i|\nu^{\otimes 2}\rangle\!\rangle(\chi_i - f^2)^2 \sum_{s=2}^{m}(s-1)\chi_i^{m-s} f^{2(s-2)}. \tag{6.63b}$$

This equation contains two terms, eq. (6.63a) and eq. (6.63b) which we will bound separately. We now proceed to upper bound the first term, that is eq. (6.63a). For this we will split the the input and output operators Q, ν into their ideal parts (that is, the Pauli operator $\sigma_{\mathbf{P}} := \mathbf{P}/\sqrt{d}$) and deviations from that ideal. We define the functions

$$H_i(Q,\nu) := \langle\!\langle Q^{\otimes 2}|\mathcal{P}_i|\nu^{\otimes 2}\rangle\!\rangle - Q_{\mathbf{P}}^2\nu_{\mathbf{P}}^2\langle\!\langle \sigma_{\mathbf{P}}^{\otimes 2}|\mathcal{P}_i|\sigma_{\mathbf{P}}^{\otimes 2}\rangle\!\rangle \tag{6.64}$$

for all $i \in \mathcal{Z}$ where $Q_{\mathbf{P}} = \mathrm{Tr}(Q\sigma_{\mathbf{P}})$ and similarly for $\nu_{\mathbf{P}}$. Using this we can write eq. (6.63a) as

$$mf^{2(m-1)} \sum_{i\in\mathcal{Z}} \langle\!\langle Q^{\otimes 2}|\mathcal{P}_i|\nu^{\otimes 2}\rangle\!\rangle(\chi_i - f^2) = Q_{\mathbf{P}}^2\nu_{\mathbf{P}}^2 mf^{2(m-1)} \sum_{i\in\mathcal{Z}} \langle\!\langle \sigma_{\mathbf{P}}^{\otimes 2}|\mathcal{P}_i|\sigma_{\mathbf{P}}^{\otimes 2}\rangle\!\rangle(\chi_i - f^2) \tag{6.65a}$$

$$+ mf^{2(m-1)} \sum_{i\in\mathcal{Z}} H_i(Q,\nu)(\chi_i - f^2). \tag{6.65b}$$

Now consider the first term of the RHS, eq. (6.65a). First note from lemma 6.2 that for $i \notin \mathcal{Z}_{\mathrm{d}} = \{\mathrm{tr}, 1, 2\}$ we have $\mathcal{P}_i|\sigma_{\mathbf{P}}^{\otimes 2}\rangle\!\rangle = 0$. Hence we have

$$\begin{aligned}
Q_{\mathbf{P}}^2\nu_{\mathbf{P}}^2 mf^{2(m-1)} &\sum_{i\in\mathcal{Z}} \langle\!\langle \sigma_{\mathbf{P}}^{\otimes 2}|\mathcal{P}_i|\sigma_{\mathbf{P}}^{\otimes 2}\rangle\!\rangle(\chi_i - f^2) \\
&= Q_{\mathbf{P}}^2\nu_{\mathbf{P}}^2 mf^{2(m-1)} \sum_{i\in\mathcal{Z}_{\mathrm{d}}} \langle\!\langle \sigma_{\mathbf{P}}^{\otimes 2}|\mathcal{P}_i|\sigma_{\mathbf{P}}^{\otimes 2}\rangle\!\rangle(\chi_i - f^2) \\
&= Q_{\mathbf{P}}^2\nu_{\mathbf{P}}^2 mf^{2(m-1)} \sum_{i\in\mathcal{Z}_{\mathrm{d}}} \frac{\mathrm{Tr}(\mathcal{P}_i)}{d^2-1}\left(\frac{\mathrm{Tr}(\mathcal{P}_i\mathcal{E}^{\otimes 2})}{\mathrm{Tr}(\mathcal{P}_i)} - f^2\right) \\
&= Q_{\mathbf{P}}^2\nu_{\mathbf{P}}^2 mf^{2(m-1)} \left[\frac{1}{d^2-1}\,\mathrm{Tr}\left[\sum_{i\in\mathcal{Z}_{\mathrm{d}}} \mathcal{P}_i\mathcal{E}^{\otimes 2}\right] - f^2\right] \\
&= Q_{\mathbf{P}}^2\nu_{\mathbf{P}}^2 mf^{2(m-1)} \left[\frac{1}{d^2-1}\sum_{\tau\in\boldsymbol{\sigma}_q} \langle\!\langle \tau^{\otimes 2}|\mathcal{E}^{\otimes 2}|\tau^{\otimes 2}\rangle\!\rangle - f^2\right]
\end{aligned} \tag{6.66}$$

where we used lemma 6.2 in the first and second equalities and the fact that

$$\sum_{i\in\mathcal{Z}_{\mathrm{d}}} \mathcal{P}_i = \sum_{\tau\in\boldsymbol{\sigma}_q} |\tau^{\otimes 2}\rangle\!\rangle\langle\!\langle \tau^{\otimes 2}| \tag{6.67}$$

in the last equality (this can be seen from lemma 6.1). Now we use lemma 6.3 and the fact that $Q_{\mathbf{P}}\nu_{\mathbf{P}} \leq 1/4$ to obtain an upper bound

$$Q_{\mathbf{P}}^2\nu_{\mathbf{P}}^2 m f^{2(m-1)} \sum_{i\in\mathcal{Z}} \langle\!\langle \sigma^{\otimes 2}|\mathcal{P}_i|\sigma^{\otimes 2}\rangle\!\rangle(\chi_i - f^2) \leq m f^{2(m-1)} \frac{d^2-2}{4(d-1)^2} r^2. \tag{6.68}$$

This leaves us with the second term in the RHS, eq. (6.65b). Here we cannot attain a bound that is quadratic in r. Instead we will attempt a bound that is linear in r using lemma 6.5. We can write

$$\begin{aligned} m f^{2(m-1)} \sum_{i\in\mathcal{Z}} H_i(Q,\nu)(\chi_i - f^2) &\leq m f^{2(m-1)} \sum_{i\in\mathcal{Z}} |H_i(Q,\nu)||\chi_i - f^2| \\ &\leq m f^{2(m-1)} \frac{2dr}{d-1} \sum_{i\in\mathcal{Z}} |H_i(Q,\nu)| \end{aligned} \tag{6.69}$$

subject to the condition $r \leq \frac{1}{3}$. Writing $\eta(Q,\nu) := \sum_{i\in\mathcal{Z}} |H_i(Q,\nu)|$ we have a bound on eq. (6.63a).
We continue by upper bounding the second term in the variance, that is eq. (6.63b). We again split off the ideal components of Q and ν and write

$$\begin{aligned} &\sum_{i\in\mathcal{Z}} \langle\!\langle Q^{\otimes 2}|\mathcal{P}_i|\nu^{\otimes 2}\rangle\!\rangle(\chi_i - f^2)^2 \sum_{s=2}^{m}(s-1)\chi_i^{m-s} f^{2(s-2)} \\ &\qquad = Q_{\mathbf{P}}^2\nu_{\mathbf{P}}^2 \sum_{i\in\mathcal{Z}} \langle\!\langle \sigma_{\mathbf{P}}^{\otimes 2}|\mathcal{P}_i|\sigma_{\mathbf{P}}^{\otimes 2}\rangle\!\rangle(\chi_i - f^2)^2 \sum_{s=2}^{m}(s-1)\chi_i^{m-s} f^{2(s-2)} \\ &\qquad\quad + \sum_{i\in\mathcal{Z}} H_i(Q,\nu)(\chi_i - f^2)^2 \sum_{s=2}^{m}(s-1)\chi_i^{m-s} f^{2(s-2)} \\ &\qquad \leq \frac{1}{4} \sum_{i\in\mathcal{Z}_{\mathrm{d}}} \frac{\mathrm{Tr}(\mathcal{P}_i)}{d^2-1}(\chi_i - f^2)^2 \sum_{s=2}^{m}(s-1)\chi_i^{m-s} f^{2(s-2)} \\ &\qquad\quad + \sum_{i\in\mathcal{Z}} |H_i(Q,\nu)|(\chi_i - f^2)^2 \chi_i^{m-2} \sum_{s=2}^{m}(s-1)\chi_i^{m-s} f^{2(s-2)} \end{aligned} \tag{6.70}$$

where we have used the definition of the function $H_i(Q,\nu)$, lemma 6.2 and the triangle

inequality. Now we use lemma 6.5 to upper bound this quantity as

$$\begin{aligned}
&\sum_{i\in\mathcal{Z}} \langle\!\langle Q^{\otimes 2}|\mathcal{P}_i|\nu^{\otimes 2}\rangle\!\rangle (\chi_i - f^2)^2 \sum_{s=2}^{m}(s-1)\chi_i^{m-s} f^{2(s-2)} \\
&\qquad \leq \sum_{i\in\mathcal{Z}_{\mathrm{d}}} \frac{\mathrm{Tr}(\mathcal{P}_i)}{d^2-1}\left(\frac{dr}{d-1}\right)^2 \sum_{s=2}^{m}(s-1)\chi_i^{m-s} f^{2(s-2)} \\
&\qquad\qquad + \sum_{i\in\mathcal{Z}} |H_i(Q,\nu)| \left(\frac{2dr}{d-1}\right)^2 \sum_{s=2}^{m}(s-1)\chi_i^{m-s} f^{2(s-2)} \qquad (6.71)\\
&\qquad \leq \frac{d^2r^2}{(d-1)^2} \sum_{s=2}^{m}(s-1)\chi_i^{m-s} f^{2(s-2)} \\
&\qquad\qquad + \frac{4d^2r^2}{(d-1)^2} \sum_{i\in\mathcal{Z}} |H_i(Q,\nu)| \sum_{s=2}^{m}(s-1)\chi_i^{m-s} f^{2(s-2)}
\end{aligned}$$

where we have used the fact that $\sum_{i\in\mathcal{Z}_{\mathrm{d}}} \mathrm{Tr}(\mathcal{P}_i) = d^2-1$. It remains to deal with the last factor. This we do by using lemma 6.4 which states that $\chi_i \leq u$ for all $i \in \mathcal{Z}$, where u is the unitarity of the channel $\mathcal{E}$. Writing again $\eta(Q,\nu) := \sum_{i\in\mathcal{Z}} |H_i(Q,\nu)|$ we then have

$$\begin{aligned}
&\sum_{i\in\mathcal{Z}} \langle\!\langle Q^{\otimes 2}|\mathcal{P}_i|\nu^{\otimes 2}\rangle\!\rangle (\chi_i - f^2)^2 \sum_{s=2}^{m}(s-1)\chi_i^{m-s} f^{2(s-2)} \\
&\qquad \leq \frac{d^2r^2}{(d-1)^2} \sum_{s=2}^{m}(s-1)u^{m-s} f^{2(s-2)} \qquad (6.72)\\
&\qquad\qquad + \frac{4d^2r^2}{(d-1)^2} \sum_{i\in\mathcal{Z}} |H_i(Q,\nu)| \sum_{s=2}^{m}(s-1)u^{m-s} f^{2(s-2)}
\end{aligned}$$

We can further make sense of this quantity by using the well known series identity

$$\sum_{k=1}^{m}(k-1)x^{k-2} = \frac{(m-1)x^m - mx^{m-1} + 1}{(1-x)^2}, \qquad m \in \mathrm{N}, \qquad (6.73)$$

Factoring out a factor of u^{m-2} and setting $x = f^2/u$ we obtain the following

$$\begin{aligned}
&\sum_{i\in\mathcal{Z}} \langle\!\langle Q^{\otimes 2}|\mathcal{P}_i|\nu^{\otimes 2}\rangle\!\rangle (\chi_i - f^2)^2 \sum_{s=2}^{m}(s-1)\chi_i^{m-s} f^{2(s-2)} \\
&\qquad \leq \frac{d^2r^2}{(d-1)^2}(1+4\eta(Q,\nu))u^{m-2} \frac{(m-1)(f^2/u)^m - m(f^2/u)^{m-1} + 1}{(1-(f^2/u))^2}.
\end{aligned} \qquad (6.74)$$

This finishes the upper bounding of eq. (6.63b). Gathering all terms we come to a final

bound

$$\mathbb{V} \leq mf^{m-1}\frac{d^2-2}{4(d+1)^2}r^2+\frac{d^2}{(d-1)^2}r^2u^{m-2}\frac{(m-1)\left(f^2/u\right)^m-m\left(f^2/u\right)^{m-1}+1}{(1-(f^2/u))^2}$$
$$+\eta(Q,\nu)mf^{m-1}r+\eta(Q,\nu)r^2u^{m-2}\frac{(m-1)\left(f^2/u\right)^m-m\left(f^2/u\right)^{m-1}+1}{(1-(f^2/u))^2} \tag{6.75}$$

which is the bound we set out to find. ■

Noting that $f^2 \leq u$ and that the factor

$$\frac{(m-1)\left(f^2/u\right)^m-m\left(f^2/u\right)^{m-1}+1}{(1-(f^2/u))^2}, \tag{6.76}$$

is monotonically decreasing in u we can upper bound this factor by taking the limit $u \to f^2$. This gives

$$\lim_{u\to f^2}\frac{(m-1)\left(f^2/u\right)^m-m\left(f^2/u\right)^{m-1}+1}{(1-(f^2/u))^2}=\frac{m(m-1)}{2}. \tag{6.77}$$

which can be confirmed by an application of l'Hôpital's rule. Plugging this in to eq. (6.75) we obtain eq. (6.5).

6.6.3. State preparation and measurement (SPAM) terms

In the central bound on the variance (theorem 6.1) we had to account for the fact that the variance can depend on how well the input states $\rho, \hat{\rho}$ and the output POVM Q can be implemented. The ideal behavior of $\nu = \frac{1}{2}(\rho - \hat{\rho})$ and Q are given by

$$Q_{\text{id}} = \frac{1}{2}(\mathbb{1} + \mathbf{P}) \tag{6.78}$$
$$\nu_{\text{id}} = \frac{\mathbf{P}}{2d} \tag{6.79}$$

where $\mathbf{P}$ is a pre-specified element of the Pauli group (see fig. 6.1). The deviation of Q and ν from this ideal can be captured by writing

$$Q = Q_{\text{id}} + Q_{\text{spam}} \tag{6.80}$$
$$\nu = \nu_{\text{id}} + \nu_{\text{spam}} \tag{6.81}$$

where $\langle Q_{\text{id}}, Q_{\text{spam}}\rangle = \langle \nu_{\text{id}}, \nu_{\text{spam}}\rangle = 0$.
In the variance bound the deviation from the ideal has an effect which is measured by the parameter $\eta(Q,\nu)$. This parameter $\eta(Q,\nu)$ was defined as

$$\eta(Q,\nu) = \sum_{i\in\mathcal{Z}} H_i(Q,\nu) = \sum_{i\in\mathcal{Z}} |\langle\!\langle Q^{\otimes 2}|\mathcal{P}_i|\nu^{\otimes 2}\rangle\!\rangle - \langle\!\langle Q_{\text{id}}^{\otimes 2}|\mathcal{P}_i|\nu_{\text{id}}^{\otimes 2}\rangle\!\rangle| \tag{6.82}$$

where $\mathcal{Z}$ indexes the irreducible representations of the traceless symmetric representation of the Clifford group and the $\mathcal{P}_i$ are projectors onto the spaces carrying these subrepresentations (lemma 6.1). Let us now analyze these terms further. For $i \in \mathcal{Z}_{\mathrm{d}}$ we have

$$\begin{aligned} H_i(Q,\nu) &= |\langle\!\langle (Q_{\mathrm{id}} + Q_{\mathrm{spam}})^{\otimes 2} | \mathcal{P}_i | (\nu_{\mathrm{id}} + \nu_{\mathrm{spam}})^{\otimes 2} \rangle\!\rangle - \langle\!\langle Q_{\mathrm{id}}^{\otimes 2} | \mathcal{P}_i | \nu_{\mathrm{id}}^{\otimes 2} \rangle\!\rangle| \\ &= |\langle\!\langle Q_{\mathrm{id}}^{\otimes 2} | \mathcal{P}_i | \nu_{\mathrm{spam}}^{\otimes 2} \rangle\!\rangle + \langle\!\langle Q_{\mathrm{id}}^{\otimes 2} | \mathcal{P}_i | \nu_{\mathrm{id}}^{\otimes 2} \rangle\!\rangle + \langle\!\langle Q_{\mathrm{spam}}^{\otimes 2} | \mathcal{P}_i | \nu_{\mathrm{spam}}^{\otimes 2} \rangle\!\rangle| \end{aligned} \tag{6.83}$$

where we have used that $\langle Q_{\mathrm{id}}, Q_{\mathrm{spam}} \rangle = \langle \nu_{\mathrm{id}}, \nu_{\mathrm{spam}} \rangle = 0$ which implies that $\langle\!\langle Q_{\mathrm{id}} \otimes Q_{\mathrm{spam}} | \mathcal{P}_i = \mathcal{P}_i | \nu_{\mathrm{id}} \otimes \nu_{\mathrm{spam}} \rangle\!\rangle = 0$ for $i \in \mathcal{Z}_{\mathrm{d}}$. Using the triangle inequality and the Cauchy-Schwarz inequality we can get

$$\begin{aligned} H_i(Q,\nu) \leq & |\langle\!\langle Q_{\mathrm{id}}^{\otimes 2} | \mathcal{P}_i | \nu_{\mathrm{spam}}^{\otimes 2} \rangle\!\rangle| + |\langle\!\langle Q_{\mathrm{spam}}^{\otimes 2} | \mathcal{P}_i | \nu_{\mathrm{id}}^{\otimes 2} \rangle\!\rangle| + |\langle\!\langle Q_{\mathrm{spam}}^{\otimes 2} | \mathcal{P}_i | \nu_{\mathrm{spam}}^{\otimes 2} \rangle\!\rangle| \\ \leq & \left\| Q_{\mathrm{id}}^{\otimes 2} \right\|_2 \left\| \mathcal{P}_i (\nu_{\mathrm{spam}}^{\otimes 2}) \right\|_2 + \left\| Q_{\mathrm{spam}}^{\otimes 2} \right\|_2 \left\| \mathcal{P}_i (\nu_{\mathrm{id}}^{\otimes 2}) \right\|_2 + \left\| Q_{\mathrm{spam}}^{\otimes 2} \right\|_2 \left\| \mathcal{P}_i (\nu_{\mathrm{spam}}^{\otimes 2}) \right\|_2 \\ \leq & \left\| \mathcal{P}_i \right\|_{2\to 2} \left(\|Q_{\mathrm{id}}\|_2^2 \|\nu_{\mathrm{spam}}\|_2^2 + \|Q_{\mathrm{spam}}\|_2^2 \|\nu_{\mathrm{id}}\|_2^2 + \|Q_{\mathrm{spam}}\|_2^2 \|\nu_{\mathrm{spam}}\|_2^2 \right) \end{aligned} \tag{6.84}$$

where $\|\mathcal{P}_i\|_{2\to 2}$ is the induced 2-norm of the superoperator $\mathcal{P}_i$. It is well known that this norm is equal to the largest singular value of the Liouville representation of $\mathcal{P}_i$ [2], which since the Liouville representation of $\mathcal{P}_i$ is an orthonormal projection, is equal to one. This means we have for $i \in \mathcal{Z}_{\mathrm{d}}$ that

$$\begin{aligned} H_i(Q,\nu) &\leq \|Q_{\mathrm{id}}\|_2^2 \|\nu_{\mathrm{spam}}\|_2^2 + \|Q_{\mathrm{spam}}\|_2^2 \|\nu_{\mathrm{id}}\|_2^2 + \|Q_{\mathrm{spam}}\|_2^2 \|\nu_{\mathrm{spam}}\|_2^2 \\ &= \|Q_{\mathrm{id}}\|_2^2 \|\nu - \nu_{\mathrm{id}}\|_2^2 + \|Q - Q_{\mathrm{id}}\|_2^2 \|\nu_{\mathrm{id}}\|_2^2 + \|Q - Q_{\mathrm{id}}\|_2^2 \|\nu - \nu_{\mathrm{id}}\|_2^2 . \end{aligned} \tag{6.85}$$

Note that this expression is zero when both Q and ν are ideally implemented but is non-zero when either of them is not. This behavior is in general unavoidable as we argue in section 6.5.7. But first we will consider the functions $H_i(Q,\nu)$ for $i \in Z_{[S]} \cup Z_{\{S\}}$. Note first that since $\mathrm{supp}(\mathcal{P}_i) \subset \mathrm{Span}\{S_{\sigma,\sigma'} \parallel \sigma, \sigma' \in \boldsymbol{\sigma}_q,\ \sigma \neq \sigma'\}$ we must have that $\mathcal{P}_i | \nu_{\mathrm{id}}^{\otimes 2} \rangle\!\rangle = \langle\!\langle Q_{\mathrm{id}}^{\otimes 2} | \mathcal{P}_i = 0$. This means we can write

$$H_i(Q,\nu) = |\langle\!\langle Q^{\otimes 2} | \mathcal{P}_i | \nu^{\otimes 2} \rangle\!\rangle - \langle\!\langle Q_{\mathrm{id}}^{\otimes 2} | \mathcal{P}_i | \nu_{\mathrm{id}}^{\otimes 2} \rangle\!\rangle| \tag{6.86}$$

$$\begin{aligned} &= |\langle\!\langle Q_{\mathrm{spam}}^{\otimes 2} | \mathcal{P}_i | \nu_{\mathrm{id}} \otimes \nu_{\mathrm{spam}} + \nu_{\mathrm{spam}} \otimes \nu_{\mathrm{id}} \rangle\!\rangle \\ &\quad + \langle\!\langle Q_{\mathrm{id}} \otimes Q_{\mathrm{spam}} + Q_{\mathrm{spam}} \otimes Q_{\mathrm{id}} | \mathcal{P}_i | \nu_{\mathrm{spam}}^{\otimes 2} \rangle\!\rangle + \langle\!\langle Q_{\mathrm{spam}}^{\otimes 2} | \mathcal{P}_i | \nu_{\mathrm{spam}}^{\otimes 2} \rangle\!\rangle \\ &\quad + \langle\!\langle Q_{\mathrm{id}} \otimes Q_{\mathrm{spam}} + Q_{\mathrm{spam}} \otimes Q_{\mathrm{id}} | \mathcal{P}_i | \nu_{\mathrm{id}} \otimes \nu_{\mathrm{spam}} + \nu_{\mathrm{spam}} \otimes \nu_{\mathrm{id}} \rangle\!\rangle| \end{aligned} \tag{6.87}$$

$$\begin{aligned} \leq \|\mathcal{P}_i\|_{2\to 2} \Big(& \left\| Q_{\mathrm{spam}}^{\otimes 2} \right\|_2 \left\| \nu_{\mathrm{spam}}^{\otimes 2} \right\|_2 + 2 \|Q_{\mathrm{spam}}\|_2 \|Q_{\mathrm{id}}\|_2 \left\| \nu_{\mathrm{spam}}^{\otimes 2} \right\|_2 \\ & + 2 \|\nu_{\mathrm{spam}}\|_2 \|\nu_{\mathrm{id}}\|_2 \left\| Q_{\mathrm{spam}}^{\otimes 2} \right\|_2 + 4 \|\nu_{\mathrm{spam}}\|_2 \|\nu_{\mathrm{id}}\|_2 \|Q_{\mathrm{spam}}\|_2 \|Q_{\mathrm{id}}\|_2 \Big) \end{aligned} \tag{6.88}$$

which we can rewrite as

$$H_i(Q,\nu) \leq \|Q-Q_{\mathrm{id}}\|_2 \|\nu-\nu_{\mathrm{id}}\|_2 \Big(\|Q-Q_{\mathrm{id}}\|_2 \|\nu-\nu_{\mathrm{id}}\|_2 + 2\|\nu-\nu_{\mathrm{id}}\|_2 \|Q_{\mathrm{id}}\|_2 + 2\|Q-Q_{\mathrm{id}}\|_2 \|\nu_{\mathrm{id}}\|_2 + 4\|\nu_{\mathrm{id}}\|_2 \|Q_{\mathrm{id}}\|_2 \Big) \tag{6.89}$$

which makes manifest that $H_i(Q,\nu) = 0$ if Q and ν are ideal and moreover that this term actually scales with the product of the deviations in Q and ν (as measured in the 2-norm). Hence we see that to lowest order in Q_{spam} and ν_{spam} the SPAM parameter $\eta(Q,\nu)$ is proportional to

$$\eta \approx \|Q-Q_{\mathrm{id}}\|_2 \|\nu-\nu_{\mathrm{id}}\|_2 + \|Q-Q_{\mathrm{id}}\|_2^2 + \|\nu-\nu_{\mathrm{id}}\|_2^2 \tag{6.90}$$

with the exact expression being

$$\eta(Q,\nu) \leq 3\Big[\|Q_{\mathrm{id}}\|_2^2 \|\nu-\nu_{\mathrm{id}}\|_2^2 + \|Q-Q_{\mathrm{id}}\|_2^2 \|\nu_{\mathrm{id}}\|_2^2 + \|Q-Q_{\mathrm{id}}\|_2^2 \|\nu-\nu_{\mathrm{id}}\|_2^2\Big] + 5\Big[\|Q-Q_{\mathrm{id}}\|_2 \|\nu-\nu_{\mathrm{id}}\|_2 \Big(\|Q-Q_{\mathrm{id}}\|_2 \|\nu-\nu_{\mathrm{id}}\|_2 + 2\|\nu-\nu_{\mathrm{id}}\|_2 \|Q_{\mathrm{id}}\|_2 + 2\|Q-Q_{\mathrm{id}}\|_2 \|\nu_{\mathrm{id}}\|_2 + 4\|\nu_{\mathrm{id}}\|_2 \|Q_{\mathrm{id}}\|_2 \Big)\Big] \tag{6.91}$$

where the factors 3 and 5 arise from the fact that $|\mathcal{Z}_{\mathrm{d}}| = 3$ and $|Z_{[S]} \cup Z_{\{S\}}| = 5$ respectively (this is for $q \geq 3$, for $q = 1$ we get the significantly better $|\mathcal{Z}_{\mathrm{d}}| = 2$ and $|Z_{[S]} \cup Z_{\{S\}}| = 1$ instead [35]).

6.6.4. SAMPLE COMPLEXITY OF ITERATIVELY REWEIGHTED LEAST SQUARES

In this section we analyze the sample complexity of the RB fitting procedure using iteratively reweighted least squares, as outlined in section 6.5.3. Given a set of sequence lengths $\mathbb{M}$ we will assume that N random sequences are sampled for each sequence length. It is possible to let N be a function of the sequence length m and prove a more general version of the theorem presented here but we will not pursue this here. We will also only be interested in the uncertainty around the estimate for the depolarizing parameter f, it is straightforward to extend our analysis to also include the uncertainty around estimate for the pre-factor A. The methods we use are all standard and can be found in [12, 22]. See also [3] for an earlier calculation of this form in the context of randomized benchmarking (not taking into account the heteroskedasticity of randomized benchmarking data).

Theorem 6.2. Let $\mathbb{M}$ be a set of integers denoting sequence lengths and let $\{k_{m,N}\}_{m\in\mathbb{M}}$ be a set of RB data points obtained by sampling N random sequences for each sequence length $m \in \mathbb{M}$. Denote by f^*, A^* the true values for the RB fitting parameters and denote by $f_{\mathrm{est}}, A_{\mathrm{est}}$ their estimates as obtained by the iteratively reweighted least squares

procedure outlined in algorithm 1. We then have that

$$\Pr\left[|f^* - f_{\text{est}}| \leq \epsilon\right] \geq 1 - \delta \tag{6.92}$$

where δ is upper bounded by

$$\delta \leq 2H[\mathbb{V}_{\text{fit}}, \epsilon_{\text{fit}}]^{N|\mathbb{M}|} \tag{6.93}$$

with H defined in eq. (8.30) and

$$\mathbb{V}_{\text{fit}} = \frac{1}{|\mathbb{M}|}\sum_{m\in\mathbb{M}} \mathbb{V}_m(f^*)w(f_{\text{est}}, m) \tag{6.94}$$

$$\epsilon_{\text{fit}} = \frac{\epsilon[J^T J]}{J_{f_{\text{est}}}} \tag{6.95}$$

and

$$J = \left[-\frac{1}{|\mathbb{M}|}\sum_{m\in\mathbb{M}} mA^* {f^*}^{m-1} w(f^*) \;\;,\;\; \frac{1}{|\mathbb{M}|}\sum_{m\in\mathbb{M}} {f^*}^m w(f^*, m)\right] \tag{6.96}$$

and $J_{f_{\text{est}}}$ is the first entry of this vector.

Proof. The starting off point for this proof is given by Eq. 1.6.19 in [22, Page 45] which states that the outcome of the IRLS procedure satisfies the following equality

$$\frac{1}{|\mathbb{M}|}\sum_{m\in\mathbb{M}} (k_{m,N} - A_{\text{est}} f_{\text{est}}^m)w(f_{\text{est}}, m) = 0 \tag{6.97}$$

where $w(f, m)$ is the weight function given by the inverse of eq. (6.24) (we suppress the dependency on the unitarity here for notational simplicity). We can rewrite eq. (6.97) as

$$\frac{1}{|\mathbb{M}|}\sum_{m\in\mathbb{M}} (k_{m,N} + A^* {f^*}^m - A^* {f^*}^m - A_{\text{est}} f_{\text{est}}^m)w(f_{\text{est}}, m) = 0 \tag{6.98}$$

$$\iff \frac{1}{|\mathbb{M}|}\sum_{m\in\mathbb{M}} (A^* {f^*}^m - A_{\text{est}} f_{\text{est}}^m)w(f_{\text{est}}, m) = \frac{1}{|\mathbb{M}|}\sum_{m\in\mathbb{M}} (k_{m,N} - A^* {f^*}^m)w(f_{\text{est}}, m). \tag{6.99}$$

We can now think of the LHS of eq. (6.99) as a function of the vector $[f_{\text{est}}, A_{\text{est}}]$. Assuming $[f_{\text{est}}, A_{\text{est}}]$ is close to $[f^*, A^*]$ we can expand the LHS of eq. (6.99) to first order to get

$$J^T\left[f^* - f_{\text{est}}, A^* - A_{\text{est}}\right] \approx \frac{1}{|\mathbb{M}|}\sum_{m\in\mathbb{M}} (A^* {f^*}^m - A_{\text{est}} f_{\text{est}}^m)w(f_{\text{est}}, m) \tag{6.100}$$

where J is the Jacobian associated to the LHS of eq. (6.99), that is:

$$J = \left[-\frac{1}{|\mathbb{M}|}\sum_{m\in\mathbb{M}} mA^* {f^*}^{m-1} w(f^*) \;\;,\;\; \frac{1}{|\mathbb{M}|}\sum_{m\in\mathbb{M}} {f^*}^m w(f^*, m)\right]. \tag{6.101}$$

Taking the Moore-Penrose inverse $J^{\mathrm{MP}} = (J^T J)^{-1} J$ of J and inserting this in the first entry of eq. (6.100) we can say that

$$f^* - f_{\mathrm{est}} \approx (J^T J)^{-1} J_{\mathrm{est}} \frac{1}{|\mathbb{M}|} \sum_{m \in \mathbb{M}} (A^* f^{*m} - A_{\mathrm{est}} f_{\mathrm{est}}^m) w(f_{\mathrm{est}}, m) \tag{6.102}$$

where $J_{\hat{f}}$ is the first entry of J. Now we can say that

$$\mathrm{Prob}\left[|f^* - f_{\mathrm{est}}| \geq \epsilon\right] \approx \mathrm{Prob}\left[\left| [J^T J]^{-1} J_{\hat{f}} \frac{1}{|\mathbb{M}|} \sum_{m \in \mathbb{M}} (A^* f^{*m} - A_{\mathrm{est}} f_{\mathrm{est}}^m) w(f_{\mathrm{est}}, m) \right| \geq \epsilon \right] \tag{6.103}$$

$$= \mathrm{Prob}\left[\left| [J^T J]^{-1} J_{\hat{f}} \frac{1}{|\mathbb{M}|} \sum_{m \in \mathbb{M}} (k_{m,N} - A^* f^{*m}) w(f_{\mathrm{est}}, m) \right| \geq \epsilon \right] \tag{6.104}$$

Now note that $k_{m,N}$ can be seen as a number drawn from a random variable K_m with mean $A^*(f^*)^m$ and variance $\mathbb{V}_m(f^*)/N^2$ where N is the number of random sequences drawn for each data-point $k_{m,N}$. Moreover $k_{m,N}$ and $k_{N,m'}$ for $m \neq m'$ are drawn from independent random variables K_m and $K_{m'}$. Hence we can apply the concentration inequality given in eq. (8.30) to eq. (6.104) to get

$$\mathrm{Prob}\left[|f^* - f_{\mathrm{est}}| \geq \epsilon\right] \leq 2H[\mathbb{V}_{\mathrm{fit}}, \epsilon_{\mathrm{fit}}]^{N|\mathbb{M}|} \tag{6.105}$$

with $\mathbb{V}_{\mathrm{fit}}, \epsilon_{\mathrm{fit}}$ given by

$$\mathbb{V}_{\mathrm{fit}} = \frac{1}{|\mathbb{M}|} \sum_{m \in \mathbb{M}} \mathbb{V}_m(f^*) w(f_{\mathrm{est}}, m) \tag{6.106}$$

$$\epsilon_{\mathrm{fit}} = \frac{\epsilon [J^T J]}{J_{f_{\mathrm{est}}}} \tag{6.107}$$

which completes the proof. ■

Using eq. (6.4) or eq. (6.6) then gives an upper bound on total amount of data that needs to be gathered for rigorous RB.

6.6.5. Technical lemmas

In this section state all technical lemmas used in the main result theorem 6.1.

Projectors in the traceless symmetric subspace

In lemma 6.2 we prove a series of useful upper bounds on the trace overlap between the superoperator-projectors associated to the traceless-symmetric representation of the Clifford group and the normalized Pauli matrices. The saturated versions of these inequalities are critical to establishing the quadratic scaling with infidelity of the variance bound in the case of SPAM-free RB.

Lemma 6.2. Let $\mathcal{E} : \mathcal{M}_d \to \mathcal{M}_d$ be a quantum channel and consider the twirled operator $\mathcal{T}_{\phi_{\mathrm{TS}}}(\mathcal{E}^{\otimes 2})$ with respect to the traceless-symmetric representation. This operator can then be written as (lemma 6.1)

$$\mathcal{T}_{\phi_{\mathrm{TS}}}(\mathcal{E}^{\otimes 2}) = \sum_{i \in \mathcal{Z}} \frac{\mathrm{Tr}(\mathcal{E}\mathcal{P}_i)}{\mathrm{Tr}(\mathcal{P}_i)} \mathcal{P}_i \tag{6.108}$$

with $\mathcal{Z} = \{\mathrm{tr}, 1, 2, [1], [2], [3], \{1\}, \{2\}\}$ and $\mathcal{P}_i$ the projector onto the spaces $V_i \subset \mathcal{M}_d^{\otimes 2}$. Let $I(x \in A)$ be the indicator function for the set A (that is $I(x \in A) = 1$ if $x \in A$ and $I(x \in A) = 0$ otherwise). We have the following statements

- For $i \in \mathcal{Z}$ and $\sigma, \sigma' \in \boldsymbol{\sigma}_q$ we have that

$$|\langle\!\langle \sigma^{\otimes 2} | \mathcal{P}_i | \sigma'^{\otimes 2} \rangle\!\rangle| = |\langle\!\langle \sigma^{\otimes 2} | \mathcal{P}_i | \sigma'^{\otimes 2} \rangle\!\rangle| I(i \in \mathcal{Z}_{\mathrm{d}}) \leq \frac{\mathrm{Tr}(\mathcal{P}_i) I(i \in \mathcal{Z}_{\mathrm{d}})}{d^2 - 1}$$

  with equality when $\sigma = \sigma'$.

- For $i \in \mathcal{Z}, \tau, \tau' \in \boldsymbol{\sigma}_q$ and $\sigma \in \mathbf{C}_\tau, \sigma' \in \mathbf{C}_{\tau'}$ we have that

$$\begin{aligned} |\langle\!\langle S_{\sigma,\sigma\cdot\tau} | \mathcal{P}_i | S_{\sigma',\sigma'\cdot\tau'} \rangle\!\rangle| &= |\langle\!\langle S_{\sigma,\sigma\cdot\tau} | \mathcal{P}_i | S_{\sigma',\sigma'\cdot\tau} \rangle\!\rangle| I(i \in \mathcal{Z}_{[S]}) \delta_{\tau,\tau'} \\ &\leq \frac{2\,\mathrm{Tr}(\mathcal{P}_i) I(i \in \mathcal{Z}_{[S]}) \delta_{\tau,\tau'}}{(d^2-1)(d^2/2-2)} \end{aligned}$$

  with equality when $\sigma = \sigma'$.

- For $i \in \mathcal{Z}, \tau, \tau' \in \boldsymbol{\sigma}_q$ and $\sigma \in \mathbf{N}_\tau, \sigma' \in \mathbf{N}_{\tau'}$ we have that

$$\begin{aligned} |\langle\!\langle S_{\sigma,i\sigma\cdot\tau} | \mathcal{P}_i | S_{\sigma',i\sigma'\cdot\tau'} \rangle\!\rangle| &= |\langle\!\langle S_{\sigma,i\sigma\cdot\tau} | \mathcal{P}_i | S_{\sigma',i\sigma'\cdot\tau} \rangle\!\rangle| I(i \in \mathcal{Z}_{\{S\}}) \delta_{\tau,\tau'} \\ &\leq \frac{2\,\mathrm{Tr}(\mathcal{P}_i) I(i \in \mathcal{Z}_{\{S\}}) \delta_{\tau,\tau'}}{(d^2-1)(d^2/2)} \end{aligned}$$

  with equality when $\sigma = \sigma'$.

where the sets $Z_{\mathrm{d}}, \mathcal{Z}_{[S]}, \mathcal{Z}_{\{S\}}$ are defined in lemma 6.1.

Proof. We begin by proving the first claim. Let $\mathcal{P}_i$ be a projector as defined in the lemma statement with $i \in \mathcal{Z}$ and take $\sigma, \sigma' \in \boldsymbol{\sigma}_q$. From lemma 6.1 we have immediately that

$$\langle\!\langle \sigma^{\otimes 2} | \mathcal{P}_i | \sigma'^{\otimes 2} \rangle\!\rangle = \langle\!\langle \sigma^{\otimes 2} | \mathcal{P}_i | \sigma'^{\otimes 2} \rangle\!\rangle I(i \in \mathcal{Z}_{\mathrm{d}}). \tag{6.109}$$

Now consider $i \in \mathcal{Z}_{\mathrm{d}}$. Note that since $\mathcal{P}_i$ is a projector it is a real matrix and we have that $\mathcal{P}_i \geq 0$, that is $\mathcal{P}_i$ is a positive semidefinite matrix. This means that we have, by the Sylvester principal minor conditions, that

$$|\langle\!\langle \sigma^{\otimes 2} | \mathcal{P}_i | \sigma'^{\otimes 2} \rangle\!\rangle| \leq \sqrt{\langle\!\langle \sigma'^{\otimes 2} | \mathcal{P}_i | \sigma'^{\otimes 2} \rangle\!\rangle \langle\!\langle \sigma^{\otimes 2} | \mathcal{P}_i | \sigma^{\otimes 2} \rangle\!\rangle} \tag{6.110}$$

for all $\sigma, \sigma' \in \boldsymbol{\sigma}_q$. Now consider the case $\sigma = \sigma'$. Note that for all $\tau, \sigma \in \boldsymbol{\sigma}_q$ there is a $G_\tau^\sigma \in \mathsf{C}_q$ such that $\mathcal{G}_\tau^\sigma(\tau) = \pm\sigma$. That is, the Clifford group acts transitively on $\boldsymbol{\sigma}_q$ [36]. This means we can write

$$\begin{aligned}
\langle\!\langle \sigma^{\otimes 2} | \mathcal{P}_i | \sigma^{\otimes 2} \rangle\!\rangle &= \frac{1}{d^2-1} \sum_{\tau \in \boldsymbol{\sigma}_q} \langle\!\langle \mathcal{G}_\tau^\sigma(\tau)^{\otimes 2} | \mathcal{P}_i | \mathcal{G}_\tau^\sigma(\tau) \rangle\!\rangle \\
&= \frac{1}{d^2-1} \sum_{\tau \in \boldsymbol{\sigma}_q} \langle\!\langle \tau^{\otimes 2} | \left(\mathcal{G}_\tau^\sigma\right)^{\dagger \otimes 2} \mathcal{P}_i \left(\mathcal{G}_\tau^\sigma\right)^{\otimes 2} | \tau^{\otimes 2} \rangle\!\rangle \\
&= \frac{1}{d^2-1} \sum_{\tau \in \boldsymbol{\sigma}_q} \langle\!\langle \tau^{\otimes 2} | \mathcal{P}_i | \tau^{\otimes 2} \rangle\!\rangle \\
&= \frac{\mathrm{Tr}(\mathcal{P}_i)}{d^2-1}
\end{aligned} \tag{6.111}$$

where we used the fact that $\mathcal{P}_i$ commutes with $\mathcal{G}^{\otimes 2}$ for all $G \in \mathsf{C}_q$ and the fact that $V_i \subset V_{\mathrm{d}}$ (where V_{d} is defined in lemma 6.1). This proves the first claim of the lemma.
Next we consider the second claim of the lemma. Let $\tau, \tau' \in \boldsymbol{\sigma}_q$ and take $\sigma \in \mathbf{C}_\tau$ and $\sigma' \in \mathbf{C}_{\tau'}$. Again from lemma 6.1 we have immediately that

$$\langle\!\langle S_{\sigma,\sigma\cdot\tau} | \mathcal{P}_i | S_{\sigma',\sigma'\cdot\tau'} \rangle\!\rangle = \langle\!\langle S_{\sigma,\sigma\cdot\tau} | \mathcal{P}_i | S_{\sigma',\sigma'\cdot\tau'} \rangle\!\rangle I(i \in \mathcal{Z}_{[S]}). \tag{6.112}$$

Now consider $i \in \mathcal{Z}_{[S]}$. From lemma 6.1 we have that we can write

$$\mathcal{P}_i = \sum_{\tau \in \boldsymbol{\sigma}_q} \mathcal{P}_i^\tau \tag{6.113}$$

where $\mathcal{P}_i^\tau$ has support in the space

$$V^{[\tau]} = \{ S_{\sigma,\sigma\cdot\tau} \parallel \sigma \in \mathbf{C}_\tau \}. \tag{6.114}$$

From this we immediately get

$$\langle\!\langle S_{\sigma,\sigma\cdot\tau} | \mathcal{P}_i | S_{\sigma',\sigma'\cdot\tau'} \rangle\!\rangle = \langle\!\langle S_{\sigma,\sigma\cdot\tau} | \mathcal{P}_i | S_{\sigma',\sigma'\cdot\tau'} \rangle\!\rangle \delta_{\tau,\tau'}. \tag{6.115}$$

Now consider $\tau = \tau'$. Again from the Sylvester minor conditions we get for all $\sigma, \sigma' \in \mathbf{C}_\tau$ that

$$|\langle\!\langle S_{\sigma,\sigma\cdot\tau} | \mathcal{P}_i | S_{\sigma',\sigma'\cdot\tau} \rangle\!\rangle| \leq \sqrt{\langle\!\langle S_{\sigma',\sigma'\cdot\tau} | \mathcal{P}_i | S_{\sigma',\sigma'\cdot\tau} \rangle\!\rangle \langle\!\langle S_{\sigma,\sigma\cdot\tau} | \mathcal{P}_i | S_{\sigma,\sigma\cdot\tau} \rangle\!\rangle}. \tag{6.116}$$

Now consider the case $\sigma = \sigma'$. From [36] we can see that the action of the Clifford group on the set $A = \{(\sigma, \sigma\cdot\tau) \parallel \tau \in \boldsymbol{\sigma}_q,\ \sigma \in \mathbf{C}_\tau\}$ is 2-transitive. That is, for all pairs $(\nu, \mu) \in A$ there is a $G_{\nu,\mu}^{\sigma,\tau} \in \mathsf{C}_q$ such that

$$\mathcal{G}_{\nu,\mu}^{\sigma,\tau\otimes 2}\left(S_{\sigma,\sigma\cdot\tau}\right) = S_{\nu,\nu\cdot\mu}. \tag{6.117}$$

This implies we can make essentially the same argument as before, that is

$$\begin{aligned}\langle\!\langle S_{\sigma,\sigma\cdot\tau}|\mathcal{P}_i|S_{\sigma,\sigma\cdot\tau}\rangle\!\rangle &= \frac{1}{|A|}\sum_{(\mu,\nu)\in A}\langle\!\langle S_{\nu,\nu\cdot\mu}|\left(\mathcal{G}_{\nu,\mu}^{\sigma,\tau}\right)^{\dagger\otimes 2}\mathcal{P}_i\mathcal{G}_{\nu,\mu}^{\sigma,\tau\otimes 2}|S_{\nu,\nu\cdot\mu}\rangle\!\rangle \\ &= \frac{1}{|A|}\sum_{(\mu,\nu)\in A}\langle\!\langle S_{\nu,\nu\cdot\mu}|\mathcal{P}_i|S_{\nu,\nu\cdot\mu}\rangle\!\rangle \\ &= \frac{2\,\mathrm{Tr}(\mathcal{P}_i)}{(d^2-1)(d^2/2-2)}\end{aligned} \tag{6.118}$$

where we have used the fact that $\mathcal{G}^{\otimes 2}$ commutes with $\mathcal{P}_i$ for all $G \in \mathsf{C}_q$ and also the definition of the space $V_{[S]}$ (given in lemma 6.1). The factor of two appears from the fact that the set A counts the basis of $V_{[S]}$ twice since $S_{\nu,\nu\cdot\mu} = S_{\nu\cdot\mu,\nu}$ for all $(\mu,\nu\cdot\mu)\in A$. We have also used that $|A| = |\boldsymbol{\sigma}_q||\mathbf{C}_\tau| = (d^2-1)(d^2/2-2)$. This proves the second claim of the lemma.
The proof of the third claim of the lemma proceeds in the same way as the proof of the second claim with the difference that anti-commuting, rather than commuting elements of the Pauli group must considered. We will not write it down explicitly. ■

BOUND ON SUM OF SQUARES OF THE DIAGONAL ELEMENTS OF A QUANTUM CHANNEL
This lemma (lemma 6.3) proves that the diagonal elements of a CPTP map are generically quite close to their mean. The key technique used here is the fact that the diagonal elements of a CPTP map are invariant under Pauli twirling. This is a structural result about quantum channels on arbitrarily many qubits and might this be of independent interest. We use it to establish the quadratic scaling of the variance in the infidelity in the case of SPAM-free RB.

Lemma 6.3. Let $\mathcal{E}:\mathcal{M}_d\to\mathcal{M}_d$ be a quantum channel with infidelity r and depolarizing parameter $f = 1-\frac{dr}{d-1}$. The quantity

$$\frac{1}{d^2-1}\sum_{\tau\in\boldsymbol{\sigma}_q}\mathcal{E}_{\tau,\tau}^2, \tag{6.119}$$

where $\mathcal{E}_{\tau,\tau} = \langle\tau,\mathcal{E}(\tau)\rangle$, has the following upper and lower bounds in terms of the infidelity r

$$f^2 = 1-\frac{2d}{d-1}r+\frac{d^2}{(d-1)^2}r^2 \le \frac{1}{d^2-1}\sum_{\tau\in\boldsymbol{\sigma}_q}\mathcal{E}_{\tau,\tau}^2 \le 1-\frac{2d}{d-1}r+\frac{2(d+1)}{(d-1)}r^2. \tag{6.120}$$

Proof. We begin by noting that upper and lower bounds of the quantity eq. (6.119) can be found by maximizing and minimizing respectively the following optimization

$$\begin{aligned}&\max\ (\min) \quad \sum_{\tau\in\boldsymbol{\sigma}_q}\mathcal{E}_{\tau,\tau}^2 \\ &\{\mathcal{E}_{\tau\tau}\}_\tau \\ &\text{subject to} \quad \sum_{\tau\in\boldsymbol{\sigma}_q}\mathcal{E}_{\tau,\tau} = (d^2-1)f \\ &\qquad\qquad \mathcal{E}\text{ a CPTP map.}\end{aligned} \tag{6.121}$$

Here we maximize (minimize) the quantity eq. (6.119) over all possible CPTP maps which have depolarizing parameter f . Solving this optimization problem is not easy since it not clear how to express the CP condition in terms of the optimization parameters $\mathcal{E}_{\tau\tau}$. We will therefore relax this problem to an easier one which we can solve. We begin by noting that the optimization variables $\mathcal{E}_{\tau\tau}$ are invariant under the action of a Pauli channel, i.e. for all $G \in \mathsf{P}_q$ with P_q the Pauli group, we have that

$$\begin{aligned}(\mathcal{G}^\dagger\mathcal{E}\mathcal{G})_{\tau,\tau} = \langle \tau, G\mathcal{E}(G^\dagger\tau G)G^\dagger\rangle &= \langle G^\dagger\tau G, \mathcal{E}(G^\dagger\tau G)\rangle \\ &= [\mathrm{sgn}(\tau, G)]^2 \langle G^\dagger G\tau, \mathcal{E}(G^\dagger G\tau)\rangle \\ &= \langle \tau, \mathcal{E}(\tau)\rangle \\ &= \mathcal{E}_{\tau,\tau},\end{aligned} \tag{6.122}$$

for all $\tau \in \boldsymbol{\sigma}_q \cup \sigma_0$ where $\mathrm{sgn}(\tau, G)$ is defined as

$$\mathrm{sgn}(\tau, G) = \begin{cases} -1 & \text{if} \quad \{\tau, G\} = 0, \\ +1 & \text{if} \quad [G, \tau] = 0, \end{cases} \tag{6.123}$$

which, since $\tau \in \boldsymbol{\sigma}_q \cup \sigma_0$ is a normalized element of the Pauli group, is well defined because elements of the multi-qubit Pauli group can either commute ($[.,.]$) or anti-commute ($\{.,.\}$) with each other [37]. By eq. (6.122) and linearity we can now note that the optimization variables in the optimization eq. (6.121) are invariant under twirling over the Pauli group P_q, i.e.

$$\mathcal{T}_P(\mathcal{E})_{\tau,\tau} = \frac{1}{|\mathsf{P}_q|}\sum_{G\in\mathsf{P}_q} \langle G^\dagger\tau G, \mathcal{E}(G^\dagger\tau G)\rangle = \frac{1}{|\mathsf{P}_q|}\sum_{G\in\mathsf{P}_q} \mathcal{E}_{\tau,\tau} = \mathcal{E}_{\tau,\tau}. \tag{6.124}$$

Note also that the "twirl" operation, for any group, preserves complete positivity [25]. This means we can relax the optimization eq. (6.121) to

$$\begin{aligned} \max\ (\min)_{\{\mathcal{T}_{\mathsf{P}_q}(\mathcal{E})_{\tau,\tau}\}_\tau} \quad & \sum_{\tau\in\boldsymbol{\sigma}_q} \mathcal{T}_{\mathsf{P}_q}(\mathcal{E})^2_{\tau,\tau} \\ \text{subject to} \quad & \sum_{\tau\in\boldsymbol{\sigma}_q} \mathcal{T}_{\mathsf{P}_q}(\mathcal{E})_{\tau,\tau} = (d^2-1)f \\ & \mathcal{T}_{\mathsf{P}_q}(\mathcal{E}) \text{ a CPTP map.} \end{aligned} \tag{6.125}$$

Note that this is a relaxation of the previous optimization because while the twirl of a CP map will always be CP the opposite need no be true. Now we use the following result due to Holevo [38] which states that any CPTP map $\mathcal{E}$, twirled over the Pauli group, is of the form

$$\mathcal{T}_{\mathsf{P}_q}(\mathcal{E})(X) = \sum_{G\in\mathsf{P}_q} p_G GXG^\dagger \qquad \forall X \in \mathcal{M}_d, \tag{6.126}$$

where $\{p_G\}_G$ is a probability distribution, i.e. $p_G \geq 0, \forall G \in \mathsf{P}_q$ and $\sum_{G\in\mathsf{P}_q} p_G = 1$. Let us now rewrite the optimization eq. (6.125) in terms of this probability distribution. We

begin by noting that since $\mathcal{E}$ is TP we have that $\mathcal{E}_{\sigma_0\sigma_0} = 1$ and hence we can write the depolarizing constraint in eq. (6.125) as

$$\sum_{\tau\in\boldsymbol{\sigma}_q} \mathcal{T}_{\mathsf{P}_q}(\mathcal{E})_{\tau,\tau} = (d^2-1)f \iff \sum_{\tau\in\boldsymbol{\sigma}_q\cup\sigma_0} \mathcal{T}_{\mathsf{P}_q}(\mathcal{E})_{\tau,\tau} = (d^2-1)f + 1. \tag{6.127}$$

Now, using the form of the Pauli-twirled channel, we can write the RHS of this equivalence as

$$\begin{aligned}\sum_{\tau\in\boldsymbol{\sigma}_q\cup\sigma_0} \mathcal{T}_{\mathsf{P}_q}(\mathcal{E})_{\tau,\tau} &= \sum_{\tau\in\boldsymbol{\sigma}_q\cup\sigma_0}\sum_{G\in\mathsf{P}_q} p_G\langle\tau, G\tau G^\dagger\rangle \\ &= \sum_{G\in\mathsf{P}_q} p_G \sum_{\tau\in\boldsymbol{\sigma}_q\cup\sigma_0} \mathrm{sgn}(\tau, G) \\ &= p_I d^2,\end{aligned} \tag{6.128}$$

where in the last line we used that the identity Pauli element I commutes with all Pauli matrices $\tau \in \boldsymbol{\sigma}_q \cup \sigma_0$, whereas every non-identity Pauli G commutes with exactly of the elements of $\boldsymbol{\sigma}_q \cup \sigma_0$ and anti-commutes with the other half. We also used that $|\boldsymbol{\sigma}_q \cup \sigma_0| = d^2$. We can make a similar calculation for the objective of eq. (6.125) which gives

$$\begin{aligned}\sum_{\tau\in\boldsymbol{\sigma}_q} \mathcal{T}_{\mathsf{P}_q}(\mathcal{E})^2_{\tau,\tau} &= \sum_{\tau\in\boldsymbol{\sigma}_q\cup\sigma_0} \mathcal{T}_{\mathsf{P}_q}(\mathcal{E})^2_{\tau,\tau} - 1 \\ &= (-1) + \sum_{\tau\in\boldsymbol{\sigma}_q\cup\sigma_0} \left(\sum_{G\in\mathsf{P}_q} p_G\langle\tau, G\tau G^\dagger\rangle\right)^2 \\ &= (-1) + \sum_{G,\hat{G}\in\mathsf{P}_q} p_G p_{\hat{G}} \sum_{\tau\in\boldsymbol{\sigma}_q\cup\sigma_0} \mathrm{sgn}(\tau, G)\mathrm{sgn}(\tau, \hat{G}^\dagger) \\ &= (-1) + \sum_{G\in\mathsf{P}_q} p_G^2 \sum_{\tau\in\boldsymbol{\sigma}_q\cup\sigma_0} \mathrm{sgn}(\tau, GG^\dagger) \\ &\qquad + \sum_{\substack{G,\hat{G}\in\mathsf{P}_q \\ G\neq\hat{G}}} p_G p_{\hat{G}} \sum_{\tau\in\boldsymbol{\sigma}_q\cup\sigma_0} \mathrm{sgn}(\tau, G\hat{G}^\dagger) \\ &= (-1) + d^2 \sum_{G\in\mathsf{P}_q} p_G^2,\end{aligned} \tag{6.129}$$

where we have used that $\mathrm{sgn}(\tau, G)\mathrm{sgn}(\tau, \hat{G}) = \mathrm{sgn}(\tau, G\hat{G})$, that $GG^\dagger = I, \forall G \in \mathsf{P}_q$ and again that the Pauli identity I commutes with all elements of $\boldsymbol{\sigma}_q \cup \sigma_0$ while every non-identity Pauli $G\hat{G}^\dagger, G \neq G\hat{G}$ commutes with exactly half of the elements of $\boldsymbol{\sigma}_q \cup \sigma_0$ and anti-commutes with the other half. We have now rewritten the optimization eq. (6.125)

completely in terms of the probability distribution $\{p_G\}_G$. This becomes

$$\begin{aligned} \max_{\{p_G\}_G} \text{(min)} \quad & (-1) + d^2 \sum_{G \in \mathsf{P}_q} p_G^2 \\ \text{subject to} \quad & d^2 p_I = (d^2-1)f + 1 \\ & \sum_{G \in \mathsf{P}_q} p_G = 1 \\ & p_G \geq 0 \qquad G \in \mathsf{P}_q. \end{aligned} \tag{6.130}$$

Noting that the element p_I is essentially fixed we can eliminate this element from the optimization and obtain an even simpler optimization

$$\begin{aligned} \max_{\{p_G\}_G} \text{(min)} \quad & (-1) + d^2 \sum_{G \in \mathsf{P}_q/\{I\}} p_G^2 + d^2 \left(\frac{d^2-1}{d^2} f + \frac{1}{d^2} \right)^2 \\ \text{subject to} \quad & \sum_{G \in \mathsf{P}_q/\{I\}} p_G = 1 - \frac{d^2-1}{d^2} f - \frac{1}{d^2} \\ & p_G \geq 0 \qquad G \in \mathsf{P}_q/\{I\}. \end{aligned} \tag{6.131}$$

The above optimization is a well studied instance of a class of optimization problems called quadratic programs [39]. This problem has the minimum [39, Chapter 4, Section 4]:

$$p_{G,min} = \frac{1}{d^2-1} \left(1 - \frac{d^2-1}{d^2} f - \frac{1}{d^2} \right) \qquad \forall G \in \mathsf{P}_q/\{I\}, \tag{6.132}$$

and has d^2-1 degenerate maxima indexed by the non-identity Pauli elements $\tilde{G}$ of the form

$$p_{G,max} = \begin{cases} 1 - \frac{d^2-1}{d^2} f - \frac{1}{d^2} & \text{if} \quad G = \tilde{G} \\ 0 & \text{otherwise.} \end{cases} \tag{6.133}$$

This means we can lower bound the quantity eq. (6.119), for any CPTP map $\mathcal{E}$, by:

$$\frac{1}{d^2-1} \sum_{\tau \in \boldsymbol{\sigma}_q} \mathcal{E}_{\tau,\tau}^2 \geq \frac{d^2}{d^2-1} \left(\frac{d^2-1}{d^2} f + \frac{1}{d^2} \right)^2 + \frac{d^2}{(d^2-1)^2} \left(1 - \frac{d^2-1}{d^2} f - \frac{1}{d^2} \right)^2 - \frac{1}{d^2-1}.$$

By now using the relation $f = 1 - \frac{dr}{d-1}$ we can rewrite this lower bound in terms of the infidelity r. This process is straightforward but rather tedious so we will not write it down. At the end of the calculation we obtain

$$\frac{1}{d^2-1} \sum_{\tau \in \boldsymbol{\sigma}_q} \mathcal{E}_{\tau,\tau}^2 \geq 1 - \frac{2dr}{d-1} + \frac{d^2 r^2}{(d-1)^2}. \tag{6.134}$$

Similarly we can write for the upper bound

$$\frac{1}{d^2-1} \sum_{\tau \in \boldsymbol{\sigma}_q} \mathcal{E}_{\tau,\tau}^2 \leq \frac{d^2}{d^2-1} \left(\frac{d^2-1}{d^2} f + \frac{1}{d^2} \right)^2 + \frac{d^2}{d^2-1} \left(1 - \frac{d^2-1}{d^2} f - \frac{1}{d^2} \right)^2 - \frac{1}{d^2-1},$$

which, by essentially the same tedious but straightforward calculation yields

$$\frac{1}{d^2-1}\sum_{\tau\in\boldsymbol{\sigma}_q}\mathcal{E}_{\tau\tau}\leq 1-2\frac{dr}{d-1}+\frac{2(d+1)}{(d-1)}r^2, \tag{6.135}$$

which completes the lemma.

■

Eigenvalues of twirled quantum channels

Lemma 6.4 proves that the unitarity upper bounds the eigenvalues of the twirled superoperator $\mathcal{T}_{\phi_{\mathrm{TS}}}(\mathcal{E}^{\otimes 2})$. This resolves an open question posed in [2] and allows us to establish the long sequence length behavior of the variance of RB.

Lemma 6.4. Let $\mathcal{E}:\mathcal{M}_d\to\mathcal{M}_d$ be a quantum channel with unitarity u and consider the twirled operator $\mathcal{T}_{\phi_{\mathrm{TS}}}(\mathcal{E}^{\otimes 2})$ with respect to the traceless-symmetric representation. This operator can then be written as (lemma 6.1)

$$\mathcal{T}_{\phi_{\mathrm{TS}}}(\mathcal{E}^{\otimes 2})=\sum_{i\in\mathcal{Z}}\chi_i\mathcal{P}_i \tag{6.136}$$

with $\mathcal{Z}=\{\mathrm{tr},1,2,[1],[2],[3],\{1\},\{2\}\}$, $\mathcal{P}_i$ the projector onto the spaces $V_i\subset\mathcal{M}_d^{\otimes 2}$ and

$$\chi_i:=\frac{\mathrm{Tr}(\mathcal{E}\mathcal{P}_i)}{\mathrm{Tr}(\mathcal{P}_i)}, \tag{6.137}$$

where the trace is taken over superoperators. We now have for all $i\in\mathcal{Z}$ that

$$\chi_i\leq u. \tag{6.138}$$

Proof. We begin by considering $i\in\mathcal{Z}_{\mathrm{d}}$. Note first that for $i=\mathrm{tr}$ we have that

$$\chi_{\mathrm{i}}=\frac{\mathrm{Tr}(\mathcal{P}_{\mathrm{tr}}\mathcal{E}^{\otimes 2})}{\mathrm{Tr}(\mathcal{P}_{\mathrm{tr}})}=\frac{1}{d^2-1}\sum_{\tau,\tau'\in\boldsymbol{\sigma}_q}\langle\!\langle\tau^{\otimes 2}|\mathcal{E}^{\otimes 2}|\tau'^{\otimes 2}\rangle\!\rangle, \tag{6.139}$$

where we have used the definition of $\mathcal{P}_{\mathrm{tr}}$ (lemma 6.1). We can calculate

$$\begin{aligned}\frac{1}{d^2-1}\sum_{\tau,\tau'\in\boldsymbol{\sigma}_q}\langle\!\langle\tau^{\otimes 2}|\mathcal{E}^{\otimes 2}|\tau'^{\otimes 2}\rangle\!\rangle&=\frac{1}{d^2-1}\sum_{\tau,\tau'\in\boldsymbol{\sigma}_q}\langle\!\langle\tau|\mathcal{E}|\tau'\rangle\!\rangle^2\\&=\frac{1}{d^2-1}\sum_{\tau,\tau'\in\boldsymbol{\sigma}_q}\langle\!\langle\tau|\mathcal{E}|\tau'\rangle\!\rangle\langle\!\langle\tau'|\mathcal{E}^\dagger|\tau\rangle\!\rangle\\&=\frac{1}{d^2-1}\sum_{\tau,\tau'\in\boldsymbol{\sigma}_q}\langle\!\langle\tau|\mathcal{E}_u\mathcal{E}_u^\dagger|\tau\rangle\!\rangle\\&=u(\mathcal{E})\end{aligned} \tag{6.140}$$

where we have used the definition of the unitarity. Now consider $i \in \mathcal{Z}_{\mathrm{d}}$. We have

$$\begin{aligned}\chi_i &= \frac{\mathrm{Tr}(\mathcal{P}_i \mathcal{E}^{\otimes 2})}{\mathrm{Tr}(\mathcal{P}_i)} \\ &= \frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau \in \boldsymbol{\sigma}_q} \langle\!\langle \tau^{\otimes 2} | \mathcal{P}_i \mathcal{E}^{\otimes 2} | \tau^{\otimes 2} \rangle\!\rangle \\ &= \frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau, \tau' \in \boldsymbol{\sigma}_q} \langle\!\langle \tau^{\otimes 2} | \mathcal{P}_i | \tau'^{\otimes 2} \rangle\!\rangle \langle\!\langle \tau'^{\otimes 2} | \mathcal{E}^{\otimes 2} | \tau^{\otimes 2} \rangle\!\rangle\end{aligned} \tag{6.141}$$

Where we have used that the support of $\mathcal{P}_i$ lies in V_{d} (defined in lemma 6.1). Now we can use lemma 6.2 to upper bound this quantity. We have

$$\begin{aligned}\chi_i &\leq \frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau, \tau' \in \boldsymbol{\sigma}_q} \frac{\mathrm{Tr}(\mathcal{P}_i)}{d^2 - 1} \langle\!\langle \tau'^{\otimes 2} | \mathcal{E}^{\otimes 2} | \tau^{\otimes 2} \rangle\!\rangle \\ &= \frac{1}{d^2 - 1} \sum_{\tau, \tau' \in \boldsymbol{\sigma}_q} \langle\!\langle \tau' | \mathcal{E} | \tau \rangle\!\rangle \langle\!\langle \tau | \mathcal{E}^{\dagger} | \tau' \rangle\!\rangle \\ &= u\end{aligned} \tag{6.142}$$

where we have again used the definition of the unitarity.
Next we consider the case of $i \in \mathcal{Z}_{[S]}$. We have

$$\chi_i = \frac{\mathrm{Tr}(\mathcal{P}_i \mathcal{E}^{\otimes 2})}{\mathrm{Tr}(\mathcal{P}_i)} = \frac{1}{4} \frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau, \tau' \in \boldsymbol{\sigma}_q} \sum_{\substack{\sigma \in \mathbf{C}_\tau \\ \sigma' \in \mathbf{C}_{\tau'}}} \langle\!\langle S_{\sigma, \sigma \cdot \tau} | \mathcal{P}_i | S_{\sigma', \sigma' \cdot \tau'} \rangle\!\rangle \langle\!\langle S_{\sigma', \sigma' \cdot \tau'} | \mathcal{E}^{\otimes 2} | S_{\sigma, \sigma \cdot \tau} \rangle\!\rangle$$

where we have used that the support of $\mathcal{P}_i$ lies in $V_{[S]}$ (defined in lemma 6.1) and the factor of $1/4$ accounts for the fact that we are double counting the basis of $V_{[S]}$ since $S_{\sigma, \sigma \cdot \tau} = S_{\sigma \cdot \tau, \sigma}$ (we double count twice: once in the definition of the trace and once in the resolution of the identity on $V_{[S]}$). From lemma 6.2 we can lose one of the sums and get

$$\begin{aligned}\chi_i &= \frac{1}{4} \frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau, \tau' \in \boldsymbol{\sigma}_q} \sum_{\substack{\sigma \in \mathbf{C}_\tau \\ \sigma' \in \mathbf{C}_{\tau'}}} \langle\!\langle S_{\sigma, \sigma \cdot \tau} | \mathcal{P}_i | S_{\sigma', \sigma' \cdot \tau'} \rangle\!\rangle \delta_{\tau, \tau'} \langle\!\langle S_{\sigma', \sigma' \cdot \tau'} | \mathcal{E}^{\otimes 2} | S_{\sigma, \sigma \cdot \tau} \rangle\!\rangle \\ &= \frac{1}{4} \frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau \in \boldsymbol{\sigma}_q} \sum_{\sigma, \sigma' \in \mathbf{C}_\tau} \langle\!\langle S_{\sigma, \sigma \cdot \tau} | \mathcal{P}_i | S_{\sigma', \sigma' \cdot \tau} \rangle\!\rangle \langle\!\langle S_{\sigma', \sigma' \cdot \tau} | \mathcal{E}^{\otimes 2} | S_{\sigma, \sigma \cdot \tau} \rangle\!\rangle.\end{aligned} \tag{6.143}$$

We can further use lemma 6.2 to upper bound this quantity as

$$\begin{aligned}\chi_i &\leq \frac{1}{4} \frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau \in \boldsymbol{\sigma}_q} \sum_{\sigma, \sigma' \in \mathbf{C}_\tau} |\langle\!\langle S_{\sigma, \sigma \cdot \tau} | \mathcal{P}_i | S_{\sigma', \sigma' \cdot \tau} \rangle\!\rangle| |\langle\!\langle S_{\sigma', \sigma' \cdot \tau} | \mathcal{E}^{\otimes 2} | S_{\sigma, \sigma \cdot \tau} \rangle\!\rangle| \\ &\leq \frac{1}{4} \frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau \in \boldsymbol{\sigma}_q} \sum_{\sigma, \sigma' \in \mathbf{C}_\tau} \frac{2 \, \mathrm{Tr}(\mathcal{P}_i)}{(d^2 - 1)(d/2 - 2)} |\langle\!\langle S_{\sigma', \sigma' \cdot \tau} | \mathcal{E}^{\otimes 2} | S_{\sigma, \sigma \cdot \tau} \rangle\!\rangle| \\ &= \frac{1}{2} \frac{1}{(d^2 - 1)(d^2/2 - 2)} \sum_{\tau \in \boldsymbol{\sigma}_q} \sum_{\sigma, \sigma' \in \mathbf{C}_\tau} |\langle\!\langle \sigma | \mathcal{E} | \sigma' \rangle\!\rangle \langle\!\langle \sigma \cdot \tau | \mathcal{E} | \sigma' \cdot \tau \rangle\!\rangle + \langle\!\langle \sigma \cdot \tau | \mathcal{E} | \sigma' \rangle\!\rangle \langle\!\langle \sigma | \mathcal{E} | \sigma' \cdot \tau \rangle\!\rangle|\end{aligned} \tag{6.144}$$

where we have also used the triangle inequality for the absolute value. Using the triangle inequality again together with the fact that $2|ab| \leq a^2 + b^2$ for all $a, b \in \mathbb{R}$ we can write

$$\begin{aligned}\chi_i &\leq \frac{1}{2}\frac{1}{(d^2-1)(d^2/2-2)} \sum_{\tau\in\boldsymbol{\sigma}_q} \sum_{\sigma,\sigma'\in\mathbf{C}_\tau} |\mathcal{E}_{\sigma,\sigma'}\mathcal{E}_{\sigma\cdot\tau,\sigma'\cdot\tau}| + |\mathcal{E}_{\sigma\cdot\tau,\sigma'}\mathcal{E}_{\sigma,\sigma'\cdot\tau}| \\ &\leq \frac{1}{4}\frac{1}{(d^2-1)(d^2/2-2)} \sum_{\tau\in\boldsymbol{\sigma}_q} \sum_{\sigma,\sigma'\in\mathbf{C}_\tau} \mathcal{E}^2_{\sigma,\sigma'} + \mathcal{E}^2_{\sigma\cdot\tau,\sigma'\cdot\tau} + \mathcal{E}^2_{\sigma\cdot\tau,\sigma'} + \mathcal{E}^2_{\sigma,\sigma'\cdot\tau}\end{aligned} \tag{6.145}$$

Now since $\sigma \in \mathbf{C}_\tau \iff \sigma\cdot\tau \in \mathbf{C}_\tau$ we can roll the four sums in the above expression into one, that is

$$\begin{aligned}\chi_i &\leq \frac{1}{(d^2-1)(d^2/2-2)} \sum_{\tau\in\boldsymbol{\sigma}_q} \sum_{\sigma,\sigma'\in\mathbf{C}_\tau} \mathcal{E}^2_{\sigma,\sigma'} \\ &= \sum_{\sigma,\sigma'\in\boldsymbol{\sigma}_q} \sum_{\tau\in\mathbf{C}_\sigma\cap\mathbf{C}_{\sigma'}} \mathcal{E}^2_{\sigma,\sigma'} \\ &\leq \frac{1}{(d^2-1)} \sum_{\sigma,\sigma'\in\boldsymbol{\sigma}_q} \mathcal{E}^2_{\sigma,\sigma'} \\ &= u\end{aligned} \tag{6.146}$$

where we used the fact that $\sigma \in \mathbf{C}_\tau \iff \tau \in \mathbf{C}_\sigma$, the fact that $|\mathbf{C}_\sigma \cap \mathbf{C}_{\sigma'}| \leq |\mathbf{C}_\sigma| = d^2/2 - 2$ and the definition of the unitarity. This means we have $\chi_i \leq u$ for all $i \in \mathcal{Z}_{[S]}$. The argument for $i \in \mathcal{Z}_{\{S\}}$ is conceptually the same as that for $i \in \mathcal{Z}_{[S]}$ so we will not write it down. ■

Lemma 6.5 proves that the eigenvalues of the twirled superoperator $\mathcal{T}_{\phi_{\mathrm{TS}}}(\mathcal{E}^{\otimes 2})$ are close to the depolarizing parameter f. This fact is key in our analysis of the variance of RB in the presence of SPAM.

Lemma 6.5. Let $\mathcal{E} : \mathcal{M}_d \to \mathcal{M}_d$ be a quantum channel with infidelity r and depolarizing parameter $f = 1 - \frac{dr}{d-1}$ and consider the twirled operator $\mathcal{T}_{\phi_{\mathrm{TS}}}(\mathcal{E}^{\otimes 2})$ with respect to the traceless-symmetric representation. This operator can then be written as (lemma 6.1)

$$\mathcal{T}_{\phi_{\mathrm{TS}}}(\mathcal{E}^{\otimes 2}) = \sum_{i\in\mathcal{Z}} \chi_i \mathcal{P}_i \tag{6.147}$$

with $\mathcal{Z} = \{\mathrm{tr}, 1, 2, [1], [2], [3], \{1\}, \{2\}\}$, $\mathcal{P}_i$ the projector onto the spaces $V_i \subset \mathcal{M}_d^{\otimes 2}$ and

$$\chi_i := \frac{\mathrm{Tr}(\mathcal{E}\mathcal{P}_i)}{\mathrm{Tr}(\mathcal{P}_i)}, \tag{6.148}$$

where the trace is taken over superoperators. We now have for all $i \in \mathcal{Z}_{\mathrm{d}}$

$$|\chi_i - f^2| \leq \frac{2dr}{d-1}, \tag{6.149}$$

and for all $i \in \mathcal{Z}_{[S]} \cup \mathcal{Z}_{\{S\}}$

$$|\chi_i - f^2| \leq \frac{2dr}{d-1}. \tag{6.150}$$

subject to the constraint $r \leq \frac{1}{3}$

Proof. From lemma 6.4 we have that $\chi_i \leq u$ for all $i \in \mathcal{Z}$. And since $u \leq 1$ for all quantum channels [13] we certainly have that

$$\chi_i - f^2 \leq 1 - \left(1 - \frac{dr}{d-1}\right)^2 \leq \frac{2dr}{d-1}. \tag{6.151}$$

Hence we are only interested in upper bounding $f^2 - \chi_i$, and thus lower bounding χ_i for all $i \in \mathcal{Z}$. First consider $i \in \mathcal{Z}_{\mathrm{d}}$. We proceed in much the same way as lemma 6.4. We have

$$\begin{aligned}
\chi_i &= \frac{\mathrm{Tr}(\mathcal{P}_i \mathcal{E}^{\otimes 2})}{\mathrm{Tr}(\mathcal{P}_i)} \\
&= \frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau,\tau' \in \boldsymbol{\sigma}_q} \langle\!\langle \tau^{\otimes 2} | \mathcal{P}_i | \tau'^{\otimes 2} \rangle\!\rangle \langle\!\langle \tau'^{\otimes 2} | \mathcal{E} | \tau^{\otimes 2} \rangle\!\rangle \\
&= \frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau \in \boldsymbol{\sigma}_q} \langle\!\langle \tau^{\otimes 2} | \mathcal{P}_i | \tau^{\otimes 2} \rangle\!\rangle \mathcal{E}_{\tau,\tau}^2 + \frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\substack{\tau,\tau' \in \boldsymbol{\sigma}_q \\ \tau \neq \tau'}} \langle\!\langle \tau^{\otimes 2} | \mathcal{P}_i | \tau'^{\otimes 2} \rangle\!\rangle \mathcal{E}_{\tau',\tau}^2
\end{aligned} \tag{6.152}$$

We begin by considering the first term in eq. (6.152). Using lemma 6.2 we can say

$$\frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau \in \boldsymbol{\sigma}_q} \langle\!\langle \tau^{\otimes 2} | \mathcal{P}_i | \tau^{\otimes 2} \rangle\!\rangle \mathcal{E}_{\tau,\tau}^2 = \frac{\mathrm{Tr}(\mathcal{P}_i)}{(d^2-1)\,\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau \in \boldsymbol{\sigma}_q} \mathcal{E}_{\tau,\tau}^2 \geq f^2 \tag{6.153}$$

where we have also used the lower bound from lemma 6.3. Now let us consider the second term in eq. (6.152). We have

$$\begin{aligned}
\frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\substack{\tau,\tau' \in \boldsymbol{\sigma}_q \\ \tau \neq \tau'}} \langle\!\langle \tau^{\otimes 2} | \mathcal{P}_i | \tau'^{\otimes 2} \rangle\!\rangle \mathcal{E}_{\tau',\tau}^2 &\geq -\frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\substack{\tau,\tau' \in \boldsymbol{\sigma}_q \\ \tau \neq \tau'}} |\langle\!\langle \tau^{\otimes 2} | \mathcal{P}_i | \tau'^{\otimes 2} \rangle\!\rangle| \mathcal{E}_{\tau',\tau}^2 \\
&\geq -\frac{1\,\mathrm{Tr}(\mathcal{P}_i)}{(d^2-1)\,\mathrm{Tr}(\mathcal{P}_i)} \sum_{\substack{\tau,\tau' \in \boldsymbol{\sigma}_q \\ \tau \neq \tau'}} \mathcal{E}_{\tau',\tau}^2 \\
&= -\frac{1}{d^2-1} \sum_{\tau,\tau' \in \boldsymbol{\sigma}_q} \mathcal{E}_{\tau',\tau}^2 + \frac{1}{d^2-1} \sum_{\tau \in \boldsymbol{\sigma}_q} \mathcal{E}_{\tau,\tau}^2 \\
&\geq -u + f^2
\end{aligned} \tag{6.154}$$

where we have again used lemma 6.2, the lower bound from lemma 6.3 and the definition of unitarity. We can now see that for $i \in \mathcal{Z}_{\mathrm{d}}$ we have

$$f^2 - \chi_i \leq f^2 - 2f^2 + u = u - f^2 \leq 1 - \left(1 - \frac{dr}{d-1}\right)^2 \leq \frac{2dr}{d-1}. \tag{6.155}$$

Now consider $i \in \mathcal{Z}_{[S]}$ (note that we are implicitly taking $d \geq 4$ for this part of the proof, this is justified as the set $\mathcal{Z}_{[S]}$ is empty for $q = 1$). From lemma 6.4 and in particular

eq. (6.143) we get

$$\chi_i = \frac{1}{4}\frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau\in\boldsymbol{\sigma}_q} \sum_{\sigma,\sigma'\in\mathbf{C}_\tau} \langle\!\langle S_{\sigma,\sigma\cdot\tau}|\mathcal{P}_i|S_{\sigma',\sigma'\cdot\tau}\rangle\!\rangle \langle\!\langle S_{\sigma',\sigma'\cdot\tau}|\mathcal{E}^{\otimes 2}|S_{\sigma,\sigma\cdot\tau}\rangle\!\rangle. \tag{6.156}$$

We can rewrite this a little bit as follows

$$\chi_i = \frac{1}{4}\frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau\in\boldsymbol{\sigma}_q} \sum_{\sigma,\sigma'\in\mathbf{C}_\tau} \langle\!\langle S_{\sigma,\sigma\cdot\tau}|\mathcal{P}_i|S_{\sigma',\sigma'\cdot\tau}\rangle\!\rangle (\mathcal{E}_{\sigma',\sigma}\mathcal{E}_{\sigma'\cdot\tau,\sigma\cdot\tau} + \mathcal{E}_{\sigma',\sigma\cdot\tau}\mathcal{E}_{\sigma'\cdot\tau,\sigma}) \tag{6.157}$$

$$= \frac{1}{4}\frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau\in\boldsymbol{\sigma}_q} \sum_{\sigma,\sigma'\in\mathbf{C}_\tau} \langle\!\langle S_{\sigma,\sigma\cdot\tau}|\mathcal{P}_i|S_{\sigma',\sigma'\cdot\tau}\rangle\!\rangle \mathcal{E}_{\sigma',\sigma}\mathcal{E}_{\sigma'\cdot\tau,\sigma\cdot\tau} + \frac{1}{4}\frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau\in\boldsymbol{\sigma}_q} \sum_{\sigma,\sigma'\in\mathbf{C}_\tau} \langle\!\langle S_{\sigma,\sigma\cdot\tau}|\mathcal{P}_i|S_{\sigma',\sigma'\cdot\tau}\rangle\!\rangle \mathcal{E}_{\sigma',\sigma\cdot\tau}\mathcal{E}_{\sigma'\cdot\tau,\sigma} \tag{6.158}$$

$$= \frac{1}{4}\frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau\in\boldsymbol{\sigma}_q} \sum_{\sigma,\sigma'\in\mathbf{C}_\tau} \langle\!\langle S_{\sigma,\sigma\cdot\tau}|\mathcal{P}_i|S_{\sigma',\sigma'\cdot\tau}\rangle\!\rangle \mathcal{E}_{\sigma',\sigma}\mathcal{E}_{\sigma'\cdot\tau,\sigma\cdot\tau} + \frac{1}{4}\frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau\in\boldsymbol{\sigma}_q} \sum_{\sigma,\sigma'\in\mathbf{C}_\tau} \langle\!\langle S_{\sigma,\sigma\cdot\tau}|\mathcal{P}_i|S_{\sigma'\cdot\tau,(\sigma'\cdot\tau)\cdot\tau}\rangle\!\rangle \mathcal{E}_{\sigma'\cdot\tau,\sigma\cdot\tau}\mathcal{E}_{(\sigma'\cdot\tau)\cdot\tau,\sigma} \tag{6.159}$$

$$= \frac{1}{2}\frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau\in\boldsymbol{\sigma}_q} \sum_{\sigma,\sigma'\in\mathbf{C}_\tau} \langle\!\langle S_{\sigma,\sigma\cdot\tau}|\mathcal{P}_i|S_{\sigma',\sigma'\cdot\tau}\rangle\!\rangle \mathcal{E}_{\sigma',\sigma}\mathcal{E}_{\sigma'\cdot\tau,\sigma\cdot\tau} \tag{6.160}$$

where we used that $\sigma' \in \mathbf{C}_\tau \iff \sigma'\cdot\tau \in \mathbf{C}_\tau$, that $(\sigma'\cdot\tau)\cdot\tau = \sigma'$ and that $S_{\sigma',\sigma'\cdot\tau} = S_{\sigma'\cdot\tau,\sigma'}$. We can again separate off the 'diagonal' terms to get

$$\chi_i = \frac{1}{2}\frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau\in\boldsymbol{\sigma}_q} \sum_{\sigma\in\mathbf{C}_\tau} \langle\!\langle S_{\sigma,\sigma\cdot\tau}|\mathcal{P}_i|S_{\sigma,\sigma\cdot\tau}\rangle\!\rangle \mathcal{E}_{\sigma,\sigma}\mathcal{E}_{\sigma\cdot\tau,\sigma\cdot\tau} \tag{6.161a}$$

$$+ \frac{1}{2}\frac{1}{\mathrm{Tr}(\mathcal{P}_i)} \sum_{\tau\in\boldsymbol{\sigma}_q} \sum_{\substack{\sigma,\sigma'\in\mathbf{C}_\tau \\ \sigma\neq\sigma'}} \langle\!\langle S_{\sigma,\sigma\cdot\tau}|\mathcal{P}_i|S_{\sigma',\sigma'\cdot\tau}\rangle\!\rangle \mathcal{E}_{\sigma',\sigma}\mathcal{E}_{\sigma'\cdot\tau,\sigma\cdot\tau}. \tag{6.161b}$$

We will analyze the terms eq. (6.161a) and eq. (6.161b) separately. We begin with eq. (6.161a). We can use lemma 6.2 to get

$$\text{eq. (6.161a)} = \frac{1}{(d^2-1)\left(\frac{d^2}{2}-2\right)} \sum_{\tau\in\boldsymbol{\sigma}_q} \sum_{\sigma\in\mathbf{C}_\tau} \mathcal{E}_{\sigma,\sigma}\mathcal{E}_{\sigma\cdot\tau,\sigma\cdot\tau}. \tag{6.162}$$

Now we use the generic statement $2ab = a^2 + b^2 - (a-b)^2$ for all $a, b \in \mathbb{R}$ to write

$$\begin{aligned}
\text{eq. (6.161a)} &= \frac{1}{2}\frac{1}{(d^2-1)\left(\frac{d^2}{2}-1\right)}\sum_{\tau\in\boldsymbol{\sigma}_q}\sum_{\sigma\in\mathbf{C}_\tau}\mathcal{E}_{\sigma,\sigma}^2+\mathcal{E}_{\sigma\cdot\tau,\sigma\cdot\tau}^2 \\
&\qquad -\frac{1}{2}\frac{1}{(d^2-1)\left(\frac{d^2}{2}-2\right)}\sum_{\tau\in\boldsymbol{\sigma}_q}\sum_{\sigma\in\mathbf{C}_\tau}(\mathcal{E}_{\sigma,\sigma}-\mathcal{E}_{\sigma\cdot\tau,\sigma\cdot\tau})^2 && (6.163)\\
&= \frac{1}{(d^2-1)\left(\frac{d^2}{2}-2\right)}\sum_{\tau\in\boldsymbol{\sigma}_q}\sum_{\sigma\in\mathbf{C}_\tau}\mathcal{E}_{\sigma,\sigma}^2 \\
&\qquad -\frac{1}{2}\frac{1}{(d^2-1)\left(\frac{d^2}{2}-2\right)}\sum_{\tau\in\boldsymbol{\sigma}_q}\sum_{\sigma\in\mathbf{C}_\tau}(\mathcal{E}_{\sigma,\sigma}-\mathcal{E}_{\sigma\cdot\tau,\sigma\cdot\tau})^2 && (6.164)\\
&= \frac{1}{(d^2-1)\left(\frac{d^2}{2}-2\right)}\sum_{\sigma\in\boldsymbol{\sigma}_q}\sum_{\tau\in\mathbf{C}_\sigma}\mathcal{E}_{\sigma,\sigma}^2 \\
&\qquad -\frac{1}{2}\frac{1}{(d^2-1)\left(\frac{d^2}{2}-2\right)}\sum_{\tau\in\boldsymbol{\sigma}_q}\sum_{\sigma\in\mathbf{C}_\tau}(\mathcal{E}_{\sigma,\sigma}-\mathcal{E}_{\sigma\cdot\tau,\sigma\cdot\tau})^2 && (6.165)\\
&= \frac{1}{d^2-1}\sum_{\sigma\in\boldsymbol{\sigma}_q}\mathcal{E}_{\sigma,\sigma}^2-\frac{1}{2}\frac{1}{(d^2-1)\left(\frac{d^2}{2}-2\right)}\sum_{\tau\in\boldsymbol{\sigma}_q}\sum_{\sigma\in\mathbf{C}_\tau}(\mathcal{E}_{\sigma,\sigma}-\mathcal{E}_{\sigma\cdot\tau,\sigma\cdot\tau})^2 && (6.166)\\
&\geq f^2-\frac{1}{2}\frac{1}{(d^2-2)\left(\frac{d^2}{2}-1\right)}\sum_{\tau\in\boldsymbol{\sigma}_q}\sum_{\sigma\in\mathbf{C}_\tau}(\mathcal{E}_{\sigma,\sigma}-\mathcal{E}_{\sigma\cdot\tau,\sigma\cdot\tau})^2 && (6.167)
\end{aligned}$$

where we again used that $\sigma \in \mathbf{C}_\tau \iff \sigma\cdot\tau \in \mathbf{C}_\tau$ and that $\sigma \in \mathbf{C}_\tau \iff \tau \in \mathbf{C}_\sigma$ and also the lower bound from lemma 6.3. It remains to bound the second term in eq. (6.167). To do this we will maximize the quantity $(\mathcal{W}_{\nu,\nu}-\mathcal{W}_{\nu\cdot\mu,\nu\cdot\mu})^2$ for $\mu \in \boldsymbol{\sigma}_q$ and $\nu \in \mathbf{C}_\mu$ subject to the constraint that $\mathcal{W}$ is a CPTP map with depolarizing parameter f. That is, we will try to solve the maximization problem

$$\begin{aligned}
&\max && (\mathcal{W}_{\nu,\nu}-\mathcal{W}_{\mu,\mu})^2 \\
&\text{subject to} && \sum_{\tau\in\boldsymbol{\sigma}_q}\mathcal{W}_{\tau\tau}=(d^2-1)f \\
& && \mathcal{W}\text{ a CPTP map.}
\end{aligned} \tag{6.168}$$

As in lemma 6.3 we can restrict ourselves to $\mathcal{W}$ being a Pauli channel (since the optimization function is a function of only the diagonal elements of $\mathcal{W}$). That is we can consider $\mathcal{W}(X) = \sum_{G\in\mathsf{P}_q} p_G GXG^\dagger$ where $\{p_G\}_G$ is a probability distribution over the Pauli

group. We can write the optimization objective as

$$\begin{aligned}(\mathcal{W}_{\nu,\nu} - \mathcal{W}_{\nu\cdot\mu,\nu\cdot\mu})^2 &= \left[\sum_{G\in\mathsf{P}_q} p_G \langle \nu, G\nu G^\dagger\rangle - \langle \nu\cdot\mu, G\nu\cdot\mu G^\dagger\rangle\right]^2 \\ &= \left[\sum_{G\in\mathsf{P}_q} p_G \langle \nu, G\nu G^\dagger\rangle - \langle \nu\cdot\mu, (G\nu G^\dagger)\cdot(G\mu G^\dagger)\rangle\right]^2 \\ &= \left[\sum_{G\in\mathsf{P}_q} p_G \mathrm{sgn}(\nu, G)\big(1-\mathrm{sgn}(\mu,G)\big)\right]^2\end{aligned} \tag{6.169}$$

where the $\mathrm{sgn}(\nu, G)$ (as defined in eq. (6.123)) encodes the commutation relations of the elements of the Pauli group. Note that the above quantity does not depended on $p_{\mathbb{1}}$ (the weight associated with the Pauli identity) since $\mathrm{sgn}(\mu, \mathbb{1}) = 1$ for all $\mu \in \boldsymbol{\sigma}_q$. Hence we can solve the optimization problem

$$\begin{aligned}\text{max} \quad & \left[\sum_{G\in\mathsf{P}_q/\{\mathbb{1}\}} p_G \mathrm{sgn}(\nu, G)\big(1-\mathrm{sgn}(\mu,G)\big)\right]^2 \\ \text{subject to} \quad & \sum_{G\in\mathsf{P}_q/\{\mathbb{1}\}} p_G = 1 - \frac{d^2-1}{d^2} f^2 - \frac{1}{d^2} \\ & p_G \geq 0 \quad \forall G \in \mathsf{P}_q.\end{aligned} \tag{6.170}$$

This problem has an easily spotted maximum in that we want to put all probability weight on a single $G \in C_\nu \cap N_\mu$ and set all other p_G to zero (subject to the constraint that the overall channel must have depolarizing parameter f, which is encoded in the first constraint of eq. (6.170)). Hence we have

$$\left[\sum_{G\in\mathsf{P}_q} p_G \mathrm{sgn}(\nu, G)\big(1-\mathrm{sgn}(\mu,G)\big)\right]^2 \leq \left[\frac{d^2-1}{d^2}(1-f^2)\right]^2. \tag{6.171}$$

We can feed this back into eq. (6.167) to obtain

$$\begin{aligned}\text{eq. (6.161a)} &\geq f^2 - \frac{1}{2}\frac{1}{(d^2-1)\left(\frac{d^2}{2}-2\right)} \sum_{\tau\in\boldsymbol{\sigma}_q}\sum_{\sigma\in\mathbf{C}_\tau} \left[\frac{d^2-1}{d^2}(1-f^2)\right]^2 \\ &= f^2 - \frac{1}{2}\left[\frac{d^2-1}{d^2}(1-f^2)\right]^2.\end{aligned} \tag{6.172}$$

This is a suitable lower bound on eq. (6.161a). Next we consider eq. (6.161b). We have

$$\begin{aligned}
\text{eq. (6.161b)} &\geq -\frac{1}{(d^2-1)\left(\frac{d^2}{2}-2\right)} \sum_{\tau\in\boldsymbol{\sigma}_q} \sum_{\substack{\sigma,\sigma'\in\mathbf{C}_\tau \\ \sigma\neq\sigma'}} |\mathcal{E}_{\sigma,\sigma'}\mathcal{E}_{\sigma\cdot\tau,\sigma'\cdot\tau}| \\
&\geq -\frac{1}{(d^2-1)\left(\frac{d^2}{2}-2\right)} \sum_{\tau\in\boldsymbol{\sigma}_q} \sum_{\substack{\sigma,\sigma'\in\mathbf{C}_\tau \\ \sigma\neq\sigma'}} \frac{1}{2}(\mathcal{E}^2_{\sigma,\sigma'} + \mathcal{E}^2_{\sigma\cdot\tau,\sigma'\cdot\tau}) \\
&= -\frac{1}{(d^2-1)\left(\frac{d^2}{2}-2\right)} \sum_{\substack{\sigma,\sigma'\in\boldsymbol{\sigma}_q \\ \sigma\neq\sigma'}} \sum_{\tau\in\mathbf{C}_\sigma\cap\mathbf{C}_{\sigma'}} \mathcal{E}^2_{\sigma,\sigma'} \qquad (6.173) \\
&= -\frac{\frac{d^4}{4}-3}{(d^2-1)\left(\frac{d^2}{2}-2\right)} \left[\sum_{\sigma,\sigma'\in\boldsymbol{\sigma}_q} \mathcal{E}^2_{\sigma,\sigma'} - \sum_{\sigma\in\boldsymbol{\sigma}_q} \mathcal{E}^2_{\sigma,\sigma'}\right] \\
&\geq -\frac{\frac{d^4}{4}-3}{\frac{d^2}{2}-2}(u-f^2)
\end{aligned}$$

where we used an array of steps that have been used before: the triangle inequality and lemma 6.2 for the first inequality, the fact that $2|ab| \leq a^2+b^2$ for all $a,b\in\mathbb{R}$ for the second inequality, the fact that $\sigma\in\mathbf{C}_\tau \iff \tau\in\mathbf{C}_\sigma$ for the third equality, the fact that $|\mathbf{C}_\sigma\cap\mathbf{C}_\sigma| = d^2/4-3$ for $\sigma\neq\sigma'$ [35] for the fourth equality and lemma 6.3 and the definition of unitarity for the last equality. This is a good lower bound on eq. (6.161b). We can now combine the lower bounds on eq. (6.161a) and eq. (6.161b) to get

$$\chi_i \geq f^2 - \frac{1}{2}\left[\frac{d^2-1}{d^2}(1-f^2)\right]^2 - \frac{\frac{d^4}{4}-3}{\left(\frac{d^2}{2}-2\right)}(u-f^2) \qquad (6.174)$$

for $i\in\mathcal{Z}_{[S]}$. This gives a final bound (using $u\leq 1$)

$$f^2-\chi_i \leq f^2 - f^2 + \frac{1}{2}\left[\frac{d^2-1}{d^2}(1-f^2)\right]^2 + \frac{\frac{d^4}{4}-3}{\frac{d^2}{2}-2}(1-f^2) \qquad (6.175)$$

which we can rewrite to yield

$$f^2-\chi_i \leq \frac{2dr}{d-1}\left(\frac{\frac{d^4}{4}-3}{\frac{d^2}{2}-2}\left(1-\frac{1}{2}\frac{dr}{d-1}\right) + \frac{1}{2}\frac{(d^2-1)^2}{d^4}\frac{2dr}{d-1}\left(1-\frac{1}{2}\frac{dr}{d-1}\right)^2\right) \qquad (6.176)$$

Setting $\left(1-\frac{1}{2}\frac{dr}{d-1}\right)\leq 1$ and working out we get

$$f^2-\chi \leq \frac{2d}{d-1}r \qquad (6.177)$$

for

$$r \leq \left(1-\frac{\frac{d^4}{4}-3}{\frac{d^2}{2}-2}\right)\frac{d^3(d-1)}{(d^2-1)^2}. \qquad (6.178)$$

This completes the proof for $i \in \mathcal{Z}_{[S]}$. The proof for $i \in \mathcal{Z}_{\{S\}}$ is conceptually the same as that of $i \in \mathcal{Z}_{[S]}$ and yields the same bound so we will not write it down here. The only notable difference is the difference in size for the sets $\mathbf{N}_\tau$ and $\mathbf{N}_\tau \cap \mathbf{N}_{\tau'}$ for $\tau, \tau' \in \boldsymbol{\sigma}_q$ which gives a different area of validity for the bound, namely

$$r \leq \frac{1}{3} \leq \left(1 - \frac{\frac{d^4}{4}}{\frac{d^2}{2}}\right) \frac{d^3(d-1)}{(d^2-1)^2}. \tag{6.179}$$

Choosing $r \leq 1/3$ satisfies both constraints for all d and thus completes the proof. ■

TELESCOPING SERIES

Lemma 6.6 and corollary 6.1 provide us with a powerful tool to break up the analysis of the variance of RB into manageable pieces.

Lemma 6.6. For two arbitrary ordered lists of m elements $\{a_1, \ldots, a_m\}$ and $\{b_1, \ldots, b_m\}$ of an algebra with associative and distributed addition and multiplication we have,

$$a_{m:1} - b_{m:1} = \sum_{j=1}^{m} a_{m:j+1}(a_j - b_j)b_{j-1:1}. \tag{6.180}$$

where $a_{j:k}$ with $j \geq k$ is defined with respect to the list $\{a_1, \ldots, a_m\}$ as

$$a_{j:k} = a_j a_{j+1} \cdots a_{k-1} a_k. \tag{6.181}$$

Proof. We will prove this by induction. For $m = 1$ the statement is trivial. For $m + 1$, we have

$$\begin{aligned} a_{m+1:1} - b_{m+1:1} &= a_{m+1}a_{m:1} - a_{m+1}b_{m:1} + a_{m+1}b_{m:1} - b_{m+1}b_{m:1} \\ &= a_{m+1}(a_{m:1} - b_{m:1}) + (a_{m+1} - b_{m+1})b_{m:1} \\ &= \sum_{j=1}^{m+1} a_{m:j+1}(a_j - b_j)b_{j-1:1} \end{aligned}$$

by induction hypothesis. This proves the lemma. ■

Corollary 6.1. For $a, b, c \in \mathbb{C}$ with $c \geq a$, we have

$$\begin{aligned} a^m - b^m &= mb^{m-1}(a-b) + (a-b)^2 a^{m-2} \frac{(m-1)(b/a)^m - m(b/a)^{m-1} + 1}{(1-(b/a))^2} \\ &\leq mb^{m-1}(a-b) + (a-b)^2 \frac{(m-1)b^m - mcb^{m-1} + c^m}{(c-b)^2} \end{aligned}$$

Proof. Note first that the statement is trivial if $a = b$. Therefore assume $a \neq b$. We begin by applying lemma 6.6 to $a^m - b^m$. This gives

$$a^m - b^m = \sum_{j=1}^{m} a^{m-j}(a-b)b^{j-1}. \tag{6.182}$$

We now perform the following manipulation

$$\begin{aligned} a^m - b^m &= \sum_{j=1}^{m} a^{m-j}(a-b)b^{j-1} \\ &= \sum_{j=1}^{m}(a^{m-j} - b^{m-j} + b^{m-j})(a-b)b^{j-1} \\ &= (a-b)\sum_{j=1}^{m} b^{m-j+j-1} + \sum_{j=1}^{m}(a^{m-j} - b^{m-j})(a-b)b^{j-1} \\ &= mb^{m-1}(a-b) + \sum_{j=1}^{m}(a^{m-j} - b^{m-j})(a-b)b^{j-1}. \end{aligned} \tag{6.183}$$

Note that be have used the fact that $a, b \in \mathbb{C}$ are commutative. Now we can apply lemma 6.6 again to the factors $(a^{m-j} - b^{m-j})$ in the second term in the above to obtain

$$\begin{aligned} a^m - b^m &= mb^{m-1}(a-b) + \sum_{j=1}^{m}\sum_{t=1}^{m-j} a^{m-j-t}(a-b)b^{j-t-1}(a-b)b^{j-1} \\ &= mb^{m-1}(a-b) + (a-b)^2\sum_{j=1}^{m}\sum_{t=1}^{m-j} a^{m-(j+t)}b^{j+t-2}. \end{aligned} \tag{6.184}$$

Performing the substitution $s = j + t$ and working out we obtain

$$\begin{aligned} a^m + b^m &= mb^{m-1}(a-b) + (a-b)^2\sum_{j=1}^{m}\sum_{t=1}^{m-j} a^{m-(j+t)}b^{j+t-2} \\ &= mb^{m-1}(a-b) + (a-b)^2\sum_{j=1}^{m}\sum_{s=j+1}^{m} a^{m-s}b^{s-2} \\ &= mb^{m-1}(a-b) + (a-b)^2\sum_{s=2}^{m}\sum_{j=1}^{s-1} a^{m-s}b^{s-2} \\ &= mb^{m-1}(a-b) + (a-b)^2\sum_{s=2}^{m}(s-1)a^{m-s}b^{s-2} \end{aligned} \tag{6.185}$$

Now we can factor out a^{m-1} from the second term to obtain

$$a^m + b^m = mb^{m-1}(a-b) + (a-b)^2 a^{m-2}\sum_{s=2}^{m}(s-1)(b/a)^{s-2}. \tag{6.186}$$

We can further rewrite this using the standard series identity

$$\sum_{k=1}^{m}(k-2)x^{k-2} = \frac{(m-1)x^m - mx^{m-1} + 1}{(1-x)^2}. \tag{6.187}$$

The upper bound follows by upper bounding each term in the sum. ■

7

The statistics of unitarity randomized benchmarking

Unitarity randomized benchmarking (URB) is an experimental procedure for estimating the coherence of implemented quantum gates independently of state preparation and measurement errors. These estimates of the coherence are measured by the unitarity. A central problem in this experiment is relating the number of data points to rigorous confidence intervals. In this work we provide a bound on the required number of data points for Clifford URB as a function of confidence and experimental parameters. This bound has favorable scaling in the regime of near-unitary noise and is asymptotically independent of the length of the gate sequences used. We also show that, in contrast to standard randomized benchmarking, a nontrivial number of data points is always required to overcome the randomness introduced by state preparation and measurement errors even in the limit of perfect gates. Our bound is sufficiently sharp to benchmark small-dimensional systems in realistic parameter regimes using a modest number of data points. For example, we show that the unitarity of single-qubit Clifford gates can be rigorously estimated using few hundred data points under the assumption of gate-independent noise. This is a reduction of orders of magnitude compared to previously known bounds.

This chapter is published, with minor changes, in B. Dirkse, J. Helsen & S. Wehner, *Efficient Unitarity Randomized Benchmarking of Few-qubit Clifford Gates*, Phys. Rev. A. A 99 (1), 012315

7.1. Introduction

In this chapter we will analyze the statistics of the unitarity randomized benchmarking protocol. The aim of this work is to contribute a solution to the following central question: How many random sequences of gates are required in the URB protocol to get a confident estimate of the unitarity from the obtained measurement data? We will proceed in a manner similar to chapter 6, by providing a sharp bound on the variance of the underlying distribution from which the URB protocol samples. This additional knowledge of the URB sampling distribution allows for more resource-efficient estimation of the unitarity from experimental data. Concretely we demonstrate how our variance bound can be used to bound the required number of random sequences as a function of desired confidence parameters.

In this chapter, we derive a bound on the variance of the distribution induced by the random sampling of gate sequences in a modified version of the Clifford URB protocol. This modification is based on the adapted RB protocol we presented in chapter 6. It requires no experimental overhead while leading to a sharper variance bound (and hence fewer required gate sequences) as well as a simpler fit model for extracting the unitarity. In addition, our statistical analysis reveals the optimal input state and output measurement for minimizing the variance and maximizing the signal strength. We then apply this variance bound using standard concentration inequalities to relate the number of random sequences to desired confidence intervals. Our result is sufficiently sharp to perform the modified URB protocol on few-qubit systems with a modest number of sequences in realistic parameter regimes. It is an improvement of several orders of magnitude in the number of sequences required for fixed confidence, compared to a concentration inequality that does not use the variance (as was first done for RB in [1]). We show that the variance, and thus number of required gate sequences, scales favorably in the regime of large unitarity, which is the relevant regime for high quality gates. We also show that, in contrast to standard RB, a non-trivial number of sequences is always required to overcome the randomness introduced by state preparation and measurement errors even in the limit of perfect gates.

This chapter is organized as follows. In the remainder of this section we introduce a modification of the protocol based on the modification made in chapter 7 for the purpose of improved statistics. Furthermore we explicitly distinguish the two different implementations of the URB protocol and emphasize their benefits and drawbacks. In section 7.2 we present our main result (eq. (7.16) and eq. (7.17)) and illustrate how to apply it using a simulated example. In section 7.3 we examine the behavior of our bound in various parameter regimes and discuss the different features of our bound. A brief overview of the proof techniques used to derive our main result is presented in section 7.4. We will however delegate most of the actual technical work to section 7.6 In section 7.5 we summarize the main conclusions of our work and provide suggestions for future research.

7.1.1. The URB protocol

This section gives an overview of the URB protocol of [2] and gives a small modification similar to modification of the RB protocol given in chapter 6. The protocol is described for any gate set G that is a unitary 2-design [3]. Note that even though the protocol works for all these gate sets, our result of the confidence analysis is only applicable to the Clifford

Algorithm 2 Outline of the modified unitarity randomized benchmarking protocol.

data: Let $\mathsf{G} \subset U(d)$ be a finite subset of the unitary group that is also a unitary 2-design on a d-dimensional Hilbert space $\mathcal{H}$. Let the noise model be $\tilde{\mathcal{G}} = \tilde{\mathcal{G}}\mathcal{E}$ for all $\mathcal{G} \in \mathsf{G}$, where $\mathcal{E}$ is a CPTP map.
input: Choose $\mathbb{M} \subset \mathbb{N}$ and a $N_m \in \mathbb{N}$ for each $m \in \mathbb{M}$. Pick states and an observable $\rho, \hat{\rho}, E \in \mathcal{M}_d$ for the two-copy implementation or pick states and an observable $\rho_{\mathcal{H}}, \hat{\rho}_{\mathcal{H}}, E_{\mathcal{H}} \in \mathcal{M}_d$ for the single-copy implementation.
output: An estimate of the unitarity of the noise map $u(\mathcal{E})$.
Fix a gate set G, choose a set of sequence lengths $\mathbb{M}$ to use and determine the number of random sequences N_m per sequence length $m \in \mathbb{M}$.

procedure URB($\mathsf{G}, \mathbb{M}, \{N_m\}$)
  for all sequence lengths $m \in \mathbb{M}$ **do**
    repeat N_m times
      Sample m random gates $\mathcal{G}_{j_1}, ..., \mathcal{G}_{j_m}$ independently and uniformly at random from G;
      Compose the sequence $\mathcal{G}_{\mathbf{j}} = \mathcal{G}_{j_m} \cdots \mathcal{G}_{j_2}\mathcal{G}_{j_1}$;
      if Two-copy implementation **then**
        Prepare states $\rho \approx \frac{\mathbb{1}+S}{d(d+1)}$ and $\hat{\rho} \approx \frac{\mathbb{1}-S}{d(d-1)}$, apply $\mathcal{G}_{\mathbf{j}}^{\otimes 2}$ to each state and measure $E \approx S$ a large number of times (where S denotes the Swap gate);
        From this data, estimate the average sequence purity as
$$q_{\mathbf{j}}^{(2)} = (\mathrm{Tr}[E\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho)] - \mathrm{Tr}[E\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\hat{\rho})]) = \mathrm{Tr}[E\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho})];$$
      end if
      if Single-copy implementation **then**
        for all non-identity Pauli's $P, Q \neq \mathbb{1}$ **do**
          Prepare states $\rho_{\mathcal{H}}^{(P)} \approx \frac{\mathbb{1}+P}{d}$ and $\hat{\rho}_{\mathcal{H}}^{(P)} \approx \frac{\mathbb{1}-P}{d}$, apply $\mathcal{G}_{\mathbf{j}}$ to each state and measure $E_{\mathcal{H}}^{(Q)} \approx Q$ a large number of times;
        end for
        From this data, estimate the average sequence purity as
$$q_{\mathbf{j}}^{(1)} = \frac{1}{d^2-1} \sum_{P,Q\neq\mathbb{1}} \left(\mathrm{Tr}[E_{\mathcal{H}}^{(Q)}\mathcal{G}_{\mathbf{j}}(\rho_{\mathcal{H}}^{(P)})] - \mathrm{Tr}[E_{\mathcal{H}}^{(Q)}\mathcal{G}_{\mathbf{j}}(\hat{\rho}_{\mathcal{H}}^{(P)})]\right)^2;$$
      end if
    end repeat
    Compute the empirical average over the sampled sequences $\bar{q}_m = \frac{1}{N_m}\sum_{\mathbf{j}} q_{\mathbf{j}}$;
  end for
  Fit $\bar{q}_m = Bu^{m-1}$, where B is a constant absorbing SPAM errors and u is the unitarity of the noise map.
end procedure

group. In algorithm 2 we present an outline of the URB protocol, where we distinguish two different implementations (discussed later in this section). The URB protocol works similar to the standard RB protocol. First one draws a uniformly distributed random sequence of gates (with length m) from the gate set G. Denote such a sequence

$$\mathcal{G}_{\mathbf{j}} = \mathcal{G}_{j_m} \cdots \mathcal{G}_{j_2} \mathcal{G}_{j_1}, \tag{7.1}$$

where each j_s denotes the randomly drawn gate from G at position s. The subscript $\mathbf{j}$ denotes the multi-index $(j_1, j_2, ..., j_m)$ and therefore indexes the entire sequence. Such a randomly sampled sequence $\mathcal{G}_{\mathbf{j}}$ is then applied to a state ρ , after which a two-outcome measurement is performed (in this work the operator E denotes the hermitian observable associated with a two-outcome measurement $\{M, I - M\}$ with outcomes ± 1). However, there are two differences here with respect to the RB protocol. First, there is no global inverse applied at the end of each sequence and second, the expectation value of the measurement outcome is squared. So the URB random variable of interest then becomes $q_{\mathbf{j}} = \mathrm{Tr}[E\mathcal{G}_{\mathbf{j}}(\rho)]^2$. Throughout this work, we shall call the URB random variable $q_{\mathbf{j}}$ the sequence purity (in standard RB, the random variable of interest is typically referred to as the survival probability). The rest of the procedure is then similar: estimate the mean of the sequence purity $q_{\mathbf{j}}$ using N random sequences of fixed length, repeat for various sequence lengths and fit to the model

$$\mathbb{E}[q_{\mathbf{j}}] = Bu^{m-1} + A \tag{7.2}$$

to obtain the unitarity. Here we analyze a slightly modified version of the protocol of [2], based on ideas of [4–6]. This protocol is outlined in algorithm 2.

Every sequence of randomly sampled gates $\mathcal{G}_{\mathbf{j}}$ is applied to two different input states ρ and $\hat{\rho}$, and half of the difference of their expectation values is taken before squaring. By linearity of quantum mechanics, this is equivalent to performing URB with the traceless input operator

$$\bar{\rho} := \frac{1}{2}(\rho - \hat{\rho}). \tag{7.3}$$

The factor $\frac{1}{2}$ is strictly not necessary but is added for better statistical comparison. The key idea behind this is that one effectively works with a traceless input operator $\bar{\rho}$. There are two main benefits of this modification. First, it improves the fitting procedure, because the modified fit model for the mean of the sequence purity becomes (see eq. (7.51) in section 7.4.2)

$$\mathbb{E}[q_{\mathbf{j}}] = Bu^{m-1}, \tag{7.4}$$

where the constant B only depends on the input operator $\bar{\rho}$ and the measurement observable E. This is a linear fitting problem in u by taking the logarithm and can therefore be performed more easily. Second, this modification narrows the distribution of the sequence purity $q_{\mathbf{j}}$, improving the confidence in our point estimate $\bar{q}_m = \frac{1}{N}\sum q_{\mathbf{j}}$ of the exact $\mathbb{E}[q_{\mathbf{j}}]$. In the next section we discuss the implementation of the protocol in more detail and emphasize that there are two possible methods to estimate $q_{\mathbf{j}}$.

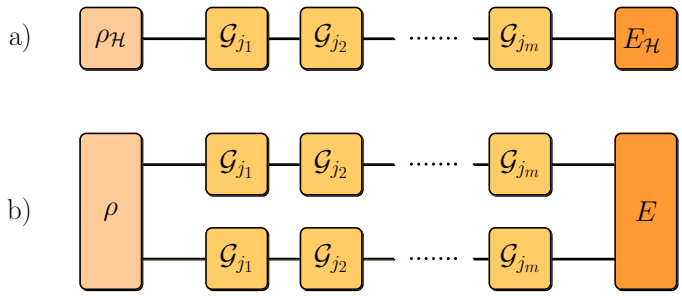


Figure 7.1: Schematic difference between the single-copy implementation (a) and the two-copy implementation (b) of the unitarity randomized benchmarking protocol. Each line represents a system on the base Hilbert space $\mathcal{H}$. In the single-copy implementation, the expected value of the measurement $\mathrm{Tr}[E_{\mathcal{H}}\mathcal{G}_{\mathbf{j}}(\bar{\rho}_{\mathcal{H}})]$ needs to be squared to obtain $q_{\mathbf{j}}$, whereas in the two-copy implementation $q_{\mathbf{j}} = \mathrm{Tr}[E\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho})]$ yields the direct outcome.

The two different implementations

In this section we discuss two different possible implementations of the URB protocol (as briefly discussed in [2] and already alluded to in chapter 4), which are illustrated in fig. 7.1. The choice of implementation depends on whether the experimenter has access to two identical copies of the system or not. The implementations differ in the way the sequence purity $q_{\mathbf{j}}$ is computed and what the ideal input operator $\bar{\rho}$ and measurement E are. By ideal operators, we mean the operators that maximize the signal strength (the proportionality factor B in the fit model eq. (7.4)) from which the unitarity is estimated. We will then show that the two implementations are closely related.

Let us start by discussing the two-copy implementation (fig. 7.1.b). As the name suggests, this requires two copies of the system $\mathcal{H}$ under investigation. The use of two copies follows from the mathematical equivalence

$$q_{\mathbf{j}} = \mathrm{Tr}[E\mathcal{G}_{\mathbf{j}}(\bar{\rho})]^2 = \mathrm{Tr}[E^{\otimes 2}\mathcal{G}_{\mathbf{j}}^{\otimes 2}\left(\bar{\rho}^{\otimes 2}\right)]. \tag{7.5}$$

If the experimenter has access to two identical copies of the system $\mathcal{H}$, the input and measurement operator can be entangled across the two copies of the system. The sequence $\mathcal{G}_{\mathbf{j}}$ is then applied to each half of the system $\mathcal{H} \otimes \mathcal{H}$. This yields the sequence purity of the two-copy implementation as

$$q_{\mathbf{j}}^{(2)} = \mathrm{Tr}[E\mathcal{G}_{\mathbf{j}}^{\otimes 2}\left(\bar{\rho}\right)], \tag{7.6}$$

where $\bar{\rho}, E \in \mathcal{M}_d$ are now operators on the two copies of the system. Since E is a two-valued measurement with outcomes (± 1) and $\bar{\rho}$ is half the difference between two physical states, it is not hard to show that the sequence purity lies in the interval $q_{\mathbf{j}}^{(2)} \in [-1, 1]$. In section 7.2.3 we show that this interval can be narrowed under mild assumptions. In the two-copy implementation it is implicitly assumed that the experimenter can operate identically on each subsystem without any cross-talk between the two subsystems. Moreover, the experimenter should be able to prepare and measure over the two copies of the system. Experimentally the input and measurement operators $\bar{\rho}, E \in \mathcal{M}_d$ should be as close to the ideal operators as possible. The ideal operators are given by (see [REF] for

more details and proof)

$$\rho_{\mathrm{id}} = \frac{\mathbb{1} + S}{d(d+1)}, \qquad \hat{\rho}_{\mathrm{id}} = \frac{\mathbb{1} - S}{d(d-1)}, \qquad E_{\mathrm{id}} = S, \tag{7.7}$$

where $\mathbb{1}$ is the identity and S is the Swap operator on $\mathcal{H} \otimes \mathcal{H}$, and d is the dimension of $\mathcal{H}$. The state ρ_{id} ($\hat{\rho}_{\mathrm{id}}$) is the maximally mixed state on the symmetric (anti-symmetric) subspace of $\mathcal{H} \otimes \mathcal{H}$. Note that the maximally mixed state on a subspace can be prepared by uniformly sampling pure states from an orthonormal basis of this subspace. The operator E_{id} is the hermitian observable associated with a two-valued measurement that discriminates between symmetric (outcome 1) and anti-symmetric states (outcome -1).
In the single-copy implementation, the experimenter must obtain an estimate of the sequence purity $q_{\mathbf{j}}$ using only a single copy of the system $\mathcal{H}$. From eq. (7.5), it can be seen that $q_{\mathbf{j}} = \mathrm{Tr}[E_{\mathcal{H}} \mathcal{G}_{\mathbf{j}}(\bar{\rho}_{\mathcal{H}})]^2$ is the sequence purity given the operators $\bar{\rho}_{\mathcal{H}}, E_{\mathcal{H}} \in \mathcal{M}_d$. Here the subscript $\mathcal{H}$ is to emphasize that the operators are on a single copy of $\mathcal{H}$. Throughout this chapter we will just write $\bar{\rho}$ and E for operators on $\mathcal{H} \otimes \mathcal{H}$ and indicate operators on a single copy explicitly by adding a subscript $\mathcal{H}$. There are two disadvantages in defining the single-copy sequence purity using one pair of input and measurement operators $\bar{\rho}_{\mathcal{H}}, E_{\mathcal{H}} \in \mathcal{M}_d$. First, the proportionality factor B in eq. (7.4) is upper bounded by $\frac{1}{d^2-1}$, where d is the dimension of $\mathcal{H}$ [2]. This means that the signal strength decreases exponentially with the system size. Second, the variance of the sequence purity is large. This leads to large uncertainty in the estimated average sequence purity $\bar{q}_m$. These disadvantages can be resolved by using multiple different pairs of input and measurement operators [2]. The ideal set of operators is chosen in such a way that summing the expectation values squared for each pair of operators leads to effectively simulating the ideal operators of eq. (7.7). Let us make this more precise. Define the single-copy sequence purity as

$$q_{\mathbf{j}}^{(1)} = \frac{1}{d^2 - 1} \sum_{P,Q \neq \mathbb{1}} \mathrm{Tr}[E_{\mathcal{H}}^{(Q)} \mathcal{G}_{\mathbf{j}}(\bar{\rho}_{\mathcal{H}}^{(P)})]^2, \tag{7.8}$$

where the sum is over all non-identity multi-qubit Pauli operators P, Q. Each $\bar{\rho}_{\mathcal{H}}^{(P)}$ and $E_{\mathcal{H}}^{(Q)}$ are different input and measurement operator settings indexed by the non-identity Pauli operators P and Q respectively. For each pair P, Q, the expectation value $\mathrm{Tr}[E_{\mathcal{H}}^{(Q)} \mathcal{G}_{\mathbf{j}}(\bar{\rho}_{\mathcal{H}}^{(P)})]$ is to be estimated experimentally. This expectation can be shown to lie in the interval $[-1, 1]$ by definition of E and $\bar{\rho}$, so that the expectation value squared lies in the unit interval. Therefore the single-copy sequence purity can in principle lie anywhere in the interval $q_{\mathbf{j}}^{(1)} \in [0, d^2 - 1]$, since each summand lies in the unit interval and the summation runs over $(d^2 - 1)^2$ terms. However in section 7.2.3 we show that this interval can be narrowed significantly under mild assumptions. Since the sum runs twice over all non-identity Pauli operators, estimating the sequence purity $q_{\mathbf{j}}^{(1)}$ requires $(d^2 - 1)^2$ different settings. This is a number that grows exponentially in the number of qubits comprising the system. We also emphasize that simply squaring and summing up estimates of $\mathrm{Tr}[E_{\mathcal{H}}^{(Q)} \mathcal{G}_{\mathbf{j}}(\bar{\rho}_{\mathcal{H}}^{(P)})]$ to obtain an estimate of $q_{\mathbf{j}}^{(1)}$ yields a positively biased estimator for $q_{\mathbf{j}}^{(1)}$. This may lead to overestimating the unitarity. See section 7.4.1 for more details on how to correctly estimate $q_{\mathbf{j}}^{(1)}$. The states $\rho_{\mathcal{H}}^{(P)}, \hat{\rho}_{\mathcal{H}}^{(P)}$ and measurement $E_{\mathcal{H}}^{(Q)}$ should be

implemented as closely as possible to the ideal operators

$$\rho^{(P)}_{\mathcal{H},\mathrm{id}} = \frac{\mathbb{1}+P}{d}, \qquad \hat{\rho}^{(P)}_{\mathcal{H},\mathrm{id}} = \frac{\mathbb{1}-P}{d}, \qquad E^{(Q)}_{\mathcal{H},\mathrm{id}} = Q. \tag{7.9}$$

The ideal state $\rho^{(P)}_{\mathcal{H},\mathrm{id}}$ ($\hat{\rho}^{(P)}_{\mathcal{H},\mathrm{id}}$) is the maximally mixed state on the positive (negative) eigenspace of the Pauli operator P, and the measurement $E^{(Q)}_{\mathcal{H},\mathrm{id}}$ is the two-valued measurement that discriminates between the positive (outcome 1) and negative (outcome -1) eigenspace of the Pauli operator Q.

Next we show that the single-copy can be interpreted as a special case of the two-copy implementation (this is not surprising in view of eq. (7.5)). To do so, we show that in the single-copy implementation, one effectively works with two-copy operators of the form

$$\begin{aligned} \bar{\rho}_{\mathrm{eff}} &= \frac{d}{d^2-1} \sum_{P\neq\mathbb{1}} \bar{\rho}^{(P)}_{\mathcal{H}} \otimes \bar{\rho}^{(P)}_{\mathcal{H}}, \\ \bar{E}_{\mathrm{eff}} &= \frac{1}{d} \sum_{Q\neq\mathbb{1}} \bar{E}^{(Q)}_{\mathcal{H}} \otimes \bar{E}^{(Q)}_{\mathcal{H}}. \end{aligned} \tag{7.10}$$

Here $\bar{E}$ ($\bar{E}^{(Q)}_{\mathcal{H}}$) is the traceless part of the observable E ($E^{(Q)}_{\mathcal{H}}$) , defined as

$$\bar{E} := E - \mathrm{Tr}[E]\frac{\mathbb{1}}{d^2}, \qquad \bar{E} := E_{\mathcal{H}} - \mathrm{Tr}[E_{\mathcal{H}}]\frac{\mathbb{1}_{\mathcal{H}}}{d}. \tag{7.11}$$

Key point is that replacing the observable E with $\bar{E}$ makes no difference, since $\mathrm{Tr}[E\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho})] = \mathrm{Tr}[\bar{E}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho})]$. This follows directly from eq. (7.11), since $\mathrm{Tr}[\mathbb{1}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho})] = 0$ by the tracelessness of $\bar{\rho}$ and the trace-preserving property of $\mathcal{G}_{\mathbf{j}}^{\otimes 2}$. Analogously, in the single-copy implementation, the traceless measurement $\bar{E}^{(Q)}_{\mathcal{H}}$ can be used instead of the observable $E^{(Q)}_{\mathcal{H}}$. Throughout, a bar over the measurement operator will mean the traceless component as defined by eq. (7.11).

The key idea of eq. (7.10) is that $\bar{\rho}_{\mathrm{eff}}$ and $\bar{E}_{\mathrm{eff}}$ are constructed such that computing $q^{(1)}_{\mathbf{j}}$ with eq. (7.8) is mathematically equivalent to computing $q^{(2)}_{\mathbf{j}}$ with eq. (7.6) using the effective operators eq. (7.10),

$$q^{(1)}_{\mathbf{j}} = \frac{1}{d^2-1} \sum_{P,Q\neq\mathbb{1}} \mathrm{Tr}[\bar{E}^{(Q)}_{\mathcal{H}} \mathcal{G}_{\mathbf{j}}(\bar{\rho}^{(P)}_{\mathcal{H}})]^2 \tag{7.12}$$

$$= \mathrm{Tr}\left[\bar{E}_{\mathrm{eff}} \mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\mathrm{eff}})\right] = q^{(2)}_{\mathbf{j}}. \tag{7.13}$$

In particular the ideal effective operators $\bar{\rho}_{\mathrm{eff,id}}$ and $\bar{E}_{\mathrm{eff,id}}$ (defined by eq. (7.10) for the ideal single-copy operators eq. (7.9)) are equal to the ideal two-copy operators eq. (7.7),

$$\bar{\rho}_{\mathrm{eff,id}} = \bar{\rho}_{\mathrm{id}} \qquad \text{and} \qquad \bar{E}_{\mathrm{eff,id}} = \bar{E}_{\mathrm{id}}. \tag{7.14}$$

This follows from the fact that [2]

$$S = \frac{1}{d} \sum_{P} P \otimes P. \tag{7.15}$$

Note that the sum is here over all Pauli matrices including the identity. As a result of this, the rest of the chapter will exclusively deal with the two-copy operators $\bar{\rho}$, $E \in \mathcal{M}_d$. The results can be interpreted for the single-copy protocol by considering the effective operators eq. (7.10).

The two-copy implementation of the protocol as previously discussed, can only be implemented if the experimenter has access to two different, but identical copies of the system under examination. These two systems must be simultaneously accessible for entangled state preparation and measurements, but the unitary control on each subsystem needs to be fully disjoint (i.e. without crosstalk) and identical (meaning noise must be identical on each subsystem). These assumptions are hard if not impossible to fulfill in any experimental system. We emphasize however that the two-copy implementation is introduced as a mathematical tool for the analysis of the URB protocol and its equivalence to the more realistic single-copy protocol was shown.

This concludes our review of the URB protocol, including the proposed modification of traceless input operators and emphasizing the two different implementations (which we have named the single-copy and two-copy implementation respectively). Next, we will present our main result. We will show how a concentration inequality can be used to relate the required resources (the number of sequences N) to parameters that quantify the confidence in the estimate of the average sequence purity $\bar{q}_m$. To do so, we will present a sharp bound σ^2 on the variance of the sequence purity $\mathbb{V}[q_{\mathbf{j}}^{(K)}]$ and present a bound L on the length of the interval in which the sequence purity $q_{\mathbf{j}}^{(K)}$ lies. These bounds are independent of K (the choice between single or two-copy implementation). Therefore, if no implementation-specific details are discussed, the sequence purity is just denoted $q_{\mathbf{j}}$.

7.2. Summary of results

In this section the main contribution of this chapter is summarized. The main result is a sharp bound on the number of sequences N required to obtain the average sequence purity $\bar{q}_m$ given fixed sequence length m with a certain a priori determined confidence. In section 7.2.1 we review a result from statistics to quantify the relation between the number of sequences N and the confidence. This relation requires some knowledge on the distribution of the sequence purity $q_{\mathbf{j}}$. A bound on the variance and a bound on the interval length of the sequence purity are needed. In section 7.2.2 we present a bound on the variance of the URB sequence purity $q_{\mathbf{j}}$ for benchmarking the Clifford gate set. This is the main contribution of this work. In section 7.2.3 we present a bound on the length of the interval in which $q_{\mathbf{j}}$ must lie. Finally in section 7.2.4 we give some examples on how to use our results.

7.2.1. Relation between the confidence parameters and the number of sequences

Using concentration inequalities from statistics, the confidence in the estimate $\bar{q}_m$ can be expressed as the probability that it deviates at most ϵ from the exact mean $\mathbb{E}[q_{\mathbf{j}}]$. If this probability $\text{Prob}[|\bar{q}_m - \mathbb{E}[q_{\mathbf{j}}]| \geq \epsilon] \leq \delta$ is to be bounded by δ, then the number of required

data points N is related to the confidence parameters ϵ, δ by [7]

$$2\left(\left(\frac{L}{L-\epsilon}\right)^{\frac{L^2-\epsilon L}{\sigma^2+L^2}}\left(\frac{\sigma^2}{\sigma^2+\epsilon L}\right)^{\frac{\sigma^2+\epsilon L}{\sigma^2+L^2}}\right)^N \leq \delta. \tag{7.16}$$

In this expression σ^2 is a bound on the variance $\mathbb{V}[q_{\mathbf{j}}]$ and L is a bound on the length of the interval in which $q_{\mathbf{j}}$ lies. Given σ^2 and L, there are two ways to apply this inequality. It can either be solved (numerically) for ϵ, given fixed N and δ, or it can be solved for N given ϵ, δ. In any case, it provides a direct relation between the number of required sequences N and the confidence parameters ϵ, δ, given L and σ^2. So in order to apply eq. (7.16), the bounds L and σ^2 are needed.
In the next section we will present a sharp bound σ^2 on the variance of the sequence purity $\mathbb{V}[q_{\mathbf{j}}]$. This bound is the key ingredient in using eq. (7.16) and it is the main contribution of this chapter.

7.2.2. BOUND ON THE VARIANCE OF THE SEQUENCE PURITY

In this section we present a bound σ^2 on the variance of the sequence purity $\mathbb{V}[q_{\mathbf{j}}]$ that is valid under the following assumptions:

1. The gate set under investigation is the d-dimensional Clifford group, denoted C_q. Here $d = 2^q$ for a q-qubit system. This assumption is necessary for deriving a variance bound. Even though the expected value $\mathbb{E}[q_{\mathbf{j}}]$ of the URB sequence purity is independent of the chosen gate set (as long as it is a unitary 2-design), the variance is not. The Clifford group was chosen as the default gate set.

2. Gate errors are independent of the gate. This is known as the gate-independent error model. In this model, the implemented noisy gate is $\tilde{\mathcal{G}} = \mathcal{G}\mathcal{E}$, where $\mathcal{G} \in \mathsf{C}_q$ is the ideal Clifford gate and $\mathcal{E}$ is an arbitrary quantum channel describing the noise. Crucially, $\mathcal{E}$ does not depend on the specific gate $\mathcal{G} \in \mathsf{C}_q$. This is assumption is necessary for deriving the fit model for URB [2]. Consequently our variance bound also employs this assumption. The URB protocol has not been analyzed in a gate dependent noise setting.

3. The noise map $\mathcal{E}$ is assumed to be unital if $q \geq 2$ (or equivalently if $d \geq 4$). A quantum channel $\mathcal{E}$ is unital if the maximally mixed state is a fixed point of the map, $\mathcal{E}(\mathbb{1}) = \mathbb{1}$. If the system under investigation $\mathcal{H}$ is a single-qubit system ($d = 2$), than this assumption is not necessary. Our result thus holds for any single-qubit quantum channel $\mathcal{E}$. This assumption enters in our derivation of the variance bound. It is not a fundamental assumption but rather a condition under which we were able to derive a useful, sharp bound.

   At this point, we emphasize that $\mathbb{V}[q_{\mathbf{j}}]$ is the between-sequence variance, i.e. the variance of $q_{\mathbf{j}}$ due to the randomly sampled sequence indexed by $\mathbf{j}$. In particular this means that given a sequence $\mathbf{j}$, we assume that $q_{\mathbf{j}}$ can be determined with arbitrary precision. In reality $q_{\mathbf{j}}$ can only be estimated due to the probabilistic nature of quantum mechanics by taking many single-shot measurements of the same sequence $\mathbf{j}$. In section 7.4.1 we relax this assumption by splitting the total variance into the sum of the between-sequence variance (the variance due to randomly sampled $\mathbf{j}$) and the within-sequence variance (the variance due to uncertainty in $q_{\mathbf{j}}$ for fixed $\mathbf{j}$).

Under the assumptions stated above, the following bound on the variance $\mathbb{V}[q_{\mathbf{j}}]$ is derived (see theorem 1 in section 7.6.3)

$$\mathbb{V}[q_{\mathbf{j}}^{(K)}] \leq \sigma^2 = \frac{1-u^{2(m-1)}}{1-u^2}(1-u)^2\Big(c_1(d) + c_2(d)\|\bar{E}_{\mathrm{err}}\|_\infty^2 + c_3(d)\|\bar{\rho}_{\mathrm{err}}\|_1^2\Big) + \|\bar{\rho}_{\mathrm{err}}\|_1^2\|\bar{E}_{\mathrm{err}}\|_\infty^2, \quad (7.17)$$

which is independent of the used implementation (single or two-copy, corresponding to $K = 1, 2$). Here u is the unitarity of $\mathcal{E}$, m is the sequence length, $\|\bar{E}_{\mathrm{err}}\|_\infty^2$, $\|\bar{\rho}_{\mathrm{err}}\|_1^2$ are quantities depending on the quality of state preparation and measurement and c_i are constants that solely depend on the dimension d. The values of c_i for small d are tabulated in table 7.1. For precise definitions of these quantities, see theorem 1 in section 7.6.3. The error operators have the following definitions:

$$\begin{aligned}\bar{\rho}_{\mathrm{err}} &= \bar{\rho} - \frac{\mathrm{Tr}[\bar{\rho}_{\mathrm{id}}\bar{\rho}]}{\|\bar{\rho}_{\mathrm{id}}\|_2^2}\bar{\rho}_{\mathrm{id}} = \bar{\rho} - (d^2-1)\,\mathrm{Tr}[\bar{\rho}_{\mathrm{id}}\bar{\rho}]\bar{\rho}_{\mathrm{id}},\\ \bar{E}_{\mathrm{err}} &= \bar{E} - \frac{\mathrm{Tr}[\bar{E}_{\mathrm{id}}\bar{E}]}{\|\bar{E}_{\mathrm{id}}\|_2^2}\bar{E}_{\mathrm{id}} = \bar{E} - \frac{\mathrm{Tr}[\bar{E}_{\mathrm{id}}\bar{E}]}{d^2-1}\bar{E}_{\mathrm{id}},\end{aligned} \quad (7.18)$$

where the ideal operators $\bar{\rho}_{\mathrm{id}}, \bar{E}_{\mathrm{id}}$ are defined in eq. (7.7) and a bar over the measurement operator indicates its traceless component $\bar{E} = E - \frac{\mathrm{Tr}[E]}{d^2}\mathbb{1}$ (as defined in eq. (7.11)). Recall that $\bar{\rho}$ was defined as the difference between two states (eq. (7.3)). The error operators are defined in such a way that they are orthogonal to the ideal operators with respect to the Hilbert-Schmidt inner product, i.e.

$$\mathrm{Tr}[\bar{\rho}_{\mathrm{err}}\bar{\rho}_{\mathrm{id}}] = \mathrm{Tr}[\bar{E}_{\mathrm{err}}\bar{E}_{\mathrm{id}}] = 0. \quad (7.19)$$

The norms on the error operators are the trace norm and operator norm respectively, defined for all $A \in \mathcal{M}_d$ as

$$\begin{aligned}\|A\|_1 &= \mathrm{Tr}[\sqrt{A^\dagger A}] = \sum_i s_i(A),\\ \|A\|_\infty &= \sup_{0\neq x\in\mathcal{H}^{\otimes 2}} \frac{\|Ax\|_2}{\|x\|_2} = \max_i\{s_i(A)\},\end{aligned} \quad (7.20)$$

with $s_i(A)$ the i-th singular value of A and $\|x\|_2$ the euclidean norm on $\mathcal{H}^{\otimes 2}$. Note that in the single-copy case the quantities $\|\bar{\rho}_{\mathrm{err}}\|_1^2, \|\bar{E}_{\mathrm{err}}\|_\infty^2$ as defined in eq. (7.18) are to be estimated using $\bar{\rho}_{\mathrm{eff}}$ and $\bar{E}_{\mathrm{eff}}$ as defined in eq. (7.10).

The variance bound of eq. (7.17) has some appealing qualitative features. The first feature is that the first term is proportional to $(1-u)^2$. This means that the first term goes to zero quadratically as the unitarity u of the error map $\mathcal{E}$ approaches 1. The fact that the second term is constant with respect to both u and m is unavoidable, as will be discussed in section 7.3.2. The second appealing feature is the fact that the bound is asymptotically independent of the sequence length m. Thus the variance bound is useful in any regime

Table 7.1: Evaluation of the constants $c_i(d)$ for various small-dimensional systems. The last row indicates the asymptotic behavior.

d	$c_1(d)$	$c_2(d)$	$c_3(d)$
2	$\frac{11}{12}$	$\frac{13}{9}$	$\frac{5}{2}$
4	$\frac{179}{60}$	54.675	48.053
8	1.6322	81.445	119.31
16	1.1443	110.64	296.88
32	1.0354	173.80	891.69
$\to \infty$	$\mathcal{O}(1)$	$\mathcal{O}(d)$	$\mathcal{O}(d^2)$

of m. In section 7.3 the dependence of the variance bound and the resulting number of sequences on various parameters is discussed in greater detail.

In the next section we present a bound L in the length of the interval in which the sequence purity $q_{\mathbf{j}}$ lies. This is the final ingredient needed in order to apply eq. (7.16).

7.2.3. BOUND ON THE INTERVAL OF THE SEQUENCE PURITY

In this section we present the improved bound L on the length of the interval in which the sequence purity $q_{\mathbf{j}}^{(K)}$ lies. Even though the actual interval depends on K, the length of these intervals is the same. Thus the bound L on the interval length of the sequence purity is independent of the implementation indexed by K. The improved bound is derived under the mild assumption that the experimental control is sufficiently good such that $\mathrm{Tr}[\bar{\rho}_{\mathrm{id}}\bar{\rho}] \geq 0$ and $\mathrm{Tr}[\bar{E}_{\mathrm{id}}\bar{E}] \geq 0$ (analogous assumption holds for the single-copy input and measurement operators). These conditions are satisfied only if the conditions

$$\mathrm{Tr}[\rho_{\mathrm{id}}\rho] \geq \mathrm{Tr}[\hat{\rho}_{\mathrm{id}}\rho], \qquad \mathrm{Tr}[\hat{\rho}_{\mathrm{id}}\hat{\rho}] \geq \mathrm{Tr}[\rho_{\mathrm{id}}\hat{\rho}], \tag{7.21}$$

$$\mathrm{Tr}[E\bar{\rho}_{\mathrm{id}}] \geq 0 \tag{7.22}$$

are satisfied. eq. (7.21) can be interpreted as requiring that the implemented states ρ, $\hat{\rho}$ have more overlap with their corresponding ideal state than with the non-corresponding ideal states. eq. (7.22) is equivalent to $\mathrm{Tr}[\bar{E}\bar{E}_{\mathrm{id}}] \geq 0$ since $\bar{E}_{\mathrm{id}} = (d^2-1)\bar{\rho}_{\mathrm{id}}$ and $\mathrm{Tr}[\bar{\rho}_{\mathrm{id}}\bar{E}] = \mathrm{Tr}[E\bar{\rho}_{\mathrm{id}}]$. eq. (7.22) has the interpretation that the measurement $\{M, \mathbb{1}-M\}$ associated with the observable $E = 2M - \mathbb{1}$ assigns the correct outcome ($+1$ for ρ_{id} and -1 for $\hat{\rho}_{\mathrm{id}}$) with at least probability $\frac{1}{2}$, or alternatively, that the measurement can correctly discriminate the maximally mixed state on the symmetric subspace (ρ_{id}) from the maximally mixed state on the anti-symmetric subspace ($\hat{\rho}_{\mathrm{id}}$). These are very reasonable assumptions for any practical quantum information device.

In lemma 7.8 of section 7.6.3 we show that under the stated assumption, the sequence

purity lies in the interval

$$q_{\mathbf{j}}^{(1)} \in [0, 1 + \|\bar{\rho}_{\mathrm{err}}\|_1 + \|\bar{E}_{\mathrm{err}}\|_\infty + \|\bar{\rho}_{\mathrm{err}}\|_1 \|\bar{E}_{\mathrm{err}}\|_\infty], \tag{7.23}$$

$$q_{\mathbf{j}}^{(2)} \in [-\|\bar{\rho}_{\mathrm{err}}\|_1 - \|\bar{E}_{\mathrm{err}}\|_\infty - \|\bar{\rho}_{\mathrm{err}}\|_1 \|\bar{E}_{\mathrm{err}}\|_\infty, 1]. \tag{7.24}$$

Therefore it follows that

$$L = 1 + \|\bar{\rho}_{\mathrm{err}}\|_1 + \|\bar{E}_{\mathrm{err}}\|_\infty + \|\bar{\rho}_{\mathrm{err}}\|_1 \|\bar{E}_{\mathrm{err}}\|_\infty \tag{7.25}$$

for both implementations. The idea of the proof of lemma 7.8 is to decompose the input and measurement operators $\bar{\rho}$ and $\bar{E}$ into their ideal and error components according to eq. (7.18). This gives rise to four terms. The ideal term $\mathrm{Tr}[E_{\mathrm{id}} \mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\mathrm{id}})]$ can be bounded in the interval $[0, 1]$. The other terms are then bounded in magnitude using Hölder's inequality, which contributes the last three terms in eq. (7.25).

7.2.4. EXAMPLES

Perhaps the best way to gain insight in the use of eq. (7.16), eq. (7.17) and eq. (7.25) is by example. In example 1 we calculate the required number of sequences for a fixed choice of all relevant parameters. In example 2 we simulate a URB experiment using fixed number of sequences and compute the confidence interval around each estimate $\bar{q}_m$. We compare the results of these examples with a previously known bound (first used in [1]). This bound does not use the variance, but just uses the boundedness of the sequence purity $q_{\mathbf{j}}$. It claims that $\mathrm{Prob}[|\bar{q}_m - \mathbb{E}[q_{\mathbf{j}}]| \geq \epsilon] \leq \delta$, whenever [7]

$$2e^{-2N\frac{\epsilon^2}{L^2}} \leq \delta. \tag{7.26}$$

The number of sequences N is merely a function of the confidence parameters ϵ, δ and the interval length L. In particular it does not depend on the variance of $q_{\mathbf{j}}$.

Example 1. Suppose that a URB experiment is performed on the single-qubit Clifford group ($d = 2$). The choice of implementation (single-copy or two-copy) is irrelevant for this example since both the variance bound eq. (7.17) and the interval length bound eq. (7.25) are independent of the choice of implementation. The only difference in practice is how to estimate the SPAM parameters $\|\bar{\rho}_{\mathrm{err}}\|_1^2, \|E_{\mathrm{err}}\|_\infty^2$. Furthermore suppose that an priori estimate of the unitarity is $u = 0.98$ and an estimate for the SPAM parameters is $\|\bar{\rho}_{\mathrm{err}}\|_1^2 = \|E_{\mathrm{err}}\|_\infty^2 = 0.02$. Then, after choosing appropriate sequence lengths to use in the experiment, an upper bound on the variance as a function of the sequence length can be computed using eq. (7.17). The interval length can be bounded using eq. (7.25). Using $\|\bar{\rho}_{\mathrm{err}}\|_1^2 = \|E_{\mathrm{err}}\|_\infty^2 = 0.02$, this yields $L = 1.02 + 0.2\sqrt{2} \approx 1.303$. Finally, choosing an interval ϵ and confidence δ, eq. (7.16) gives the required number of sequences N (at fixed length m). Concretely, setting $\epsilon = 0.02$, $\delta = 0.01$ and all other parameters as discussed, the number of sequences required for sequences of length $m = 10$, is $N = 242$. For sequence length $m = 30$, the required number is $N = 366$, whereas $m = 100$ requires $N = 452$. The long sequence length limit (when $u^{2(m-1)} \ll 1$), yields $N = 457$.
Let us compare these numbers with the previously known bound eq. (7.26) that does not use the variance of $q_{\mathbf{j}}$. Given our choices of $\epsilon = 0.02$, $\delta = 0.01$ and $\|\bar{\rho}_{\mathrm{err}}\|_1^2 = \|E_{\mathrm{err}}\|_\infty^2 =$

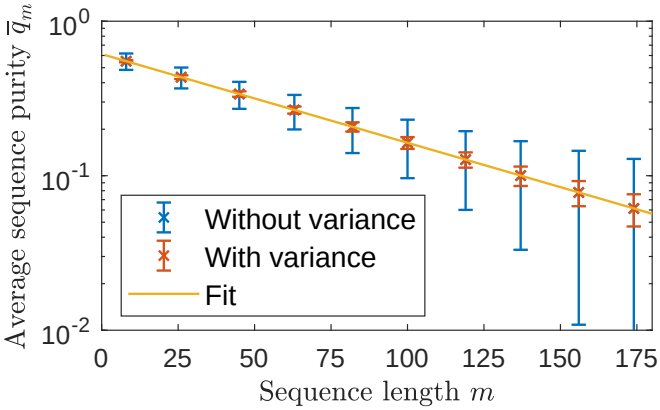


Figure 7.2: Comparison of the 99% confidence intervals around the average sequence purity $\bar{q}_m$ calculated with and without our variance bound at several different sequence lengths. The plot is based on a simulated URB experiment of the single-qubit Clifford group with $N = 250$ samples per sequence length m. The empirical average sequence purity $\bar{q}_m$ (marked with a cross) is plotted versus the sequence length m on a semilogarithmic scale. The larger (blue) bars indicate the 99% confidence interval without variance (eq. (7.26)) and the smaller (red) bars indicate the 99% confidence interval of eq. (7.16) based on our sharp variance bound eq. (7.17). Here we used an a priori estimates of the unitarity and SPAM parameters of $u = 0.98$ and $\|\bar{\rho}_{\text{err}}\|_1^2 = \|E_{\text{err}}\|_\infty^2 = 0.02$ respectively. Then eq. (7.25) yields $L = 1.02 + 0.2\sqrt{2}$. For completeness, a least-squares fit according to the model $\bar{q}_m = Bu^{m-1}$ (see eq. (7.4)) is shown in the yellow solid line. This yields $u \approx 0.987$.

0.02 (from which $L = 1.02 + 0.2\sqrt{2} \approx 1.303$ is computed using eq. (7.25)), the bound eq. (7.26) yields $N = 11242$ required sequences. We emphasize that this number is independent of u or m. In this scenario, our bound gives approximately two orders of magnitude improvement. □

Example 2. In fig. 7.2 we compare the 99% confidence intervals ϵ (for fixed $N = 250$ and $\delta = 0.01$) around the empirical average sequence purity $\bar{q}_m$ calculated with and without our variance bound at several different sequence lengths. The empirical average sequence purity $\bar{q}_m$ data is based on a simulated single-qubit Clifford URB experiment. The length of the confidence interval ϵ without variance (larger blue bars) is computed from eq. (7.26). Then the choice of $N = 250$ and $\delta = 0.01$ yields $\epsilon = 0.134$. On the other hand, the length of the confidence interval ϵ with variance (smaller red bars in the plot) is computed from eq. (7.16) by solving the equation for ϵ, using our sharp variance bound eq. (7.17). In the evaluation of eq. (7.17), the a priori estimates $u = 0.98$ and $\|\bar{\rho}_{\text{err}}\|_1^2 = \|E_{\text{err}}\|_\infty^2 = 0.02$ were used. Then eq. (7.25) yields $L = 1.02 + 0.2\sqrt{2}$. Using our sharp variance bound, the values of the confidence interval vary between $\epsilon = 0.019$ (for $m = 8$) and $\epsilon = 0.029$ (for $m = 174$). This is approximately an order of magnitude larger than the confidence interval without variance $\epsilon = 0.134$.
In this simulated experiment the Clifford gates are implemented with a fixed error channel $\mathcal{E}$ that is generated by taking a convex combination of the identity channel (with high weight) and a random CPTP map (sampled using QETLAB [8]). Similarly, the noisy input states and measurement operator are simulated by taking a convex combination of the ideal operators and randomly generated operators (generated using QETLAB). For this

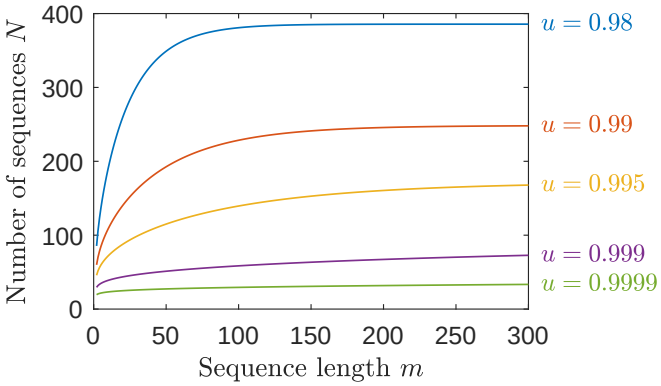


Figure 7.3: Number of sequences N versus the sequence length m for various values of the unitarity u when benchmarking the single-qubit Clifford group ($d = 2$). Confidence parameters are $\epsilon = 0.02$ and $\delta = 0.01$. The SPAM parameters are $\|\bar{\rho}_{\text{err}}\|_1^2 = \|E_{\text{err}}\|_\infty^2 = 0$. By eq. (7.25) then $L = 1$ is used. The number of sequences is asymptotically independent of the sequence length. This is consistent with our variance bound eq. (7.17).

particular realization of an error map $\mathcal{E}$, the data points seem to be even more accurate than our confidence interval might suggest based on their proximity to the fit. This is due to the fact that this particular error channel is well-behaved. We emphasize that our bound is valid for any unital or single-qubit error map. In particular this means that our bound is valid for the worst case realizations of $\mathcal{E}$. It is unclear what error map $\mathcal{E}$ maximizes the variance of the sequence purity.

We emphasize that the point of this simulated example is not to prescribe a direct method for extracting the confidence in the unitarity, as this generally depends on the fitting model and the way the uncertainty in the average sequence purity are propagated into the uncertainty of the unitarity. Moreover, more advanced statistical tools may be used to extract the unitarity from the obtained (in this case simulated) data, like e.g. [9, 10]. The goal of this example is to illustrate the significant gain in confidence of the average sequence purity when the simple concentration inequalities of Hoeffding are applied [7]. The point is that the additional knowledge of a variance bound on the underlying distribution of the sequence purity $q_{\mathbf{j}}$ can be used by statistical tools to extract the unitarity with improved confidence. □

In the next section we explore the behavior of our bound in various parameter regimes.

7.3. Discussion

This section is devoted to discussing the variance bound and the interval length of the sequence purity in more detail. In particular we discuss the variance bound in several different parameter regimes in more detail and aim to provide a better understanding of the parameters that ultimately determine the statistical confidence of the measurements. In section 7.3.1 we discuss the dependence of the variance bound eq. (7.17) on the unitarity u and the sequence length m. In section 7.3.2 we discuss the dependence on the SPAM parameters $\|\bar{\rho}_{\text{err}}\|_1^2$ and $\|\bar{E}_\infty\|_1^2$. Here we also show by example that the variance of the

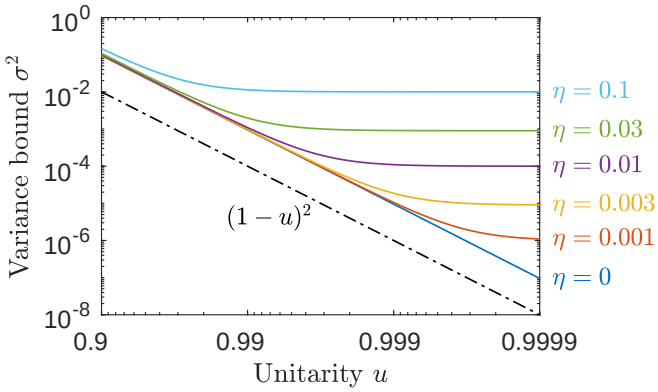


Figure 7.4: Semilogarithmic plot of the variance bound σ^2 as a function of the unitarity u for various magnitudes of SPAM errors in the large sequence limit for single-qubit Clifford URB ($d = 2$). The black dash-dotted line is a reference line plotting $\sigma^2 = (1-u)^2$. The differently colored solid lines indicate the various magnitudes of SPAM errors, where $\|\bar{\rho}_{\text{err}}\|_1^2 = \|\bar{E}_{\text{err}}\|_\infty^2 = \eta$. There are two regimes. For small SPAM errors and small u, the variance scales as $(1-u)^2$, whereas for nonzero SPAM errors and large u, the variance approaches a constant.

sequence purity does not go to zero in the presence of SPAM errors. In section 7.3.3 the dependence of the variance bound on the system size is discussed.

7.3.1. Dependence on unitarity and sequence length

First, we discuss the dependence of the number of required sequences N on the sequence length m. In fig. 7.3 this dependence is plotted for various values of u in the absence of SPAM errors (i.e. $\|\bar{\rho}_{\text{err}}\|_1^2 = \|E_{\text{err}}\|_\infty^2 = 0$). The confidence parameters were fixed at $\delta = 0.01$ and $\epsilon = 0.02$. It can be seen from the figure that N approaches a constant as m increases. This is consistent with our variance bound eq. (7.17), where the factor depending on m is

$$\frac{1-u^{2(m-1)}}{1-u^2}(1-u)^2. \tag{7.27}$$

This approaches a constant in the limit of large sequence lengths. This limit is approximately achieved when $u^{2(m-1)} \ll 1$. The exact limit is given by

$$\lim_{m\to\infty} \frac{1-u^{2(m-1)}}{1-u^2}(1-u)^2 = \frac{1-u}{1+u}. \tag{7.28}$$

In the presence of SPAM errors, the asymptotic constant is larger than in its absence, but the behavior is similar. Since the variance approaches a constant, so does the required number of sequences for fixed values of the confidence parameters. From here on out, the 'large sequence limit' means the regime of m where $u^{2(m-1)} \ll 1$ so that the variance bound (and thus the number of sequences) is approximately independent of m.

Second we discuss the dependence of the variance bound on the unitarity u. In fig. 7.4 the variance bound σ^2 is plotted as a function of the unitarity u for various values of SPAM errors in the long sequence length limit. This figure shows two regimes. In the

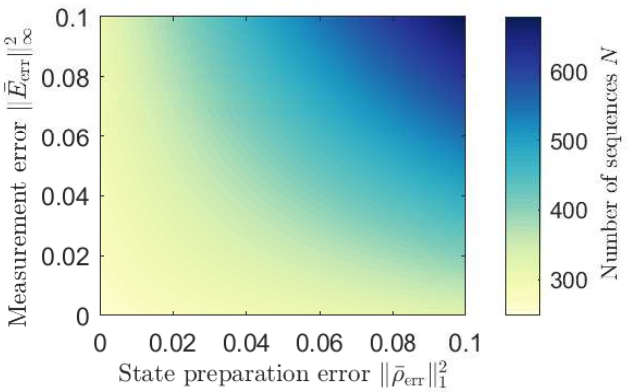


Figure 7.5: Color plot of the number of sequences N as a function of the SPAM parameters $\|\bar{\rho}_{\rm err}\|_1^2$ and $\|E_{\rm err}\|_\infty^2$ in the large sequence length limit for single-qubit Clifford URB ($d = 2$). The parameters $u = 0.99$ and $\epsilon = 0.02$, $\delta = 0.01$ were used. This plot illustrates the sensitivity of our result to SPAM errors. In particular, the number of sequences increases most significantly when both state preparation and measurement errors are large.

regime of low unitarity and small SPAM error, the variance is proportional to $(1-u)^2$. This is consistent with eq. (7.17), where the variance is dominated by the first term in this regime. However, for nonzero SPAM error and large unitarity, this behavior transitions into a constant variance. In this regime, the variance is dominated by the second, constant term (independent of u) in eq. (7.17).

The number of required sequences N shows qualitatively similar behavior, but there are differences. This is due to the fact that N is a nonlinear function of σ^2. In the regime of constant variance, the number of sequences is also constant. In the regime where the variance bound is proportional to $(1-u)^2$, the number of sequences also decreases as N increases, but the rate depends also on the choice of ϵ.

7.3.2. Dependence on SPAM parameters

In fig. 7.5 we show a color plot of the number of sequences N as a function of the SPAM parameters $\|\bar{\rho}_{\rm err}\|_1^2$ and $\|E_{\rm err}\|_\infty^2$ for fixed unitarity u in the limit of large sequences. The plot illustrates the qualitative dependence of N on the magnitude of these SPAM parameters. There are two ways that the SPAM parameters contribute to the number of required sequences N. First, the variance bound σ^2 depends on the SPAM parameters $\|\bar{\rho}_{\rm err}\|_1^2$ and $\|E_{\rm err}\|_\infty^2$ (see eq. (7.17)). Second, the interval length bound L depends on the square root of these parameters, $\|\bar{\rho}_{\rm err}\|_1$ and $\|E_{\rm err}\|_\infty$ (see eq. (7.25). Both these bounds increase as the SPAM parameters increase. From the concentration inequality eq. (7.16), it follows that the required number of sequences N for fixed confidence parameters grows with increasing variance and interval length. Both these effects have qualitatively similar behavior. This translate into the illustrated dependence of the number of sequences N on the SPAM parameters in fig. 7.5. In particular, the number of sequences most strongly depends on the product between the two, showing a larger required number in the area where the product $\|\bar{\rho}_{\rm err}\|_1^2\|E_{\rm err}\|_\infty^2$ is largest.

The variance bound of eq. (7.17) has a constant term $\|\bar{\rho}_{\rm err}\|_1^2\|E_{\rm err}\|_\infty^2$, independent of the

unitarity u and sequence length m. In particular this means that the variance bound is nonzero in the presence of SPAM error for all sequence lengths m even in the limit of ideal gates $\mathcal{E} \to \mathcal{I}$. This behavior is also seen in fig. 7.4. We argue that this is fundamental to the URB protocol, by showing that the actual variance of the sequence purity $\mathbb{V}[q_{\mathbf{j}}]$ also has this behavior even when ideal gates are considered. This is done in example 3. In this example we construct noisy operators $\bar{\rho}$ and $\bar{E}$ such that the average sequence purity $q_{\mathbf{j}}$ is not constant over all possible ideal gate sequences $\mathcal{G}_{\mathbf{j}}$ (i.e. sequences with $\mathcal{E} = \mathcal{I}$). Thus there exists an error channel (namely $\mathcal{E} = \mathcal{I}$) and noisy operators (namely those constructed in example 3) such that the variance, and thus the required number of sequences, is nonzero. This behavior is in contrast with standard RB, where all RB gate sequences compose to the identity when $\mathcal{E} = \mathcal{I}$ (in the RB protocol, a global inverse gate is applied after each sequence). Therefore in standard RB, the survival probability does not depend on the sequence in the absence of gate errors and hence the variance is zero.

Example 3. Consider a URB experiment where the gate set under investigation is the single-qubit Clifford group C_q. Suppose that the gates are implemented perfectly, i.e. $\mathcal{E} = \mathcal{I}$. Furthermore assume that the state and measurement operators are given by

$$\rho, \hat{\rho} = \frac{\mathbb{1} \otimes \mathbb{1} \pm X \otimes X}{4}, \qquad \text{and} \qquad E = X \otimes X, \tag{7.29}$$

where $\mathbb{1}$ is the identity and X is the Pauli-X matrix on the single-qubit Hilbert space $\mathcal{H} \simeq \mathbb{C}^2$. Since $\mathcal{E} = \mathcal{I}$, the sequence $\mathcal{G}_{\mathbf{j}}$ of m independently and uniformly distributed Clifford gates reduces to a single Clifford gate $\mathcal{G}_i$ uniformly drawn from C_q. The group C_q has 24 elements, 8 of which map $X \mapsto \pm X$. Whether such a map sends X to $+X$ or $-X$ is irrelevant, since if $\mathcal{G}$ maps $X \mapsto \pm X$ then $\mathcal{G}^{\otimes 2}$ maps $X^{\otimes 2} \mapsto X^{\otimes 2}$ in either case. The other 16 Clifford gates send $X \mapsto \pm Y$ or $X \mapsto \pm Z$, where again the sign is irrelevant. Thus, given that $\bar{\rho} = \frac{X \otimes X}{4}$, a fraction $\frac{8}{24}$ of all sequences $\mathcal{G}_{\mathbf{j}}$ will satisfy $\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}) = \frac{X \otimes X}{4}$ while the others will send $\bar{\rho}$ either to $\frac{Y \otimes Y}{4}$ or $\frac{Z \otimes Z}{4}$. Since $\mathrm{Tr}[E(\frac{X \otimes X}{4})] = 1$ and $\mathrm{Tr}[E(\frac{Y \otimes Y}{4})] = \mathrm{Tr}[E(\frac{Z \otimes Z}{4})] = 0$, the following probability distribution on $q_{\mathbf{j}}^{(2)}$ is obtained:

$$\text{Prob}\left[q_{\mathbf{j}}^{(2)} = 1\right] = \frac{1}{3} \qquad \text{and} \qquad \text{Prob}\left[q_{\mathbf{j}}^{(2)} = 0\right] = \frac{2}{3}. \tag{7.30}$$

Clearly then $\mathbb{E}[q_{\mathbf{j}}^{(2)}] = \frac{1}{3}$ and $\mathbb{V}[q_{\mathbf{j}}^{(2)}] = \frac{2}{9} > 0$. This example shows that the variance $\mathbb{V}[q_{\mathbf{j}}]$ of the sequence purity can not go to zero as the unitarity $u \to 1$. □

Given noisy implementations $\bar{\rho}$ and E in the two-copy implementation, the SPAM parameters $\|\rho_{\text{err}}\|_1^2$ and $\|\bar{E}_{\text{err}}\|_\infty^2$ defined in eq. (7.18) can in principle be estimated by relating them to the ideal states and measurements of eq. (7.7). In practice, this requires (partial) knowledge of the noisy operators $\bar{\rho}$ and E. If a full (tomographic) description of $\rho, \hat{\rho}, E$ is available, then $\|\bar{\rho}_{\text{err}}\|_1^2$ and $\|\bar{E}_{\text{err}}\|_\infty^2$ can be calculated from the definition eq. (7.18). However, if only partial knowledge is available (e.g. a lower bound on state preparation fidelity), then the SPAM quantities need to be bounded. For example $\|\bar{\rho}_{\text{err}}\|_1^2$ can be upper bounded if the fidelity between ρ ($\hat{\rho}$) and ρ_{id} ($\hat{\rho}_{\text{id}}$) is known, by application of the Fuchs-Van de Graaff inequality [11]. In the single-copy implementation, slightly more work is needed. The SPAM parameters are then defined with respect to $\bar{\rho}_{\text{eff}}$ and $\bar{E}_{\text{eff}}$ (eq. (7.10)).

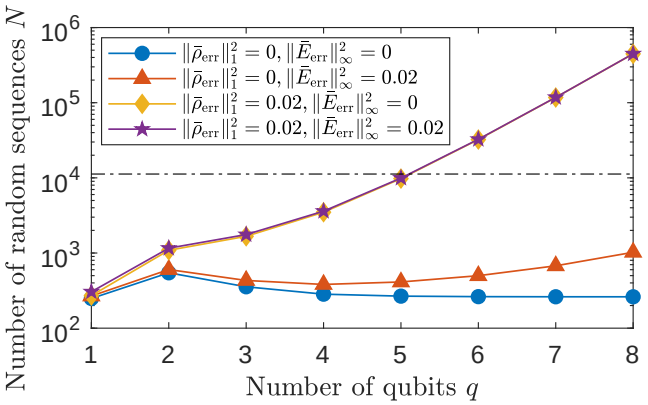


Figure 7.6: Number of sequences N as a function of the number of qubits q comprising the system for different values of the SPAM parameters. A fixed unitarity $u = 0.99$ and the large sequence length limit are used. The interval bound L is computed using eq. (7.25) as a function of the SPAM quantities (see legend). The confidence parameters $\epsilon = 0.02$, $\delta = 0.01$ were used. The dashed line indicates the first-order bound (eq. (7.26)) corresponding to $\|\bar{\rho}_{\rm err}\|_1^2 = \|\bar{E}_{\rm err}\|_\infty^2 = 0.02$. For the given confidence and SPAM parameters, our bound gives an improvement of the required number of sequences up to 5-qubit systems.

However, only (partial) knowledge of the physical operators $\rho_{\mathcal{H}}$ and $E_{\mathcal{H}}$ are available. Noise on these physical operators needs to be translated to noise on the effective operators $\bar{\rho}_{\rm eff}$ and $\bar{E}_{\rm eff}$.

7

7.3.3. DIMENSION-DEPENDENT CONSTANTS

In this section, the dependence of the variance bound eq. (7.17) and consequently the number of sequences on the system size is examined. An undesirable feature of the variance bound is the asymptotic growth of the constants $c_2(d)$ and $c_3(d)$ with the dimension $d = 2^q$ of the q-qubit system. This means that for large systems, the bound becomes loose and ultimately vacuous. This is illustrated in fig. 7.6, where the number of sequences N is plotted as a function of the system size q on a semilogarithmic scale (for fixed unitarity u and large sequence length m). The number of sequences is plotted in the absence of SPAM error, with state preparation or measurement error only and with both errors simultaneously. This is done to distinguish the different contributions of the constants c_1, c_2 and c_3 in eq. (7.17). In the absence of SPAM error, only c_1 is relevant. This constant takes its maximum at $q = 2$ and asymptotically goes to 1. However with measurement error, the number of sequences needed grows exponentially with the system size. With state preparation error, this expectational growth is even faster. This is consistent with the asymptotic limits of the constants $c_2 = \mathcal{O}(d)$ and $c_3 = \mathcal{O}(d^2)$, since $d = 2^q$. In particular, this figure shows that our variance bound is prohibitively loose for $q \geq 6$ (assuming $u = 0.99$ and large m), since the first order bound eq. (7.26) yields a smaller number of sequences N as indicated by the black dash-dotted line in the figure.

We believe that the unbounded growth of our variance bound with the system size is an artifact of the proof rather than a fundamental property. The sequence purity $q_{\mathbf{j}}$ is a bounded, discrete random variable, where the bound L does not depend on the dimension

d. Therefore the exact variance $\mathbb{V}[q_{\mathbf{j}}]$ can not asymptotically grow with the system dimension d. The bound of eq. (7.17) is however sharp enough for practical use in few-qubit systems.

7.4. METHODS

This section gives an high level overview of the methods used for deriving our main result eq. (7.16), eq. (7.17). In section 7.4.1 we focus on the statistical aspect of our result related to eq. (7.16). We also relate the between-sequence variance $\mathbb{V}[q_{\mathbf{j}}]$ (the quantity which we bounded in this work) to the within-sequence variance that arises due to the fact that $q_{\mathbf{j}}$ can only be estimated by collecting a finite sample of single shot measurements for a given sequence. In section 7.4.2 we discuss the derivation of the fit model (as derived in [2]) and derive an expression for the variance $\mathbb{V}[q_{\mathbf{j}}]$. In section 7.4.3 we give an outline of the proof of our variance bound eq. (7.17).

7.4.1. ESTIMATION THEORY

Ultimately, the URB protocol leads to the complex statistical estimation problem of determining u and the confidence thereof, given a large set of realizations of the sequence purity $q_{\mathbf{j}}$ (for multiple sequence lengths m). There are several ways one can go about this problem (see e.g. [10] for a Bayesian inference approach). Here we we take a frequentist approach and determine a confidence interval for the point estimates $\bar{q}_m$ of $\mathbb{E}[q_{\mathbf{j}}]$. These confidence intervals (for different values of m) can then be taken into account when fitting the point estimates $\bar{q}_m = Bu^{m-1}$ to the fit model. The main contribution of this work is improving the confidence interval of $\bar{q}_m$ by bounding the variance of the sequence purity $q_{\mathbf{j}}$. This variance bound provides strictly more information on the distribution of $q_{\mathbf{j}}$ than what was known before [2] and could therefore also be of value when using other estimation techniques to extract the unitarity u from the set of measurement outcomes.

The intuitive idea is that estimating the mean of a bounded distribution of random variables requires fewer samples when the distribution is narrowly peaked around the mean. Since the variance is a measure of the spread of the distribution, it is intuitive that having knowledge of the variance improves the confidence in the estimate of the mean. This idea is made precise in statistics by concentration inequalities. Here we use a concentration inequality due to Hoeffding [7]. Given a collection of N independent and identically distributed (i.i.d.) random variables X_i, sampled from a distribution on a length L interval with mean μ and variance σ^2, the following statement holds for all $0 \leq \epsilon \leq L$

$$\text{Prob}\left[|\bar{X} - \mu| \geq \epsilon\right] \leq 2\left[\left[\frac{L}{L-\epsilon}\right]^{\frac{L^2-\epsilon L}{\sigma^2+L^2}}\left[\frac{\sigma^2}{\sigma^2+\epsilon L}\right]^{\frac{\sigma^2+\epsilon L}{\sigma^2+L^2}}\right]^N, \tag{7.31}$$

where $\bar{X} = \frac{1}{N}\sum_i X_i$ is the empirical mean. This is essentially eq. (7.16) using the fact that $q_{\mathbf{j}}$ are i.i.d. random variables. The point is that if one wishes to bound this probability by δ, then upper bounding the right-hand-side by δ gives a means to relate N, δ and ϵ. Instead of the exact (unknown) variance of the distribution of $q_{\mathbf{j}}$, an upper bound is used. The fact that our variance bound eq. (7.17) depends on the unitarity u, the quantity that one ultimately attempts to estimate, may seems strange and circular. But this is actually

a feature of statistics, which is more apparent in the Bayesian view. One may have an a priori distribution of the unitarity u of the gate set and given some experimental data (the complete URB data set) one can construct a more concentrated a posteriori distribution on the unitarity. In the frequentist view, an a priori lower bound to the unitarity can be known with very high confidence. Then performing URB will improve the estimate of the unitarity and increase the confidence in this estimate. In principle this procedure can be done by doing several successive URB experiments, further increasing the confidence in the outcome. Note that a first lower bound can always be obtained from the average gate fidelity (see chapter 2), which is estimated using standard RB.

Finally there is one subtlety that deserves some attention. The protocol requires the experimenter to measure $\mathrm{Tr}[E\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho)]$, but actually this is an expectation value of the measurement operator E (a hermitian observable) given the state $\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho)$. This expectation value must be learned from multiple single-shot measurements of preparing the state, apply gates and measure. The outcome is inherently probabilistic (with a Bernoulli distribution) by the laws of quantum mechanics and either a click or no click is observed with the probability given by Born's rule. To estimate the expectation value $\mathrm{Tr}[E\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho)]$, a large number of single shot measurements must be taken and the proportion of clicks is an estimate $\mathrm{Tr}[E\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho)]$. In reality then, there is also some uncertainty in each data point $q_{\mathbf{j}}$, which propagates into increased uncertainty in the average $\bar{q}_m$. So far we have assumed the uncertainty in $\bar{q}_m$ is dominated by the uncertainty due to the randomly sampled sequences and not due to the uncertainty in determining each sequence purity $q_{\mathbf{j}}$. This assumption is motivated by experiments in which it is hard to store many sequences, but easy to repeat single shot measurements of the same sequence. In these experiments it is then easy to do enough single-shot measurements of each $q_{\mathbf{j}}$, such that the uncertainty in $\bar{q}_m$ is dominated by the uncertainty due to the randomly sampled sequences. This assumption is however not fundamental but is related to classical hardware control of the experimenter. In the next section we will discuss the validity of this assumption, estimate the number of required single-shot measurements and show how this assumption can be dropped if one wishes to explicitly take into account finite sampling uncertainty.

FINITE SAMPLING STATISTICS

In the previous section it was discussed that the quantity $q_{\mathbf{j}}$ is actually not directly accessible, but must be estimated by performing a large number of single shot measurements. Born's rule states that given a (two-valued) POVM measurement $\{M, \mathbb{1} - M\}$ and a state ρ, the probability of getting outcome 1 (associated with M) is given by $\mathrm{Tr}[M\rho]$ and outcome 0 (associated with $\mathbb{1} - M$) is $1 - \mathrm{Tr}[M\rho]$. This can be used to construct a probability distribution for a single shot measurement of $q_{\mathbf{j}}^{(K)}$, given a fixed sequence indexed by $\mathbf{j}$. The distribution is determined by the definition of $q_{\mathbf{j}}^{(K)}$ and depends on the choice of implementation. Recall that $q_{\mathbf{j}}$ is calculated using the difference of two states $\bar{\rho} = \frac{1}{2}(\rho - \hat{\rho})$. Let us denote $\bar{q}_{\mathbf{j}}$ an unbiased estimator for the exact $q_{\mathbf{j}}$ given a fixed sequence indexed by $\mathbf{j}$. Then there is uncertainty in $\bar{q}_{\mathbf{j}}$ due to the uniformly distributed random sequences $\mathbf{j}$ and due to the fact that $\bar{q}_{\mathbf{j}}$ is itself a random variable for fixed $\mathbf{j}$ (since it is an estimator for the exact $q_{\mathbf{j}}$). The contribution of each source of uncertainty can be quantified by the law of

total variance [12], which states that

$$\begin{aligned}\mathbb{V}[\bar{q}_{\mathbf{j}}] &= \mathbb{E}[\mathbb{V}[\bar{q}_{\mathbf{j}}|\mathbf{j}]] + \mathbb{V}[\mathbb{E}[\bar{q}_{\mathbf{j}}|\mathbf{j}]] \\ &= \mathbb{E}[\mathbb{V}[\bar{q}_{\mathbf{j}}|\mathbf{j}]] + \mathbb{V}[q_{\mathbf{j}}].\end{aligned} \tag{7.32}$$

Here the quantity $\mathbb{V}[\bar{q}_{\mathbf{j}}|\mathbf{j}]$ is referred to as the within-sequence variance (for the given sequence $\mathbf{j}$). It is the variance of the sequence purity $\bar{q}_{\mathbf{j}}$ given fixed $\mathbf{j}$ solely due to the finite sampling statistics. The quantity $\mathbb{V}[q_{\mathbf{j}}]$ is the between-sequence variance of $q_{\mathbf{j}}$ and is solely due to the fact that the sequences $\mathbf{j}$ are sampled from a uniform distribution. This equation expresses that the total variance is the sum of the expected within-sequence variance (expected over the uniformly distributed random sequences) and the between-sequence variance. The quantity $\mathbb{V}[q_{\mathbf{j}}]$ was bounded in this work (eq. (7.17)).
To examine the term $\mathbb{E}[\mathbb{V}[\bar{q}_{\mathbf{j}}|\mathbf{j}]]$ in eq. (7.32), an expression or bound on the within-sequence variance $\mathbb{V}[\bar{q}_{\mathbf{j}}|\mathbf{j}]$ as a function of the number of single shot repetitions is required. We will show how this is done for the two-copy implementation, leaving the more cumbersome (but in principle not more difficult) single-copy implementation as an open problem. Define the single shot random variable by x_r, where the subscript r indexes the different single shot realizations (for $r = 1, ..., R$ for some large R), by the following distribution

$$\text{Prob}[x_r = y|\mathbf{j}] = \begin{cases} a(1-b), & \text{if } y = 1, \\ ab + (1-a)(1-b), & \text{if } y = 0, \\ (1-a)b, & \text{if } y = -1. \end{cases} \tag{7.33}$$

Here $a = \text{Tr}[M\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho)]$, $b = \text{Tr}[M\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\hat{\rho})]$ and $M = \frac{1}{2}(\mathbb{1} + E)$ is the POVM element associated with the two-valued measurement E. The outcome $x_r = 1$ is interpreted as measuring a click only for ρ, outcome $x_r = 0$ corresponds to a click for both or neither states and outcome $x_r = -1$ is associated with a click only for $\hat{\rho}$. This is indeed the single-shot outcome measurement outcome of a $q_{\mathbf{j}}^{(2)}$ measurement, since

$$q_{\mathbf{j}}^{(2)} = \mathbb{E}[x_r|\mathbf{j}] = a - b = \text{Tr}[E\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho})]. \tag{7.34}$$

The natural unbiased estimator of $q_{\mathbf{j}}^{(2)}$ is then given by

$$\bar{q}_{\mathbf{j}}^{(2)} = \frac{1}{R}\sum_{r=1}^{R} x_r. \tag{7.35}$$

The within-sequence variance $\mathbb{V}[\bar{q}_{\mathbf{j}}^{(2)}|\mathbf{j}]$ is related to the variance of x_r (which can be computed given the probability distribution eq. (7.33)) using the fact that x_r are i.i.d. and mutually uncorrelated random variables

$$\mathbb{V}[\bar{q}_{\mathbf{j}}^{(2)}|\mathbf{j}] = \mathbb{V}\left[\frac{1}{R}\sum_{r=1}^{R} x_r \middle| \mathbf{j}\right] = \frac{1}{R^2}\sum_{r=1}^{R} \mathbb{V}\left[x_r|\mathbf{j}\right] = \frac{1}{R}\mathbb{V}[x_r|\mathbf{j}].$$

This follows the definition of the variance and linearity of the expected value. The variance of x_r (computed from the distribution eq. (7.33)) is then

$$\mathbb{V}[x_r|\mathbf{j}] = (a(1-a) + b(1-b)) \leq \frac{1}{2}, \tag{7.36}$$

where the upper bound is trivially obtained by maximizing over $0 \leq a, b \leq 1$. The within-sequence variance thus satisfies

$$\mathbb{V}[\bar{q}_{\mathbf{j}}^{(2)}|\mathbf{j}] = \frac{1}{R}(a(1-a) + b(1-b)) \leq \frac{1}{2R}. \tag{7.37}$$

Hence for the two-copy implementation, the total variance is bounded by

$$\mathbb{V}[\bar{q}_{\mathbf{j}}^{(2)}] \leq \sigma^2 + \frac{1}{2R}, \tag{7.38}$$

where R is the number of single shot measurements taken per sequence and σ^2 is the variance bound of eq. (7.17).
It may seem that the modification of the protocol to use the difference of two states $\bar{\rho}$ means that twice as many single shot measurements must be taken. This is however not the case, as already argued for RB in chapter 6. To see this, let $\mathbb{V}_\rho$ be the variance associated with a single measurement setting on the state ρ. Then for the difference of two states, the variance associated with that measurement satisfies

$$\mathbb{V}_{\bar{\rho}} = \mathbb{V}_{\frac{1}{2}(\rho-\hat{\rho})} \leq \frac{1}{4}(\mathbb{V}_\rho + \mathbb{V}_{\hat{\rho}}) \leq \frac{1}{2}\max(\mathbb{V}_\rho, \mathbb{V}_{\hat{\rho}}). \tag{7.39}$$

So to the contrary, fewer sequences are required to get an accurate estimate of $\mathrm{Tr}[E\mathcal{G}_{\mathbf{j}}(\bar{\rho})]$ than of $\mathrm{Tr}[E\mathcal{G}_{\mathbf{j}}(\rho)]$. This can explicitly be seen in the two-copy implementation, where the within-sequence variance $\mathbb{V}_{\bar{\rho}}[\bar{q}_{\mathbf{j}}|\mathbf{j}]$ was computed in eq. (7.37). However, if only a single state ρ were used, then $\mathrm{Prob}[x_r = 1] = a$ and $\mathrm{Prob}[x_r = -1] = 1 - a$. Therefore the variance $\mathbb{V}_\rho[\bar{q}_{\mathbf{j}}|\mathbf{j}] = \frac{1}{R}\mathbb{V}_\rho[x_r|\mathbf{j}] = \frac{4a(1-a)}{R} \leq \frac{1}{R}$, which is indeed a factor 2 larger than in eq. (7.37).

7

The unbiased estimator of the sequence purity in the single-copy implementation

In the single-copy implementation care must be taken in defining an appropriate estimator of $q_{\mathbf{j}}^{(1)}$. Analogously to the above, one can define a random variable x_r^{PQ} associated with a single shot measurement of $\mathrm{Tr}[E_{\mathcal{H}}^{(Q)}\mathcal{G}_{\mathbf{j}}(\bar{\rho}_{\mathcal{H}}^{(P)})]$ for a fixed sequence indexed by $\mathbf{j}$, depending on the Pauli's P and Q. Then

$$\mathbb{E}[x_r^{PQ}|\mathbf{j}] = \mathrm{Tr}[E_{\mathcal{H}}^{(Q)}\mathcal{G}_{\mathbf{j}}(\bar{\rho}_{\mathcal{H}}^{(P)})], \tag{7.40}$$

so that

$$q_{\mathbf{j}}^{(1)} = \frac{1}{d^2-1}\sum_{P,Q\neq\mathbb{1}} \mathbb{E}[x_r^{PQ}|\mathbf{j}]^2. \tag{7.41}$$

If we denote $\bar{x}_{PQ} = \frac{1}{R}\sum_{r=1}^{R} x_r^{PQ}$, then one could try to estimate $q_{\mathbf{j}}^{(1)}$ by $\bar{q}_{\mathbf{j}}^{(1)} = \frac{1}{d^2-1}\sum_{P,Q\neq\mathbb{1}} \bar{x}_{PQ}^2$. This estimate is however biased, and overestimates the actual value of $q_{\mathbf{j}}^{(1)}$, since

$$\begin{aligned}\mathbb{E}[\bar{x}_{PQ}^2|\mathbf{j}] &= \mathbb{E}[\bar{x}_{PQ}|\mathbf{j}]^2 + \mathbb{V}[\bar{x}_{PQ}|\mathbf{j}] \\ &= \mathbb{E}[\bar{x}_{PQ}|\mathbf{j}]^2 + \frac{1}{R}\mathbb{V}[x_r^{PQ}|\mathbf{j}].\end{aligned} \tag{7.42}$$

To remedy this, one can make use of the unbiased estimator

$$\bar{q}_{\mathbf{j}}^{(1)} = \frac{1}{d^2-1} \sum_{P,Q \neq \mathbb{1}} \bar{x}_{PQ}^2 - \frac{1}{R} s_{PQ}^2, \tag{7.43}$$

where

$$s_{PQ}^2 = \frac{1}{R-1} \sum_{r=1}^{R} (x_r^{PQ} - \bar{x}_{PQ}) \tag{7.44}$$

is the unbiased estimate of $\mathbb{V}[x_r^{PQ}|\mathbf{j}]$. It is important to take this into consideration when performing a Clifford URB experiment using the single-copy implementation, since overestimating $q_{\mathbf{j}}^{(1)}$ can lead to an overestimate of the unitarity obtained from the experiment.

7.4.2. FIT MODEL AND VARIANCE EXPRESSION

In this section we first briefly review the derivation of the fit model of URB (as derived in [2]), slightly adapted with our modification of a traceless input operator $\bar{\rho}$. Then we derive an expression for the variance of the sequence purity. For this it will be more convenient to work in the Liouville representation, see chapter 2.

Using this notation, the expected value of the sequence purity $\mathbb{E}[q_{\mathbf{j}}]$ can be written as

$$\begin{aligned} \mathbb{E}[q_{\mathbf{j}}] &= \frac{1}{|\mathsf{C}_q|^m} \sum_{\mathbf{j}} \langle\langle E | \mathcal{G}_{\mathbf{j}}^{\otimes 2} | \bar{\rho} \rangle\rangle \\ &= \langle\langle E | \left(\mathcal{G}_{\text{avg}}^{(2)} \mathcal{E}^{\otimes 2} \right)^m | \bar{\rho} \rangle\rangle, \end{aligned} \tag{7.45}$$

where

$$\mathcal{G}_{\text{avg}}^{(n)} = \frac{1}{|\mathsf{C}_q|} \sum_{\mathcal{G} \in \mathsf{C}_q} \mathcal{G}^{\otimes n}. \tag{7.46}$$

The key idea behind deriving the fitting model is that $\mathcal{G}_{\text{avg}}^{(2)}$ is the orthogonal projection onto the vector space $W = \text{Span}\{\mathbb{1}, S\} \subset \mathcal{M}_d$. This is a result from representation theory of finite groups, see lemma 3.2 in chapter 3 for details. It is for this reason that the ideal state and measurement operators of eq. (7.7) are elements of the subspace W. The operators I and S do not form an orthogonal basis for W, but the following orthonormal basis can be constructed

$$B_1 = \frac{\mathbb{1}}{d} = \sigma_0 \otimes \sigma_0, \tag{7.47}$$

$$B_2 = \frac{S - \frac{\mathbb{1}}{d}}{\sqrt{d^2-1}} = \frac{1}{\sqrt{d^2-1}} \sum_{\sigma \in \boldsymbol{\sigma}_q} \sigma \otimes \sigma, \tag{7.48}$$

where σ_0 is the Hilbert-Schmidt normalized identity on $\mathcal{H}$ and $\sigma \in \boldsymbol{\sigma}_q$ are the d^2-1 traceless normalized Pauli operators on $\mathcal{H}$. Since $\mathcal{G}_{\text{avg}}^{(2)}$ is an orthogonal projection, it follows that $(\mathcal{G}_{\text{avg}}^{(2)})^2 = \mathcal{G}_{\text{avg}}^{(2)}$. Therefore we can rewrite

$$\mathbb{E}[q_{\mathbf{j}}] = \langle\langle E | \mathcal{M}^{m-1} \mathcal{E}^{\otimes 2} | \bar{\rho} \rangle\rangle, \tag{7.49}$$

where $\mathcal{M} = \mathcal{G}^{(2)}_{\text{avg}}\mathcal{E}^{\otimes 2}\mathcal{G}^{(2)}_{\text{avg}}$. It can be shown that $\mathcal{M}$ (which as only support on W) has the following matrix entries [2]

$$\mathcal{M} = \begin{bmatrix} 1 & 0 \\ \frac{\|\alpha(\mathcal{E})\|^2}{\sqrt{d^2-1}} & u(\mathcal{E}) \end{bmatrix}, \tag{7.50}$$

in the basis $\{B_1, B_2\}$, with α the non-unitality vector of $\mathcal{E}$. In particular this means that $u(\mathcal{E}) = \langle\langle B_2|\mathcal{E}^{\otimes 2}|B_2\rangle\rangle$, which might not be too surprising in the view of definition 13. Since the input state $\bar{\rho}$ is traceless and quantum channels are trace preserving, eq. (7.49) is evaluated as

$$\mathbb{E}[q_{\mathbf{j}}] = \langle\langle E|B_2\rangle\rangle\langle\langle B_2|\bar{\rho}\rangle\rangle u^{m-1} = Bu^{m-1}, \tag{7.51}$$

where the final channel $\mathcal{E}^{\otimes 2}$ has been absorbed into the state as state preparation error. The robustness to state preparation and measurement errors stems from the fact that every component of $\bar{\rho}$ and E outside the subspace W is projected out by the procedure.
In very similar fashion the variance, defined as $\mathbb{V}[q_{\mathbf{j}}] = \mathbb{E}[q^2_{\mathbf{j}}] - \mathbb{E}[q_{\mathbf{j}}]^2$, can be computed. Using $\text{Tr}[A]^2 = \text{Tr}[A^{\otimes 2}]$, the mixed-product property of the tensor product (i.e. $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$) and linearity, we write

$$\begin{aligned} \mathbb{E}[q^2_{\mathbf{j}}] &= \frac{1}{|\mathsf{C}_q|^m}\sum_{\mathbf{j}}\langle\langle E^{\otimes 2}|\mathcal{G}^{\otimes 4}_{\mathbf{j}}|\bar{\rho}^{\otimes 2}\rangle\rangle \\ &= \langle\langle E^{\otimes 2}|\left(\mathcal{G}^{(4)}_{\text{avg}}\mathcal{E}^{\otimes 4}\right)^m|\bar{\rho}^{\otimes 2}\rangle\rangle \\ &= \langle\langle E^{\otimes 2}|\mathcal{N}^{m-1}\mathcal{E}^{\otimes 4}|\bar{\rho}^{\otimes 2}\rangle\rangle, \end{aligned} \tag{7.52}$$

where $\mathcal{N} = \mathcal{G}^{(4)}_{\text{avg}}\mathcal{E}^{\otimes 4}\mathcal{G}^{(4)}_{\text{avg}}$, using the fact that $\mathcal{G}^{(4)}_{\text{avg}}$ is also an orthogonal projection (lemma 6.2 in chapter 3), and

$$\mathbb{E}[q_{\mathbf{j}}]^2 = \langle\langle E^{\otimes 2}|(\mathcal{M}^{\otimes 2})^{m-1}\mathcal{E}^{\otimes 4}|\bar{\rho}^{\otimes 2}\rangle\rangle. \tag{7.53}$$

Putting it together yields the following expression for the variance

$$\mathbb{V}[q_{\mathbf{j}}] = \langle\langle E^{\otimes 2}|\mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1}|\bar{\rho}^{\otimes 2}\rangle\rangle, \tag{7.54}$$

where the final channel $\mathcal{E}^{\otimes 4}$ has again been absorbed into the state as state preparation error. One of the key ingredients of understanding this expression is finding the subspace onto which $\mathcal{G}^{(4)}_{\text{avg}}$ projects. The next section elaborates on this idea.

7.4.3. SKETCH OF PROOF ON VARIANCE BOUND

In this section we discuss and sketch the main ideas for the proof of our variance bound eq. (7.17). A complete proof is given in section 7.6.3, theorem 1. We actually prove a slightly stronger statement

$$\begin{aligned} \mathbb{V}[q_{\mathbf{j}}] \leq \|\bar{\rho}_{\text{err}}\|_1^2\|\bar{E}_{\text{err}}\|_\infty^2 + \frac{1-u^{2(m-1)}}{1-u^2}(1-u)^2\times \\ \left(\alpha^2\beta^2 c_1(d) + \alpha^2 c_2(d)\|\bar{E}_{\text{err}}\|_\infty^2 + \beta^2 c_3(d)\|\bar{\rho}_{\text{err}}\|_1^2\right), \end{aligned} \tag{7.55}$$

where

$$\alpha = \frac{\mathrm{Tr}[\bar{\rho}_{\mathrm{id}}\bar{\rho}]}{\|\bar{\rho}_{\mathrm{id}}\|_2^2} = (d^2-1)\,\mathrm{Tr}[\bar{\rho}_{\mathrm{id}}\bar{\rho}] \tag{7.56}$$

$$\beta = \frac{\mathrm{Tr}[\bar{E}_{\mathrm{id}}\bar{E}]}{\|\bar{E}_{\mathrm{id}}\|_2^2} = \frac{\mathrm{Tr}[\bar{E}_{\mathrm{id}}\bar{E}]}{d^2-1}. \tag{7.57}$$

These quantities arise in the decomposition of the operators $\bar{\rho}, \bar{E}$ into an ideal and error parts as

$$\bar{\rho} = \alpha\bar{\rho}_{\mathrm{id}} + \bar{\rho}_{\mathrm{err}} \qquad \text{and} \qquad \bar{E} = \beta\bar{E}_{\mathrm{id}} + \bar{E}_{\mathrm{err}}. \tag{7.58}$$

It can be shown that $-1 \leq \alpha, \beta \leq 1$ (see section 7.6.3, lemma 7.7), so that eq. (7.55) indeed implies eq. (7.17). The quantities α, β are generally unknown to the experimenter and therefore easily eliminated from the variance bound. Finally we remark that the bound on the interval length L (given in eq. (7.25)) can also be slightly improved if additional information on α or β is known. See section 7.6.3, lemma 7.8 for precise statement.

Our analysis departs from the expression of the variance eq. (7.54). First let us note that fully characterizing the operator $\mathcal{N}$ seems infeasible. This was possible for the operator $\mathcal{M}$, since it only has support on the 2-dimensional subspace W. The dimension of the support of $\mathcal{N}$ (the dimension of the space onto which $\mathcal{G}_{\mathrm{avg}}^{(4)}$ projects) is however given by [13–15]

$$|\,\mathrm{Rge}(\mathcal{N})| = \begin{cases} 15 & \text{if } d = 2; \\ 29 & \text{if } d = 4; \\ 30 & \text{otherwise.} \end{cases} \tag{7.59}$$

Therefore calculating the $|\,\mathrm{Rge}(\mathcal{N})|^2$ matrix entries of $\mathcal{N}$ seems infeasible. A different approach is thus needed. We use a telescoping series expansion, also used in chapter 6 (see lemma 6.6 therein)

$$\mathcal{N}^m - (\mathcal{M}^{\otimes 2})^m = \sum_{s=1}^{m} \mathcal{N}^{m-s}[\mathcal{N} - \mathcal{M}^{\otimes 2}](\mathcal{M}^{\otimes 2})^{s-1} \tag{7.60}$$

in eq. (7.54). The main idea of this is to study the middle operator $\mathcal{N} - \mathcal{M}^{\otimes 2}$ carefully and sharply bound the relevant matrix entries of this operator. The action of $(\mathcal{M}^{\otimes 2})^{s-1}$ is well understood because the full 2-by-2 matrix description of $\mathcal{M}$ is known (given in eq. (7.50)). Finally the action of the remaining higher powers $\mathcal{N}^{m-s-1}$ are bounded more trivially, since less information in computed about $\mathcal{N}$. Let us make these ideas more precise now.

In the previous it was discussed that the operator $\mathcal{M}$ only has support on the subspace $W = \mathrm{Span}\{\mathbb{1}, S\} = \mathrm{Span}\{B_1, B_2\}$. Therefore the analysis of the variance expression is quite different for the components of $\bar{\rho}$ and E on the subspace W and its orthogonal complement. In fact, this lead to the decomposition of the operators $\bar{\rho}, \bar{E}$ into an ideal and error parts as

$$\bar{\rho} = \alpha\bar{\rho}_{\mathrm{id}} + \bar{\rho}_{\mathrm{err}} \qquad \text{and} \qquad \bar{E} = \beta\bar{E}_{\mathrm{id}} + \bar{E}_{\mathrm{err}}, \tag{7.61}$$

where the bar over E indicates its traceless component. In fact, the identity component of E does not contribute at all to $q_{\mathbf{j}}$ (and therefore to its mean and variance), because the input operator is traceless and all applied maps $\mathcal{G}_{\mathbf{j}}$ are trace preserving. So the traceless ideal

components are in the traceless subspace of W (spanned by B_2) and the error components are in the orthogonal complement $W^\perp$. In principle, plugging the above expansion into eq. (7.54) yields 16 different terms after distributing the tensor powers in $\bar{\rho}$ and E over the sum. However, 12 factors containing mixed tensor products of ideal and error components (e.g. $\bar{E}_{\text{id}} \otimes \bar{E}_{\text{err}}$) vanish. This is due to the structure of the space onto which $\mathcal{G}_{\text{avg}}^{(4)}$ projects (see section 7.6.2 for more details). Thus we expand eq. (7.54) as

$$\mathbb{V}[q_{\mathbf{j}}] = \alpha^2\beta^2\langle\langle \bar{E}_{\text{id}}^{\otimes 2}|\mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1}|\bar{\rho}_{\text{id}}^{\otimes 2}\rangle\rangle \tag{7.62}$$

$$+\alpha^2\langle\langle \bar{E}_{\text{err}}^{\otimes 2}|\mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1}|\bar{\rho}_{\text{id}}^{\otimes 2}\rangle\rangle \tag{7.63}$$

$$+\beta^2\langle\langle \bar{E}_{\text{id}}^{\otimes 2}|\mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1}|\bar{\rho}_{\text{err}}^{\otimes 2}\rangle\rangle \tag{7.64}$$

$$+\langle\langle \bar{E}_{\text{err}}^{\otimes 2}|\mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1}|\bar{\rho}_{\text{err}}^{\otimes 2}\rangle\rangle. \tag{7.65}$$

Each of these terms is bounded separately. Here we will demonstrate the ideas of our proof using the term of eq. (7.63). The two terms eq. (7.62) and eq. (7.64) are similar (only a few technical details are different, see the theorem 1 in section 7.6.3 for precise treatment of all terms). Using the telescoping series eq. (7.60) term eq. (7.63) can be written as

$$\begin{aligned}(7.63) &= \alpha^2 \sum_{s=1}^{m-1} \langle\langle \bar{E}_{\text{err}}^{\otimes 2}|\mathcal{N}^{m-s-1}[\mathcal{N} - \mathcal{M}^{\otimes 2}](\mathcal{M}^{\otimes 2})^{s-1}|\bar{\rho}_{\text{id}}^{\otimes 2}\rangle\rangle \\ &= \alpha^2 \sum_{s=1}^{m-1} u^{2(s-1)} \langle\langle \bar{E}_{\text{err}}^{\otimes 2}|\mathcal{N}^{m-s-1}[\mathcal{N} - \mathcal{M}^{\otimes 2}]|\bar{\rho}_{\text{id}}^{\otimes 2}\rangle\rangle,\end{aligned} \tag{7.66}$$

where the second line follows from the fact that $\mathcal{M}\,|B_2\rangle\rangle = u\,|B_2\rangle\rangle$ and $\bar{\rho}_{\text{id}} = \frac{1}{\sqrt{d^2-1}} B_2$ (see eq. (7.105) in section 7.6.3). The next step is analyzing

$$\mathcal{N} - \mathcal{M}^{\otimes 2}\,|\frac{1}{d^2-1}B_2^{\otimes 2}\rangle\rangle = \frac{1}{d^2-1}\sum_i a_i\,|A_i\rangle\rangle \tag{7.67}$$

where $a_i = \langle\langle A_i|\mathcal{N} - \mathcal{M}^{\otimes 2}|B_2^{\otimes 2}\rangle\rangle$ and $|A_i\rangle\rangle$ is a basis for the space $\text{Rge}(\mathcal{G}_{\text{avg}}^{(4)})$ on which $\mathcal{N}$ has support. To find the basis $|A_i\rangle\rangle$ explicitly, the following ideas from representation theory are used.

The map $\mathcal{G} \mapsto \mathcal{G}^{\otimes n}$ is a group representation of the Clifford group C_q for any n. From the discussion on projection formulas in chapter 3 we know that $\mathcal{G}_{\text{avg}}^{(n)}$ is the orthogonal projection onto the trivial subspace of the representation $\mathcal{G} \mapsto \mathcal{G}^{\otimes n}$. For $n = 2$, the trivial subspace was found to be the space W [2], giving rise to the fit model of eq. (7.51). The task at hand here is to find the trivial subspace for $n = 4$. To do so, the following is used. If ϕ is an irreducible, real representation of a group C on a space V, then per the discussion on tensor powers of representations in chapter 3,

$$\text{Span}\{\sum_{v\in V} v \otimes v\} \tag{7.68}$$

carries the only trivial subrepresentation of $\phi \otimes \phi$. This allows us to calculate all trivial subrepresentations of $\mathcal{G} \mapsto \mathcal{G}^{\otimes 4}$, using a complete description of the irreducible representations of $\mathcal{G} \mapsto \mathcal{G}^{\otimes 2}$. These were derived in chapter 5. Therefore eq. (7.68) provides

a method to compute the $|A_i\rangle\!\rangle$ using the explicit description of the irreducible spaces of $\mathcal{G} \mapsto \mathcal{G}^{\otimes 2}$ found in chapter 5.

Hence, the following expression is obtained for eq. (7.63), using the expansion eq. (7.67):

$$(7.63) = \frac{\alpha^2}{d^2-1} \sum_{s=1}^{m-1} u^{2(s-1)} \sum_i a_i \langle\!\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-s-1} | A_i \rangle\!\rangle, \tag{7.69}$$

where $a_i = \langle\!\langle A_i | \mathcal{N} - \mathcal{M}^{\otimes 2} | B_2^{\otimes 2} \rangle\!\rangle$ are the coefficients of the expansion. The factor $\frac{1}{d^2-1}$ is later absorbed into the constant $c_2(d)$ in the final result. Up until this point, equality still holds. Now we are finally in a position to start bounding the term eq. (7.63). To do so, we upper bound each a_i. These bounds involve constants depending on the dimension d (which are all absorbed into $c_2(d)$) and are proportional to $(1-u)^2$. Finally the inner product containing $\mathcal{N}^{m-s-1}$ is upper bounded by a constant depending on the dimension and proportional to $\|\bar{E}_{\text{err}}\|_\infty^2$ (and in particular independent of m or s). This then gives a total bound on the term eq. (7.63),

$$(7.63) \leq \frac{1-u^{2(m-1)}}{1-u^2} (1-u)^2 \alpha^2 c_2(d) \|\bar{E}_{\text{err}}\|_\infty^2, \tag{7.70}$$

where we used the geometric series

$$\sum_{s=1}^{m-1} u^{2(s-1)} = \frac{1-u^{2(m-1)}}{1-u^2}. \tag{7.71}$$

The terms eq. (7.62) and eq. (7.64) can be bounded by repeating all these steps, using a different telescoping series expansion where the factors $(\mathcal{M}^{\otimes 2})^{s-1}$ and $\mathcal{N}^{m-s}$ are interchanged in eq. (7.60). The analysis is then performed by simplifying the inner product from left to right. This involves a few technicalities, but no new ideas. In the end, only the bound on the final inner product with $\mathcal{N}^{m-s-1}$ and the proportionality constants $c_1(d), c_3(d)$ differ, as can be seen from the result eq. (7.55). Finally for the final term eq. (7.65), there is not much more to do than

$$(7.65) = \langle\!\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-1} | \bar{\rho}_{\text{err}}^{\otimes 2} \rangle\!\rangle \leq \|\bar{E}_{\text{err}}\|_\infty^2 \|\bar{\rho}_{\text{err}}\|_1^2, \tag{7.72}$$

using Hölder's inequality and the fact that $\mathcal{N}$ is contractive in the induced trace norm [16], i.e. $\|\mathcal{N}\|_{1\to 1} \leq 1$ (see proposition 7.7 in section 7.6.4).

7.5. CONCLUSION AND FUTURE WORK

In this work we have shown a significant reduction in the required number of random sequences for unitarity randomized benchmarking (URB) than previously could be justified. This reduction is achieved by analyzing the statistics of the protocol. In particular, we have provided a bound on the variance of the sequence purity. Application of a concentration inequality yields the reduction in number of sequences, provided that the variance bound is sharp enough. We have shown that in realistic parameter regimes, the required number of sequences is in the order of hundreds, when benchmarking few-qubit Clifford

gates. This brings benchmarking the unitarity of few-qubit Clifford gates into the realm of experimental feasibility.
The main ingredient of this result was a sharp bound on the variance of the sequence purity. The analysis was done for a slightly modified version of the protocol. This modification leads to better guarantees on the confidence and additionally yields a linear fitting problem. Our variance bound has the attractive property that it scales quadratically in $1 - u$, where u is the unitarity, up to constant contribution due to state preparation and measurement (SPAM) errors. This implies that fewer sequences are required to estimate highly coherent gates. We show that the constant contribution due to SPAM errors is a fundamental property of URB (and therefore not an artifact of our bound). Furthermore our bound is asymptotically independent of the sequence length and is therefore applicable in both short and long sequence lengths. Finally our bound grows exponentially in the number of qubits comprising the system. We argue that this is an artifact of the bound, which could be improved upon. As a result, our bound becomes vacuous for large systems. However, we have shown that our bound is sharp enough to benchmark few-qubit systems (say, up to 5 qubits).
During the analysis of the URB protocol, we have emphasized two different implementation techniques. We have explicitly shown their optimal state preparation and measurement settings for practical implementation. We highlighted the benefits and drawbacks of each implementation and showed the statistical difference between the two.

Future work. There are a few caveats in the analysis of this work, which arise from the assumptions under which the bound holds. Each of these assumptions as summarized in section 7.2 is an open avenue for future research. First and foremost, the assumption of the gate independent error model is rather strong and never completely satisfied in practical implementations of gates. The analysis of the URB protocol so far has been restricted to the gate-independent noise model [2]. There are three somewhat independent open problems with the URB protocol when one wants to generalize the model to (Markovian) gate-dependent errors. First, the behavior of the protocol must be studied. This means that the validity and deviation of the fit model must be studied under this more general noise model. Second, the statistics of the protocol can be studied in the gate-dependent error model. This aims to provide an answer to the question how many resources are required to extract the unitarity from measurement data in this more general noise model, provided that a generalized fit model is found. Finally one can attempt to relate the URB decay rate(s) in the gate-dependent setting to physically relevant quantities (like the unitarity) of the gates comprising the gate set. All three of these problems relating to gate-dependent errors are tough problems and many research focused on answering analogous questions for standard RB. For standard RB, progress has been made in terms of understanding the fit model and relating the decay rate to a physically interpretable infidelity in the gate-dependent error model [17–19]. However, statistical analyses of standard RB only apply to the gate-independent error model [4, 5, 20]. We suspect that some of the progress made in analyzing gate-dependent RB can be modified and applied to URB, but we have left this for future work.
A second interesting avenue is exploring how unitarity randomized benchmarking behaves when the assumption of unitary 2-design is relaxed [21]. This would give rise to a

protocol that can benchmark the unitarity of different gate sets that do not form a 2-design. Interesting examples are the Dihedral group [22, 23], subgroups of monomial unitary matrices [24] and subgroups of the Clifford group [25, 26], where progress have been made for standard RB. Note that the first two of these gate sets are particularly interesting since they contain the T-gate. A general framework for standard RB given an arbitrary gate set is provided in [27]. An interesting open question is whether these techniques can be applied to URB.
Finally it is interesting if the current limitations of our bound can be improved upon. In particular an open question is how to improve this bound to be asymptotically independent of the dimension, a caveat that currently renders our bound impractical for large system ($q \gg 5$). Similarly we wonder if our bound can be generalized to general multi-qubit noise models that need not be unital. These lines of future work could improve the applicability of our bound.

7.6. TECHNICAL STATEMENTS

In this technical section we will prove eq. (7.17) in detail (actually we prove eq. (7.55), which implies eq. (7.17)). This section contains mostly technical statements and their proofs, and gives no insights beyond those already given in previous section. Therefore readers uninterested in technical details may skip this section.
This section is organized as follows. In section 7.6.1 we recall some lemmas which we will need later. In section 7.6.3 then the variance bound of eq. (7.17) is proven. It also contains the proof of the interval of the sequence purity (eq. (7.25)). Finally, all technical lemma's used in the proof of the variance bound are collected in section 7.6.4.

7.6.1. TECHNICAL LEMMA'S FROM LITERATURE

In this section we review a few lemma's from literature that are required for our variance bound. Some lemma's are stated without proof and the reader is then referred to the reference for a proof.
First we present a lemma that bounds the induced Schatten $p \to p$ norm of a quantum channel. We recall the definition of the Schatten p-norms (for $1 \leq p \leq \infty$), which are defined as

$$\|A\|_p^p = \mathrm{Tr}\left[(A^\dagger A)^{\frac{p}{2}}\right] = \|s(A)\|_p^p = \sum_i s_i(A)^p. \tag{7.73}$$

Here $s(A)$ denotes the vector of singular values $s_i(A)$ of A. The Hilbert-Schmidt norm corresponds to $p = 2$. Other important special cases are the trace norm ($p = 1$) and the operator norm to ($p = \infty$). The space of superoperators is typically equipped with the induced Schatten-norms, defined as

$$\|\mathcal{E}\|_{p\to q} = \sup_{A\in\mathcal{M}_d} \{\|\mathcal{E}(A)\|_q : \|A\|_p = 1\}. \tag{7.74}$$

Important special cases are $p = q = 1$, which yields the induced trace norm and $p = q = 2$ which results in the operator norm ($\|\mathcal{E}\|_\infty = \|\mathcal{E}\|_{2\to 2}$).

Lemma 7.1 (Perez-Garcia et al. [16])**.** Let $\mathcal{E}$ be a CPTP quantum channel on a d-dimensional Hilbert space $\mathcal{H}$, with $d = 2^q$ for a q-qubit system. Then for all $p \in [1, \infty]$,

$$\|\mathcal{E}\|_{p\to p} = \max_{A\in\mathcal{M}_d} \{\|\mathcal{E}(A)\|_p : \|A\|_p = 1\} \leq d^{1-\frac{1}{p}} \tag{7.75}$$

and

$$\|\mathcal{E}\|^H_{p\to p} := \max_{A\in\mathcal{M}_d} \left\{\|\mathcal{E}(A)\|_p : \|A\|_p = 1, \mathrm{Tr}[A] = 0, A = A^\dagger\right\} \leq \left(\frac{d}{2}\right)^{1-\frac{1}{p}}. \tag{7.76}$$

If in addition $\mathcal{E}$ is unital ($\mathcal{E}(\mathbb{1}) = \mathbb{1}$), then $\|\mathcal{E}\|_{p\to p} \leq 1$ for all $p \in [1, \infty]$.

The following three lemma's are used to bound the quantities a_i (eq. (7.67)). First we restate lemma 7.2 from chapter 6 in a slightly different form. The proof is the same.

Lemma 7.2. Let $\mathcal{E}$ be a CPTP map on a d-dimensional Hilbert space. Then

$$0 \leq \frac{1}{d^2-1} \sum_{\sigma\in\boldsymbol{\sigma}_q} \langle\langle\sigma|\mathcal{E}|\sigma\rangle\rangle^2 - f^2 \leq \frac{d^2-2}{d^2}(1-f)^2, \tag{7.77}$$

where

$$f = \frac{1}{d^2-1} \sum_{\sigma\in\boldsymbol{\sigma}_q} \langle\langle\sigma|\mathcal{E}|\sigma\rangle\rangle \tag{7.78}$$

is the randomized benchmarking decay parameter of $\mathcal{E}$.

Here this lemma is applied to channels of the form

$$\mathcal{E}_1 = \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{E}_\mathrm{u}\mathcal{E}_\mathrm{u}^\dagger, \end{bmatrix} \qquad \text{and} \qquad \mathcal{E}_2 = \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{E}_\mathrm{u}^\dagger\mathcal{E}_\mathrm{u}, \end{bmatrix} \tag{7.79}$$

where $\mathcal{E}_\mathrm{u}$ is the unital block of the error map $\mathcal{E}$ under investigation, since then $f(\mathcal{E}_1) = f(\mathcal{E}_2) = u(\mathcal{E})$. It is not clear that these superoperators are even a quantum channel (i.e. that they are CPTP). Therefore the following lemma provides a necessary condition on $\mathcal{E}$ for which eq. (7.79) are CPTP maps.

Lemma 7.3. Let $\mathcal{E}$ be a CPTP quantum channel on a d-dimensional Hilbert space. Then the channels $\mathcal{E}_1, \mathcal{E}_2$ defined in eq. (7.79) are CPTP if either $d = 2$ or if $\mathcal{E}$ is unital (or both). Moreover $\|\mathcal{E}_1\|_{2\to2}, \|\mathcal{E}_2\|_{2\to2} \leq 1$.

Proof. If $d = 2$ (that is, if $\mathcal{E}$ is a single-qubit channel), then the unital part of $\mathcal{E}$, defined as

$$\hat{\mathcal{E}} = \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{E}_\mathrm{u} \end{bmatrix}, \tag{7.80}$$

is CPTP [30, Theorem IV.1]. For the general d-dimensional case, it is assumed that $\mathcal{E}$ is unital, so that $\mathcal{E} = \hat{\mathcal{E}}$. So in either case, $\hat{\mathcal{E}}$ is CPTP and unital. It can be shown that the adjoint of a CPTP and unital map is also CPTP and unital [29, Proposition 2.18 and Theorem 2.26], i.e. $\hat{\mathcal{E}}^\dagger$ is CPTP and unital. Therefore $\mathcal{E}_1 = \hat{\mathcal{E}}\hat{\mathcal{E}}^\dagger$ and $\mathcal{E}_2 = \hat{\mathcal{E}}^\dagger\hat{\mathcal{E}}$ are also CPTP and unital. lemma 7.1 then ensures that $\|\mathcal{E}_1\|_{2\to2} \leq 1$ and $\|\mathcal{E}_2\|_{2\to2} \leq 1$. ∎

Third is a lemma from matrix analysis. It is a characterization of positive semi-definite matrices in terms of its principal minors. This lemma was used on $\mathcal{I} - \hat{\mathcal{E}}\hat{\mathcal{E}}^\dagger$ to bound its off-diagonal terms.

Lemma 7.4 (Sylvester's criterion)**.** Let $A \in \mathcal{M}_d$ be a hermitian matrix. Then A is positive semi-definite if and only if all of its principal minors are nonnegative.

Proof. See e.g. [31, Corollary 7.1.5 and Theorem 7.2.5] ■

Next we present two results, also from matrix analysis, that are used several times to bound inner products. The first is a trace inequality and the second is Hölder's inequality.

Lemma 7.5. Let $A, B \in \mathcal{M}_d$ be two linear operators on a d-dimensional Hilbert space $\mathcal{H}$. Denote their singular values as $s_i(A), s_i(B)$ respectively with $i = 1, ..., d$, both in decreasing order. Finally let $s(A)$ and $s(B)$ denote vectors with entries $s_i(A)$ and $s_i(B)$. Then

1. (Von Neumann's trace inequality) $\mathrm{Re}(\mathrm{Tr}[AB]) \leq \sum_{i=1}^d s_i(A)s_i(B)$, and
2. (Hölder's inequality) $\sum_{i=1}^d |s_i(A)s_i(B)| \leq \|\sigma(A)\|_p \|\sigma(B)\|_q = \|A\|_p \|B\|_q$, for any pair $p, q \in [1, \infty]$ such that $p^{-1} + q^{-1} = 1$.

Since singular values are positive, combining the statements yields $\mathrm{Re}(\mathrm{Tr}[AB]) \leq \|A\|_p \|B\|_q$ for any pair $p, q \in [1, \infty]$ such that $p^{-1} + q^{-1} = 1$.

Proof. Statement 1 is proven for example in [31, Theorem 8.7.6] and statement 2 is proven in [32, Theorem 31.3]. ■

Corollary 7.1. If $A, B \in \mathcal{M}_d$ are hermitian, then $\mathrm{Tr}[AB]^* = \mathrm{Tr}[(AB)^\dagger] = \mathrm{Tr}[B^\dagger A^\dagger] = \mathrm{Tr}[BA] = \mathrm{Tr}[AB]$, so that $\mathrm{Tr}[AB]$ is real. Therefore $\mathrm{Tr}[AB] \leq \|A\|_p \|B\|_q$ for any $p, q \in [1, \infty]$ satisfying $p^{-1} + q^{-1} = 1$.

Finally some of our bounds use the fact that the mean of squares is larger than the square of the mean. We show this well-known fact below.

Lemma 7.6 (Mean of squares is larger than square of mean)**.** Let $\{x_i\} \subset \mathbb{R}$ be a collection of N real numbers. Then

$$\left(\frac{1}{N}\sum_{i=1}^N x_i\right)^2 \leq \frac{1}{N}\sum_{i=1}^N x_i^2. \tag{7.81}$$

Proof. By direct computation, it follows that

$$\begin{aligned}
\left(\frac{1}{N}\sum_{i=1}^N x_i^2\right) - \left(\frac{1}{N}\sum_{i=1}^N x_i\right)^2 &= \left(\frac{1}{N}\sum_{i=1}^N x_i^2\right) - 2\left(\frac{1}{N}\sum_{i=1}^N x_i\right)\left(\frac{1}{N}\sum_{k=1}^N x_k\right) + \left(\frac{1}{N}\sum_{k=1}^N x_k\right)^2 \\
&= \frac{1}{N}\sum_{i=1}^N \left(x_i^2 - 2x_i\left(\frac{1}{N}\sum_{k=1}^N x_k\right) + \left(\frac{1}{N}\sum_{k=1}^N x_k\right)^2\right) \\
&= \frac{1}{N}\sum_{i=1}^N \left(x_i - \left(\frac{1}{N}\sum_{k=1}^N x_k\right)\right)^2 \geq 0,
\end{aligned} \tag{7.82}$$

since it is the sum of real numbers squared, proving the result. ■

7.6.2. Trivial subrepresentations of the tensor-4 Liouville representation of the Clifford group

This section is concerned with presenting the trivial subrepresentations of the representation $G \mapsto \mathcal{G}^{\otimes 4}$ of the Clifford group C_q. This representation is equivalent to $G \mapsto \mathcal{G}^{\otimes 4}$ by the intertwining isomorphism $|\cdot\rangle\!\rangle$. Therefore both are considered the same and with slight abuse of notation we refer to them both as the same representation, which we will call the four-copy representation or tensor-4 Liouville representation.

The key idea is to apply lemma 3.2 and its corollaries to find the trivial subrepresentations of the tensor-4 representation $G \mapsto \mathcal{G}^{\otimes 4}$. This requires a full description of the Liouville tensor-2 representation (or two-copy representation) $G \mapsto \mathcal{G}^{\otimes 2}$ in terms of its irreducible components, which we discussed in chapter 5 Let us denote $V = \mathcal{M}_d$ as the space that carries the tensor-2 representation. The present problem is therefore to find the trivial subrepresentations of $\mathcal{G}^{\otimes 4}$, given a decomposition of $\mathcal{G}^{\otimes 2}$ into irreducible representations. In an earlier result [13] the multiplicity of the trivial representation in $\mathcal{G}^{\otimes 4}$ was calculated. They found that

$$|(\mathcal{G}^{\otimes 4})_q^{\mathsf{C}}| = \begin{cases} 15 & \text{if } d = 2; \\ 29 & \text{if } d = 4; \\ 30 & \text{otherwise,} \end{cases} \tag{7.83}$$

which is a justification of eq. (7.59) in the main text.

We continue with a quick review on the work done in chapter 5 as we will need these results here. The two-copy representation of the Clifford group $\mathcal{G}^{\otimes 2}$ decomposes into a number of irreducible representations, carried by subspaces summarized in table 7.2. First, the representation $\mathcal{G}^{\otimes 2}$ decomposes into subrepresentations carried by the spaces

$$\begin{aligned} V_S &:= \mathrm{Span}\{\frac{\sigma\tau + \tau\sigma}{\sqrt{2}} : \sigma, \tau \in \boldsymbol{\sigma}_q, \sigma \neq \tau\}, \\ V_A &:= \mathrm{Span}\{\frac{\sigma\tau - \tau\sigma}{\sqrt{2}} : \sigma, \tau \in \boldsymbol{\sigma}_q, \sigma \neq \tau\}, \\ V_d &:= \mathrm{Span}\{\sigma\sigma : \sigma \in \boldsymbol{\sigma}_q\}, \\ V_{r,l} &:= \mathrm{Span}\{\sigma_0\sigma, \sigma\sigma_0 : \sigma \in \boldsymbol{\sigma}_q\}, \\ V_{\mathrm{id}} &:= \mathrm{Span}\{B_1 = \sigma_0\sigma_0\}. \end{aligned} \tag{7.84}$$

Recall that the tensor symbol is omitted for brevity (so $\sigma\tau$ means $\sigma \otimes \tau$ here). Each of these spaces carries a subrepresentation and furthermore $V = V_{\mathrm{id}} \oplus V_{r,l} \oplus V_d \oplus V_S \oplus V_A$. Finally let us define the traceless, symmetric subspace as

$$V_{TS} := V_S \oplus V_d. \tag{7.85}$$

Since the ideal input and measurement operators for the URB protocol $\bar{\rho}_{\mathrm{id}}, \bar{E}_{\mathrm{id}}$ (as defined in eq. (7.7), see also eq. (7.105)) are elements of V_{TS} and since $\mathcal{E}^{\otimes 2}(V_{TS}) \subseteq V_{TS}$ by the trace-preserving property of $\mathcal{E}$ and the symmetry with respect to swapping the two copies of $\mathcal{H}$, the only relevant subspace of V is V_{TS}. Therefore we continue our analysis of V_{TS}.

Table 7.2: Hierarchy of subspaces contained within the traceless, symmetric subspace V_{TS}, carrying the relevant subrepresentation of the Liouville tensor-4 representation $G \mapsto \mathcal{G}^{\otimes 4}$. In the second and third row, the spaces are broken into a direct sum of subspaces (each of which also carry a subrepresentation), summing to the complete parent space in the row above it. Definitions of all of these spaces are given in the main text (eq. (7.84), eq. (7.85) and eq. (7.86)). The third row spaces V_0, $V_{1,2}$ and V_S are irreducible if $q = 1$. The fourth row gives the final decomposition into irreducible representations for $q \geq 2$. These spaces are not explicitly defined in this text (see [15] for their definitions). The last row gives the dimensions of the irreducible representations. If $|V_i| = 0$ for certain $d = 2$ and/or $d = 4$, this means that the space is empty, i.e. not present in the decomposition. Adding the sizes of the decompositions together, yields the following sizes for the decomposable spaces: $|V_{1,2}| = d^2 - 2$, $|V_d| = d^2 - 1$, $|V_S| = \frac{1}{2}(d^2-1)(d^2-2)$ and $|V_{TS}| = \frac{1}{2}d^2(d^2-1)$.

V_{TS}							
V_d			V_S				
V_0	$V_{1,2}$						
V_0	V_1	V_2	$V_{\{1\}}$	$V_{\{2\}}$	$V_{[1]}$	$V_{[2]}$	$V_{[\text{adj}]}$
1	$\frac{d(d+1)-2}{2}$	$\frac{d(d-1)-2}{2}$	$\frac{(d^2-1)d(d+2)}{8}$	$\frac{(d^2-1)d(d-2)}{8}$	$(d^2-1)\left(\frac{d(d+2)}{8}-1\right)$	$(d^2-1)\left(\frac{d(d-2)}{8}-1\right)$	d^2-1

The space V_d can be broken up into the two subrepresentations

$$V_0 := \text{Span}\left\{ B_2 = \frac{1}{d^2-1}\sum_{\sigma\in\boldsymbol{\sigma}_q} \sigma\sigma \right\} \qquad \text{and} \qquad V_{1,2} := V_d \setminus V_0. \tag{7.86}$$

In the single-qubit case ($q = 1$), the spaces V_S and $V_{1,2}$ are irreducible, therefore fully characterizing $V_{TS} = V_0 \oplus V_{1,2} \oplus V_S$. However, if $q \geq 2$ the space $V_{1,2}$ breaks into 2 irreps, indexed by the index set $\mathcal{Z}_{1,2}$. For $q = 2$, V_S breaks into 4 irreps, while for $q \geq 3$ it breaks into 5 irreps, which will be indexed by $\mathcal{Z}_S$. So the space V_{TS} breaks up into the following number of irreps

$$|\mathcal{Z}_{TS}| = \begin{cases} 3 & \text{if } q = 1; \\ 7 & \text{if } q = 2; \\ 8 & \text{if } q \geq 3, \end{cases} \tag{7.87}$$

where $\mathcal{Z}_{TS} := \mathcal{Z}_d \cup \mathcal{Z}_S = \{0\} \cup \mathcal{Z}_{1,2} \cup \mathcal{Z}_S$. A summary of all the subspaces of V_{TS} that carry subrepresentations is given in table 7.2, together with the dimensions of the spaces. Now we will use lemma 3.2 to connect the irreducible representations of the tensor-2 Liouville representation given above with the trivial representations of the tensor-4 trivial subrepresentations. Let $\mathcal{B}_i$ denote an orthonormal basis for V_i, for $i \in \mathcal{Z}_{TS}$. Then since all irreps indexed by $\mathcal{Z}_{TS}$ are mutually inequivalent, lemma 3.2 gives an explicit way to compute the trivial subrepresentations the representation carried by $(V_{TS} \otimes V_{TS})$ as

$$A_i = \frac{1}{\sqrt{|V_i|}} \sum_{v_i \in \mathcal{B}_i} v_i v_i, \quad \forall i \in \mathcal{Z}_{TS}, \tag{7.88}$$

where the normalization constant is to normalize A_i with respect to the Hilbert-Schmidt norm $\|A_i\|_2 = 1$. In the multi-qubit case where $V_{1,2}$ and V_S are not irrep, it is still useful to define

$$A_j = \frac{1}{\sqrt{|V_j|}} \sum_{i \in \mathcal{Z}_j} \sqrt{|V_i|} A_i, \qquad j \in \{S;\, d;\, 1, 2\}. \tag{7.89}$$

In fact, this allows us to explicitly find $A_{1,2}$ from A_d and A_0. Using the basis for V_0, V_d and V_S (in eq. (7.84) and eq. (7.86)), we therefore explicitly find

$$A_0 := B_2 B_2 = \frac{1}{d^2-1} \sum_{\sigma,\tau \in \boldsymbol{\sigma}_q} \sigma\sigma\tau\tau, \tag{7.90}$$

$$A_{1,2} := \frac{1}{\sqrt{d^2-2}} \left(\sum_{\sigma \in \boldsymbol{\sigma}_q} \sigma^{\otimes 4} - A_0 \right), \tag{7.91}$$

$$A_S := \sqrt{\frac{1}{2(d^2-1)(d^2-2)}} \sum_{\substack{\sigma,\tau \in \boldsymbol{\sigma}_q \\ \sigma \neq \tau}} \sigma\tau\sigma\tau + \sigma\tau\tau\sigma. \tag{7.92}$$

No explicit expression is needed for any $i \in \mathcal{Z}_S$ or $i \in \mathcal{Z}_{1,2}$ if V_S and $V_{1,2}$ are reducible (which happens in the multi-qubit case), because bounds are defined in terms of A_S and $A_{1,2}$. The only exception to this is $i = [\mathrm{adj}] \in \mathcal{Z}_S$. The space $V_{[\mathrm{adj}]} \subset V_S$, which carries an irrep, is defined by (see chapter 5)

$$V_{[\mathrm{adj}]} = \mathrm{Span} \left\{ v_\tau^{[\mathrm{adj}]} = \frac{1}{2\sqrt{|\mathbf{C}_\tau|}} \sum_{\sigma \in \mathbf{C}_\tau} \sigma(\sigma \cdot \tau) + (\sigma \cdot \tau)\sigma \Big| \tau \in \boldsymbol{\sigma}_q \right\}, \tag{7.93}$$

where $\cdot$ indicates the normalized matrix product and where $\mathbf{C}_\tau$ is the set of all elements of $\boldsymbol{\sigma}_q$ that commute with τ as defined in eq. (5.2) in chapter 5 The corresponding trivial subrepresentation, as computed using eq. (7.88), is

$$A_{[\mathrm{adj}]} = \frac{1}{2(d^2-4)\sqrt{d^2-1}} \sum_{\tau \in \boldsymbol{\sigma}_q} \left(\sum_{\sigma \in \mathbf{C}_\tau} (\sigma \cdot \tau)\sigma + \sigma(\sigma \cdot \tau) \right)^{\otimes 2}. \tag{7.94}$$

In the next section, we use the trivial subrepresentations of the Liouville tensor-4 representation to prove our variance bound.

7.6.3. STATEMENT AND PROOF OF THE VARIANCE BOUND AND INTERVAL LENGTH BOUND

In this section we will state and prove our main theorem on the variance bound and prove the interval in which the average sequence purity is found. We also show the optimality of the ideal input and measurement operators. First, we will recapture some of the most important definitions and results discussed in the main text. The point of departure is the expression for the variance of eq. (7.54) derived in the main text

$$\mathbb{V}[q_{\mathbf{j}}] = \langle\!\langle \bar{E}^{\otimes 2} | \mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1} | \bar{\rho}^{\otimes 2} \rangle\!\rangle, \tag{7.95}$$

where the operators are defined as

$$\mathcal{M} := \mathcal{G}_{\mathrm{avg}}^{(2)} \mathcal{E}^{\otimes 2} \mathcal{G}_{\mathrm{avg}}^{(2)}, \qquad \mathcal{N} := \mathcal{G}_{\mathrm{avg}}^{(4)} \mathcal{E}^{\otimes 4} \mathcal{G}_{\mathrm{avg}}^{(4)}, \qquad \mathcal{G}_{\mathrm{avg}}^{(n)} := \frac{1}{|\mathsf{C}_q|} \sum_{\mathcal{G} \in \mathsf{C}_q} \mathcal{G}^{\otimes n}. \tag{7.96}$$

Here $q_{\mathbf{j}}$ is the sequence purity due to the sequence $\mathbf{j}$. As discussed in the main text, $\mathcal{M}$ only has support on the space $W = \mathrm{Span}\{B_1, B_2\} \subset \mathcal{M}_d$ [2], where

$$B_1 = \frac{\mathbb{1}}{d} = \sigma_0\sigma_0, \tag{7.97}$$

$$B_2 = \frac{S - B_1}{\sqrt{d^2-1}} = \frac{1}{\sqrt{d^2-1}} \sum_{\sigma \in \boldsymbol{\sigma}_q} \sigma\sigma. \tag{7.98}$$

In particular the matrix elements of $\mathcal{M}$ with respect to this basis (see also eq. (7.50) in the main text) as

$$\mathcal{M} = \begin{bmatrix} 1 & 0 \\ \frac{\|\alpha(\mathcal{E})\|^2}{\sqrt{d^2-1}} & u(\mathcal{E}) \end{bmatrix}. \tag{7.99}$$

From this it follows that

$$\mathcal{M}\,|B_2\rangle\!\rangle = u\,|B_2\rangle\!\rangle, \tag{7.100}$$

which implies that $\langle\!\langle B_2|\mathcal{M}|B_2\rangle\!\rangle = \langle\!\langle B_2|\mathcal{E}^{\otimes 2}|B_2\rangle\!\rangle = u$, since $\mathcal{G}^{(2)}_{\mathrm{avg}}\,|B_2\rangle\!\rangle = |B_2\rangle\!\rangle$ and B_2 is normalized. This is used in the analysis of eq. (7.95).
In eq. (7.95) the measurement E is replaced with its the traceless counterpart $\bar{E}$, which is defined as

$$\bar{E} := E - \frac{\mathrm{Tr}[E]}{d^2}\mathbb{1} = E - \langle\!\langle B_1|E\rangle\!\rangle B_1. \tag{7.101}$$

Since $\bar{\rho}$ is traceless by construction and $\mathcal{G}_{\mathbf{j}}$ is trace-preserving, it follows that $q_{\mathbf{j}} = \langle\!\langle E|\mathcal{G}_{\mathbf{j}}^{\otimes 2}|\bar{\rho}\rangle\!\rangle = \langle\!\langle \bar{E}|\mathcal{G}_{\mathbf{j}}^{\otimes 2}|\bar{\rho}\rangle\!\rangle$. This justifies the replacement of E by $\bar{E}$ is all expectation value and variance expressions. In our analysis it is advantageous to think of $\bar{E}$ instead of E, since then $\bar{E}_{\mathrm{id}}, \bar{\rho}_{\mathrm{id}} \propto B_2$. The ideal state and measurement operators were defined in eq. (7.7). For completeness, they are

$$E_{\mathrm{id}} = S = B_1 + \sqrt{d^2-1}B_2, \tag{7.102}$$

$$\rho_{\mathrm{id}} = \frac{\mathbb{1}+S}{d(d+1)} = \frac{1}{d}B_1 + \frac{\sqrt{d^2-1}}{d(d+1)}B_2, \tag{7.103}$$

$$\hat{\rho}_{\mathrm{id}} = \frac{\mathbb{1}-S}{d(d-1)} = \frac{1}{d}B_1 - \frac{\sqrt{d^2-1}}{d(d-1)}B_2, \tag{7.104}$$

from which it follows that

$$\bar{E}_{\mathrm{id}} = \sqrt{d^2-1}B_2 \qquad \text{and} \qquad \bar{\rho}_{\mathrm{id}} = \frac{\rho_{\mathrm{id}} - \hat{\rho}_{\mathrm{id}}}{2} = \frac{1}{\sqrt{d^2-1}}B_2. \tag{7.105}$$

The implemented operators $\bar{\rho}$ and E can then be decomposed into an ideal part and an error part as

$$\alpha := \frac{\langle\!\langle \bar{\rho}_{\mathrm{id}}|\bar{\rho}\rangle\!\rangle}{\langle\!\langle \bar{\rho}_{\mathrm{id}}|\bar{\rho}_{\mathrm{id}}\rangle\!\rangle} = (d^2-1)\langle\!\langle \bar{\rho}_{\mathrm{id}}|\bar{\rho}\rangle\!\rangle, \qquad \bar{\rho}_{\mathrm{err}} := \bar{\rho} - \alpha\bar{\rho}_{\mathrm{id}}, \tag{7.106}$$

$$\beta := \frac{\langle\!\langle \bar{E}_{\mathrm{id}}|\bar{E}\rangle\!\rangle}{\langle\!\langle \bar{E}_{\mathrm{id}}|\bar{E}_{\mathrm{id}}\rangle\!\rangle} = \frac{1}{d^2-1}\langle\!\langle \bar{E}_{\mathrm{id}}|\bar{E}\rangle\!\rangle, \qquad \bar{E}_{\mathrm{err}} := \bar{E} - \beta\bar{E}_{\mathrm{id}}. \tag{7.107}$$

This decomposition is chosen such that $\mathrm{Tr}[\bar{\rho}_{\mathrm{id}}\bar{\rho}_{\mathrm{err}}] = \mathrm{Tr}[\bar{E}_{\mathrm{id}}\bar{E}_{\mathrm{err}}] = 0$. It can be shown that the ideal operators $\bar{\rho}_{\mathrm{id}}$, $\bar{E}_{\mathrm{id}}$ are in fact ideal, in the sense that they maximize the prefactor B in the fit model $\mathbb{E}[q_{\mathbf{j}}] = Bu^{m-1}$ (and also minimize the variance as we will see). The prefactor B is given by (see eq. (7.51) of the main text)

$$B = \langle\!\langle E|\mathcal{G}^{(2)}_{\mathrm{avg}}|\bar{\rho}\rangle\!\rangle = \langle\!\langle \bar{E}|B_2\rangle\!\rangle\langle\!\langle B_2|\bar{\rho}\rangle\!\rangle = \alpha\beta. \tag{7.108}$$

The ideal operators $\bar{\rho}_{\mathrm{id}}$, $\bar{E}_{\mathrm{id}}$ will yield $B = 1$. The following lemma shows that this is in fact optimal.

Lemma 7.7 (Optimality of ideal operators)**.** The prefactor B in the fit model for URB as given in eq. (7.108) satisfies $|B| \leq 1$ for all input and measurement operators $\bar{\rho}, E$.

Proof. Let us write the two-valued measurement E with outcomes ± 1 in terms of its POVM elements $\{M, \mathbb{1} - M\}$, so that $E = M - (\mathbb{1} - M) = 2M - \mathbb{1}$. By definition M satisfies $0 \leq M \leq \mathbb{1}$. Since $\mathcal{G}^{(2)}_{\mathrm{avg}}$ is a CPTP map and $\rho, \hat{\rho} \geq 0$ are quantum states, it follows that $\mathcal{G}^{(2)}_{\mathrm{avg}}(\rho), \mathcal{G}^{(2)}_{\mathrm{avg}}(\hat{\rho}) \geq 0$. Using the fact that $\mathrm{Tr}[AB] \geq 0$ for all positive semi-definite operators $A, B \geq 0$, it follows that

$$0 = \langle\!\langle 0|\mathcal{G}^{(2)}_{\mathrm{avg}}|\rho\rangle\!\rangle \leq \langle\!\langle M|\mathcal{G}^{(2)}_{\mathrm{avg}}|\rho\rangle\!\rangle \leq \langle\!\langle \mathbb{1}|\mathcal{G}^{(2)}_{\mathrm{avg}}|\rho\rangle\!\rangle = 1. \tag{7.109}$$

In terms of the measurement E, this means that $-1 \leq \langle\!\langle E|\mathcal{G}^{(2)}_{\mathrm{avg}}|\rho\rangle\!\rangle \leq 1$. Analogously, this holds for $\hat{\rho}$. Since $\bar{\rho} = \frac{1}{2}(\rho - \hat{\rho})$ is follows that $-1 \leq B = \langle\!\langle E|\mathcal{G}^{(2)}_{\mathrm{avg}}|\bar{\rho}\rangle\!\rangle \leq 1$. ∎

Corollary 7.2. The quantities α, β as defined in eq. (7.106) and eq. (7.107) satisfy $-1 \leq \alpha, \beta \leq 1$.

Proof. lemma 7.7 and eq. (7.108) show that $-1 \leq \alpha\beta \leq 1$ for all $\bar{\rho}, E$. Note that α only depends on $\bar{\rho}$ and β only on E. Therefore if we fix $\bar{\rho} = \bar{\rho}_{\mathrm{id}}$ (which implies $\alpha = 1$), then we have $-1 \leq \beta \leq 1$. Analogously fixing $E = E_{\mathrm{id}}$ (which implies $\beta = 1$) yields $-1 \leq \alpha \leq 1$. ∎

Very similar reasoning also gives the bound on the interval in which the sequence purity $q^{(K)}_{\mathbf{j}}$ lies (see eq. (7.25) of the main text). This bound will be proven in the following lemma.

Lemma 7.8 (Bound on interval lengths)**.** Let $q^{(K)}_{\mathbf{j}}$ denote the sequence purity of the K-copy implementation due to the random sequence $\mathbf{j}$ as defined in eq. (7.8) and eq. (7.6):

$$q^{(1)}_{\mathbf{j}} = \frac{1}{d^2-1}\sum_{P,Q\neq\mathbb{1}} \langle\!\langle E^{(Q)}_{\mathcal{H}}|\mathcal{G}_{\mathbf{j}}|\bar{\rho}^{(P)}_{\mathcal{H}}\rangle\!\rangle^2 \qquad \text{and} \qquad q^{(2)}_{\mathbf{j}} = \langle\!\langle E|\mathcal{G}^{\otimes 2}_{\mathbf{j}}|\bar{\rho}\rangle\!\rangle. \tag{7.110}$$

Assume that $\alpha, \beta \geq 0$ (equivalent to $\mathrm{Tr}[\bar{\rho}_{\mathrm{id}}\bar{\rho}] \geq 0$ and $\mathrm{Tr}[\bar{E}_{\mathrm{id}}\bar{E}] \geq 0$ stated in section 7.2.3). Then for all operators $\bar{\rho}, E$ (which are the effective operators in the single-copy implementation, see eq. (7.10)), all CPTP error maps $\mathcal{E}$ and all sequences of Clifford gates indexed by $\mathbf{j}$,

$$q^{(1)}_{\mathbf{j}} \in [0, \alpha\beta + \beta\|\bar{\rho}_{\mathrm{err}}\|_1 + \alpha\|\bar{E}_{\mathrm{err}}\|_\infty + \|\bar{\rho}_{\mathrm{err}}\|_1\|\bar{E}_{\mathrm{err}}\|_\infty], \tag{7.111}$$

$$q^{(2)}_{\mathbf{j}} \in [-\beta\|\bar{\rho}_{\mathrm{err}}\|_1 - \alpha\|\bar{E}_{\mathrm{err}}\|_\infty - \|\bar{\rho}_{\mathrm{err}}\|_1\|\bar{E}_{\mathrm{err}}\|_\infty, 1]. \tag{7.112}$$

Corollary 7.3. The interval length for $q_{\mathbf{j}}^{(1)}$ and $q_{\mathbf{j}}^{(2)}$ can be bounded independent of α, β by using that $\alpha, \beta \leq 1$ (lemma 7.7) as $L = 1 + \|\bar{\rho}_{\text{err}}\|_1 + \|\bar{E}_{\text{err}}\|_\infty + \|\bar{\rho}_{\text{err}}\|_1 \|\bar{E}_{\text{err}}\|_\infty$.

Proof. Starting with the two-copy implementation, let us write $E = M - (\mathbb{1} - M) = 2M - \mathbb{1}$, where $0 \leq M \leq \mathbb{1}$ is a POVM element (the measurement E is described by the POVM set $\{M, \mathbb{1} - M\}$, assigning outcome 1 to M and -1 to $\mathbb{1} - M$). Then using the fact that $\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho) \geq 0$ is positive semi-definite, it follows that

$$0 = \text{Tr}[0\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho)] \leq \text{Tr}[M\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho)] \leq \text{Tr}[\mathbb{1}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho)] = 1,$$

expressing that $\text{Tr}[M\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho)]$ is indeed the probability associated with obtaining outcome M. Therefore $-1 \leq \text{Tr}[E\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho)] \leq 1$. Exactly the same argument holds for $\hat{\rho}$, so that (recall that $\bar{\rho} = \frac{1}{2}(\rho - \hat{\rho})$)

$$-1 \leq q_{\mathbf{j}}^{(2)} = \text{Tr}[E\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho})] \leq 1. \tag{7.113}$$

The lower bound can be improved by using the decomposition eq. (7.106) and eq. (7.107) to write $\bar{\rho} = \alpha\bar{\rho}_{\text{id}} + \bar{\rho}_{\text{err}}$ and $\bar{E} = \beta\bar{E}_{\text{id}} + \bar{E}_{\text{err}}$. Then

$$q_{\mathbf{j}}^{(2)} = \alpha\beta \,\text{Tr}[\bar{E}_{\text{id}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{id}})] + \alpha \,\text{Tr}[\bar{E}_{\text{err}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{id}})] + \beta \,\text{Tr}[\bar{E}_{\text{id}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{err}})] + \text{Tr}[\bar{E}_{\text{err}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{err}})]. \tag{7.114}$$

The first term satisfies $\text{Tr}[\bar{E}_{\text{id}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{id}})] \leq 1$ by eq. (7.113) (which holds for all E, $\bar{\rho}$ so in particular for E_{id}, $\bar{\rho}_{\text{id}}$). However, we also find that

$$\text{Tr}[\bar{E}_{\text{id}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{id}})] = \text{Tr}[B_2\mathcal{G}_{\mathbf{j}}^{\otimes 2}(B_2)] = \frac{1}{d^2 - 1} \sum_{\sigma,\tau \in \boldsymbol{\sigma}_q} \text{Tr}[\sigma\mathcal{G}_{\mathbf{j}}(\tau)]^2 \geq 0. \tag{7.115}$$

The remaining three terms in eq. (7.114) are bounded using proposition 7.7, which yields (using $\alpha, \beta \geq 0$)

$$\begin{aligned} \alpha|\,\text{Tr}[\bar{E}_{\text{err}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{id}})]| &\leq \alpha\|\bar{E}_{\text{err}}\|_\infty\|\bar{\rho}_{\text{id}}\| = \alpha\|\bar{E}_{\text{err}}\|_\infty \\ \beta|\,\text{Tr}[\bar{E}_{\text{id}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{err}})]| &\leq \beta\|\bar{E}_{\text{id}}\|_\infty\|\bar{\rho}_{\text{err}}\|_1 = \beta\|\bar{\rho}_{\text{err}}\|_1 \\ |\,\text{Tr}[\bar{E}_{\text{err}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{err}})]| &\leq \|\bar{E}_{\text{err}}\|_\infty\|\bar{\rho}_{\text{err}}\|_1 \end{aligned} \tag{7.116}$$

So by combining eq. (7.114), eq. (7.115) and eq. (7.116), we find that

$$q_{\mathbf{j}}^{(2)} \geq 0 - \alpha\|\bar{E}_{\text{err}}\|_\infty - \beta\|\bar{\rho}_{\text{err}}\|_1 - \|\bar{\rho}_{\text{err}}\|_1\|\bar{E}_{\text{err}}\|_\infty. \tag{7.117}$$

The above argument also holds in the single-copy implementation if we let $E = E_{\text{eff}}$ and $\bar{\rho} = \bar{\rho}_{\text{eff}}$ as defined in eq. (7.10) of the main text. However, now we use it to upper bound $q_{\mathbf{j}}^{(1)}$. It follows that

$$\begin{aligned} q_{\mathbf{j}}^{(1)} &= \alpha\beta \,\text{Tr}[\bar{E}_{\text{id}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{id}})] + \alpha \,\text{Tr}[\bar{E}_{\text{err}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{id}})] + \beta \,\text{Tr}[\bar{E}_{\text{id}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{err}})] + \text{Tr}[\bar{E}_{\text{err}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{err}})] \\ &\leq \alpha\beta + \beta\|\bar{\rho}_{\text{err}}\|_1 + \alpha\|\bar{E}_{\text{err}}\|_\infty + \|\bar{\rho}_{\text{err}}\|_1\|\bar{E}_{\text{err}}\|_\infty. \end{aligned} \tag{7.118}$$

The lower bound $q_{\mathbf{j}}^{(1)} \geq 0$ follows directly from the fact that it is defined as the sum of real numbers squared. ∎

So far we have recaptured the essential definitions and notations, shown optimality of the ideal operators and proven a bound in the interval in which the sequence purity $q_{\mathbf{j}}$ lies. Next we will state our variance bound eq. (7.17) and give the complete proof.

Theorem 1 (Variance bound)**.** *Let $\mathcal{H}$ be a d-dimensional Hilbert space, with $d = 2^q$ for a q-qubit system. Let $E \in \mathcal{M}_d^{\otimes 2}$ be the hermitian observable associated with a two-valued measurement with outcomes ± 1 and $\rho, \hat{\rho} \in \mathcal{M}_d^{\otimes 2}$ be two quantum states on two copies of the system. Consider the URB experiment (using the states and measurement $\rho, \hat{\rho}, E$) of the Clifford group C_q, assuming that a noisy implementation of $\mathcal{G} \in \mathsf{C}_q$ is given by $\tilde{\mathcal{G}} = \mathcal{G}\mathcal{E}$, where $\mathcal{E}$ is a CPTP map. In this experiment the sequence purity is $q_{\mathbf{j}} = \langle\!\langle \bar{E} | (\mathcal{M}^{\otimes 2})^{m-1} | \bar{\rho} \rangle\!\rangle$, with $\mathcal{M}$ defined in eq.* (7.96)*.*
Under the assumption that $d = 2$ or $\mathcal{E}$ is unital (that is, $\mathcal{E}(\mathbb{1}) = \mathbb{1}$), the following bound on the variance $\mathbb{V}[q_{\mathbf{j}}]$ holds

$$\mathbb{V}[q_{\mathbf{j}}] \leq \sigma^2 = \frac{1 - u^{2(m-1)}}{1 - u^2}(1-u)^2 \Big(\alpha^2 \beta^2 c_1(d) + \alpha^2 c_2(d) \|\bar{E}_{\mathrm{err}}\|_\infty^2 + \beta^2 c_3(d) \|\bar{\rho}_{\mathrm{err}}\|_1^2 \Big) + \|\bar{\rho}_{\mathrm{err}}\|_1^2 \|\bar{E}_{\mathrm{err}}\|_\infty^2, \tag{7.119}$$

where u is the unitarity of $\mathcal{E}$, m is the length of the sequence indexed by $\mathbf{j}$, $c_i(d)$ are functions only of the dimension d and α, β, $\bar{\rho}_{\mathrm{err}}$ and $\bar{E}_{\mathrm{err}}$ are defined in eq. (7.106) *and eq.* (7.107)*. Precise definitions of the dimension-dependent functions $c_i(d)$ will be given in the proof, but closed form expressions are messy and therefore not written down explicitly. Asymptotically, these functions satisfy*

$$c_1(d) = \mathcal{O}(1), \qquad c_2(d) = \mathcal{O}(d), \qquad c_3(d) = \mathcal{O}(d^2). \tag{7.120}$$

Proof. We start from the derived expression for the variance eq. (7.95). First, let us decompose the state and measurement operators in ideal and error components as (see eqs. (7.105)-(7.107))

$$\bar{\rho} = \alpha \bar{\rho}_{\mathrm{id}} + \bar{\rho}_{\mathrm{err}} \qquad \text{and} \qquad \bar{E} = \beta \bar{E}_{\mathrm{id}} + \bar{E}_{\mathrm{err}}. \tag{7.121}$$

Define again $W = \mathrm{Span}\{B_1, B_2\} \subset \mathcal{M}_d^{\otimes 2}$, with B_1, B_2 defined in eq. (7.97) and eq. (7.98) respectively. Then the ideal components $\bar{\rho}_{\mathrm{id}}$ and $\bar{E}_{\mathrm{id}}$ are in W and the error components $\bar{\rho}_{\mathrm{err}}$ and $\bar{E}_{\mathrm{err}}$ are in the orthogonal complement $W^\perp$. Plugging this expansion into eq. (7.95) in principle yields 16 terms. However, the 12 terms with an ideal component tensor error component (e.g. $\bar{\rho}_{\mathrm{err}} \otimes \bar{\rho}_{\mathrm{id}}$) vanish. This is because both

$$(\mathcal{G}_{\mathrm{avg}}^{(2)})^{\otimes 2}(W \otimes W^\perp) = (\mathcal{G}_{\mathrm{avg}}^{(2)})^{\otimes 2}(W^\perp \otimes W) = \emptyset, \tag{7.122}$$

$$\mathcal{G}_{\mathrm{avg}}^{(4)}(W \otimes W^\perp) = \mathcal{G}_{\mathrm{avg}}^{(4)}(W^\perp \otimes W) = \emptyset. \tag{7.123}$$

eq. (7.122) is easy to see because $\mathcal{G}_{\mathrm{avg}}^{(2)}$ is the orthogonal projection onto W. eq. (7.123) follows from the fact that W carries the trivial subrepresentations of the Liouville tensor-2 representation and $W^\perp$ carries all other necessarily non-trivial subrepresentations. By lemma 3.2 the spaces $W^\perp \otimes W$ and $W \otimes W^\perp$ (which are representations of the Liouville

tensor-4 representation) do not carry trivial subrepresentations. Hence $\mathcal{G}^{(4)}_{\text{avg}}$, the projector onto the trivial subrepresentations of the Liouville tensor-4 representation does not project onto any subspace of $W^\perp \otimes W$ and $W \otimes W^\perp$. This justifies the following expression for the variance

$$\begin{aligned}
\mathbb{V}[q_{\mathbf{j}}] = &\alpha^2\beta^2 \langle\!\langle B_2^{\otimes 2} | \mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1} | B_2^{\otimes 2} \rangle\!\rangle && (7.124)\\
&+ \frac{\alpha^2}{d^2-1} \langle\!\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1} | B_2^{\otimes 2} \rangle\!\rangle && (7.125)\\
&+ (d^2-1)\beta^2 \langle\!\langle B_2^{\otimes 2} | \mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1} | \bar{\rho}_{\text{err}}^{\otimes 2} \rangle\!\rangle && (7.126)\\
&+ \langle\!\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1} | \bar{\rho}_{\text{err}}^{\otimes 2} \rangle\!\rangle, && (7.127)
\end{aligned}$$

where the expressions of eq. (7.105) are used for the ideal operators $\bar{\rho}_{\text{id}}, \bar{E}_{\text{id}}$. We will analyze each of the four terms separately. The term we start with is eq. (7.125), since this term most clearly conveys the idea of our analysis. Then the terms eq. (7.124) and eq. (7.126) are treated in similar fashion, but with a small additional technicality. Finally the term eq. (7.127) is treated in a totally different fashion.

The analysis of eq. (7.125) starts by using lemma 6.6 (telescoping series lemma), so that we can write this term as

$$\begin{aligned}
eq.\ (7.125) &= \frac{\alpha^2}{d^2-1} \sum_{s=1}^{m-1} \langle\!\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-s-1} [\mathcal{N} - \mathcal{M}^{\otimes 2}] (\mathcal{M}^{\otimes 2})^{s-1} | B_2^{\otimes 2} \rangle\!\rangle \\
&= \frac{\alpha^2}{d^2-1} \sum_{s=1}^{m-1} u^{2(s-1)} \langle\!\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-s-1} [\mathcal{N} - \mathcal{M}^{\otimes 2}] | B_2^{\otimes 2} \rangle\!\rangle.
\end{aligned} \tag{7.128}$$

In the second line we used that $\mathcal{M}\,|B_2\rangle\!\rangle = u\,|B_2\rangle\!\rangle$. The idea is to expand $\mathcal{N} - \mathcal{M}^{\otimes 2}\,|B_2^{\otimes 2}\rangle\!\rangle$ in the basis $\{A_i : i \in \mathcal{Z}_{TS}\}$ of the subspace $V_{TS} \otimes V_{TS} \subset \text{Rge}(\mathcal{G}^{(4)}_{\text{avg}}) \subset \mathcal{M}_d^{\otimes 4}$. V_{TS} is the trace-preserving, symmetric subspace of $\mathcal{M}_d^{\otimes 2}$, as defined in section 7.6.2. The restriction of $\mathcal{G}^{(4)}_{\text{avg}}$ to $V_{TS} \otimes V_{TS}$ is justified by the fact that $\mathcal{E}^{\otimes 2}(B_2) \in V_{TS}$. Hence we expand

$$\mathcal{N} - \mathcal{M}^{\otimes 2}\,|B_2^{\otimes 2}\rangle\!\rangle = \sum_{i \in \mathcal{Z}_{TS}} a_i\,|A_i\rangle\!\rangle, \qquad \text{where} \qquad a_i := \langle\!\langle A_i | \mathcal{N} - \mathcal{M}^{\otimes 2} | B_2^{\otimes 2} \rangle\!\rangle. \tag{7.129}$$

Therefore eq. (7.125) can be written as

$$eq.\ (7.125) = \frac{\alpha^2}{d^2-1} \sum_{s=1}^{m-1} u^{2(s-1)} \sum_{i \in \mathcal{Z}_{TS}} a_i \langle\!\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-s-1} | A_i \rangle\!\rangle. \tag{7.130}$$

For the terms eq. (7.124) and eq. (7.126), something similar is done. The telescoping series

(lemma 6.6) is now written in the other way. Therefore we can write eq. (7.124) as

$$eq.\ (7.124) = \alpha^2\beta^2 \sum_{s=1}^{m-1} \langle\langle B_2^{\otimes 2}|(\mathcal{M}^{\otimes 2})^{s-1}[\mathcal{N}-\mathcal{M}^{\otimes 2}]\mathcal{N}^{m-s-1}|B_2^{\otimes 2}\rangle\rangle \tag{7.131}$$

$$= \alpha^2\beta^2 \sum_{s=1}^{m-1} u^{2(s-1)} \langle\langle B_2^{\otimes 2}|[\mathcal{N}-\mathcal{M}^{\otimes 2}]\mathcal{N}^{m-s-1}|B_2^{\otimes 2}\rangle\rangle \tag{7.132}$$

$$\tag{7.133}$$

The step from eq. (7.131) to eq. (7.132) is not immediately clear, since

$$\langle\langle B_2 B_2|\,(\mathcal{M}^{\otimes 2})^{s-1} = x_{11}^{(s)}\langle\langle B_1 B_1| + x_{12}^{(s)}\langle\langle B_1 B_2| + x_{21}^{(s)}\langle\langle B_2 B_1| + u^{2(s-1)}\langle\langle B_2 B_2|\,, \tag{7.134}$$

for some coefficients $x_{11}^{(s)}, x_{12}^{(s)}, x_{21}^{(s)} \in \mathbb{R}$. However we show that eq. (7.132) is justified, since

$$\langle\langle B_k B_l|[\mathcal{N}-\mathcal{M}^{\otimes 2}]\mathcal{N}^{m-s-1}|B_2^{\otimes 2}\rangle\rangle = 0, \qquad \text{if } k=1 \text{ or } l=1. \tag{7.135}$$

This follows from the trace-preserving properties of $\mathcal{N}, \mathcal{M}$, the tracelessness of B_2 and the fact that $B_1 = \frac{\mathbb{1}}{d}$. In particular,

$$\langle\langle B_k B_l|\mathcal{N}^{m-s}|B_2^{\otimes 2}\rangle\rangle = \frac{1}{|\mathsf{C}_q|^{m-s}} \sum_{\mathbf{j}} \langle\langle B_k|\mathcal{G}_{\mathbf{j}}^{\otimes 2}|B_2\rangle\rangle\langle\langle B_l|\mathcal{G}_{\mathbf{j}}^{\otimes 2}|B_2\rangle\rangle = 0, \tag{7.136}$$

$$\langle\langle B_k B_l|\mathcal{M}^{\otimes 2}\mathcal{N}^{m-s-1}|B_2^{\otimes 2}\rangle\rangle = \frac{1}{|\mathsf{C}_q|^{m-s-1}} \sum_{\mathbf{j}} \langle\langle B_k|\mathcal{M}\mathcal{G}_{\mathbf{j}}^{\otimes 2}|B_2\rangle\rangle\langle\langle B_l|\mathcal{M}\mathcal{G}_{\mathbf{j}}^{\otimes 2}|B_2\rangle\rangle = 0, \tag{7.137}$$

if $l = 1$ or $k = 1$, since $\langle\langle B_1|\mathcal{M}\mathcal{G}_{\mathbf{j}}^{\otimes 2}|B_2\rangle\rangle = 0$ and $\langle\langle B_1|\mathcal{G}_{\mathbf{j}}^{\otimes 2}|B_2\rangle\rangle = 0$. This justifies eq. (7.132). Next we use a similar expansion

$$\langle\langle B_2^{\otimes 2}|\,\mathcal{N}-\mathcal{M}^{\otimes 2} = \sum_{i\in\mathcal{Z}_{TS}} b_i\langle\langle A_i|\,, \qquad \text{where} \qquad b_i := \langle\langle B_2^{\otimes 2}|\mathcal{N}-\mathcal{M}^{\otimes 2}|A_i\rangle\rangle. \tag{7.138}$$

Therefore we arrive at

$$eq.\ (7.124) = \alpha^2\beta^2 \sum_{s=1}^{m-1} u^{2(s-1)} \sum_{i\in\mathcal{Z}_{TS}} b_i\langle\langle A_i|\mathcal{N}^{m-s-1}|B_2^{\otimes 2}\rangle\rangle. \tag{7.139}$$

Similarly to the analysis eq. (7.124), we can write eq. (7.126) as

$$eq.\ (7.126) = (d^2-1)\beta^2 \sum_{s=1}^{m-1} u^{2(s-1)} \sum_{i\in\mathcal{Z}_{TS}} b_i\langle\langle A_i|\mathcal{N}^{m-s-1}|\bar{\rho}_{\mathrm{err}}^{\otimes 2}\rangle\rangle. \tag{7.140}$$

Finally, we slightly rewrite eq. (7.127) by noting that $eq.\ (7.127) = \langle\langle \bar{E}_{\mathrm{err}}^{\otimes 2}|\mathcal{N}^{m-1}|\bar{\rho}_{\mathrm{err}}^{\otimes 2}\rangle\rangle$,

because $\mathcal{M}\,|\rho_{\text{err}}\rangle\!\rangle = 0$. We therefore arrive at the following expression of the variance

$$\mathbb{V}[q_{\mathfrak{j}}] = \alpha^2\beta^2 \sum_{s=1}^{m-1} u^{2(s-1)} \sum_{i\in\mathcal{Z}_{TS}} a_i \langle\!\langle B_2^{\otimes 2}|\mathcal{N}^{m-s-1}|A_i\rangle\!\rangle \tag{7.141}$$

$$+ \frac{1}{d^2-1}\alpha^2 \sum_{s=1}^{m-1} u^{2(s-1)} \sum_{i\in\mathcal{Z}_{TS}} a_i \langle\!\langle \bar{E}_{\text{err}}^{\otimes 2}|\mathcal{N}^{m-s-1}|A_i\rangle\!\rangle \tag{7.142}$$

$$+ (d^2-1)\beta^2 \sum_{s=1}^{m-1} u^{2(s-1)} \sum_{i\in\mathcal{Z}_{TS}} b_i \langle\!\langle A_i|\mathcal{N}^{m-s-1}|\bar{\rho}_{\text{err}}^{\otimes 2}\rangle\!\rangle \tag{7.143}$$

$$+ \langle\!\langle \bar{E}_{\text{err}}^{\otimes 2}|\mathcal{N}^{m-1}|\bar{\rho}_{\text{err}}^{\otimes 2}\rangle\!\rangle. \tag{7.144}$$

This expression is still exact, as we have only expanded each term in the equation.
The variance bound is obtained by bounding the remaining inner products and the quantities a_i, b_i in this expression. This technical task is delegated to section 7.6.4, with a number of technical propositions that compute bounds on the quantities above. We summarize the results here. The bounds on a_i and b_i for $i \in \{0; [\text{adj}]; S; 1, 2\}$ are obtained under the assumption that $d = 2$ or that $\mathcal{E}$ is unital in Propositions 7.1, 7.2, 7.3, 7.4, 7.5 and 7.6. In summary

$$0 = a_0 = b_0, \tag{7.145}$$

$$0 \leq a_{1,2}, b_{1,2} \leq \frac{\sqrt{d^2-2}}{d^2}(1-u)^2, \tag{7.146}$$

$$0 \leq a_S, b_S \leq \sqrt{\frac{d^2-2}{d^2-1}}\sqrt{2}(1-u)^2, \tag{7.147}$$

$$0 \leq a_{[\text{adj}]}, b_{[\text{adj}]} \leq \sqrt{d^2-1}(1-u)^2. \tag{7.148}$$

In the case of $d \geq 4$, bounds on a_i are needed for $i \in \mathcal{Z}_{1,2} \cup \mathcal{Z}_S \setminus \{[\text{adj}]\}$ in terms of the above bounds on a_S and $a_{1,2}$. To do so, we use eq. (7.89), which states

$$\sqrt{|V_{1,2}|}A_{1,2} = \sum_{i\in\mathcal{Z}_{1,2}} \sqrt{|V_i|}A_i \qquad \sqrt{|V_S|}A_S = \sum_{i\in\mathcal{Z}_S} \sqrt{|V_i|}A_i. \tag{7.149}$$

From this it follows that

$$\sqrt{|V_{1,2}|}a_{1,2} = \sum_{i\in\mathcal{Z}_{1,2}} \sqrt{|V_i|}a_i, \qquad \sqrt{|V_S|}a_S = \sum_{i\in\mathcal{Z}_S} \sqrt{|V_i|}a_i, \tag{7.150}$$

$$\sqrt{|V_{1,2}|}b_{1,2} = \sum_{i\in\mathcal{Z}_{1,2}} \sqrt{|V_i|}b_i \qquad \sqrt{|V_S|}b_S = \sum_{i\in\mathcal{Z}_S} \sqrt{|V_i|}b_i. \tag{7.151}$$

Thus, since $a_i, b_i \geq 0$ by proposition 7.1, these equations imply the following bounds

$$a_i \leq \sqrt{\frac{|V_{1,2}|}{|V_i|}}a_{1,2}, \qquad b_i \leq \sqrt{\frac{|V_{1,2}|}{|V_i|}}b_{1,2}, \qquad \forall i \in \mathcal{Z}_{1,2} \tag{7.152}$$

$$a_i \leq \sqrt{\frac{|V_S|}{|V_i|}}a_S, \qquad b_i \leq \sqrt{\frac{|V_S|}{|V_i|}}b_S, \qquad \forall i \in \mathcal{Z}_S \setminus \{[\text{adj}]\}. \tag{7.153}$$

The size of the relevant spaces (as derived in [15]) was summarized in table 7.2. The inner products in eq. (7.141)-eq. (7.144) are bounded using Propositions 7.7, 7.8 and 7.9. proposition 7.7 is applicable since $\mathcal{N}^m$ is a CPTP map for any $m \in \mathbb{N}$, since CPTP maps are closed under composition. Now $\mathcal{N}$ is CPTP because $\mathcal{N}$ is the convex combination of the CPTP sequences $\mathcal{G}_{\mathbf{j}}$ and a convex combination of CPTP maps is CPTP. The results of Propositions 7.7, 7.8 and 7.9 are summarized as follows:

$$\langle\langle A_i|\mathcal{N}^{m-s-1}|B_2^{\otimes 2}\rangle\rangle \leq \frac{1}{\sqrt{|V_i|}}, \tag{7.154}$$

$$\langle\langle \bar{E}_{\mathrm{err}}^{\otimes 2}|\mathcal{N}^{m-s-1}|A_i\rangle\rangle \leq d^2\|\bar{E}_{\mathrm{err}}\|_\infty^2, \tag{7.155}$$

$$\langle\langle A_i|\mathcal{N}^{m-s-1}|\bar{\rho}_{\mathrm{err}}^{\otimes 2}\rangle\rangle \leq \sqrt{\frac{6}{(d-2)(d-1)}}\|\bar{\rho}_{\mathrm{err}}\|_1^2, \tag{7.156}$$

$$\langle\langle \bar{E}_{\mathrm{err}}^{\otimes 2}|\mathcal{N}^{m-1}|\bar{\rho}_{\mathrm{err}}^{\otimes 2}\rangle\rangle \leq \|\bar{E}_{\mathrm{err}}\|_\infty^2\|\bar{\rho}_{\mathrm{err}}\|_1^2, \tag{7.157}$$

where we have used that $\|A^{\otimes k}\|_p = \|A\|_p^k$ for any $k \in \mathbb{N}$ and $p \in [1,\infty]$. eq. (7.155) and eq. (7.156) have single-qubit specific ($d = 2$) improvements (derived in proposition 7.8), using the fact that $V_{1,2}$ and V_S actually carry irreducible subrepresentations of $\mathcal{G}^{\otimes 2}$. Since we have explicit expressions for $A_{1,2}$ and A_S (eq. (7.91) and eq. (7.92) respectively), their norms can be computed directly. Using this gives the improved single-qubit bounds,

$$\langle\langle \bar{E}_{\mathrm{err}}^{\otimes 2}|\mathcal{N}^{m-s-1}|A_S\rangle\rangle \leq \frac{5}{\sqrt{3}}\|\bar{E}_{\mathrm{err}}\|_\infty^2, \quad \langle\langle \bar{E}_{\mathrm{err}}^{\otimes 2}|\mathcal{N}^{m-s-1}|A_{1,2}\rangle\rangle \leq 2\sqrt{2}\|\bar{E}_{\mathrm{err}}\|_\infty^2, \tag{7.158}$$

$$\langle\langle A_S|\mathcal{N}^{m-s-1}|\bar{\rho}_{\mathrm{err}}^{\otimes 2}\rangle\rangle \leq \frac{1}{\sqrt{3}}\|\bar{\rho}_{\mathrm{err}}\|_1^2, \qquad \langle\langle A_{1,2}|\mathcal{N}^{m-s-1}|\bar{\rho}_{\mathrm{err}}^{\otimes 2}\rangle\rangle \leq \frac{\sqrt{2}}{3}\|\bar{\rho}_{\mathrm{err}}\|_1^2. \tag{7.159}$$

Plugging all of these bounds into eq. (7.141)-eq. (7.144) and using the geometric series

$$\sum_{s=1}^{m-1} u^{2(s-1)} = \frac{1-u^{2(m-1)}}{1-u^2} \tag{7.160}$$

will yield the bound eq. (7.119)

$$\mathbb{V}[q_{\mathbf{j}}] \leq \sigma^2 = \frac{1-u^{2(m-1)}}{1-u^2}(1-u)^2\Big(\alpha^2\beta^2 c_1(d) + \alpha^2 c_2(d)\|\bar{E}_{\mathrm{err}}\|_\infty^2 + \beta^2 c_3(d)\|\bar{\rho}_{\mathrm{err}}\|_1^2\Big) + \|\bar{\rho}_{\mathrm{err}}\|_1^2\|\bar{E}_{\mathrm{err}}\|_\infty^2, \tag{7.161}$$

where

$$c_1(d) = \begin{cases} \frac{\sqrt{2}}{4}\frac{1}{\sqrt{2}} + \sqrt{\frac{2}{3}}\sqrt{2}\frac{1}{\sqrt{3}} = \frac{11}{12}, & \text{if } d = 2, \\ \frac{\sqrt{d^2-2}}{d^2}\sum_{i\in\mathcal{Z}_{1,2}}\frac{\sqrt{|V_{1,2}|}}{|V_i|} + \sqrt{2}\sqrt{\frac{d^2-2}{d^2-1}}\sum_{i\in\mathcal{Z}_S\setminus\{[\mathrm{adj}]\}}\frac{\sqrt{|V_S|}}{|V_i|} + \frac{\sqrt{d^2-1}}{\sqrt{|V_{[\mathrm{adj}]}|}}, & \text{if } d \geq 4, \end{cases}$$

$$c_2(d) = \begin{cases} \frac{1}{3}\left(\frac{\sqrt{2}}{4}2\sqrt{2} + \sqrt{\frac{2}{3}}\sqrt{2}\frac{5}{\sqrt{3}}\right) = \frac{13}{9}, & \text{if } d = 2, \\ \frac{d^2}{d^2-1}\Bigg(\frac{\sqrt{d^2-2}}{d^2}\sum_{i\in\mathcal{Z}_{1,2}}\frac{\sqrt{|V_{1,2}|}}{\sqrt{|V_i|}} \\ \qquad +\sqrt{2}\sqrt{\frac{d^2-2}{d^2-1}}\sum_{i\in\mathcal{Z}_S\setminus\{[\mathrm{adj}]\}}\frac{\sqrt{|V_S|}}{\sqrt{|V_i|}} + \sqrt{d^2-1}\Bigg), & \text{if } d \geq 4, \end{cases}$$

$$c_3(d) = \begin{cases} 3\left(\frac{\sqrt{2}}{4}\frac{\sqrt{2}}{3} + \sqrt{\frac{2}{3}}\sqrt{2}\frac{1}{\sqrt{3}}\right) = \frac{5}{2}, & \text{if } d = 2, \\ (d^2-1)\sqrt{\frac{6}{(d-2)(d-1)}}\Bigg(\frac{\sqrt{d^2-2}}{d^2}\sum_{i\in\mathcal{Z}_{1,2}}\frac{\sqrt{|V_{1,2}|}}{\sqrt{|V_i|}} \\ \qquad +\sqrt{2}\sqrt{\frac{d^2-2}{d^2-1}}\sum_{i\in\mathcal{Z}_S\setminus\{[\mathrm{adj}]\}}\frac{\sqrt{|V_S|}}{\sqrt{|V_i|}} + \sqrt{d^2-1}\Bigg), & \text{if } d \geq 4. \end{cases}$$

The size of the spaces V_i in these equations are found in table 7.2. The asymptotic behavior of the dimension-dependent functions $c_i(d)$ can be found if all relevant dimensions of the spaces are plugged into the above equations. ■

7.6.4. BOUNDS ON INDIVIDUAL QUANTITIES IN THE PROOF

This section provides the technical lemma's and propositions referred to in the previous section. They are collected here together in an attempt not to clutter the main line of the proof. Most of these technical lemma's put a bound on quantities arising in the proof of theorem 1.

We start by bounding the a_i. Only bounds on a_0, a_S, $a_{1,2}$ and $a_{[\mathrm{adj}]}$ are provided. In the multi-qubit case where V_S and $V_{1,2}$ are not irreducible representations, the quantities a_i for $i \in \mathcal{Z}_S, \mathcal{Z}_d$ are bounded by a_S and $a_{1,2}$. The only exception is $i = [\mathrm{adj}]$, for which we provide a separate bound. Let us start with showing that all a_i and b_i are nonnegative.

Proposition 7.1 (Lower bound on a_i and b_i)**.** For all CPTP $\mathcal{E}$ and all $i \in \mathcal{Z}_{TS}$, one has

$$a_i = \langle\langle A_i|\mathcal{N} - \mathcal{M}^{\otimes 2}|B_2B_2\rangle\rangle \geq 0, \qquad b_i = \langle\langle B_2B_2|\mathcal{N} - \mathcal{M}^{\otimes 2}|A_i\rangle\rangle \geq 0. \tag{7.162}$$

Proof. If $i = 0$, then proposition 7.2 will show that $a_0 = 0$, which includes this lower bound. For all other $i \in \mathcal{Z}_{TS} \setminus \{0\}$, we show that $\mathcal{M}^{\otimes 2}|A_i\rangle\rangle = 0$. This is because $\mathcal{M}^{\otimes 2}$ is supported on $W^{\otimes 2} = \mathrm{Span}\{B_1B_1, B_1B_2, B_2B_1, B_2B_2\}$, where $A_0 = B_2B_2$. But $B_1B_1, B_1B_2, B_2B_1 \in (V_{TS}^{\otimes 2})^\perp$. Since $A_i \in V_{TS}^{\otimes 2} \setminus \mathrm{Span}\{A_0\}$ the claim follows. Therefore $a_i = \langle\langle B_2B_2|\mathcal{N}|A_i\rangle\rangle$. Using the definitions of $\mathcal{N}$ (eq. (7.96)) and A_i (eq. (7.88)), it follows that

$$\begin{aligned} a_i &= \frac{1}{|V_i||\mathsf{C}_q|^2}\sum_{\mathcal{G},\mathcal{G}'\in\mathsf{C}_q}\sum_{v_i\in\mathcal{B}_i}\langle\langle v_iv_i|\mathcal{G}^{\otimes 4}\mathcal{E}^{\otimes 4}\mathcal{G}'^{\otimes 4}|B_2B_2\rangle\rangle \\ &= \frac{1}{|V_i||\mathsf{C}_q|^2}\sum_{\mathcal{G},\mathcal{G}'\in\mathsf{C}_q}\sum_{v_i\in\mathcal{B}_i}\langle\langle v_i|\mathcal{G}^{\otimes 2}\mathcal{E}^{\otimes 2}\mathcal{G}'^{\otimes 2}|B_2\rangle\rangle^2 \geq 0, \end{aligned} \tag{7.163}$$

which is nonnegative as it is the sum of real numbers squared. Analogously,

$$b_i = \frac{1}{|V_i||\mathsf{C}_q|^2} \sum_{\mathcal{G},\mathcal{G}'\in\mathsf{C}_q} \sum_{v_i\in\mathcal{B}_i} \langle\!\langle B_2|\mathcal{G}^{\otimes 2}\mathcal{E}^{\otimes 2}\mathcal{G}'^{\otimes 2}|v_i\rangle\!\rangle^2 \geq 0. \qquad \blacksquare$$

Next we show that a_0 vanishes.

Proposition 7.2 (Bound on a_0)**.** Let a_0 be defined by eq. (7.129). Then for all CPTP quantum channels $\mathcal{E}$, $a_0 = 0$.

Proof. By definition of eq. (7.129) it follows that (using that $A_0 = B_2B_2$ by definition of eq. (7.90))

$$a_0 = \langle\!\langle A_0|\mathcal{N} - \mathcal{M}^{\otimes 2}|B_2B_2\rangle\!\rangle = \langle\!\langle B_2B_2|\mathcal{E}^{\otimes 4} - \mathcal{E}^{\otimes 4}|B_2B_2\rangle\!\rangle = 0, \tag{7.164}$$

since $\mathcal{G}^{(4)}_{\text{avg}}\,|B_2B_2\rangle\!\rangle = (\mathcal{G}^{(2)}_{\text{avg}})^{\otimes 2}\,|B_2B_2\rangle\!\rangle = |B_2B_2\rangle\!\rangle$. $\blacksquare$

The next proposition gives a bound on $a_{1,2}$.

Proposition 7.3 (Bound on $a_{1,2}$)**.** Let $a_{1,2}$ be defined as in eq. (7.129) and let $\mathcal{E}$ be a CPTP map. If $\mathcal{E}$ is a single-qubit channel (i.e. if $d = 2$) or if $\mathcal{E}$ is unital (i.e. $\mathcal{E}(\mathbb{1}) = \mathbb{1}$), then

$$a_{1,2} = \frac{1}{\sqrt{d^2-2}}\left(\frac{1}{d^2-1}\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\!\langle\sigma|\mathcal{E}_{\text{u}}\mathcal{E}_{\text{u}}^\dagger|\sigma\rangle\!\rangle^2 - u^2\right) \leq \frac{\sqrt{d^2-2}}{d^2}(1-u)^2. \tag{7.165}$$

Proof. By the definition eq. (7.129), $a_{1,2} = \langle\!\langle A_{1,2}|\mathcal{N} - \mathcal{M}^{\otimes 2}|B_2B_2\rangle\!\rangle$, where

$$B_2B_2 = \frac{1}{d^2-1}\sum_{\sigma,\tau\in\boldsymbol{\sigma}_q}\sigma\sigma\tau\tau \qquad \text{and} \qquad A_{1,2} = \frac{1}{\sqrt{d^2-2}}\left(\sum_{\sigma\in\boldsymbol{\sigma}_q}\sigma^{\otimes 4} - A_0\right) \tag{7.166}$$

were defined in eq. (7.90) and eq. (7.91) respectively. Therefore $a_{1,2}$ is computed as (recalling that $A_0 = B_2B_2$ and using eq. (7.100))

$$\begin{aligned}
a_{1,2} &= \langle\!\langle A_{1,2}|\mathcal{N} - \mathcal{M}^{\otimes 2}|B_2B_2\rangle\!\rangle \\
&= \frac{1}{(d^2-1)\sqrt{d^2-2}}\left(\sum_{\sigma,\hat{\sigma},\hat{\tau}\in\boldsymbol{\sigma}_q}\langle\!\langle\sigma\sigma\sigma\sigma|\mathcal{E}^{\otimes 4}|\hat{\sigma}\hat{\sigma}\hat{\tau}\hat{\tau}\rangle\!\rangle - \langle\!\langle B_2B_2|\mathcal{E}^{\otimes 4}|B_2B_2\rangle\!\rangle\right) \\
&= \frac{1}{\sqrt{d^2-2}}\left(\frac{1}{d^2-1}\sum_{\sigma,\hat{\sigma},\hat{\tau}\in\boldsymbol{\sigma}_q}\langle\!\langle\sigma|\mathcal{E}|\hat{\sigma}\rangle\!\rangle^2\langle\!\langle\sigma|\mathcal{E}|\hat{\tau}\rangle\!\rangle^2 - u^2\right) \\
&= \frac{1}{\sqrt{d^2-2}}\left(\frac{1}{d^2-1}\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\!\langle\sigma|\mathcal{E}_{\text{u}}\mathcal{E}_{\text{u}}^\dagger|\sigma\rangle\!\rangle^2 - u^2\right),
\end{aligned} \tag{7.167}$$

where in the last step, the following was used

$$\sum_{\hat{\sigma}\in\boldsymbol{\sigma}_q} \langle\!\langle \sigma|\mathcal{E}|\hat{\sigma}\rangle\!\rangle\langle\!\langle \tau|\mathcal{E}|\hat{\sigma}\rangle\!\rangle = \sum_{\hat{\sigma}\in\boldsymbol{\sigma}_q} \langle\!\langle \sigma|\mathcal{E}|\hat{\sigma}\rangle\!\rangle\langle\!\langle \hat{\sigma}|\mathcal{E}^\dagger|\tau\rangle\!\rangle = \langle\!\langle \sigma|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\tau\rangle\!\rangle, \qquad \forall \sigma,\tau\in\boldsymbol{\sigma}_q, \tag{7.168}$$

abusing notation slightly by writing $\mathcal{E}_{\mathrm{u}}$ instead of $1\oplus\mathcal{E}_{\mathrm{u}}$ and using the fact that $\sum_{\hat{\sigma}\in\boldsymbol{\sigma}_q}|\hat{\sigma}\rangle\!\rangle\langle\!\langle\hat{\sigma}|$ is the projection onto the unital block.
The bound of eq. (7.165) is then shown as follows. The idea is to apply lemma 7.2 to the map

$$\mathcal{E} := \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger \end{bmatrix}, \tag{7.169}$$

since this map is constructed such that

$$f(\mathcal{E}) = \frac{1}{d^2-1}\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\!\langle\sigma|\mathcal{E}|\sigma\rangle\!\rangle = \frac{1}{d^2-1}\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\!\langle\sigma|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\rangle\!\rangle = u(\mathcal{E}) \tag{7.170}$$

and

$$\frac{1}{d^2-1}\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\!\langle\sigma|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\rangle\!\rangle^2 - u(\mathcal{E})^2 = \frac{1}{d^2-1}\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\!\langle\sigma|\mathcal{E}|\sigma\rangle\!\rangle^2 - f(\mathcal{E})^2. \tag{7.171}$$

Application of lemma 7.2 requires the map $\mathcal{E}$ to be CPTP. This is guaranteed by lemma 7.3, using the assumption that $\mathcal{E}$ is a single-qubit or unital channel. Therefore lemma 7.2 applied to the channel $\mathcal{E}$ defined above, yields

$$\frac{1}{d^2-1}\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\!\langle\sigma|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\rangle\!\rangle^2 - u(\mathcal{E})^2 = \frac{1}{d^2-1}\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\!\langle\sigma|\mathcal{E}|\sigma\rangle\!\rangle^2 - f(\mathcal{E})^2 \tag{7.172}$$

$$\leq \frac{d^2-2}{d^2}(1-f(\mathcal{E}))^2 = \frac{d^2-2}{d^2}(1-u(\mathcal{E}))^2. \tag{7.173}$$

Plugging this into eq. (7.167) yields the result. ■

The next proposition bounds the quantity a_S.

Proposition 7.4 (Bound on a_S)**.** Let a_S be defined as in eq. (7.129) and let $\mathcal{E}$ be a CPTP map. If $\mathcal{E}$ is a single-qubit channel (i.e. if $d=2$) or if $\mathcal{E}$ is unital (i.e. $\mathcal{E}(\mathbb{1})=\mathbb{1}$), then

$$a_S = \frac{\sqrt{2}}{(d^2-1)^{\frac{3}{2}}(d^2-2)^{\frac{1}{2}}}\sum_{\substack{\sigma,\tau\in\boldsymbol{\sigma}_q\\ \sigma\neq\tau}}\langle\!\langle\sigma|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\tau\rangle\!\rangle^2 \leq \sqrt{\frac{d^2-2}{d^2-1}}\sqrt{2}(1-u)^2. \tag{7.174}$$

Proof. First, let us show the evaluation of a_S. By the definition eq. (7.129) we have $a_S = \langle\!\langle A_S|\mathcal{N}-\mathcal{M}^{\otimes 2}|B_2B_2\rangle\!\rangle$, where

$$B_2B_2 = \frac{1}{d^2-1}\sum_{\sigma,\tau\in\boldsymbol{\sigma}_q}\sigma\sigma\tau\tau, \tag{7.175}$$

$$A_S = \sqrt{\frac{1}{2(d^2-1)(d^2-2)}}\sum_{\substack{\sigma,\tau\in\boldsymbol{\sigma}_q\\ \sigma\neq\tau}}\sigma\tau\sigma\tau + \sigma\tau\tau\sigma \tag{7.176}$$

were defined in eq. (7.90) and eq. (7.91) respectively. Therefore a_S is computed as

$$\begin{aligned} a_S &= \langle\langle A_S|\mathcal{N}-\mathcal{M}^{\otimes 2}|B_2B_2\rangle\rangle \\ &= \frac{1}{\sqrt{2}(d^2-1)^{\frac{3}{2}}(d^2-2)^{\frac{1}{2}}} \sum_{\substack{\sigma,\tau,\hat{\sigma},\hat{\tau}\in\boldsymbol{\sigma}_q \\ \sigma\neq\tau}} \langle\langle \sigma\tau\sigma\tau+\sigma\tau\tau\sigma|\mathcal{E}^{\otimes 4}|\hat{\sigma}\hat{\sigma}\hat{\tau}\hat{\tau}\rangle\rangle \\ &= \frac{1}{\sqrt{2}(d^2-1)^{\frac{3}{2}}(d^2-2)^{\frac{1}{2}}} \sum_{\substack{\sigma,\tau,\hat{\sigma},\hat{\tau}\in\boldsymbol{\sigma}_q \\ \sigma\neq\tau}} 2\langle\langle \sigma|\mathcal{E}|\hat{\sigma}\rangle\rangle\langle\langle \tau|\mathcal{E}|\hat{\sigma}\rangle\rangle\langle\langle \sigma|\mathcal{E}|\hat{\tau}\rangle\rangle\langle\langle \tau|\mathcal{E}|\hat{\tau}\rangle\rangle \\ &= \frac{\sqrt{2}}{(d^2-1)^{\frac{3}{2}}(d^2-2)^{\frac{1}{2}}} \sum_{\substack{\sigma,\tau\in\boldsymbol{\sigma}_q \\ \sigma\neq\tau}} \langle\langle \sigma|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}|\tau\rangle\rangle^2 \\ &= \frac{\sqrt{2}}{(d^2-1)^{\frac{3}{2}}(d^2-2)^{\frac{1}{2}}} \sum_{\substack{\sigma,\tau\in\boldsymbol{\sigma}_q \\ \sigma\neq\tau}} \langle\langle \sigma|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}|\tau\rangle\rangle^2. \end{aligned} \tag{7.177}$$

In the fourth step, the trick of eq. (7.168) was again used. In the final step, it is used that $\langle\langle \sigma|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}|\tau\rangle\rangle^2$ is the square of off-diagonal matrix elements of $\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}$, so that $\langle\langle \sigma|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}|\tau\rangle\rangle^2 = \langle\langle \sigma|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}|\tau\rangle\rangle^2$.

7

The bound is derived as follows. Under the stated assumption that $\mathcal{E}$ is a single-qubit or unital channel, lemma 7.3 guarantees that $\|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}\|_{2\to 2}\leq 1$. Here $\|\cdot\|_{2\to 2}$ is the induced Schatten 2-norm (see eq. (7.74)). Since $\langle\langle A|B\rangle\rangle = \mathrm{Tr}[A^{\dagger}B]$ for any $A,B\in\mathcal{M}_d$ (and therefore $\|A\|_2 = \||A\rangle\rangle\|_2$ for all $A\in\mathcal{M}_d$), it follows that $\|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}\|_{2\to 2} = \|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}\|_{2\to 2}$. But the operator norm (Schatten ∞-norm) on matrices is just the induced $2\to 2$ norm, so that it can be concluded that $\|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}\|_{\infty} = \|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}\|_{2\to 2} = \|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}\|_{2\to 2}\leq 1$. Together with the fact that a matrix of the form $\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}$ is itself positive semi-definite, this implies that the matrix $\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}\geq 0$ is also positive semi-definite as a matrix (not to be confused with being a positive superoperator). Now the key idea is to bound the off-diagonal elements of the symmetric positive semi-definite matrix $\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}$ by the diagonal elements using the Sylvester's Criterion for positive semi-definite matrices (lemma 7.4). This criterion states that a hermitian matrix is positive semi-definite if and only if all of its principal minors are non-negative. Here we use the only if part, since it has been established that $\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}$ is positive semi-definite. In particular we use that the positive semi-definiteness of $\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}$ implies that all of its second order minors are non-negative. This means that

$$\langle\langle \sigma|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}|\sigma\rangle\rangle\langle\langle \tau|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}|\tau\rangle\rangle - \langle\langle \sigma|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}|\tau\rangle\rangle^2 \geq 0, \qquad \forall\sigma,\tau\in\boldsymbol{\sigma}_q,\ \sigma\neq\tau. \tag{7.178}$$

Plugging this into eq. (7.177) yields

$$a_S \leq \frac{\sqrt{2}}{(d^2-1)^{\frac{3}{2}}(d^2-2)^{\frac{1}{2}}} \sum_{\substack{\sigma,\tau\in\boldsymbol{\sigma}_q \\ \sigma\neq\tau}} \langle\!\langle \sigma|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}|\sigma\rangle\!\rangle \langle\!\langle \tau|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}|\tau\rangle\!\rangle$$
$$= \frac{\sqrt{2}}{(d^2-1)^{\frac{3}{2}}(d^2-2)^{\frac{1}{2}}} \left(\left(\sum_{\sigma\in\boldsymbol{\sigma}_q} \langle\!\langle \sigma|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}|\sigma\rangle\!\rangle \right)^2 - \sum_{\sigma\in\boldsymbol{\sigma}_q} \langle\!\langle \sigma|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}|\sigma\rangle\!\rangle^2 \right). \tag{7.179}$$

The final step is to use that the mean of squares is larger than the square of the mean (lemma 7.6). This means in our setting that

$$\frac{1}{d^2-1} \sum_{\sigma\in\boldsymbol{\sigma}_q} \langle\!\langle \sigma|Ic-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}|\sigma\rangle\!\rangle^2 \geq \left(\frac{1}{d^2-1} \sum_{\sigma\in\boldsymbol{\sigma}_q} \langle\!\langle \sigma|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}|\sigma\rangle\!\rangle \right)^2. \tag{7.180}$$

Multiplying by $-(d^2-1)$ and plugging into eq. (7.179) yields the bound:

$$a_S \leq \frac{\sqrt{2}}{(d^2-1)^{\frac{3}{2}}(d^2-2)^{\frac{1}{2}}} \left(1-\frac{1}{d^2-1}\right) \left(\sum_{\sigma\in\boldsymbol{\sigma}_q} \langle\!\langle \sigma|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}|\sigma\rangle\!\rangle \right)^2 \tag{7.181}$$
$$= \frac{\sqrt{2}}{(d^2-1)^{\frac{3}{2}}(d^2-2)^{\frac{1}{2}}} \frac{d^2-2}{d^2-1} \left((d^2-1)(1-u)\right)^2 \tag{7.182}$$
$$= \sqrt{\frac{d^2-2}{d^2-1}}\sqrt{2}(1-u)^2, \tag{7.183}$$

using the definition of u (definition 13) and the fact that $u(\mathcal{I})=1$. ■

Finally, a bound on $a_{[\mathrm{adj}]}$ is presented.

Proposition 7.5 (Bound on $a_{[\mathrm{adj}]}$)**.** Let $a_{[\mathrm{adj}]}$ be defined as in eq. (7.129) and let $\mathcal{E}$ be a CPTP map. If $\mathcal{E}$ is a single-qubit channel (i.e. if $d=2$) or if $\mathcal{E}$ is unital (i.e. $\mathcal{E}(\mathbb{1})=\mathbb{1}$), then

$$a_{[\mathrm{adj}]} = \frac{2}{(d^2-4)(d^2-1)^{\frac{3}{2}}} \sum_{\tau\in\boldsymbol{\sigma}_q} \left(\sum_{\sigma\in\mathbf{C}_\tau} \langle\!\langle \sigma\cdot\tau|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^{\dagger}|\sigma\rangle\!\rangle \right)^2 \leq \sqrt{d^2-1}(1-u)^2, \tag{7.184}$$

where $\mathbf{C}_\tau$ is the set of all normalized Pauli's that commute with τ (except for τ and σ_0), as defined in eq. (5.2).

Proof. By the definition eq. (7.129), $a_{[\mathrm{adj}]} = \langle\!\langle A_S|\mathcal{N}-\mathcal{M}^{\otimes 2}|B_2B_2\rangle\!\rangle$, where

$$B_2B_2 = \frac{1}{d^2-1} \left(\sum_{\hat{\sigma}\in\boldsymbol{\sigma}_q} \hat{\sigma}\hat{\sigma} \right)^{\otimes 2}, \tag{7.185}$$

$$A_{[\mathrm{adj}]} = \frac{1}{2(d^2-4)\sqrt{d^2-1}} \sum_{\tau\in\boldsymbol{\sigma}_q} \left(\sum_{\sigma\in\mathbf{C}_\tau} (\sigma\cdot\tau)\sigma + \sigma(\sigma\cdot\tau) \right)^{\otimes 2} \tag{7.186}$$

were defined in eq. (7.90) and eq. (7.94) respectively. Therefore $a_{[\mathrm{adj}]}$ is computed as

$$\begin{aligned} a_{[\mathrm{adj}]} &= \frac{1}{2(d^2-4)(d^2-1)^{\frac{3}{2}}} \sum_{\tau\in\boldsymbol{\sigma}_q} \left\langle\!\!\left\langle \left(\sum_{\sigma\in\mathbf{C}_\tau} (\sigma\cdot\tau)\sigma + \sigma(\sigma\cdot\tau)\right)^{\otimes 2} \left|\mathcal{E}^{\otimes 4}\right| \left(\sum_{\hat{\sigma}\in\boldsymbol{\sigma}_q} \hat{\sigma}\hat{\sigma}\right)^{\otimes 2} \right\rangle\!\!\right\rangle \\ &= \frac{1}{2(d^2-4)(d^2-1)^{\frac{3}{2}}} \sum_{\tau\in\boldsymbol{\sigma}_q} \left(\sum_{\sigma\in\mathbf{C}_\tau}\sum_{\hat{\sigma}\in\boldsymbol{\sigma}_q} \langle\!\langle (\sigma\cdot\tau)\sigma + \sigma(\sigma\cdot\tau)|\mathcal{E}^{\otimes 2}|\hat{\sigma}\hat{\sigma}\rangle\!\rangle\right)^2 \\ &= \frac{1}{2(d^2-4)(d^2-1)^{\frac{3}{2}}} \sum_{\tau\in\boldsymbol{\sigma}_q} \left(\sum_{\sigma\in\mathbf{C}_\tau}\sum_{\hat{\sigma}\in\boldsymbol{\sigma}_q} 2\langle\!\langle \sigma\cdot\tau|\mathcal{E}|\hat{\sigma}\rangle\!\rangle\langle\!\langle \sigma|\mathcal{E}|\hat{\sigma}\rangle\!\rangle\right)^2 \\ &= \frac{2}{(d^2-4)(d^2-1)^{\frac{3}{2}}} \sum_{\tau\in\boldsymbol{\sigma}_q} \left(\sum_{\sigma\in\mathbf{C}_\tau} \langle\!\langle \sigma\cdot\tau|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\rangle\!\rangle\right)^2, \end{aligned} \tag{7.187}$$

where in the final line we used again the trick of eq. (7.168). Our bound on this quantity again starts with using the fact that the mean of the squares is larger than the square of the mean (lemma 7.6), yielding for all $\tau \in \boldsymbol{\sigma}_q$

$$\left(\frac{2}{d^2-4}\sum_{\sigma\in\mathbf{C}_\tau} \langle\!\langle \sigma\cdot\tau|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\rangle\!\rangle\right)^2 \leq \frac{2}{d^2-4}\sum_{\sigma\in\mathbf{C}_\tau} \langle\!\langle \sigma\cdot\tau|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\rangle\!\rangle^2. \tag{7.188}$$

Multiplying with $\frac{d^2-4}{2}$ and plugging into the above yields

$$a_S \leq \frac{1}{(d^2-1)^{\frac{3}{2}}} \sum_{\tau\in\boldsymbol{\sigma}_q}\sum_{\sigma\in\mathbf{C}_\tau} \langle\!\langle \sigma\cdot\tau|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\rangle\!\rangle^2. \tag{7.189}$$

Now we use the facts that $\sigma\cdot\tau \neq \sigma$ to write this as

$$a_S \leq \frac{1}{(d^2-1)^{\frac{3}{2}}} \sum_{\tau\in\boldsymbol{\sigma}_q}\sum_{\sigma\in\mathbf{C}_\tau} \langle\!\langle \sigma\cdot\tau|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\rangle\!\rangle^2, \tag{7.190}$$

where $\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger$ is a positive semi-definite matrix, since $\|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger\|_\infty = \|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger\|_{2\to 2} \leq 1$ under the stated assumptions on $\mathcal{E}$ by lemma 7.3 and the fact that a matrix of the form $\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger$ is itself positive semi-definite. This allows us again to use Sylvester's criterion (lemma 7.4) to bound off-diagonal terms by diagonal terms by using the fact that all minors of degree 2 of $\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger$ must be nonnegative:

$$\langle\!\langle \sigma|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\rangle\!\rangle\langle\!\langle \sigma\cdot\tau|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\cdot\tau\rangle\!\rangle - \langle\!\langle \sigma\cdot\tau|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\rangle\!\rangle^2 \geq 0, \qquad \forall\tau\in\boldsymbol{\sigma}_q,\ \forall\sigma\in\mathbf{C}_\tau. \tag{7.191}$$

Therefore, we arrive at

$$\begin{aligned} a_S &\leq \frac{1}{(d^2-1)^{\frac{3}{2}}} \sum_{\tau\in\boldsymbol{\sigma}_q} \sum_{\sigma\in\mathbf{C}_\tau} \langle\!\langle \sigma|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\rangle\!\rangle \langle\!\langle \sigma\cdot\tau|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\cdot\tau\rangle\!\rangle \\ &\leq \frac{1}{(d^2-1)^{\frac{3}{2}}} \sum_{\tau,\sigma\in\boldsymbol{\sigma}_q} \langle\!\langle \sigma|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\rangle\!\rangle \langle\!\langle \tau|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\tau\rangle\!\rangle \\ &= \sqrt{d^2-1}(1-u)^2, \end{aligned} \tag{7.192}$$

where in the second line the sum over $\sigma\in\mathbf{C}_\tau$ was completed to the sum over $\sigma\in\boldsymbol{\sigma}_q$ by adding
all the nonnegative terms $\langle\!\langle \sigma|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\rangle\!\rangle \langle\!\langle \sigma\cdot\tau|\mathcal{I}-\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\cdot\tau\rangle\!\rangle$ with $\sigma\in\boldsymbol{\sigma}_q\setminus\mathbf{C}_\tau$ for each $\tau\in\boldsymbol{\sigma}_q$. All these terms are nonnegative because they are the product of diagonal elements of positive-semidefinite matrices, which must be nonnegative. ■

This completes the set of propositions to bound the quantities a_i. The quantities b_i are strongly related to the quantities a_i, and we will show that they satisfy the same upper bounds. More precisely, the next proposition establishes that all bounds on a_i also hold for b_i, for $i\in\{1,2;S;0;[\mathrm{adj}]\}$.

Proposition 7.6 (Bounds on b_i)**.** Let $\mathcal{E}$ be a CPTP map. Assume that $d=2$ or that $\mathcal{E}$ is unital. Let $a_i=\langle\!\langle A_i|\mathcal{N}-\mathcal{M}^{\otimes 2}|B_2B_2\rangle\!\rangle$ and $b_i=\langle\!\langle B_2B_2|\mathcal{N}-\mathcal{M}^{\otimes 2}|A_i\rangle\!\rangle$ as above. Then

$$b_0=a_0=0, \tag{7.193}$$

$$b_{1,2}=a_{1,2}\leq\frac{\sqrt{d^2-2}}{d^2}(1-u)^2, \tag{7.194}$$

$$b_S=a_S\leq\sqrt{\frac{d^2-2}{d^2-1}}\sqrt{2}(1-u)^2, \tag{7.195}$$

$$b_{[\mathrm{adj}]}\leq\sqrt{d^2-1}(1-u)^2. \tag{7.196}$$

Proof. The equality $b_0=a_0=\langle\!\langle B_2B_2|\mathcal{N}-\mathcal{M}^{\otimes 2}|B_2B_2\rangle\!\rangle$ immediately follows from the fact that $A_0=B_2B_2$. Thus $b_0=0$ by proposition 7.2. In general, b_i can be written as

$$b_i=\langle\!\langle B_2B_2|\mathcal{N}-\mathcal{M}^{\otimes 2}|A_i\rangle\!\rangle=\langle\!\langle A_i|\mathcal{N}^\dagger-(\mathcal{M}^{\otimes 2})^\dagger|B_2B_2\rangle\!\rangle. \tag{7.197}$$

Now since $\mathcal{G}_{\mathrm{avg}}^{(n)}$ are orthogonal projections, $(\mathcal{G}_{\mathrm{avg}}^{(n)})^\dagger=\mathcal{G}_{\mathrm{avg}}^{(n)}$. Therefore $\mathcal{N}^\dagger=\mathcal{G}_{\mathrm{avg}}^{(4)}(\mathcal{E}^\dagger)^{\otimes 4}\mathcal{G}_{\mathrm{avg}}^{(4)}$ and $\mathcal{M}^\dagger=\mathcal{G}_{\mathrm{avg}}^{(2)}(\mathcal{E}^\dagger)^{\otimes 2}\mathcal{G}_{\mathrm{avg}}^{(2)}$. Thus, b_i and a_i are related by $b_i(\mathcal{E})=a_i(\mathcal{E}^\dagger)$. That is, b_i can be obtained from a_i by replacing $\mathcal{E}$ with $\mathcal{E}^\dagger$ in the exact expressions.
We first show that this implies $b_{1,2}=a_{1,2}$ and $b_S=a_S$. This follows from the two identities (using only the trick of eq. (7.168) over and over again)

$$\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\!\langle \sigma|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\sigma\rangle\!\rangle^2=\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\!\langle \sigma|\mathcal{E}_{\mathrm{u}}|\hat{\sigma}\rangle\!\rangle^2\langle\!\langle \sigma|\mathcal{E}_{\mathrm{u}}|\hat{\sigma}\rangle\!\rangle^2=\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\!\langle \hat{\sigma}|\mathcal{E}_{\mathrm{u}}^\dagger\mathcal{E}_{\mathrm{u}}|\hat{\sigma}\rangle\!\rangle^2, \tag{7.198}$$

$$\sum_{\sigma,\tau\in\boldsymbol{\sigma}_q}\langle\!\langle \sigma|\mathcal{E}_{\mathrm{u}}\mathcal{E}_{\mathrm{u}}^\dagger|\tau\rangle\!\rangle^2=\sum_{\sigma,\tau,\hat{\sigma},\hat{\tau}\in\boldsymbol{\sigma}_q}\langle\!\langle \sigma|\mathcal{E}_{\mathrm{u}}|\hat{\sigma}\rangle\!\rangle\langle\!\langle \tau|\mathcal{E}_{\mathrm{u}}|\hat{\sigma}\rangle\!\rangle\langle\!\langle \sigma|\mathcal{E}_{\mathrm{u}}|\hat{\tau}\rangle\!\rangle\langle\!\langle \tau|\mathcal{E}_{\mathrm{u}}|\hat{\tau}\rangle\!\rangle=\sum_{\hat{\sigma},\hat{\tau}\in\boldsymbol{\sigma}_q}\langle\!\langle \hat{\sigma}|\mathcal{E}_{\mathrm{u}}^\dagger\mathcal{E}_{\mathrm{u}}|\hat{\tau}\rangle\!\rangle^2. \tag{7.199}$$

Now eq. (7.198) implies that $a_{1,2}(\mathcal{E}) = a_{1,2}(\mathcal{E}^\dagger) = b_{1,2}(\mathcal{E})$. Subtracting eq. (7.198) from eq. (7.199) implies that $a_S(\mathcal{E}) = a_S(\mathcal{E}^\dagger) = b_S(\mathcal{E})$. This shows the second and third claim of this proposition (eq. (7.194) and eq. (7.195)), using the bounds and expressions for $a_{1,2}$ and a_S from proposition 7.3 and proposition 7.4

For $b_{[\mathrm{adj}]}$ it is not clear that $b_{[\mathrm{adj}]}$ equals $a_{[\mathrm{adj}]}$. However, by copying the technique of the proof of proposition 7.5 we show that the same bounds hold. Since $b_{[\mathrm{adj}]}(\mathcal{E}) = a_{[\mathrm{adj}]}(\mathcal{E}^\dagger)$, proposition 7.5 implies that

$$b_{[\mathrm{adj}]} = \frac{2}{(d^2-4)(d^2-1)^{\frac{3}{2}}} \sum_{\tau \in \boldsymbol{\sigma}_q} \left(\sum_{\sigma \in \mathbf{C}_\tau} \langle\langle \sigma \cdot \tau | \mathcal{E}_{\mathrm{u}}^\dagger \mathcal{E}_{\mathrm{u}} | \sigma \rangle\rangle \right)^2 . \tag{7.200}$$

The bound is proven in exactly the same spirit as proposition 7.5. We first bound the square of the mean by the mean of the squares (lemma 7.6) and then use that $\langle\langle \sigma \cdot \tau | \mathcal{E}_{\mathrm{u}}^\dagger \mathcal{E}_{\mathrm{u}} | \sigma \rangle\rangle^2 = \langle\langle \sigma \cdot \tau | \mathcal{I} - \mathcal{E}_{\mathrm{u}}^\dagger \mathcal{E}_{\mathrm{u}} | \sigma \rangle\rangle^2$ (since $\sigma \cdot \tau \neq \pm\sigma$). The matrix $\mathcal{I} - \mathcal{E}_{\mathrm{u}}^\dagger \mathcal{E}_{\mathrm{u}}$ is then shown to be positive semi-definite using $\|\mathcal{E}_{\mathrm{u}}^\dagger \mathcal{E}_{\mathrm{u}}\|_\infty \leq 1$ (by the assumptions on $\mathcal{E}$ and lemma 7.3) together with the fact that $\mathcal{E}_{\mathrm{u}}^\dagger \mathcal{E}_{\mathrm{u}} \geq 0$ is positive semi-definite. Thus Sylvester's criterion can be applied (lemma 7.4)Therefore

$$\begin{aligned}
b_{[\mathrm{adj}]} &= \frac{2}{(d^2-4)(d^2-1)^{\frac{3}{2}}} \sum_{\tau \in \boldsymbol{\sigma}_q} \left(\sum_{\sigma \in \mathbf{C}_\tau} \langle\langle \sigma \cdot \tau | \mathcal{E}_{\mathrm{u}}^\dagger \mathcal{E}_{\mathrm{u}} | \sigma \rangle\rangle \right)^2 \\
&\leq \frac{1}{(d^2-1)^{\frac{3}{2}}} \sum_{\tau \in \boldsymbol{\sigma}_q} \sum_{\sigma \in \mathbf{C}_\tau} \langle\langle \sigma \cdot \tau | \mathcal{E}_{\mathrm{u}}^\dagger \mathcal{E}_{\mathrm{u}} | \sigma \rangle\rangle^2 \\
&= \frac{1}{(d^2-1)^{\frac{3}{2}}} \sum_{\tau \in \boldsymbol{\sigma}_q} \sum_{\sigma \in \mathbf{C}_\tau} \langle\langle \sigma \cdot \tau | \mathcal{I} - \mathcal{E}_{\mathrm{u}}^\dagger \mathcal{E}_{\mathrm{u}} | \sigma \rangle\rangle^2 \\
&\leq \frac{1}{(d^2-1)^{\frac{3}{2}}} \sum_{\tau \in \boldsymbol{\sigma}_q} \sum_{\sigma \in \mathbf{C}_\tau} \langle\langle \sigma | \mathcal{I} - \mathcal{E}_{\mathrm{u}}^\dagger \mathcal{E}_{\mathrm{u}} | \sigma \rangle\rangle \langle\langle \sigma \cdot \tau | \mathcal{I} - \mathcal{E}_{\mathrm{u}}^\dagger \mathcal{E}_{\mathrm{u}} | \sigma \cdot \tau \rangle\rangle \\
&\leq \frac{1}{(d^2-1)^{\frac{3}{2}}} \sum_{\tau, \sigma \in \boldsymbol{\sigma}_q} \langle\langle \sigma | \mathcal{I} - \mathcal{E}_{\mathrm{u}}^\dagger \mathcal{E}_{\mathrm{u}} | \sigma \rangle\rangle \langle\langle \tau | \mathcal{I} - \mathcal{E}_{\mathrm{u}}^\dagger \mathcal{E}_{\mathrm{u}} | \tau \rangle\rangle \\
&= \sqrt{d^2-1}(1-u)^2,
\end{aligned} \tag{7.201}$$

where in the last inequality the sum is completed by adding the nonnegative terms $\langle\langle \sigma | \mathcal{I} - \mathcal{E}_{\mathrm{u}}^\dagger \mathcal{E}_{\mathrm{u}} | \sigma \rangle\rangle \langle\langle \tau | \mathcal{I} - \mathcal{E}_{\mathrm{u}}^\dagger \mathcal{E}_{\mathrm{u}} | \tau \rangle\rangle$ for all $\tau \in \boldsymbol{\sigma}_q$ and $\sigma \in \boldsymbol{\sigma}_q \setminus \mathbf{C}_\tau$. Note that this is the same bound as on $a_{[\mathrm{adj}]}$. ■

Finally two more propositions are needed to bound the inner products in the expanded variance expression. The tool for this is the following. This proposition is formulated for any general CPTP map $\mathcal{E}$ and hermitian operators $X, Y \in \mathcal{M}_d$. This theorem is applicable to inner products involving the map $\mathcal{N}^{m-s-1}$, since this is a CPTP map.

Proposition 7.7. Let $\mathcal{E}$ be a CPTP map on a general Hilbert space $\mathcal{H}$. Then for any pair of hermitian operators $X, Y \in \mathcal{M}_d$

$$\langle\langle X | \mathcal{E} | Y \rangle\rangle \leq \|X\|_\infty \|Y\|_1. \tag{7.202}$$

Proof. By Von Neumann's trace inequality and Hölders inequality (lemma 7.5) it follows that

$$\langle\langle X|\mathcal{E}|Y\rangle\rangle = \mathrm{Tr}[X\mathcal{E}(Y)] \leq \|X\|_\infty \|\mathcal{E}(Y)\|_1, \tag{7.203}$$

using that X and $\mathcal{E}(Y)$ are hermitian. We then use the induced trace norm (the $1 \to 1$ norm) and the fact that the map $\mathcal{E}$ is a CPTP map so that $\|\mathcal{E}\|_{1\to 1} \leq 1$ (lemma 7.1). Therefore

$$\|X\|_\infty \|\mathcal{E}(Y)\|_1 \leq \|X\|_\infty \|\mathcal{E}\|_{1\to 1} \|Y\|_1 \leq \|X\|_\infty \|Y\|_1. \tag{7.204}$$

Putting this together proves the bound. ■

In order to apply the above proposition to the inner products occurring in the variance proof, a bound on the norms of the operators A_i with $i \in \mathcal{Z}_{TS}$ is needed.

Proposition 7.8 (Norm bounds on A_i)**.** Let $\{A_i : i \in \mathcal{Z}_{TS}\}$ be defined as in eq. (7.88). Then for $d \geq 4$ the following bounds hold

$$\|A_i\|_1 \leq d^2 \qquad \text{and} \qquad \|A_i\|_\infty \leq \sqrt{\frac{6}{(d-2)(d-1)}}, \qquad \forall i \in \mathcal{Z}_{TS}. \tag{7.205}$$

If $d = 2$, then $\mathcal{Z}_{TS} = \{S;\, 1, 2\}$, and

$$\|A_S\|_1 = \frac{5}{\sqrt{3}}, \qquad \|A_S\|_\infty = \frac{1}{\sqrt{3}}, \tag{7.206}$$

$$\|A_{1,2}\|_1 = 2\sqrt{2}, \qquad \|A_{1,2}\|_\infty = \frac{\sqrt{2}}{3}. \tag{7.207}$$

Proof. For the $d = 2$ case, the norms can be computed directly, since A_S and $A_{1,2}$ are explicitly defined in eq. (7.91)-eq. (7.92). By direct computation the result follows. For $d \geq 4$, the trace norm bound is trivial, since

$$\|A_i\|_1 \leq \sqrt{d^4}\|A_i\|_2 = d^2, \tag{7.208}$$

by Hölder's inequality. The last equality uses the fact that A_i are Hilbert-Schmidt normalized ($\|A_i\|_2 = 1$). The effort of the proof is in the bound on $\|A_i\|_\infty$.
The proof of this statement uses the description of the tensor-2 Liouville representation of [13] over [15], since their description is basis-free. Ref. [13] considers the action of the Clifford group C_q on $\mathcal{H}^{\otimes 4}$. The representation $\mathcal{H}^{\otimes 4}$ of the Clifford group C_q decomposes as

$$\mathcal{H}^{\otimes 4} = \bigoplus_k W_k \otimes \mathbb{C}^{d_k} \tag{7.209}$$

where W_k are irreducible, pairwise inequivalent representations of the Clifford group that occur with multiplicity d_k. Here k is just an index for the irreducible, inequivalent representations. Descriptions of these spaces and explicit expressions for their dimensions are given in [13] (there the index k runs over Young Diagrams λ and signs s). We will show that

$$\|A_i\|_\infty \leq \max_k \frac{1}{\sqrt{|W_k|}}. \tag{7.210}$$

Since the dimensions of all W_k are given, the maximization can easily be done.
Using the intertwining isomorphism $\mathcal{M}_d \simeq \mathcal{H} \otimes \mathcal{H}^*$ the tensor-4 Liouville representation on $\mathcal{M}_d^{\otimes 4}$ can be written in terms of the decomposition eq. (7.209):

$$\mathcal{M}_d^{\otimes 4} = \bigoplus_{k,l} \mathsf{L}(W_l, W_k) \otimes \mathsf{L}(\mathbb{C}^{d_l}, \mathbb{C}^{d_k}). \tag{7.211}$$

where $\mathsf{L}(\mathbb{C}^{d_l}, \mathbb{C}^{d_k})$ denotes the linear operators from $\mathbb{C}^{d_l}$ to $\mathbb{C}^{d_k}$ and $\mathsf{L}(W_l, W_k)$ denotes the linear operators from W_l to W_k. In principle $\mathsf{L}(W_l, W_k)$ do not carry irreducible representations. However, only the trivial subrepresentations of $\mathsf{L}(W_l, W_k)$ (denoted $(\mathsf{L}(W_l, W_k))_q^{\mathsf{C}}$) are relevant, since

$$(\mathcal{M}_d^{\otimes 4})_q^{\mathsf{C}} = \bigoplus_{k,l} (\mathsf{L}(W_l, W_k))_q^{\mathsf{C}} \otimes \mathsf{L}(\mathbb{C}^{d_l}, \mathbb{C}^{d_k}). \tag{7.212}$$

The key point is that every element $\varphi \in (\mathsf{L}(W_l, W_k))_q^{\mathsf{C}}$ is an intertwining operator between the representations W_k and W_l [33]. By Schur's Lemma [33] and the fact that W_k are mutually inequivalent irreducible representations it follows that $\varphi \propto \delta_{k,l} \mathbb{1}_{W_k}$. Therefore

$$(\mathcal{M}_d^{\otimes 4})_q^{\mathsf{C}} = \bigoplus_k \mathrm{Span}\{\mathbb{1}_{W_k}\} \otimes \mathcal{M}_{d_k}). \tag{7.213}$$

This description provides a simple orthogonal basis for the space $(\mathcal{M}_d^{\otimes 4})_q^{\mathsf{C}}$, namely

$$\mathcal{A} = \{P_{W_k} \otimes E_{m,n} | k; m, n = 1, ..., d_k\}, \tag{7.214}$$

where P_{W_k} is the orthogonal projection onto W_k and $\{E_{m,n} | m, n = 1, ..., d_k\}$ is the canonical (or any other) orthonormal basis of $\mathcal{M}_{d_k}$. Normalizing with respect to the Hilbert-Schmidt norm yields the orthonormal basis operators

$$A_{k,m,n} = \frac{P_{W_k}}{\sqrt{|W_k|}} \otimes E_{m,n}. \tag{7.215}$$

Note that our basis operators $\{A_i : i \in \mathcal{Z}_{TS}\}$ might be different than these $A_{k,m,n}$. However, these A_i also span trivial subrepresentations of $\mathcal{M}_d^{\otimes 4}$, so $A_i \in (\mathcal{M}_d^{\otimes 4})_q^{\mathsf{C}}$. We now show that $\|A\| \leq \max_k |W_k|^{-\frac{1}{2}}$ for all $A \in (\mathcal{M}_d^{\otimes 4})_q^{\mathsf{C}}$ such that $\|A\|_2 = 1$. Therefore this bound holds in particular for our A_i of interest. To do so, A is written in the basis $\mathcal{A}$ as

$$A = \sum_k \sum_{m,n=1}^{d_k} \alpha_{k,m,n} A_{k,m,n}, \qquad \text{s.t.} \qquad \sum_k \sum_{m,n=1}^{d_k} |\alpha_{k,m,n}|^2 = 1. \tag{7.216}$$

Now we use that the operator $A \in (\mathcal{M}_d^{\otimes 4})_q^{\mathsf{C}}$ is block diagonal with respect to the spaces $\mathrm{Span}\{\mathbb{1}_{W_k}\} \otimes \mathcal{M}_{d_k}$ (see eq. (7.213)). Therefore the infinity norm can be computed as the

maximum over k of the infinity norm of A restricted to $\mathrm{Span}\{\mathbb{1}_{W_k}\} \otimes \mathcal{M}_{d_k}$), yielding

$$\|A\|_\infty = \left\| \sum_k \sum_{m,n=1}^{d_k} \alpha_{k,m,n} A_{k,m,n} \right\|_\infty = \max_k \left\| \sum_{m,n=1}^{d_k} \alpha_{k,m,n} A_{k,m,n} \right\|_\infty = \max_k \left\| \frac{P_{W_k}}{\sqrt{|W_k|}} \otimes \sum_{m,n=1}^{d_k} \alpha_{k,m,n} E_{m,n} \right\|_\infty . \tag{7.217}$$

Using some basic properties of the Schatten p-norms, this is bounded as follows

$$\|A\|_\infty = \max_k \left\| \frac{P_{W_k}}{\sqrt{|W_k|}} \right\|_\infty \left\| \sum_{m,n=1}^{d_k} \alpha_{k,m,n} E_{m,n} \right\|_\infty = \max_k \frac{\|P_{W_k}\|_\infty}{\sqrt{|W_k|}} \left\| \sum_{m,n=1}^{d_k} \alpha_{k,m,n} E_{m,n} \right\|_\infty \leq \max_k \frac{1}{\sqrt{|W_k|}}, \tag{7.218}$$

using that $\|P_{W_k}\|_\infty = 1$ and

$$\left\| \sum_{m,n=1}^{d_k} \alpha_{k,m,n} E_{m,n} \right\|_\infty \leq \left\| \sum_{m,n=1}^{d_k} \alpha_{k,m,n} E_{m,n} \right\|_2 \leq \sum_{m,n=1}^{d_k} |\alpha_{k,m,n}|^2 \left\| E_{m,n} \right\|_2 = 1. \tag{7.219}$$

By Lemma 1 of [13], which gives all dimensions $|W_k|$, it follows that

$$\|A\|_\infty \leq \max_k \frac{1}{\sqrt{|W_k|}} = \sqrt{\frac{6}{(d-1)(d-2)}}, \tag{7.220}$$

provided that $d = 2^q \geq 4$, $q \in \mathbb{N}$. This proves the last bound. ■

Finally, there is one inner product in the proof of theorem 1 for which a sharper bound can be found than using proposition 7.7 and proposition 7.8. This sharper bound is given in the following proposition.

Proposition 7.9. Let $\mathcal{N}$ be defined as in eq. (7.96), with $\mathcal{E}$ a single-qubit or unital quantum channel. Then for any $m \in \mathbb{N}$ the following bound holds

$$\langle\langle A_i | \mathcal{N}^m | B_2 B_2 \rangle\rangle \leq \frac{1}{\sqrt{|V_i|}}, \qquad \forall i \in \mathcal{Z}_{TS}. \tag{7.221}$$

Proof. Slightly rewriting the inner product yields

$$\langle\langle A_i | \mathcal{N}^m | B_2 B_2 \rangle\rangle = \langle\langle A_i | \mathcal{N}^m (B_2 B_2) \rangle\rangle. \tag{7.222}$$

From the definition of $\mathcal{N}$ eq. (7.96) it follows that

$$\mathcal{N}^m (B_2 B_2) = \frac{1}{|\mathsf{C}_q|^m} \sum_{\mathbf{j}} \mathcal{G}_{\mathbf{j}}^{\otimes 4} (B_2 B_2) = \frac{1}{|\mathsf{C}_q|^m} \sum_{\mathbf{j}} [\mathcal{G}_{\mathbf{j}}^{\otimes 2} (B_2)]^{\otimes 2}, \tag{7.223}$$

where the sum is over all noisy sequences of length m indexed by $\mathbf{j}$ (i.e. $\mathbf{j}$ is a multi-index of length m). We will show that $\|\mathcal{G}_{\mathbf{j}}^{\otimes 2}(B_2)\|_2 \leq 1$. we treat the multi-qubit and single-qubit case separately. In the multi-qubit case, we have

$$\|\mathcal{G}_{\mathbf{j}}^{\otimes 2}(B_2)\|_2 \leq \|\mathcal{G}_{\mathbf{j}}^{\otimes 2}\|_{2\to 2}\|B_2\|_2 = \|\mathcal{G}_{\mathbf{j}}\|_{2\to 2}^2. \tag{7.224}$$

The inequality follows from the definition of the induced Schatten norms (see eq. (7.74)). The equality is due to the fact that $\|B_2\|_2 = 1$ is normalized. Under the assumption that $\mathcal{E}$ is unital, the entire sequence $\mathcal{G}_{\mathbf{j}}$ is unital. Therefore by lemma 7.1 (Perez-Garcia), $\|\mathcal{G}_{\mathbf{j}}\|_{2\to 2}^2 \leq 1$. This shows that $\|\mathcal{G}_{\mathbf{j}}^{\otimes 2}(B_2)\|_2 \leq 1$.
In case of a single-qubit, non-unital error channels $\mathcal{E}$, some extra care must be taken. Let us denote $\mathcal{M}_d^H := \{A \in \mathcal{M}_d : \mathrm{Tr}[A] = 0, A = A^\dagger\} = \mathrm{Span}_{\mathbb{R}}\{\sigma : \sigma \in \boldsymbol{\sigma}_q\}$ as the traceless hermitian subspace of $\mathcal{M}_d$. This space is a vector space over $\mathbb{R}$, with an orthonormal basis $\boldsymbol{\sigma}_q$. Since $\mathcal{G}_{\mathbf{j}}$ is positive (and thus maps hermitian operators to hermitian operators) and trace-preserving, it maps the traceless hermitian subspace $\mathcal{M}_d^H$ to itself. Observe that $B_2 \in (\mathcal{M}_d^H)^{\otimes 2}$. Therefore restrict $\mathcal{G}_{\mathbf{j}}^{\otimes 2}$ to $(\mathcal{M}_d^H)^{\otimes 2}$. This results in

$$\|\mathcal{G}_{\mathbf{j}}^{\otimes 2}(B_2)\|_2 = \|\mathcal{G}_{\mathbf{j}}^{\otimes 2}\big|_{(\mathcal{M}_d^H)^{\otimes 2}}(B_2)\|_2 \leq \|\mathcal{G}_{\mathbf{j}}^{\otimes 2}\big|_{(\mathcal{M}_d^H)^{\otimes 2}}\|_{2\to 2}\|B_2\|_2 = \|\mathcal{G}_{\mathbf{j}}\big|_{\mathcal{M}_d^H}\|_{2\to 2}^2. \tag{7.225}$$

The first equality is the restriction of $\mathcal{G}_{\mathbf{j}}$ to the traceless hermitian subspace. The inequality follows from the definition of the induced Schatten norm (eq. (7.74)). The final equality is due to the fact that $\|B_2\|_2 = 1$. The key point of restricting to the traceless hermitian subspace $\|\mathcal{G}_{\mathbf{j}}\big|_{\mathcal{M}_d^H}\|_{2\to 2}$ allows for the application of statement eq. (7.76) of lemma 7.1 (Perez-Garcia). By the lemma (where $\|\mathcal{G}_{\mathbf{j}}\big|_{\mathcal{M}_d^H}\|_{2\to 2}$ is denoted $\|\mathcal{G}_{\mathbf{j}}\|_{2\to 2}^H$), we have

$$\|\mathcal{G}_{\mathbf{j}}\big|_{\mathcal{M}_d^H}\|_{2\to 2} \leq \sqrt{\frac{d}{2}}, \tag{7.226}$$

which in the single-qubit case means $\|\mathcal{G}_{\mathbf{j}}\big|_{\mathcal{M}_d^H}\|_{2\to 2} \leq 1$. Therefore, we also have $\|\mathcal{G}_{\mathbf{j}}^{\otimes 2}(B_2)\|_2 \leq 1$ in the single-qubit, non-unital case.
We have thus established that $\|\mathcal{G}_{\mathbf{j}}^{\otimes 2}(B_2)\|_2 \leq 1$ for single-qubit or unital noise maps $\mathcal{E}$. Therefore, the following upper bound is valid

$$\begin{aligned}
\langle\!\langle A_i|\mathcal{N}^m|B_2B_2\rangle\!\rangle &= \frac{1}{|\mathsf{C}_q|^m}\sum_{\mathbf{j}}\langle\!\langle A_i|[\mathcal{G}_{\mathbf{j}}^{\otimes 2}(B_2)]^{\otimes 2}\rangle\!\rangle \\
&\leq \frac{1}{|\mathsf{C}_q|^m}\sum_{\mathbf{j}}\max_{\substack{Q\in(\mathcal{M}_d^H)^{\otimes 2}\\ \|Q\|_2\leq 1}}\langle\!\langle A_i|Q^{\otimes 2}\rangle\!\rangle \\
&= \max_{\substack{Q\in(\mathcal{M}_d^H)^{\otimes 2}\\ \|Q\|_2\leq 1}}\langle\!\langle A_i|Q^{\otimes 2}\rangle\!\rangle.
\end{aligned} \tag{7.227}$$

In the second line, we have replaced the particular operator $\mathcal{G}_{\mathbf{j}}^{\otimes 2}(B_2) \in (\mathcal{M}_d^H)^{\otimes 2}$ which satisfies $\|\mathcal{G}_{\mathbf{j}}^{\otimes 2}(B_2)\|_2 \leq 1$ with the maximization over all operators $Q \in (\mathcal{M}_d^H)^{\otimes 2}$ that

satisfy $\|Q\|_2 \leq 1$. To continue, we use the definition of A_i (eq. (7.88)), which is given by

$$A_i = \frac{1}{\sqrt{|V_i|}} \sum_{s=1}^{|V_i|} v_s^{(i)} v_s^{(i)}, \quad \forall i \in \mathcal{Z}_{TS}, \tag{7.228}$$

where $\{v_s^{(i)}\}$ is an orthonormal basis of $V_i \subset (\mathcal{M}_d^H)^{\otimes 2}$. Let us expand Q is this basis,

$$Q = q_\perp v_\perp^{(i)} + \sum_{s=1}^{|V_i|} q_s v_s^{(i)} \quad \text{s.t.} \quad |q_\perp|^2 + \sum_{s=1}^{|V_i|} |q_s|^2 \leq 1, \quad q_\perp, q_s \in \mathbb{C}, \ \forall s = 1, ..., |V_i|. \tag{7.229}$$

Here $q_\perp v_\perp^{(i)}$ is the component of Q in the space orthogonal to V_i, i.e. $q_\perp v_\perp^{(i)} \in (\mathcal{M}_d^H)^{\otimes 2} \setminus V_i$. The condition on $q_\perp$ and the q_s follow from the requirement that $\|Q\|_2 \leq 1$. Actually, there are additional constraints on $q_\perp$ and the q_s needed to ensure that Q is traceless and hermitian, but these constraints are not necessary to prove the result. Using the expansion eq. (7.229) it follows that

$$\max_{\substack{Q \in (\mathcal{M}_d^H)^{\otimes 2} \\ \|Q\|_2 \leq 1}} \langle\langle A_i | Q^{\otimes 2} \rangle\rangle \leq \max_{\substack{\{q_s\} \\ \sum_s |q_s|^2 \leq 1}} \frac{1}{\sqrt{|V_i|}} \sum_{s,t,k=1}^{|V_i|} |q_s q_t| |\langle\langle v_k^{(i)} v_k^{(i)} | v_s^{(i)} v_t^{(i)} \rangle\rangle| \tag{7.230}$$

$$= \max_{\substack{\{q_s\} \\ \sum_s |q_s|^2 \leq 1}} \frac{1}{\sqrt{|V_i|}} \sum_{k=1}^{|V_i|} |q_k|^2 \leq \frac{1}{\sqrt{|V_i|}}, \tag{7.231}$$

using the fact that $\langle\langle v_k^{(i)} v_k^{(i)} | v_s^{(i)} v_t^{(i)} \rangle\rangle = \delta_{sk}\delta_{tk}$ by orthonormality of the basis. This completes the proof. ■

8

Character randomized benchmarking

Randomized benchmarking is a technique for estimating the average fidelity of a set of quantum gates. However, if this gate-set is not the multi-qubit Clifford group, robustly extracting the average fidelity is difficult. Here we propose a new method based on representation theory that has little experimental overhead and robustly extracts the average fidelity for a broad class of gate-sets. We apply our method to a multi-qubit gate-set that includes the T-gate, and propose a new interleaved benchmarking protocol that extracts the average fidelity of a two-qubit Clifford gate using only single-qubit Clifford gates as reference.

This chapter is based on J. Helsen, X. Xue, L.M.K. Vandersypen & S. Wehner, *A new class of efficient randomized benchmarking protocols*, hopefullyfancyjournalhere

8.1. Introduction

In chapter 4 we discussed randomized benchmarking. There we limited our discussion to randomized benchmarking of gatesets that are 2-designs, in particular the Clifford group. In this case, as we saw in chapter 4, it can be shown (under the assumption of gate-independent noise) that the data $\{p_m\}_m$ yielded by randomized benchmarking can be fitted to a single exponential decay of the form

$$p_m \approx_{\text{fit}} A + Bf^m \tag{8.1}$$

where A, B only depend on how well the state ρ was prepared and measured and the quality parameter f only depends on how well the gates in the gate-set G are implemented. This parameter f can be related to the average fidelity F_{avg} [1].

However it is possible, and desirable, to perform randomized benchmarking on gate-sets that are not the Clifford group and a wide array of proposals for randomized benchmarking using non-Clifford gate-sets appear in the literature [2–7]. The most prominent use-case is benchmarking a gate-set G that includes the vital T-gate [2, 3, 6] which, together with the Clifford group, forms a universal set of gates for quantum computing. Another use-case is simultaneous randomized benchmarking [7], which extracts information about crosstalk and unwanted coupling between neighboring qubits by performing randomized benchmarking on the gate-set consisting of single qubit Clifford gates on all qubits. In these cases, and in other examples of randomized benchmarking with non-Clifford gate-sets [4, 6, 7], the fitting relation eq. (8.1) does not hold. From lemma 4.1 in chapter 4 we see that it must instead be generalized to

$$p_m \approx_{\text{fit}} \sum_{\lambda \in R_{\mathsf{G}}} A_\lambda f_\lambda^m, \tag{8.2}$$

where R_{G} is an index set that only depends on the chosen gate-set G, the f_λ are general 'quality parameters' that only depend on the gates being implemented and the A_λ prefactors depend only on SPAM. The above holds because for non-Clifford groups averaging does not fully smear out the noise. Rather the system state space will split into several 'sectors' labeled by λ such that states within the same sector experience the same noise but the noise varies from sector to sector. The interpretation of the parameters f_λ varies depending on the gate-set G. In the case of simultaneous randomized benchmarking [7] they can be interpreted as a measure of crosstalk and unwanted coupling between neighboring qubits. For other gate-sets an interpretation is not always available. However, as was pointed out for specific gate-sets in [2–4, 6] and for general finite groups in [5] the parameters f_λ can always be jointly related (see eq. (8.5)) to the average fidelity F_{avg} of the gate-set G. This means that randomized benchmarking can extract the average fidelity of a gate-set even when it is not the Clifford group.

However in practice the multi-parameter fitting problem given by eq. (8.2) is difficult to perform, with poor confidence intervals around the parameters f_λ unless impractically large amounts of data are gathered. More fundamentally it is, even in the limit of infinite data, impossible to associate the estimates from the fitting procedure to the correct decay

channel in eq. (8.2) and thus to the correct f_λ, making it impossible to reliably reconstruct the average fidelity of the gate-set.

In the current literature on non-Clifford randomized benchmarking, with the notable exception of [6], this issue is sidestepped by performing the experiment several times using different input states ρ_λ that are carefully tuned to maximize one of the prefactors A_λ while minimizing the others. This is unsatisfactory for several reasons: (1) the accuracy of the fit now depends on the preparation of ρ_λ, undoing one of the main advantages of randomized benchmarking over other methods such as direct fidelity estimation [8], (2) it is, for more general gate-sets, not always possible to efficiently find such a maximizing state ρ_λ and (3) both previous problems become more pronounced as the number of quality parameters f_λ increases. These problems limit the practical applicability of current non-Clifford randomized benchmarking protocols and more generally restrict which groups can practically be benchmarked.

In this chapter we propose an adaptation of the randomized benchmarking procedure, which we call character randomized benchmarking, which solves the above problems and allows reliable and efficient extraction of average fidelities for gate-sets that are not the Clifford group. We begin, in section 8.2, by discussing the general method, before applying it to specific examples in section 8.3. In section 8.4 we discuss using character randomized benchmarking in practice and argue the new method does not impose significant experimental overhead. Finally in section 8.5 we argue that character randomized benchmarking is robust against gate-dependent fluctuations. Previous adaptations of randomized benchmarking, as discussed in [6, 9–11], can be regarded as special cases of our method.

8.2. CHARACTER RANDOMIZED BENCHMARKING

In this section we will introduce the character randomized benchmarking protocol. We will use the Liouville representation of quantum channels, see chapter 2, denoting a unitary CPTP map by $\mathcal{G}$. For convenience we will assume *gate-independent noise*. This means we assume the existence of a CPTP map $\mathcal{E}$ such that the implementation of a unitary G is given by $\tilde{\mathcal{G}} = \mathcal{E}\mathcal{G}$ for all $G \in \mathsf{G}$. Our results however also hold in the case of gate-dependent noise, see section 8.5.3.

As explained in chapter 4, under the assumption of gate independent noise the average sequence probability p_m of the randomized benchmarking procedure with a gate-set G (with input state ρ and measurement POVM $\{Q, \mathbb{1} - Q\}$) with sequence length m can be written as:

$$p_m = \langle\langle Q| \left(\mathop{\mathbb{E}}_{G\in\mathsf{G}} \mathcal{G}^\dagger \mathcal{E} \mathcal{G} \right)^m |\rho\rangle\rangle. \tag{8.3}$$

where $\mathbb{E}_{G\in\mathsf{G}}$ denotes the uniform average over G.

As discussed in chapter 4, the key insight to randomized benchmarking is that $\mathcal{G}$ is a *representation* of $G \in \mathsf{G}$. This representation will not in general be irreducible but will rather decompose into irreducible subrepresentations, that is $\mathcal{G} = \bigoplus_{\lambda\in R_\mathsf{G}} \varphi_\lambda(G)$ where R_G is

an index set and φ_λ are irreducible representations of G. We will assume for convenience that all the representations φ_λ are mutually inequivalent, i.e. $\mathcal{G}$ is a multiplicity-free representation. As was proven in lemma 4.1 in chapter 4 we can write eq. (8.3) as

$$p_m = \sum_\lambda \langle\langle Q|\mathcal{P}_\lambda|\rho\rangle\rangle f_\lambda^m, \tag{8.4}$$

where $\mathcal{P}_\lambda$ is the orthogonal projector onto the support of φ_λ (note that this is a superoperator) and $f_\lambda := \mathrm{Tr}(\mathcal{P}_\lambda\mathcal{E})/\mathrm{Tr}(\mathcal{P}_\lambda)$ is the quality parameter associated to the representation φ_λ (note that the trace is taken over superoperators). This reproduces eq. (8.2). The average fidelity of the gate-set G can then be related to the parameters f_λ as

$$\frac{2^q F_{\mathrm{avg}} + 1}{2^q + 1} = \sum_{\lambda \in R_{\mathsf{G}}} \frac{\mathrm{Tr}(\mathcal{P}_\lambda)}{2^q} f_\lambda. \tag{8.5}$$

In order to estimate the parameters f_λ we will make use of methods from the character theory of representations. As discussed in chapter 3, associated to any representation $\hat{\varphi}$ of a group $\hat{\mathsf{G}}$ is a character function $\chi_{\hat{\varphi}} : \hat{\mathsf{G}} \to \mathbb{R}$, from the group to the real numbers [*]. Associated to this function is the following projection formula (see lemma 3.4 in chapter 3):

$$\mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} \chi_{\hat{\varphi}}(\hat{G})\hat{\mathcal{G}} = \frac{1}{|\hat{\varphi}|}\mathcal{P}_{\hat{\varphi}}, \tag{8.6}$$

where $\mathcal{P}_{\hat{\varphi}}$ is the projector onto the support of $\hat{\varphi}$ (and any subrepresentation of $\hat{\mathcal{G}}$ equivalent to $\hat{\varphi}$) and $|\hat{\varphi}|$ is the dimension of the representation $\hat{\varphi}$.

We will leverage this formula to adapt the randomized benchmarking procedure in a way that singles out a particular exponential decay f_λ^m in eq. (8.2). To see this consider a group G such that $\mathcal{G}$ is multiplicity free. We will estimate the average fidelity of this group and will henceforth refer to it as the 'benchmarking group'. Now fix a $\lambda' \in R_{\mathsf{G}}$. $f_{\lambda'}$ is the quality parameter associated to a specific subrepresentation $\varphi_{\lambda'}$ of $\mathcal{G}$.

Now we must find a way to leverage eq. (8.6). We do this by introducing another group $\hat{\mathsf{G}} \subset \mathsf{G}$ which we will refer to as the 'character group' going forward. It is important to choose $\hat{\mathsf{G}}$ in such a way that the Liouville representation $\hat{\mathcal{G}}$ of $\hat{\mathsf{G}}$ has a subrepresentation $\hat{\varphi}$, with character function $\chi_{\hat{\varphi}}$, which has support inside the representation $\varphi_{\lambda'}$ of G. This means that we want that $\mathcal{P}_{\hat{\varphi}}\mathcal{P}_{\lambda'} = \mathcal{P}_{\hat{\varphi}}$. Note that such a pair $\hat{\mathsf{G}}, \hat{\varphi}$ always exists; we can always choose $\hat{\mathsf{G}} = \mathsf{G}$ and $\hat{\varphi} = \varphi_{\lambda'}$. However other natural choices often exist, as we shall see in the examples in section 8.3. Now we can consider the following adapted randomized benchmarking protocol which we call character randomized benchmarking. This protocol will estimate the quality parameter $f_{\lambda'}$ associated to the benchmarking group G.

The major difference between the standard and character randomized benchmarking protocols is the introduction of an extra average over the character group $\hat{\mathsf{G}}$. This extra gate

[*] Generally the character function is a map to the complex numbers, but in our case it is enough to only consider real representations.

1. Choose a state ρ and a two-component POVM $\{Q, \mathbb{1} - Q\}$ such that $\mathrm{Tr}(Q\mathcal{P}_{\hat{\varphi}}(\rho))$ is large.
2. Sample $\vec{G} = G_1, \ldots, G_m$ uniformly at random from G.
3. Sample $\hat{G}$ uniformly at random from $\hat{\mathsf{G}}$.
4. Prepare the state ρ and apply the gates $(G_1\hat{G}), G_2, \ldots G_m$.
5. Compute the inverse $G_{\mathrm{inv}} = (G_m \cdots G_1)^\dagger$ and apply it (note that $\hat{G}$ is not inverted).
6. Estimate the weighted 'survival probability'
$$k_m^{\lambda'}(\vec{G}, \hat{G}) = |\hat{\varphi}|\chi_{\hat{\varphi}}(\hat{G})\langle\langle Q|\widetilde{\mathcal{G}}_{\mathrm{inv}}\widetilde{\mathcal{G}}_m \cdots \widetilde{(\mathcal{G}_1\hat{\mathcal{G}})}\,|\rho\rangle\rangle.$$
7. Repeat for suitably many $\hat{G} \in \hat{\mathsf{G}}$ to estimate the average $k_m^{\lambda'}(\vec{G}) = \mathbb{E}_{\hat{G}}(k_m^{\lambda'}(\vec{G}, \hat{G}))$.
8. Repeat for suitably many $\vec{G}$ to estimate the average $k_m^{\lambda'} = \mathbb{E}_{\vec{G}}(k_m^{\lambda'}(\vec{G}))$.
9. Repeat for various m and fit to the exponential function $Af_{\lambda'}^m$ to obtain $f_{\lambda'}$.

Figure 8.1: The character randomized benchmarking protocol

$\hat{G} \in \hat{\mathsf{G}}$ is not included when computing the global inverse $G_{\mathrm{inv}} = (G_1 \ldots G_m)^\dagger$. Note that this extra gate is compiled into the sequence of gates $(G_1, \ldots, G_m)$ and thus does not result in extra noise. The average over the elements of $\hat{\mathsf{G}}$ is also weighted by the character function $\chi_{\hat{\varphi}}$ associated to the representation $\hat{\varphi}$ of $\hat{\mathsf{G}}$. This means that we are in effect 'constructing' the projector in eq. (8.6). Similar to eq. (8.3) we can rewrite the uniform average over all $\vec{G} \in \mathsf{G}^{\times m}$ and $\hat{G} \in \hat{\mathsf{G}}$ as

$$k_m^{\lambda'} = |\hat{\varphi}|\langle\langle Q| \left[\mathop{\mathbb{E}}_{G\in\mathsf{G}} \mathcal{G}^\dagger\mathcal{E}\mathcal{G}\right]^m \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} (\chi_{\hat{\varphi}}(\hat{G})\hat{\mathcal{G}})|\rho\rangle\rangle.$$

Using the character projection formula (eq. (8.6)) and the standard randomized benchmarking representation theory formula (eq. (8.4)) we can write this as

$$k_m^{\lambda'} = \sum_{\lambda\in R_{\mathsf{G}}} \langle\langle Q|\mathcal{P}_\lambda\mathcal{P}_{\hat{\varphi}}|\rho\rangle\rangle f_\lambda^m = \langle\langle Q|\mathcal{P}_{\hat{\varphi}}|\rho\rangle\rangle f_{\lambda'}^m, \tag{8.7}$$

since we have chosen the character group $\hat{\mathsf{G}}$ and $\hat{\varphi}$ such that the support of $\mathcal{P}_{\hat{\varphi}}$ is a subspace of the support of $\mathcal{P}_{\lambda'}$. We can state the above formula more formally as the following lemma.

Lemma 8.1. Consider the character randomized benchmarking procedure as described in fig. 8.1 with a benchmarking group G (with Liouville representation $\mathcal{G} = \oplus_{\lambda\in R_{\mathsf{G}}}\varphi_\lambda(G)$), a character group $\hat{\mathsf{G}}$ such that the Liouville representation of $\hat{\mathcal{G}}$ has an irreducible subrepresentation $\hat{\varphi}$ s.t. $\mathcal{P}_{\hat{\varphi}}\mathcal{P}_{\lambda'} = \mathcal{P}_{\hat{\varphi}}$, a parameter $\lambda' \in R_{\mathsf{G}}$ and a set of sequence lengths $\mathbb{M}$. Let $\mathcal{E}$ be a CPTP map such that $\tilde{\mathcal{G}} = \mathcal{E}\mathcal{G}$ for all $G \in \mathsf{G}$. Then the output data $\{k_m^{\lambda'}\}_{m\in\mathbb{M}}$

can be fitted to an exponential decay of the form

$$k_m^{\lambda'} =_{\text{fit}} A_{\lambda'} f_{\lambda'}^m, \tag{8.8}$$

with $f_{\lambda'} = \mathrm{Tr}(\mathcal{E}\mathcal{P}_{\lambda'})/\mathrm{Tr}(\mathcal{P}_{\lambda'})$ where $\mathcal{P}_{\lambda'}$ is the projection onto the irreducible subrepresentation $\varphi_{\lambda'}$ of the Liouville representation of G.

Proof. Choose $m \in \mathbb{M}$ and consider the expression for $k_m^{\lambda'}$. Using the fact that $\tilde{\mathcal{G}} = \mathcal{E}\mathcal{G}$ for all $G \in \mathsf{G}$ we have

$$k_m^{\lambda'} = |\hat{\varphi}| \mathop{\mathbb{E}}_{G_1,\ldots,G_m \in \mathsf{G}} \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} \chi_{\hat{\varphi}}(\hat{G}) \langle\langle Q|\mathcal{E}\mathcal{G}_{\text{inv}}\mathcal{E}\mathcal{G}_m\mathcal{E}\ldots\mathcal{E}\mathcal{G}\hat{\mathcal{G}}|\rho\rangle\rangle. \tag{8.9}$$

Using linearity we move the average over $\hat{\mathsf{G}}$ inside the inner product and use the character projection formula eq. (8.6). This gives

$$k_m^{\lambda'} = \mathop{\mathbb{E}}_{G_1,\ldots,G_m \in \mathsf{G}} \langle\langle Q|\mathcal{E}\mathcal{G}_{\text{inv}}\mathcal{E}\mathcal{G}_m\mathcal{E}\ldots\mathcal{E}\mathcal{G}\mathcal{P}_{\hat{\varphi}}|\rho\rangle\rangle, \tag{8.10}$$

where $\mathcal{P}_{\hat{\varphi}}$ is as before the projector onto the irreducible subrepresentation $\hat{\varphi}$ of the Liouville representation $\hat{\mathcal{G}}$ of $\hat{\mathsf{G}}$. Now we use the reasoning of lemma 4.1 to rewrite the above expression further to

$$k_m^{\lambda'} = \langle\langle Q|\mathcal{E}\left(\sum_{\lambda\in R_{\mathsf{G}}} f_\lambda \mathcal{P}_\lambda\right)^m \mathcal{P}_{\hat{\varphi}}|\rho\rangle\rangle. \tag{8.11}$$

Now we use the fact that $\mathcal{P}_{\hat{\varphi}}\mathcal{P}_\lambda = \delta_{\lambda\lambda'}\mathcal{P}_{\hat{\varphi}}$ by construction and thus we get

$$k_m^{\lambda'} = f_{\lambda'}^m \langle\langle Q|\mathcal{P}_{\hat{\varphi}}|\rho\rangle\rangle, \tag{8.12}$$

where we set $Q \to \mathcal{E}^\dagger(Q)$. This means we can fit the data $\{k_m^{\lambda'}\}_{m\in\mathbb{M}}$ to an exponential decay of the form

$$k_m^{\lambda'} =_{\text{fit}} A_{\lambda'} f_{\lambda'}^m. \tag{8.13}$$

This completes the proof. ∎

This means the character randomized benchmarking protocol isolates the exponential decay associated to the quality parameter $f_{\lambda'}$ independent of state preparation and measurement. In practice one should take care to choose Q and ρ in a way that makes $\langle\langle Q|\mathcal{P}_{\hat{\varphi}}|\rho\rangle\rangle$ large enough. Repeating this procedure for all $\lambda' \in R_{\mathsf{G}}$ (choosing $\hat{\mathsf{G}}$ and representations $\hat{\varphi}$ of $\hat{\mathsf{G}}$ such that $\mathcal{P}_{\hat{\varphi}} \subset \mathcal{P}_{\lambda'}$) we can reliably reconstruct all quality parameters f_λ associated with randomized benchmarking over the benchmarking group G. Once we have all these parameters we can use eq. (8.5) to obtain the average fidelity of the gate-set G.

8.3. EXAMPLES OF CHARACTER BENCHMARKING

We will now discuss several examples of randomized benchmarking experiments where the character randomized benchmarking approach is beneficial. The first example, benchmarking T-gates, is taken from the literature [2] while the second one, performing interleaved benchmarking on a 2-qubit gate using only single qubit gates a reference, is a new protocol.

8.3.1. BENCHMARKING T-GATES

The most common universal gate-set considered in the literature is the Clifford+T gate-set. The average fidelity of the Clifford gates can be extracted using standard randomized benchmarking over the Clifford group but to extract the average fidelity of the T gate a different approach is needed. One choice is to perform randomized benchmarking over the group T_q generated by the CNOT, Pauli X and T gates on q-qubits. (Another choice would be to use dihedral randomized benchmarking [6] but this is limited to single qubit systems). This group is an example of a CNOT-dihedral group and its use for randomized benchmarking was investigated in [2]. There it was derived that the Liouville representation of the group T_q decomposes into 3 irreducible subrepresentations $\varphi_1, \varphi_2, \varphi_3$ with associated quality parameters f_1, f_2, f_3 and projectors

$$\mathcal{P}_1 = |\sigma_0\rangle\!\rangle\langle\!\langle\sigma_0| \tag{8.14}$$

$$\mathcal{P}_2 = \sum_{\sigma\in\mathcal{Z}_q} |\sigma\rangle\!\rangle\langle\!\langle\sigma| \tag{8.15}$$

$$\mathcal{P}_2 = \sum_{\sigma\in\boldsymbol{\sigma_q}\setminus\mathcal{Z}} |\sigma\rangle\!\rangle\langle\!\langle\sigma|, \tag{8.16}$$

where σ_0 is the normalized identity, $\boldsymbol{\sigma_q}$ is the set of normalized non-identity Pauli matrices and $\mathcal{Z}_q$ is the subset of the normalized Pauli matrices composed only of tensor products of Z and $\mathbb{1}$. Since there are three representations, we must estimate three quality parameters f_1, f_2, f_3 in order to estimate the average fidelity. However, assuming the noisy gates are CPTP maps it is easy to see that $f_1 = 1$. This leaves us with estimating the parameters f_2, f_3 using character randomized benchmarking. In order to perform character randomized benchmarking we must first choose a group $\hat{\mathsf{G}}$. A good choice for $\hat{\mathsf{G}}$ is in this case the Pauli group P_q. Note that $\mathsf{P}_q \subset \mathsf{T}_q$ since $T^4 = Z$ the Pauli Z matrix.

Having chosen $\hat{\mathsf{G}} = \mathsf{P}_q$ we must also choose irreducible subrepresentations $\hat{\varphi}$ of the Liouville representation of the Pauli group P_q such that $\mathcal{P}_{\hat{\varphi}}\mathcal{P}_\lambda = \mathcal{P}_{\hat{\varphi}}$ for $\lambda \in \{2, 3\}$. As explained in greater detail in section 8.5.1, the Liouville representation of the Pauli group has 2^{2q} irreducible inequivalent subrepresentations φ_σ labeled by the Pauli basis elements $\sigma \in \sigma_0 \cap \boldsymbol{\sigma}_q$. Concretely we have that the projector onto the support of φ_σ is given by $\mathcal{P}_\sigma = |\sigma\rangle\!\rangle\langle\!\langle\sigma|$. The character associated to the representation φ_σ is $\chi_\sigma(P) = (-1)^{\langle P,\sigma\rangle}$ where $\langle P, \sigma\rangle = 1$ if and only if P and σ anti-commute and zero otherwise. This is again derived in greater detail in section 8.5.1.

Now we explicitly write down the experiments that must be done to estimate f_2 and f_3.

Estimating f_2:

To estimate the quality parameter f_2 we must perform the following set of steps

1. Choose $\mathsf{G} = \mathsf{T}_q$ the CNOT-dihedral group on q qubits and choose $\hat{\mathsf{G}} = \mathsf{P}_q$ the q-qubit Pauli group.
2. Choose $\{Q, \mathbb{1} - Q\}$ a two component POVM with $Q = \frac{1}{2}(\mathbb{1} + Z^{\otimes q})$ and choose $\rho = \frac{1}{2^q}(\mathbb{1} + Z^{\otimes q})$ (see section 8.4 on how to prepare this non-pure state efficiently).

3. Choose φ_σ with $\sigma = 2^{-q/2} Z^{\otimes q}$ an irreducible subrepresentation of the Liouville representation of P_q with character function χ_σ (which can be computed from lemma 8.3).

4. Perform a character randomized benchmarking experiment (as given in fig. 8.1) with $\mathsf{G} = \mathsf{T}_q$, $\hat{\mathsf{G}} = \mathsf{P}_q$ and $\hat{\varphi} = \varphi_\sigma$ with $\sigma = 2^{-q/2} Z^{\otimes q}$ to obtain the quality parameter f_2.

For completeness we have included the character function χ_σ for $\sigma = Z^{\otimes 2}/2$ in table 8.1.

Estimating f_3:

To estimate the quality parameter f_3 we must perform the following set of steps

1. Choose $\mathsf{G} = \mathsf{T}_q$ the CNOT-dihedral group on q qubits and choose $\hat{\mathsf{G}} = \mathsf{P}_q$ the q-qubit Pauli group.

2. Choose $\{Q, \mathbb{1} - Q\}$ a two component POVM with $Q = \frac{1}{2}(\mathbb{1} + X^{\otimes q})$ and choose $\rho = \frac{1}{2^q}(\mathbb{1} + X^{\otimes q})$ (see section 8.4 on how to prepare this non-pure state efficiently).

3. Choose φ_σ with $\sigma = 2^{-q/2} X^{\otimes q}$ an irreducible subrepresentation of the Liouville representation of P_q with character function χ_σ (which can be computed from lemma 8.3).

4. Perform a character randomized benchmarking experiment (as given in fig. 8.1) with $\mathsf{G} = \mathsf{T}_q$, $\hat{\mathsf{G}} = \mathsf{P}_q$ and $\hat{\varphi} = \varphi_\sigma$ with $\sigma = 2^{-q/2} X^{\otimes q}$ to obtain the quality parameter f_3.

For completeness we have included the character function χ_σ for $\sigma = X^{\otimes 2}/2$ in table 8.1.

Computing the average fidelity:

The average fidelity can now be computed from eq. (8.5), provided we know the quantities $\mathrm{Tr}(\mathcal{P}_2)$ and $\mathrm{Tr}(\mathcal{P}_3)$. These were computed in [2], giving an average fidelity formula of the form

$$F_{\text{avg}} = \frac{2^q - 1}{2^q} \left(1 - \frac{f_2 + 2^q f_3}{2^q + 1}\right). \tag{8.17}$$

8.3.2. 2-FOR-1 INTERLEAVED BENCHMARKING

The next example is a new protocol, which we call 2-for-1 interleaved randomized benchmarking. It is a way to perform interleaved randomized benchmarking [12] of a 2-qubit Clifford gate using only single-qubit Clifford gates as reference gates. The advantages of this are (1) lower experimental requirements and (2) high fidelity of the reference gates relative to the interleaved gate which allows for a tighter estimate of the average fidelity of the interleaved gate (assuming single qubit gates have higher fidelity than two qubit gates).

An interleaved benchmarking experiment consists of two stages, (1) a reference experiment and (2) an interleaved experiment. The reference experiment for 2-for-1 interleaved

randomized benchmarking consists of character randomized benchmarking using 2 copies of the single-qubit Clifford group $\mathsf{G} = \mathsf{C}_1^{\otimes 2}$ (this is also the group considered in [7]). The fitting curve of a randomized benchmarking experiment over this group involves 4 quality parameters f_w indexed by $w = (w_1, w_2) \in \{0,1\}^{\times 2}$. Deriving the above is an easy exercise in representation theory (performed earlier in [7]) but we include a proof for completeness. Concretely we have the following lemma:

Lemma 8.2. Let $\mathsf{G} = \mathsf{C}_1^{\otimes 2}$ be the two-fold tensor product of the single qubit Clifford group. The Liouville representation of this group (acting on two qubits), decomposes into four inequivalent irreducible subrepresentations φ_w indexed by $w \in \{0,1\}^{\times 2}$ with projectors onto the supports of φ_w given by

$$\mathcal{P}_{(0,0)} = |\sigma_0 \otimes \sigma_0\rangle\!\rangle\langle\!\langle \sigma_0 \otimes \sigma_0| \tag{8.18}$$

$$\mathcal{P}_{(1,0)} = \sum_{\sigma \in \boldsymbol{\sigma}_1} |\sigma \otimes \sigma_0\rangle\!\rangle\langle\!\langle \sigma \otimes \sigma_0| \tag{8.19}$$

$$\mathcal{P}_{(0,1)} = \sum_{\sigma \in \boldsymbol{\sigma}_1} |\sigma_0 \otimes \sigma\rangle\!\rangle\langle\!\langle \sigma_0 \otimes \sigma| \tag{8.20}$$

$$\mathcal{P}_{(0,1)} = \sum_{\sigma,\sigma' \in \boldsymbol{\sigma}_1} |\sigma \otimes \sigma'\rangle\!\rangle\langle\!\langle \sigma \otimes \sigma'|. \tag{8.21}$$

Proof. We begin by noting that for all $G \in \mathsf{C}_1$ we have that $\mathcal{G}|\sigma_0\rangle\!\rangle = |\sigma_0\rangle\!\rangle$. This already implies that

$$\mathcal{C}\mathcal{P}_w = \mathcal{P}_w\mathcal{C}, \qquad C \in \mathsf{C}_1^{\otimes 2}, \;\; w \in \{0,1\}^{\times 2} \tag{8.22}$$

which means all φ_w defined in the lemma statement are subrepresentations of the Liouville representation of $\mathsf{C}_1^{\otimes 2}$. To see that they are also irreducible we calculate the character inner product of the Liouville representation of $\mathsf{C}^{\otimes 2}$. We have

$$\langle \chi_{\text{Liouville}}, \chi_{\text{Liouville}} \rangle = \mathop{\mathbb{E}}_{C_1, C_2 \in \mathsf{C}_1} |\operatorname{Tr}(\mathcal{C}_1 \otimes \mathcal{C}_2)|^2 = \left(\mathop{\mathbb{E}}_{C_1 \in \mathsf{C}_1} |\operatorname{Tr}(\mathcal{C}_1)|^2 \right)^2. \tag{8.23}$$

Because the single qubit Clifford group is a two-design we know that $\mathbb{E}_{C_1 \in \mathsf{C}_1} |\operatorname{Tr}(\mathcal{C}_1)|^2 = 2$ [13], and hence that $\langle \chi_{\text{Liouville}}, \chi_{\text{Liouville}} \rangle = 4$. Since characters are additive w.r.t. taking direct sums of representations and $\langle \chi_\varphi, \chi_\varphi \rangle \geq 1$ with equality if and only if φ is irreducible (see section 3.2) we conclude that φ_w must also be irreducible for all $w \in \{0,1\}^{\times 2}$.

■

We now outline in detail the steps that must be taken to perform both the reference and interleaved experiments.

Reference experiment

To perform the reference experiment, i.e. character randomized benchmarking with $\mathsf{G} = \mathsf{C}_1^{\otimes 2}$, we choose $\hat{\mathsf{G}} = \mathsf{P}_2$ the 2-qubit Pauli group. For each $w \in \{0,1\}^{\times 2}$ we can isolate f_w by choosing a subrepresentation φ_σ of the Liouville representation of P_2. Recalling that $\mathcal{P}_\sigma = |\sigma\rangle\!\rangle\langle\!\langle\sigma|$ we can choose $\hat{\varphi} = \varphi_\sigma$ correctly (using lemma 8.2) to isolate the parameters

f_w. Once we have obtained all relevant quality parameters f_w we can compute the average reference fidelity F_{ref}. More explicitly we must perform the following sequence of steps.

1. Choose $\mathsf{G} = \mathsf{C}_1^{\otimes 2}$ the group of single qubit Cliffords on two qubits and choose $\hat{\mathsf{G}} = \mathsf{P}_2$ the two qubit Pauli group
2. Choose $\{Q, \mathbb{1} - Q\}$ a two component POVM with $Q = |00\rangle\langle 00|$ and choose $\rho = |00\rangle\langle 00|$
3. Choose φ_σ with $\sigma = (Z \otimes \mathbb{1})/2$ an irreducible subrepresentation of the Liouville representation of P_2 with character function χ_σ (given explicitly in table 8.1)
4. Perform a character randomized benchmarking experiment (as given in fig. 8.1) with $\mathsf{G} = \mathsf{C}^{\otimes 2}$, $\hat{\mathsf{G}} = \mathsf{P}_2$ and $\hat{\varphi} = \varphi_\sigma$ with $\sigma = (Z \otimes \mathbb{1})/2$ to obtain the quality parameter f_w with $w = (1, 0)$
5. Choose φ_σ with $\sigma = (\mathbb{1} \otimes Z)/2$ an irreducible subrepresentation of the Liouville representation of P_2 with character function χ_σ (given explicitly in table 8.1)
6. Perform a character randomized benchmarking experiment (as given in fig. 8.1) with $\mathsf{G} = \mathsf{C}^{\otimes 2}$, $\hat{\mathsf{G}} = \mathsf{P}_2$ and $\hat{\varphi} = \varphi_\sigma$ with $\sigma = (\mathbb{1} \otimes Z)/2$ to obtain the quality parameter f_w with $w = (0, 1)$
7. Choose φ_σ with $\sigma = (Z \otimes Z)/2$ an irreducible subrepresentation of the Liouville representation of P_2 with character function χ_σ (given explicitly in table 8.1)
8. Perform a character randomized benchmarking experiment (as given in fig. 8.1) with $\mathsf{G} = \mathsf{C}^{\otimes 2}$, $\hat{\mathsf{G}} = \mathsf{P}_2$ and $\hat{\varphi} = \varphi_\sigma$ with $\sigma = (Z \otimes Z)/2$ to obtain the quality parameter f_w with $w = (1, 1)$

Knowing that $f_w = 1$ for $w = (0, 0)$ (assuming the noise $\mathcal{E}$ affecting the gates is CPTP) we can use eq. (8.5) to obtain the average reference fidelity $F_{\text{avg}}^{\text{ref}}$ as

$$F_{\text{avg}}^{\text{ref}} = \frac{1}{5}\left(\frac{1}{4}\left(1 + 3f_{(0,1)} + 3f_{(1,0)} + 9f_{(1,1)}\right) + 1\right). \tag{8.24}$$

Interleaved experiment

The interleaved experiment similarly consists of a character randomized benchmarking experiment using $\mathsf{G} = \mathsf{C}_1^{\otimes 2}$ but for every sequence $\vec{G} = (G_1, \ldots, G_m)$ we apply the sequence $(G_1, C, G_2, \ldots, C, G_m)$ instead, where C is a 2-qubit interleaving gate (from the 2-qubit Clifford group). Note that we must then also invert this sequence (with C) to the identity [12] . Similarly choosing $\hat{\mathsf{G}} = \mathsf{P}_2$ we can again isolate the parameters f_w and from these compute the 'interleaved fidelity' F_{int}. Note that it is not immediately obvious that this interleaved RB process again yields a single exponential. In section 8.5.2 we give a proof that this is indeed the case up to a very small correction.
To perform the interleaved stage of two-for-one interleaved benchmarking we must perform the following sequence of steps:

$\sigma\backslash P$	$\mathbb{1}\mathbb{1}$	$Z\mathbb{1}$	$\mathbb{1}Z$	ZZ	$X\mathbb{1}$	$\mathbb{1}X$	XX	$Y\mathbb{1}$	$\mathbb{1}Y$	YY	ZX	XZ	ZY	YZ	XY	YX
$Z\mathbb{1}$	1	1	1	1	−1	1	−1	−1	1	−1	1	−1	1	−1	−1	−1
$\mathbb{1}Z$	1	1	1	1	1	−1	−1	1	−1	−1	−1	1	−1	1	−1	−1
ZZ	1	1	1	1	−1	−1	1	−1	−1	1	−1	−1	−1	−1	1	1
XX	1	−1	−1	1	1	1	1	−1	−1	1	−1	−1	1	1	−1	−1

Table 8.1: Values for the character function $\chi_\sigma(P)$ for $P \in \mathsf{P}_2$ and $\sigma \in \{(Z\mathbb{1})/2, (\mathbb{1}Z)/2, (ZZ)/2, (XX)/2\}$, suppressing the tensor product.

1. Choose $\mathsf{G} = \mathsf{C}_1^{\otimes 2}$ the group of single qubit Cliffords on two qubits and choose $\hat{\mathsf{G}} = \mathsf{P}_2$ the two qubit Pauli group
2. Choose $\{Q, \mathbb{1} - Q\}$ a two component POVM with $Q = |00\rangle\langle 00|$ and choose $\rho = |00\rangle\langle 00|$
3. Choose φ_σ with $\sigma = (Z \otimes \mathbb{1})/2$ an irreducible subrepresentation of the Liouville representation of P_2 with character function χ_σ (given explicitly in table 8.1)
4. Perform an interleaved character randomized benchmarking experiment (by following the steps in fig. 8.1 but interleaving the gate C) with $\mathsf{G} = \mathsf{C}^{\otimes 2}$, $\hat{\mathsf{G}} = \mathsf{P}_2$ and $\hat{\varphi} = \varphi_\sigma$ with $\sigma = (Z \otimes \mathbb{1})/2$ to obtain the quality parameter f_w with $w = (1, 0)$
5. Choose φ_σ with $\sigma = (\mathbb{1} \otimes Z)/2$ an irreducible subrepresentation of the Liouville representation of P_2 with character function χ_σ (given explicitly in table 8.1)
6. Perform an interleaved character randomized benchmarking experiment (by following the steps in fig. 8.1 but interleaving the gate C) with $\mathsf{G} = \mathsf{C}^{\otimes 2}$, $\hat{\mathsf{G}} = \mathsf{P}_2$ and $\hat{\varphi} = \varphi_\sigma$ with $\sigma = (\mathbb{1} \otimes Z)/2$ to obtain the quality parameter f_w with $w = (0, 1)$
7. Choose φ_σ with $\sigma = (Z \otimes Z)/2$ an irreducible subrepresentation of the Liouville representation of P_2 with character function χ_σ (given explicitly in table 8.1)
8. Perform an interleaved character randomized benchmarking experiment (by following the steps in fig. 8.1 but interleaving the gate C) with $\mathsf{G} = \mathsf{C}^{\otimes 2}$, $\hat{\mathsf{G}} = \mathsf{P}_2$ and $\hat{\varphi} = \varphi_\sigma$ with $\sigma = (Z \otimes Z)/2$ to obtain the quality parameter f_w with $w = (1, 1)$

Knowing that $f_w = 1$ for $w = (0, 0)$ (assuming the noise affecting the gates is CPTP) we can use eq. (8.5) to obtain the average interleaved fidelity $F_{\text{avg}}^{\text{ref}}$ as [7]

$$F_{\text{avg}}^{\text{int}} = \frac{1}{5}\left(\frac{1}{4}\left(1 + 3f_{(0,1)} + 3f_{(1,0)} + 9f_{(1,1)}\right) + 1\right). \tag{8.25}$$

Obtaining the interleaved gate average fidelity

Given values for $F_{\text{avg}}^{\text{ref}}$ and $F_{\text{avg}}^{\text{int}}$ (estimated by the protocols above) we can place upper and lower bounds on the average fidelity $F_{\text{avg}}(\widetilde{\mathcal{C}}, \mathcal{C})$ of the implementation of the gate C. We will use the optimal bounds derived in [14] which state that

$$|\psi^{(\text{int})} - \psi^{(C)}\psi^{(\text{ref})} + (1 - \psi^{(C)})(1 - \psi^{(\text{ref})})| \leq \sqrt{\psi^{(C)}(1 - \psi^{(C)})}\sqrt{\psi^{(\text{ref})}(1 - \psi^{(\text{ref})})}, \tag{8.26}$$

where

$$\psi^{(C)} = 2^{-q}((2^q - 1)F_{\text{avg}}(\widetilde{\mathcal{C}}, \mathcal{C}) - 1), \tag{8.27}$$

and similarly for $\psi^{(\text{int})}$ and $\psi^{(\text{ref})}$. We can numerically solve the above inequality to obtain lower and upper bounds on the value for $\psi^{(C)}$ given $\psi^{(\text{int})}$ and $\psi^{(\text{ref})}$ and thus for $F_{\text{avg}}(\widetilde{\mathcal{C}}, \mathcal{C})$ given $F_{\text{avg}}^{\text{ref}}$ and $F_{\text{avg}}^{\text{int}}$.

An often quoted number for the gate average fidelity $F_{\text{avg}}(\widetilde{\mathcal{C}}, \mathcal{C})$ is the 'interleaved gate fidelity estimate' F^{est}, given by [12]

$$F^{\text{est}} = 1 - \frac{(2^q - 1)}{2^q}\left(1 - \frac{2^q F_{\text{avg}}^{\text{ref}} - 1}{2^q F_{\text{avg}}^{\text{int}} - 1}\right), \tag{8.28}$$

which can also be estimated using 2-for-1 interleaved benchmarking. We however stress that this number, without further knowledge of the underlying noise process, has no interpretation as a point estimate of $F_{\text{avg}}(\widetilde{\mathcal{C}}, \mathcal{C})$ (apart from being a point in the interval given by solving eq. (8.26)).

Comparing standard interleaved randomized benchmarking and 2-for-1 interleaved randomized benchmarking

Note that in eq. (8.26) higher values for $F_{\text{avg}}^{\text{ref}}$ and $F_{\text{avg}}^{\text{int}}$ lead to sharper bounds on $F_{\text{avg}}(\widetilde{\mathcal{C}}, \mathcal{C})$. This is, apart from lower resource cost, the main advantage of 2-for-1 character randomized benchmarking. In a typical quantum computing platform the single qubit gate fidelity is much higher than the two qubit gate fidelity. Since a typical 2-qubit Clifford gate is composed of two layers of single qubit gates and a single two qubit gate [15] the expected reference fidelity in 2-for-1 interleaved randomized benchmarking is much higher than the reference fidelity in standard interleaved randomized benchmarking, thus leading to much sharper bounds on the average fidelity of the interleaved gate. To illustrate this we have simulated 2-for-1 interleaved randomized benchmarking and standard interleaved randomized benchmarking using realistic values for single qubit gate fidelities and two qubit gate fidelities [16]. In particular we have chosen the single qubit average gate fidelity to be $F_{\text{avg}}^{(1)} = 0.99$ and the two qubit gate fidelity to be $F_{\text{avg}}^{(2)} = 0.898$. In fig. 8.2 we show the result of a simulated experiment using these values. We see that the reference fidelity in 2-for-1 interleaved benchmarking is significantly higher ($F_{\text{avg}}^{\text{ref}} \approx 0.98$) than the reference fidelity of standard interleaved benchmarking ($F_{\text{avg}}^{\text{ref}} \approx 0.87$). This in turn leads to a significantly higher lower bound for the average fidelity of the interleaved gate ($F_{\text{avg}}(\widetilde{\mathcal{C}}, \mathcal{C}) \gtrsim 0.79$ for 2-for-1 interleaved benchmarking and $F_{\text{avg}}(\widetilde{\mathcal{C}}, \mathcal{C}) \gtrsim 0.62$ for standard interleaved benchmarking).

8.4. SCALABILITY AND STATISTICS

In this section we examine in more detail some aspects of the behavior of the character randomized benchmarking protocol. First we will argue that the protocol is scalable (with respect to the number of qubits q) as long as the character group $\hat{\mathsf{G}}$ is chosen properly.

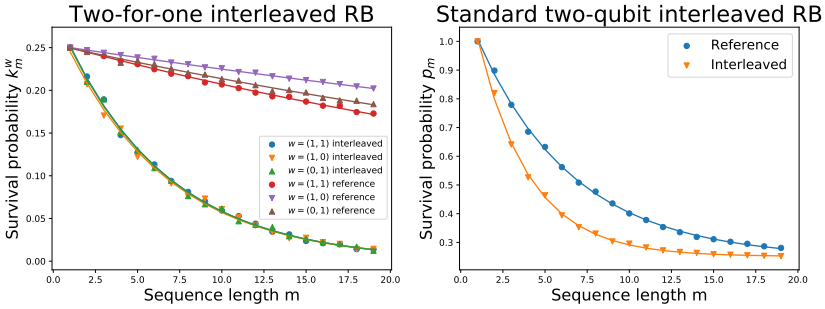


Figure 8.2: Simulation of 2-for-1 interleaved randomized benchmarking (left) and standard two-qubit interleaved randomized benchmarking (right). Inspired by the experimental results of [16] we chose single qubit gate average fidelities of $F_{\text{avg}} = 0.987$ (on both qubits) and two-qubit gate average fidelities of $F_{\text{avg}} = 0.898$, explicitly realized by a random unitary error map (corresponding to an error model dominated by calibration errors). Also following [16] we simulated a measurement fidelity of $F = 0.8$ and a state preparation fidelity of $F = 0.99$. Both experiments sampled 100 random sequences per sequence length for sequence lengths in the interval $[1, 15]$. The 2-for-1 interleaved experiment produces a reference fidelity of $F_{\text{ref}} \approx 0.98$ and an interleaved fidelity of $F_{\text{int}} \approx 0.87$. This leads to an 'interleaved gate fidelity estimate' of $F_{\text{est}} = 0.89$ with a guaranteed lower bound of $F_{\text{avg}}(\tilde{\mathcal{C}}, \mathcal{C}) \gtrsim 0.79$. On the other hand the standard interleaved randomized benchmarking experiment produces a reference fidelity of $F_{\text{ref}} \approx 0.86$ and an interleaved fidelity of $F_{\text{int}} = 0.78$. This leads to an 'interleaved gate fidelity estimate' of $F_{\text{est}} \approx 0.9$ with a guaranteed lower bound of $F_{\text{avg}}(\tilde{\mathcal{C}}, \mathcal{C}) \gtrsim 0.62$. Note that the lower bound produced by the standard interleaved randomized benchmarking experiment is significantly worse than the lower bound produced by 2-for-1 interleaved benchmarking. (Note that that we have not included error estimates for the fitted values as we are only interested in the qualitative behavior of the experiment here.) Note also that the 2-for-1 interleaved randomized benchmarking experiment yields three single exponential decays that very nearly overlap. We will explain this behavior in section 8.5.2.

Secondly we will investigate the finite sampling regime of character randomized benchmarking, arguing that character benchmarking has finite sampling properties similar to standard randomized benchmarking.

8.4.1. SCALABILITY OF CHARACTER RANDOMIZED BENCHMARKING

When performing character randomized benchmarking, especially on many qubits, care must be taken to select the group $\hat{\mathsf{G}}$ and associated representation $\hat{\varphi}$. For a group $\hat{\mathsf{G}}$ the representation $\hat{\varphi}$ could have a dimension $|\hat{\varphi}|$ that grows exponentially in the number of qubits in the system. Similarly the character function $\chi_{\hat{\varphi}}$ could have values that grow exponentially in the number of qubits. This means the quantity $k_m^{\lambda'}(\vec{G})$ can not necessarily be efficiently estimated for experiments involving more than a few qubits. A solution to this is to choose $\hat{\mathsf{G}}$ such that $\hat{\varphi}$ has small dimension. Since the maximal absolute value of the character function is bounded by the dimension of the associated representation [17], the value of the character function will also be small. This was the case in the two examples where we chose $\hat{\mathsf{G}} = \mathsf{P}_q$ which has only one-dimensional subrepresentations. When benchmarking any group G which has the Pauli group P_q as a subgroup one can always set $\mathsf{P}_q = \hat{\mathsf{G}}$ and project onto one of the one-dimensional subrepresentations of the Liouville representation of P_q.

8.4.2. FINITE SAMPLING

Here we elaborate on the statistical aspects of character randomized benchmarking. We will denote probability distributions by capital Greek letters (such as Λ) and their means by the letter μ subscripted with the corresponding distribution. The character randomized benchmarking protocol requires one to calculate the means of probability distributions. This is however impossible to do exactly using only a finite amount of samples drawn from the probability distribution. Instead one must rely on empirical estimates of these means. The reliability of these estimates is expressed by *confidence intervals*. Imagine being given a distribution with mean μ and an empirical estimate $\mu_N = \frac{1}{N}\sum_{x\in R_N} x$ where R_N is a set of N samples drawn independently from the distribution. Now a confidence interval (around μ_N) is a pair of real numbers (ϵ, δ) such that

$$\Pr(|\mu_N - \mu| \geq \epsilon) \leq 1 - \delta, \tag{8.29}$$

where the probability is taken with respect to the distribution being sampled from. Even though confidence intervals seem to require knowledge of the distribution being sampled from they can in fact be constructed using only very limited knowledge of the distribution. In particular, if one knows that the distribution being sampled from is bounded i.e., it only takes values inside an interval $[a, b]$ for $a, b \in \mathbb{R}$ then we can use Hoeffding's concentration inequality [18], given by

$$\Pr(|\mu_N - \mu| \geq \epsilon) \leq 1 - 2\exp\left(\frac{-N\epsilon^2}{(a-b)^2}\right). \tag{8.30}$$

Plugging in δ and inverting this equation we get a relation between the confidence interval (ϵ, δ) and the number of samples N from the distribution we need to construct this

interval. We have

$$N \geq \frac{\log(2/\delta)(a-b)^2}{\epsilon^2}. \tag{8.31}$$

Note that this equation is completely generic, it can be used to empirically estimate the mean of any probability distribution, as long as this distribution is bounded.

With the above we can analyze the behavior of character randomized benchmarking protocol w.r.t. finite sampling. The main question we aim to answer here is how many samples are required to accurately estimate the character average $k_m^{\lambda'}$ for fixed m and λ'. There are 3 sources of randomness in the character randomized benchmarking protocol.

1. The first source of randomness comes from sampling sequences uniformly at random from the set $\mathsf{G}^{\times m}$.

2. The second source of randomness comes from sampling an element from $\hat{\mathsf{G}}$ uniformly at random.

3. The last source of randomness is quantum mechanics itself. In general we can perform the following sequence of events:

   (a) Prepare a system in a state ρ,

   (b) Apply some quantum operation $\mathcal{E}$,

   (c) Measure using some two-component POVM $\{Q, \mathbb{1} - Q\}$.

   At the end of this sequence we will get a single bit of information x which takes the value 0 (measure Q) or 1 (measure $\mathbb{1} - Q$). We can think of x as being an instance of a random variable X which follows a Bernoulli distribution Λ_{Bern} with mean $\mu_{\Lambda_{\text{Bern}}} = \langle\langle Q|\mathcal{E}|\rho\rangle\rangle$.

We will now investigate each of these sources of randomness.

In [11, 19] (see also chapter 6) it was shown that the average $k_m^{\lambda'}$ over sequences $\vec{G} \in \mathsf{G}^{\times m}$ can be estimated with high precision and high confidence using only a few hundred sequences. These results, which were derived for standard Clifford-randomized benchmarking, can be extended to character randomized benchmarking. Whether similar results hold when performing randomized benchmarking with other groups is however an open question, making it a topic for further research in both character and standard randomized benchmarking. This is however not an issue inherent to the character benchmarking approach but is rather common to all non-Clifford randomized benchmarking protocols.

The second source of randomness is unique to character randomized benchmarking, namely estimating the mean of the distribution induced by uniform random sampling from the group $\hat{\mathsf{G}}$. Formally we have

$$k_m^{\lambda'}(\vec{G}) = \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} \chi_{\hat{\varphi}}(\hat{G})|\hat{\varphi}|\langle\langle Q|\widetilde{\mathcal{G}}_{\text{inv}}\widetilde{\vec{\mathcal{G}}}\hat{\mathcal{G}}|\rho\rangle\rangle. \tag{8.32}$$

Note that this quantity mixes two of the above types of randomness as $k_m^{\lambda'}(\vec{G})$ is an average of quantities $\langle\langle Q|\widetilde{\mathcal{G}}_{\text{inv}}\widetilde{\vec{\mathcal{G}}}\hat{\mathcal{G}}|\rho\rangle\rangle$ which are themselves means of Bernoulli distributions.

The naive way of estimating $k_m^{\lambda'}(\vec{G})$ would be to first estimate the means $\langle\langle Q|\widetilde{\mathcal{G}}_{\text{inv}}\widetilde{\vec{\mathcal{G}}}\hat{\mathcal{G}}|\rho\rangle\rangle$ one by one, by performing the associated measurement procedure N times and using the concentration inequality given above to construct an (accurate) estimate of $\langle\langle Q|\widetilde{\mathcal{G}}_{\text{inv}}\widetilde{\vec{\mathcal{G}}}\hat{\mathcal{G}}|\rho\rangle\rangle$. We can then multiply each estimate by $\chi_{\hat{\varphi}}(\hat{G})|\hat{\varphi}|$ and average them to obtain an estimate for $k_m^{\lambda'}(\vec{G})$.

However, to calculate $k_m^{\lambda'}(\vec{G})$ we would have to perform this procedure for every $\hat{G} \in \hat{\mathsf{G}}$, which would require $|\hat{\mathsf{G}}|N$ samples in total. This is not a good approach when performing character randomized benchmarking on more than a few qubits. The reason for this is that typically the size of $\hat{\mathsf{G}}$ will grow exponentially with the number of qubits. For instance, if $\hat{\mathsf{G}}$ is the Pauli group we have $|\hat{\mathsf{G}}| = |\mathsf{P}_q| = 4^q$ for q qubits.

A second method, which will be more efficient when $|\hat{\mathsf{G}}|$ is very big, is to not try to estimate all means $\langle\langle Q|\widetilde{\mathcal{G}}_{\text{inv}}\widetilde{\vec{\mathcal{G}}}\hat{\mathcal{G}}|\rho\rangle\rangle$ individually. Instead we will perform an empirical estimate of $k_m^{\lambda'}(\vec{G})$ directly by the following procedure.

1. Sample $\hat{G} \in \hat{\mathsf{G}}$ uniformly at random .
2. Prepare the state $\mathcal{G}_{\text{inv}}\mathcal{G}_m \cdots \mathcal{G}_1\hat{\mathcal{G}}|\rho\rangle\rangle$ and measure it once obtaining a result $b(\hat{G}) \in \{0, 1\}$.
3. Compute $x(\hat{G}) = \chi_{\hat{\varphi}}(\hat{G})|\hat{\varphi}|b(\hat{G}) \in \{0, \chi_{\hat{\varphi}}(\hat{G})|\hat{\varphi}|\}$.
4. Repeat sufficiently many times and compute the empirical average of $x(\hat{G})$.

8

Every time we perform steps (1)-(3) we are are drawing a single sample from a certain probability distribution. This probability distribution is a *mixture distribution*. Mixture distributions are defined as linear combinations of probability distributions. Note that there there is a difference between a mixture of distributions and an linear combination of random variables [20]. Formally the mixture distribution induced by the procedure outlined above will be defined as

$$\Lambda_{\lambda'} = \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} |\hat{\varphi}|\chi_{\hat{\varphi}}(\hat{G})\Lambda_{\text{Bern},\hat{G}}, \tag{8.33}$$

where $\Lambda_{\text{Bern},\hat{G}}$ is a Bernoulli distribution with mean $\mu_{\Lambda_{\text{Bern},\hat{G}}} = \langle\langle Q|\widetilde{\mathcal{G}}_{\text{inv}}\widetilde{\vec{\mathcal{G}}}\hat{\mathcal{G}}|\rho\rangle\rangle$. The distribution $\Lambda_{\lambda'}$ will in general be rather complex (as it is the mixture of $|\hat{\mathsf{G}}|$ Bernoulli distributions). A useful feature of mixture distributions however, is that their mean is given by the weighted average the means of the mixing distributions with the weights precisely given by the weights in the mixture [20]. In particular that means we have for

$\mu_{\Lambda_{\lambda'}}$ that

$$\mu_{\Lambda_{\lambda'}} = \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} |\hat{\varphi}|\chi_{\hat{\varphi}}(\hat{G})\mu_{\Lambda_{\mathrm{Bern},\hat{G}}}, \tag{8.34}$$

$$= \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} |\hat{\varphi}|\chi_{\hat{\varphi}}(\hat{G})\langle\langle Q|\widetilde{\mathcal{G}}_{\mathrm{inv}}\widetilde{\vec{\mathcal{G}}}\hat{\mathcal{G}}|\rho\rangle\rangle, \tag{8.35}$$

$$= k_m^{\lambda'}(\vec{G}). \tag{8.36}$$

Moreover the distribution $\Lambda_{\lambda'}$ is upper and lower bounded by $\pm|\hat{\varphi}|\chi*_{\hat{\varphi}}$ where $\chi^*_{\hat{\varphi}} = \max_{\hat{G}}|\chi_{\hat{\varphi}}(\hat{G})|$. This means that we can use the concentration inequality eq. (8.30) to bound the number of times we need to sample from $\Lambda_{\lambda'}$ (via the procedure above) in order to estimate $k_m^{\lambda'}(\vec{G})$. Note that the number of samples that need to be taken will now not depend on $|\hat{\mathsf{G}}|$ at all.

As an illustration consider the follow example. Let $\hat{\mathsf{G}}$ be the Pauli group P_q on q qubits. This group is of size $|\mathsf{P}_q| = 4^q$. However, as discussed above, the subrepresentations of of the Pauli transfer matrix representation $\mathcal{P}$ are all of dimension one and are indexed by the normalized Pauli matrices $\sigma \in \{\sigma_0\}\cup\boldsymbol{\sigma}_q$. Let's perform character randomized benchmarking where $\lambda' = \sigma$ for some normalized Pauli matrix σ. Since the representation φ_σ is one dimensional we have $|\hat{\varphi}| = |\varphi_\sigma| = 1$. Moreover we have that the character $|\chi_\sigma(P)| = 1$ for all $P \in \mathsf{P}_q$. This means that the distribution Λ_σ is upper and lower bounded by ± 1. If we now want to estimate the mean $k_m^{\lambda'}(\vec{G})$ for a particular sequence $\vec{G}$ we can perform the procedure above to sample from Λ_σ. Using the concentration inequality eq. (8.30) see that for a confidence interval of size $\epsilon = 0.02$ and confidence $\delta = 0.99$ around the mean $\mu_{\Lambda_\sigma} = k_m^{\lambda'}(\vec{G})$ we need to draw

$$N \geq \frac{\log(2/0.99)(1-(-1))^2}{0.02^2} = 1769 \tag{8.37}$$

samples. Note that this number is both 'reasonable' and completely independent of the number of qubits q. Moreover, this is not a sophisticated estimate and using more knowledge of the distribution being sampled one can probably further reduce this number.

The third source of randomness, quantum mechanics, is well studied. We would however like to make a note about a particular part of the procedure for estimating $\langle\langle Q|\mathcal{E}|\rho\rangle\rangle$, that is the preparation of the state ρ. It will often be the case that the optimal state for a character randomized benchmarking procedure is not a pure state but rather represented by a density matrix of high rank. This introduces further experimental difficulties as an experimental setup usually only gives access to pure states (by design). We can overcome this difficulty by realizing that every density matrix ρ can be written as a probability distribution over pure states, that is

$$\rho = \sum_\psi p^\rho_\psi |\psi\rangle\langle\psi|, \qquad p^\rho_\psi \geq 0, \qquad \sum_\psi p^\rho_\psi = 1. \tag{8.38}$$

This means that $\langle\langle Q|\mathcal{E}|\rho\rangle\rangle$ is also the mean of a mixture distribution that takes values in the set $\{0,1\}$ (so the mixture is still a Bernoulli distribution). In particular it is a mixture of Bernoulli distributions with mean $\langle\langle Q|\mathcal{E}|\psi\rangle\rangle$. This means that in the case of non-pure ρ we can update our sampling procedure to be

1. Fix a decomposition $\rho = \sum_\psi p_\psi^\rho |\psi\rangle\langle\psi|$.
2. Sample ψ according to $\{p_\psi^\rho\}_\psi$.
3. Sample $\hat{G} \in \hat{\mathsf{G}}$ uniformly at random .
4. Prepare the state $\mathcal{G}_{\text{inv}}\mathcal{G}_m \cdots \mathcal{G}_1\hat{\mathcal{G}}|\psi\rangle\rangle$ and measure it once obtaining a result $b(\hat{G}) \in \{0,1\}$.
5. Compute $x(\hat{G}) = \chi_{\hat{\varphi}}(\hat{G})|\hat{\varphi}|b(\hat{G}) \in \{0, \chi_{\hat{\varphi}}(\hat{G})|\hat{\varphi}|\}$.
6. Repeat sufficiently many times and compute the empirical average of $x(\hat{G})$.

This means we are now sampling from the mixture distribution

$$\Lambda_{\lambda'} = \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} \sum_\psi p_\psi^\rho |\hat{\varphi}|\chi_{\hat{\varphi}}(\hat{G})\Lambda_{\text{Bern},\hat{G},\psi}, \tag{8.39}$$

where $\Lambda_{\text{Bern},\hat{G},\psi}$ is now a Bernoulli distribution with mean $\langle\langle Q|\widetilde{\mathcal{G}_{\text{inv}}}\widetilde{\mathcal{G}_m} \cdots \widetilde{\mathcal{G}_1\hat{\mathcal{G}}}|\psi\rangle\rangle$. However the same reasoning as above holds and the number of samples (repetitions of the above procedure) required to obtain an estimate for the mean of $\Lambda_{\lambda'}$ still only depends on the interval on which $\Lambda_{\lambda'}$ is defined, yielding no increase in the number of samples needed even when the ideal input state ρ is very non-pure (has high rank).

8.5. TECHNICAL STATEMENTS

In this section we give some technical arguments that are not integral to the central idea of the character randomized benchmarking protocol but are important enough to be mentioned explicitly. Firstly we discuss the subrepresentations of the Liouville representation of the Pauli group. This representation features prominently in the examples of character randomized benchmarking discussed in section 8.3 so it is valuable to repeat here in detail some results about this representation. Secondly we discuss a technical lemma that extends the validity of character randomized benchmarking beyond the assumption of gate-independent noise. In particular we will prove a robustness result similar to the one recently proven for standard randomized benchmarking in [21].

8.5.1. REPRESENTATIONS OF THE PAULI GROUP

Probably the most useful choice for the group $\hat{\mathsf{G}}$ is the multi-qubit Pauli group. This group is defined as $\mathsf{P}_q = \langle i\mathbb{1}, X, Z\rangle^{\otimes q}$. The reason this group is useful lies in the fact that the irreducible subrepresentations of the Pauli transfer matrix representations of P_q are all of dimension one and moreover that they are all inequivalent. We have the following lemma:

Lemma 8.3. Let P_q be the Pauli group on q qubits and consider its Liouville representation. The Liouville representation decomposes as

$$\mathcal{P} = \bigoplus_{\sigma\in\{\sigma_0\}\cup\sigma_{\mathbf{q}}} \varphi_\sigma(P), \qquad \forall P \in \mathsf{P}_q, \tag{8.40}$$

with the projector $\mathcal{P}_\sigma$ onto the support of φ_σ given by

$$\mathcal{P}_\sigma = |\sigma\rangle\!\rangle\langle\!\langle\sigma|, \tag{8.41}$$

for all $\sigma \in \{\sigma_0\} \cup \sigma_{\mathbf{q}}$. Moreover all representations φ_σ are one-dimensional, mutually inequivalent and have character functions χ_σ given by

$$\chi_\sigma(P) = (-1)^{\langle\sigma,P\rangle}, \tag{8.42}$$

with

$$\langle\sigma, P\rangle = \begin{cases} 0 & \iff P \text{ and } \sigma \text{ commute} \\ 1 & \iff P \text{ and } \sigma \text{ anti-commute.} \end{cases} \tag{8.43}$$

Proof. Consider the action of $\mathcal{P}$ on the vector $|\sigma\rangle\!\rangle$ for $\sigma \in \{\sigma_0\} \cup \sigma_{\mathbf{q}}$ and $P \in \mathsf{P}_q$:

$$\mathcal{P}|\sigma\rangle\!\rangle = |P\sigma P^\dagger\rangle\!\rangle = (-1)^{\langle\sigma,P\rangle}|PP^\dagger\sigma\rangle\!\rangle = (-1)^{\langle\sigma,P\rangle}|\sigma\rangle\!\rangle. \tag{8.44}$$

This means that $|\sigma\rangle\!\rangle$ spans a subrepresentation of $\mathcal{P}$. Since the space spanned by $|\sigma\rangle\!\rangle$ is one dimensional, this subrepresentation is also irreducible. We call this subrepresentation φ_σ. By construction $\mathcal{P}_\sigma = |\sigma\rangle\!\rangle\langle\!\langle\sigma|$. Moreover the character function χ_σ is given as

$$\chi_\sigma(P) = \mathrm{Tr}(\mathcal{P}|\sigma\rangle\!\rangle\langle\!\langle\sigma|) = \langle\!\langle\sigma|\mathcal{P}|\sigma\rangle\!\rangle = (-1)^{\langle\sigma,P\rangle}. \tag{8.45}$$

It remains to prove that for $\sigma \neq \sigma'$ the representations $\varphi_\sigma, \varphi_{\sigma'}$ are inequivalent. We do this by leveraging eq. (3.13) from chapter 3 which states that the irreducible $\varphi_\sigma, \varphi_{\sigma'}$ are inequivalent if and only if the character inner product

$$\langle\chi_{\varphi_\sigma}, \chi_{\varphi_{\sigma'}}\rangle = \mathop{\mathbb{E}}_{G\in\mathsf{G}} \chi_{\varphi_\sigma}(G)\bar{\chi}_{\varphi_{\sigma'}}(G) \tag{8.46}$$

is equal to zero.
We calculate the character inner product for representations $\varphi_\sigma, \varphi_{\sigma'}$ of P_q as follows:

$$\langle\chi_\sigma, \chi_{\sigma'}\rangle = \mathop{\mathbb{E}}_{P\in\mathsf{P}_q} \chi_\sigma(P)\bar{\chi}_{\sigma'}(P) = \mathop{\mathbb{E}}_{P\in\mathsf{P}_q} (-1)^{\langle P,\sigma\rangle}(-1)^{\langle P,\sigma'\rangle}. \tag{8.47}$$

It is easy to verify by explicit computation that $(-1)^{\langle P,\sigma\rangle}(-1)^{\langle P,\sigma'\rangle} = (-1)^{\langle P,\tau\rangle}$ with $\tau \approx \sigma\sigma'$, i.e τ is equal to $\sigma\sigma'$ up to a proportionality factor. Since $\sigma\sigma' \approx \mathbb{1}$ if and only if $\sigma = \sigma'$ we have that $\tau \neq \mathbb{1}$. Since a non-identity Pauli matrix (such as τ) commutes with precisely half of the elements of the Pauli group and anti-commutes with the other half (for a proof of this fact see for instance [22, Lemma 1]) we have that $\langle\chi_\sigma, \chi_{\sigma'}\rangle = 0$, completing the lemma. ■

Note that for two Pauli matrices P, P' we can also efficiently (in the number of qubits q) decide whether they commute or anti-commute. This means that the character function $\chi_\sigma(P)$ can be efficiently computed on the fly for any σ and P. This is important because we must compute an instantiation of the character function for every random sample drawn during the character randomized benchmarking procedure. Note however that this can be done in post-processing so high speed (not just efficient) calculation of the character function is not a requirement for the success of the character randomized benchmarking procedure.

8.5.2. INTERLEAVED CHARACTER RANDOMIZED BENCHMARKING

In section 8.3.2 we proposed 2-for-1 interleaved randomized benchmarking, a form of character interleaved randomized benchmarking. More generally we can consider performing interleaved character randomized benchmarking with a benchmarking group G, a character group $\hat{\mathsf{G}}$, and an interleaving gate C. However it is not obvious that the interleaved character randomized benchmarking procedure (for arbitrary G and C) always yields data that can be fitted to a single exponential such that the average fidelity can be extracted. Here we will justify this behavior subject to an assumption on the relation between the interleaving gate C and the benchmarking group G which we expect to be quite general. This relation is phrased in terms of what we call the 'mixing matrix' of the group G and gate C. This matrix, which we denote by M, has entries

$$M_{\lambda,\hat{\lambda}} = \frac{1}{\mathrm{Tr}(\mathcal{P}_\lambda)} \mathrm{Tr}\left(\mathcal{P}_\lambda \mathcal{C} \mathcal{P}_{\hat{\lambda}} \mathcal{C}^\dagger\right) \tag{8.48}$$

for $\lambda, \lambda' \in R'_{\mathsf{G}} = R_{\mathsf{G}} \backslash \{\mathrm{id}\}$ with φ_{id} the trivial subrepresentation of the PTM representation of G carried by $|\mathbb{1}\rangle\rangle$ and where $\mathcal{P}_\lambda$ is the projector onto the subrepresentation φ_λ of $\mathcal{G}$. Note that this matrix is defined completely by C and the PTM representation of G. Note also that this matrix has only non-negative entries, that is $M_{\lambda,\hat{\lambda}} \geq 0 \ \ \forall \lambda, \hat{\lambda}$.

In the following lemma we will assume that the mixing matrix M is not only non-negative but also irreducible in the Perron-Frobenius sense [23]. Formally this means that there exists an integer L such that A^L has only strictly positive entries. This assumption will allow us to invoke the powerful Perron-Frobenius theorem [23] to prove in lemma 8.4 that interleaved character randomized benchmarking works as advertised. Below lemma 8.4 we will also explicitly verify the irreducibility condition for 2-for-1 interleaved benchmarking with the CPHASE gate. We note that the assumption of irreducibility of M can be easily relaxed to M being a direct sum of irreducible matrices with the proof of lemma 8.4 basically unchanged. It is an open question if it can be relaxed further to encompass all non-negative mixing matrices.

Lemma 8.4. Consider the outcome $k^m_{\lambda'}$ of an interleaved character randomized benchmarking experiment $(\mathsf{G}, \hat{\mathsf{G}}, \lambda', m, C)$ and assume the existence of quantum channels $\mathcal{E}_C, \mathcal{E}$ s.t. $\widetilde{\mathcal{C}} = \mathcal{C}\mathcal{E}_C$ and $\widetilde{\mathcal{G}} = \mathcal{E}\mathcal{G}$ for all $G \in \mathsf{G}$. Now consider the matrix $M(\mathcal{E}_C\mathcal{E})$ as a function of the composed channel $\mathcal{E}_C\mathcal{E}$ with entries

$$M_{\lambda,\hat{\lambda}}(\mathcal{E}_C\mathcal{E}) = \frac{1}{\mathrm{Tr}(\mathcal{P}_\lambda)} \mathrm{Tr}\left(\mathcal{P}_\lambda \mathcal{C} \mathcal{P}_{\hat{\lambda}} \mathcal{C}^\dagger \mathcal{E}_C \mathcal{E}\right) \tag{8.49}$$

for $\lambda, \lambda' \in R'_{\mathsf{G}} = R_{\mathsf{G}} \backslash \{\mathrm{id}\}$ where $\mathcal{P}_\lambda$ is again the projector onto the subrepresentation φ_λ of $\mathcal{G}$. If for $\mathcal{E} = \mathcal{E}_C = \mathcal{I}$ (the identity map) the matrix $M(\mathcal{I}) = M$ (the mixing matrix defined above) is irreducible (in the sense of Perron-Frobenius), then there exist parameters $A, f_{\lambda'}$ s.t.

$$|k^m_{\lambda'} - A f^m_{\lambda'}| \leq \delta_1 \delta_2^m \tag{8.50}$$

with $\delta_1 = O(1 - F_{\mathrm{avg}}(\mathcal{E}_C\mathcal{E}))$ and $\delta_2 = \gamma + O([1 - F_{\mathrm{avg}}(\mathcal{E}_C\mathcal{E})]^2)$ where γ is the second largest eigenvalue (in absolute value) of M. Moreover we have that (noting that $f_{\mathrm{id}} = 1$

as the map $\mathcal{E}_C\mathcal{E}$ is CPTP):

$$\left|\frac{1}{2^q}\sum_{\lambda'\in R_{\mathsf{G}}}\mathrm{Tr}(\mathcal{P}_\lambda)f_{\lambda'}-\frac{2^q(F_{\mathrm{avg}}(\mathcal{E}_C\mathcal{E})+1)}{2^q+1}\right|\leq O\left([1-F_{\mathrm{avg}}(\mathcal{E}_C\mathcal{E})]^2\right) \tag{8.51}$$

Proof. Consider the definition of $k_{\lambda'}^m$:

$$k_m^{\lambda'}=|\hat{\varphi}|\mathop{\mathbb{E}}_{G_1,\ldots,G_m\in\mathsf{G}}\mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}}\chi_{\hat{\phi}}(\hat{G})\langle\!\langle Q|\mathcal{E}_{\mathrm{inv}}\mathcal{G}_{\mathrm{inv}}\mathcal{C}\mathcal{E}_C\mathcal{E}\mathcal{G}_m\mathcal{C}\mathcal{E}_C\mathcal{E}\ldots\mathcal{C}\mathcal{E}_C\mathcal{E}\mathcal{G}_1\hat{\mathcal{G}}|\rho\rangle\!\rangle, \tag{8.52}$$

where $G_{\mathrm{inv}}=G_1^\dagger C^\dagger\cdots G_m^\dagger C^\dagger$ and $\mathcal{E}_{\mathrm{inv}}$ is the noise associated to the inverse gate (which we assume to be constant). Using the character projection formula and Schur's lemma we can write this as

$$k_m^{\lambda'}=\mathop{\mathbb{E}}_{G_1,\ldots,G_{m-1}\in\mathsf{G}}\langle\!\langle Q|\mathcal{E}_{\mathrm{inv}}\mathcal{G}_1^\dagger\mathcal{C}^\dagger\cdots\mathcal{G}_{m-1}^\dagger\mathcal{C}^\dagger\left[\sum_{\lambda_m\in R_{\mathsf{G}}}\frac{\mathrm{Tr}(P_{\lambda_m}\mathcal{E}_C\mathcal{E})}{\mathrm{Tr}(\mathcal{P}_{\lambda_m})}\mathcal{P}_{\lambda_m}\right] \tag{8.53}$$

$$\times\,\mathcal{C}\mathcal{E}_C\mathcal{E}\mathcal{G}_{m-1}\mathcal{C}\mathcal{E}_C\mathcal{E}\ldots\mathcal{C}\mathcal{E}_C\mathcal{E}\mathcal{G}_1\mathcal{P}_{\hat{\varphi}}|\rho\rangle\!\rangle. \tag{8.54}$$

Note now that in general $\mathcal{C}$ and $\mathcal{P}_{\lambda_m}$ do not commute. This means that we can not repeat the reasoning of lemma 4.1 but must instead write (using Schur's lemma again):

$$k_m^{\lambda'}=\sum_{\lambda_m\in R_{\mathsf{G}}}\frac{\mathrm{Tr}(P_{\lambda_m}\mathcal{E}_C\mathcal{E})}{\mathrm{Tr}(\mathcal{P}_{\lambda_m})}\mathop{\mathbb{E}}_{G_1,\ldots,G_{m-2}\in\mathsf{G}}\langle\!\langle Q|\mathcal{E}_{\mathrm{inv}}\mathcal{G}_1^\dagger\mathcal{C}^\dagger\cdots\mathcal{G}_{m-2}^\dagger\mathcal{C}^\dagger \tag{8.55}$$

$$\times\left[\sum_{\lambda_{m-1}\in R_{\mathsf{G}}}\frac{\mathrm{Tr}(\mathcal{P}_{\lambda_{m-1}}\mathcal{C}^\dagger\mathcal{P}_{\lambda_m}\mathcal{C}\mathcal{E}_C\mathcal{E})}{\mathrm{Tr}(\mathcal{P}_{\lambda_{m-1}})}\right] \tag{8.56}$$

$$\times\,\mathcal{P}_{\lambda_{m-1}}\mathcal{C}\mathcal{E}_C\mathcal{E}\mathcal{G}_{m-2}\mathcal{C}\mathcal{E}_C\mathcal{E}\ldots\mathcal{C}\mathcal{E}_C\mathcal{E}\mathcal{G}_1\mathcal{P}_{\hat{\varphi}}|\rho\rangle\!\rangle. \tag{8.57}$$

Here we recognize the definition of the matrix element $M_{\lambda_{m-1},\lambda_m}(\mathcal{E}_C\mathcal{E})$. Moreover we can apply the above expansion to G_{m-2}, G_{m-3} and so forth, writing the result in terms of powers of the matrix $M(\mathcal{E}_C\mathcal{E})$. After some reordering we get

$$k_m^{\lambda'}=\sum_{\lambda_1,\lambda_m\in R_{\mathsf{G}}}\frac{\mathrm{Tr}(P_{\lambda_m}\mathcal{E}_C\mathcal{E})}{\mathrm{Tr}(\mathcal{P}_{\lambda_m})}[M^{m-1}]_{\lambda_1,\lambda_m}\langle\!\langle Q|\mathcal{P}_{\lambda_1}\mathcal{P}_{\hat{\varphi}}|\rho\rangle\!\rangle \tag{8.58}$$

where we have again absorbed the noise associated with the inverse G_{inv} into the measurement POVM element Q. Now recognizing that by construction $\mathcal{P}_{\hat{\varphi}}\subset\mathcal{P}_{\lambda'}$ we can write $k_m^{\lambda'}$ as

$$k_m^{\lambda'}=e_{\lambda'}M^m v^T\langle\!\langle Q|\mathcal{P}_{\hat{\varphi}}|\rho\rangle\!\rangle \tag{8.59}$$

where $e_{\lambda'}$ is the $\lambda' th$ standard basis row vector of length R'_{G} and $v=v(\mathcal{E}_C\mathcal{E})$ is a row vector of length R'_{G} with entries $[v]_\lambda=\frac{\mathrm{Tr}(P_{\lambda_m}\mathcal{E}_C\mathcal{E})}{\mathrm{Tr}(\mathcal{P}_{\lambda_m})}$. This looks somewhat like an exponential decay but not quite. Ideally we would like that M^m has one dominant eigenvalue

and moreover that the vector v has high overlap with the corresponding eigenvector. This would guarantee that $k_m^{\lambda'}$ is close to a single exponential. The rest of the proof will argue that this is indeed the case. Now we use the assumption of the irreducibility of the mixing matrix $M = M(\mathcal{I})$. Subject to this assumption, the Perron-Frobenius theorem [23] states that the matrix M has a non-degenerate eigenvalue $\gamma_{\max}(M(\mathcal{I}))$ that is strictly larger in absolute value than all other eigenvalues of $M(\mathcal{I})$ and moreover satisfies the inequality

$$\min_{\lambda \in R'_{\mathsf{G}}} \sum_{\hat{\lambda} \in R'_{\mathsf{G}}} M_{\lambda,\hat{\lambda}} \leq \gamma_{\max}(M(\mathcal{I})) \leq \max_{\lambda \in R'_{\mathsf{G}}} \sum_{\hat{\lambda} \in R'_{\mathsf{G}}} M_{\lambda,\hat{\lambda}}. \tag{8.60}$$

It is easy to see from the definition of $M_{\lambda,\hat{\lambda}}$ that

$$\sum_{\hat{\lambda} \in R'_{\mathsf{G}}} M_{\lambda,\hat{\lambda}} = \sum_{\hat{\lambda} \in R'_{\mathsf{G}}} \frac{1}{\mathrm{Tr}(\mathcal{P}_\lambda)} \mathrm{Tr}\left(\mathcal{P}_\lambda \mathcal{C} \mathcal{P}_{\hat{\lambda}} \mathcal{C}^\dagger\right) \tag{8.61}$$

$$= \sum_{\hat{\lambda} \in R'_{\mathsf{G}}} \frac{1}{\mathrm{Tr}(\mathcal{P}_\lambda)} \left(\mathcal{P}_\lambda \mathcal{C} \sum_{\hat{\lambda} \in R'_{\mathsf{G}}} \mathcal{P}_{\hat{\lambda}} \mathcal{C}^\dagger\right) \tag{8.62}$$

$$= \frac{\mathrm{Tr}(\mathcal{P}_\lambda)}{\mathrm{Tr}(\mathcal{P}_\lambda)} = 1 \tag{8.63}$$

for all $\lambda \in R'_{\mathsf{G}}$. This means the largest eigenvalue of $M(\mathcal{I})$ is exactly 1. Moreover, as one can easily deduce by direct calculation, the associated right-eigenvector is the vector $v^R = (1, 1, \ldots, 1)$. Note that this vector is precisely $v(\mathcal{E}_C\mathcal{E})$ (as defined in eq. (8.59)) for $\mathcal{E}_C\mathcal{E} = \mathcal{I}$. Similarly the left-eigenvector of $M = M(\mathcal{I})$ is given by (in terms of its components) $v^L_\lambda = \mathrm{Tr}(\mathcal{P}_\lambda)$. This allows us to calculate that $k_m^{\lambda'} = \langle\langle Q|\mathcal{P}_{\hat{\varphi}}|\rho\rangle\rangle$ if $\mathcal{E}_C\mathcal{E} = \mathcal{I}$, which is as expected.

Now we will consider the map $\mathcal{E}_C\mathcal{E}$ as a perturbation of $\mathcal{I}$ with the perturbation parameter

$$\alpha = 1 - \frac{\mathrm{Tr}(\mathcal{P}_{\mathrm{tot}}\mathcal{E}_C\mathcal{E})}{\mathrm{Tr}(\mathcal{P}_{\mathrm{tot}})} \tag{8.64}$$

with $\mathcal{P}_{\mathrm{tot}} = \sum_{\lambda \in R'_{\mathsf{G}}} \mathcal{P}_\lambda$. We can write the quantum channel $\mathcal{E}_C\mathcal{E}$ as $\mathcal{E}_C\mathcal{E} = \mathcal{I} - \alpha\mathcal{D}$ where $\mathcal{D}$ is some superoperator (not CP, but by construction trace-annihilating). Since $M(\mathcal{E}_C\mathcal{E})$ is linear in its argument we can write $M(\mathcal{E}_C\mathcal{E}) = M(\mathcal{I}) - \alpha M(\mathcal{D})$. From standard matrix perturbation theory [24, Section 5.1] we can approximately calculate the largest eigenvalue of $M(\mathcal{E}_C\mathcal{E})$ as

$$\gamma_{\max}(M(\mathcal{E}_C\mathcal{E})) = \gamma_{\max}(M(\mathcal{I})) - \alpha \frac{v^L M(\mathcal{D}) v^{R^T}}{v^L v^{R^T}} + O(\alpha^2) \tag{8.65}$$

We can now calculate the prefactor $\frac{v^L M(\mathcal{D}) v^{R\,T}}{v^L v^{R\,T}}$ as

$$\frac{v^L A(\mathcal{D}) v^{R\,T}}{v^L v^{R\,T}} = \frac{\sum_{\lambda\in R'_{\mathcal{G}}}\sum_{\hat{\lambda}\in R'_{\mathcal{G}}} v^L_\lambda M(\mathcal{D})_{\lambda,\hat{\lambda}} v^R_{\hat{\lambda}}}{\mathrm{Tr}(\mathcal{P}_{\mathrm{tot}})} \tag{8.66}$$

$$= \frac{\sum_{\lambda\in R'_{\mathcal{G}}}\sum_{\hat{\lambda}\in R'_{\mathcal{G}}} \mathrm{Tr}(P_\lambda C^\dagger \mathcal{P}_{\hat{\lambda}} \mathcal{D})}{\mathrm{Tr}(\mathcal{P}_{\mathrm{tot}})} \tag{8.67}$$

$$= \frac{\mathrm{Tr}\left(\mathcal{P}_{\mathrm{tot}}\mathcal{D}\right)}{\mathrm{Tr}(\mathcal{P}_{\mathrm{tot}})} \tag{8.68}$$

$$= \frac{1}{\alpha}\frac{\mathrm{Tr}\left(\mathcal{P}_{\mathrm{tot}}[\mathcal{I}-\mathcal{E}_C\mathcal{E}]\right)}{\mathrm{Tr}(\mathcal{P}_{\mathrm{tot}})} \tag{8.69}$$

$$= 1 \tag{8.70}$$

where we used the definition of α in the last line. This means that $\gamma_{\max}(M(\mathcal{E}_C\mathcal{E})) = 1-\alpha$ up to $O(\alpha^2)$corrections. One could in principle calculate the prefactor of the correction term, but we will not pursue this here. Now we know that the matrix $M(\mathcal{E}_C\mathcal{E})^{m-1}$ in eq. (8.59) will be dominated by a factor $(1-\alpha+O(\alpha^2))^{m-1}$. However it could still be that the vector $v(\mathcal{E}_C\mathcal{E})$ in eq. (8.59) has small overlap with the right-eigenvector $v^R(\mathcal{E}_C\mathcal{E})$ of $M(\mathcal{E}_C\mathcal{E})$ associated to the largest eigenvalue $\gamma_{\max}(M(\mathcal{E}_C\mathcal{E}))$. We can again use a perturbation argument to see that this overlap will be big. Again from standard perturbation theory [24, Section 5.1] we have

$$\left\|v^R(\mathcal{E}_C\mathcal{E}) - v^R(\mathcal{I})\right\| = O(|\alpha|). \tag{8.71}$$

Moreover, by definition of $v^R(\mathcal{I})$ and $v(\mathcal{E}_C\mathcal{E})$ we have that $v^R v(\mathcal{E}_C\mathcal{E})^T = 1-\alpha$. By the triangle inequality we thus have

$$\left\|v^R(\mathcal{E}_C\mathcal{E}) - v(\mathcal{E}_C\mathcal{E})\right\| = O(|\alpha|). \tag{8.72}$$

One can again fill in the constant factors here if one desires a more precise statement. Finally we note from the definition of average fidelity (definition 11).

$$\alpha = 1 - \frac{\mathrm{Tr}(\mathcal{P}_{\mathrm{tot}}\mathcal{E}_C\mathcal{E})}{\mathrm{Tr}(\mathcal{P}_{\mathrm{tot}})} = \frac{2^q}{2^q-1}(F(\mathcal{E}_C\mathcal{E})-1) \tag{8.73}$$

This means that in the relevant limit of high fidelity, α will be small, justifying our perturbative analysis. Defining γ to be the second largest (in absolute value) eigenvalue of $M(\mathcal{E}_C\mathcal{E})$, which by the same argument as above will be the second largest eigenvalue of $M(\mathcal{I})$ up to $O(\alpha^2)$ corrections, we get

$$\left|k^{\lambda'}_m - \langle\langle Q|\mathcal{P}_{\hat{\varphi}}|\rho\rangle\rangle - \gamma_{\max}(M(\mathcal{E}_C\mathcal{E}))^{m-1}\langle\langle Q|\mathcal{P}_{\hat{\varphi}}|\rho\rangle\rangle\right| \leq \delta_1\delta_2^m \tag{8.74}$$

with $\delta_1 = O(1-F_{\mathrm{avg}}(\mathcal{E}_C\mathcal{E}))$ and $\delta_2 = |\gamma| + O((1-F_{\mathrm{avg}}(\mathcal{E}_C\mathcal{E}))^2)$. Moreover, we have from eqs. (8.65) and (8.73) that

$$\gamma_{\max}(A(\mathcal{E}_C\mathcal{E})) = 1 - \frac{2^q}{2^q-1}(F(\mathcal{E}_C\mathcal{E})-1) + O\left([1-F_{\mathrm{avg}}(\mathcal{E}_C\mathcal{E})]^2\right) \tag{8.75}$$

which immediately implies

$$\left| \frac{1}{2^q} \sum_{\lambda' \in R_{\mathsf{G}}} \mathrm{Tr}(\mathcal{P}_\lambda) f_{\lambda'} - \frac{2^q (F_{\mathrm{avg}}(\mathcal{E}_C \mathcal{E}) + 1)}{2^q + 1} \right| \leq O\left([1 - F_{\mathrm{avg}}(\mathcal{E}_C \mathcal{E})]^2 \right) \tag{8.76}$$

proving the lemma. ■

It is instructive to calculate the mixing matrix for a relevant example. We will calculate M for C the CPHASE gate and $\mathsf{G} = \mathsf{C}^{\otimes 2}$ two copies of the single qubit Clifford gates. Recall from the main text that the PTM representation of $\mathsf{C}^{\otimes 2}$ has three non-trivial subrepresentations. From their definitions in eq. (8.18) and the action of the CPHASE gate on the two qubit Pauli operators it is straightforward to see that the mixing matrix is of the form

$$M = \begin{pmatrix} 1/3 & 0 & 2/3 \\ 0 & 1/3 & 2/3 \\ 2/9 & 2/9 & 5/9 \end{pmatrix}. \tag{8.77}$$

Calculating M^2 one can see that M is indeed irreducible. Moreover M has eigenvalues $1, 1/3$ and $-1/9$. This means that for 2-for-1 interleaved benchmarking the interleaved experiment produces data that deviates from a single exponential no more than $(1/3)^m$ (for sufficiently high fidelity) which will be negligible for even for fairly small m. This means that for 2-for-1 interleaved benchmarking the assumption that the interleaved experiment produces data described by a single exponential is good. We saw this confirmed numerically in the simulated experiment presented in fig. 7.2.

8.5.3. GATE-DEPENDENT NOISE

Thus far we have developed the theory of character randomized benchmarking under the assumption of gate-independent noise. This is is not a very realistic assumption. Here we will generalize our framework to include gate-dependent noise. In particular we will deal with the so called 'Markovian' noise model. This noise model is formally specified by the existence of a function $\Phi : \mathsf{G} \to \mathcal{S}_{2^q}$ which assigns to each element G of the group G a quantum channel $\Phi(G) = \mathcal{E}_G$. Note that this model is not the most general, it does not take into account the possibility of time dependent effects or memory effects during the experiment. It is however much more general and realistic than the gate-independent noise model. Note that this section is significantly more technical than the rest of the chapter and it is not required to understand it to use character randomized benchmarking in practice. We have however chosen to include it for completeness. In this section we will prove two things:

1. A character randomized benchmarking experiment always yields data that can be fitted to a single exponential decay up to a small and exponentially decreasing corrective term.

2. The decay rates yielded by a character randomized benchmarking experiment can be related to the average fidelity of the noise in between gates, averaged over all gates.

Both of these statements, and their proofs, are straightforward generalizations of the work of Wallman [21] which dealt with standard randomized benchmarking. We will see that his conclusion, that randomized benchmarking measures the average fidelity of noise in between quantum gates up to a small correction, generalizes to the character benchmarking case. We begin with a technical theorem, which generalizes [21, theorem 2] to twirls over arbitrary groups (with multiplicity-free Liouville representations).

Theorem 8.1. Let G be a group such that its Liouville representation $\mathcal{G} = \bigoplus_{\lambda \in R_{\mathsf{G}}} \varphi_\lambda(G)$ is multiplicity-free. Denote for all λ by f_λ the largest eigenvalue of the operator $\mathbb{E}_{G\in\mathsf{G}}(\tilde{\mathcal{G}} \otimes \varphi_\lambda(G))$ where $\tilde{\mathcal{G}}$ is the CPTP implementation of $G \in \mathsf{G}$. There exist Hermicity-preserving linear superoperators $\mathcal{L}, \mathcal{R}$ such that

$$\mathop{\mathbb{E}}_{G\in\mathsf{G}}(\tilde{\mathcal{G}}\mathcal{L}\mathcal{G}^\dagger) = \mathcal{L}\mathcal{D}_{\mathsf{G}}, \tag{8.78}$$

$$\mathop{\mathbb{E}}_{G\in\mathsf{G}}(\mathcal{G}^\dagger\mathcal{R}\tilde{\mathcal{G}}) = \mathcal{D}_{\mathsf{G}}\mathcal{R}, \tag{8.79}$$

$$\mathop{\mathbb{E}}_{G\in\mathsf{G}}(\mathcal{G}\mathcal{R}\mathcal{L}\mathcal{G}^\dagger) = \mathcal{D}_{\mathsf{G}}, \tag{8.80}$$

where $\mathcal{D}_{\mathsf{G}}$ is defined as

$$\mathcal{D}_{\mathsf{G}} = \sum_\lambda f_\lambda \mathcal{P}_\lambda, \tag{8.81}$$

with $\mathcal{P}_\lambda$ the projector onto the representation φ_λ for all $\lambda \in R_{\mathsf{G}}$.

Proof. Using the definition of $\mathcal{G}$ and $\mathcal{D}_{\mathsf{G}}$ we can rewrite eq. (8.78) as

$$\sum_\lambda \mathop{\mathbb{E}}_{G\in\mathsf{G}}(\tilde{\mathcal{G}}(\mathcal{L}\mathcal{P}_\lambda)\varphi_\lambda(G)^\dagger) = \sum_\lambda f_\lambda \mathcal{L}\mathcal{P}_\lambda. \tag{8.82}$$

This means that, without loss of generality, we can take $\mathcal{L}$ to be of the form

$$\mathcal{L} = \sum_\lambda \mathcal{L}_\lambda, \qquad \mathcal{L}_\lambda\mathcal{P}_{\lambda'} = \delta_{\lambda\lambda'}\mathcal{L}_\lambda, \quad \forall\lambda'. \tag{8.83}$$

Similarly we can take $\mathcal{R}$ to be

$$\mathcal{R} = \sum_\lambda \mathcal{R}_\lambda, \qquad \mathcal{P}_{\lambda'}\mathcal{R}_\lambda = \delta_{\lambda\lambda'}\mathcal{R}_\lambda, \quad \forall\lambda'. \tag{8.84}$$

This means eqs. (8.78) and (8.79) decompose into independent pairs of equations for each λ:

$$\mathop{\mathbb{E}}_{G\in\mathsf{G}}(\tilde{\mathcal{G}}\mathcal{L}_\lambda\varphi_\lambda(G)^\dagger) = f_\lambda\mathcal{L}_\lambda \tag{8.85}$$

$$\mathop{\mathbb{E}}_{G\in\mathsf{G}}(\varphi_\lambda(G)^\dagger\mathcal{R}\tilde{\mathcal{G}}) = f_\lambda\mathcal{R}_\lambda. \tag{8.86}$$

Next we use the vectorization operator $\mathrm{vec} : \mathcal{M}_{2^{2q}} \to \mathbb{R}^{2^{4q}}$ mapping the Liouville representations of superoperators to vectors of length $\mathbb{R}^{2^{4q}}$. This operator has the property that for all $A, B, C \in \mathcal{M}_{2^{2q}}$ we have

$$\mathrm{vec}(ABC) = A \otimes C^T \mathrm{vec}(B) \tag{8.87}$$

where C^T is the transpose of C. Applying this to the equations eqs. (8.85) and (8.86) and noting that $\mathcal{G}^\dagger = \mathcal{G}^T$ since $\mathcal{G}$ is a real matrix we get the eigenvalue problems equivalent to eqs. (8.85) and (8.86),

$$\mathop{\mathbb{E}}_{G\in\mathsf{G}} (\tilde{\mathcal{G}} \otimes \varphi_\lambda(G))\mathrm{vec}(\mathcal{L}_\lambda) = f_\lambda \mathrm{vec}(\mathcal{L}_\lambda), \tag{8.88}$$

$$\mathop{\mathbb{E}}_{G\in\mathsf{G}} (\tilde{\mathcal{G}} \otimes \varphi_\lambda(G))^T \mathrm{vec}(\mathcal{R}_\lambda) = f_\lambda \mathrm{vec}(\mathcal{R}_\lambda). \tag{8.89}$$

Since we have defined f_λ to be the largest eigenvalue of $\mathbb{E}_{G\in\mathsf{G}}(\tilde{\mathcal{G}} \otimes \varphi_\lambda(G))$ (and equivalently of $\mathbb{E}_{G\in\mathsf{G}}(\tilde{\mathcal{G}} \otimes \varphi_\lambda(G))^T$) we can choose $\mathrm{vec}(\mathcal{L})$ and $\mathrm{vec}(\mathcal{R})$ to be the left and right eigenvectors respectively of $\mathbb{E}_{G\in\mathsf{G}}(\tilde{\mathcal{G}} \otimes \varphi_\lambda(G))$ associated to f_λ. Inverting the vectorization we obtain solutions to the equations eqs. (8.85) and (8.86) and hence also eqs. (8.78) and (8.79). To see that this solution also satisfies eq. (8.80) we note first that $\mathbb{E}_{G\in\mathsf{G}}(\mathcal{G}\mathcal{R}_\lambda\mathcal{L}_\lambda\mathcal{G}^\dagger)$ is proportional to P_λ for any $\mathcal{R}_\lambda, \mathcal{L}_\lambda$ satisfying eqs. (8.83) and (8.84) (by Schur's lemma). Since the eigenvectors of $\mathbb{E}_{G\in\mathsf{G}}(\tilde{\mathcal{G}} \otimes \varphi_\lambda(G))$ are only defined up to a constant we can for every λ choose proportionality constants such that $\mathbb{E}_{G\in\mathsf{G}}(\mathcal{G}\mathcal{R}_\lambda\mathcal{L}_\lambda\mathcal{G}^\dagger) = f_\lambda P_\lambda$ and thus that eq. (8.80) is satisfied.

■

Next we prove that if we perform a character randomized benchmarking experiment the observed data can always be fitted (up to an exponentially small correction) to a single exponential decay. The decay rate of $f_{\lambda'}$ associated to this experiment will be the largest eigenvalue of the operator $\mathbb{E}_{G\in\mathsf{G}}(\tilde{\mathcal{G}} \otimes \varphi_{\lambda'}(G))$ mentioned in the theorem above. Later we will give an operational interpretation of this number. We begin by defining, for all $G \in \mathsf{G}$ a superoperator Δ_G which captures the 'gate-dependence' of the noise implementation of $\mathcal{G}$,

$$\Delta_G := \tilde{\mathcal{G}} - \mathcal{L}\mathcal{G}\mathcal{R}, \tag{8.90}$$

where $\mathcal{R}, \mathcal{L}$ are defined as in theorem 8.1. Using this expansion we have the following theorem, which generalizes [21, theorem 4] to character randomized benchmarking over arbitrary finite groups with multiplicity-free Liouville representation.

Theorem 8.2. Let G be a group such that its Liouville representation $\mathcal{G} = \bigoplus_{\lambda\in R_\mathsf{G}} \varphi_\lambda(G)$ is multiplicity-free. Consider the outcome of a character randomized benchmarking experiment with benchmarking group G, character group $\hat{\mathsf{G}}$, a parameter $\lambda' \in R_\mathsf{G}$ a subrepresentation $\hat{\varphi}$ of $\hat{\mathcal{G}}$ s.t. $\mathcal{P}_{\hat{\varphi}} \subset \mathcal{P}_{\lambda'}$ and a sequence length m. That is, consider the real number

$$k_m^{\lambda'} = \mathop{\mathbb{E}}_{G\in\mathsf{G}} \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} \chi_{\hat{\varphi}}(\hat{G}) |\hat{\varphi}| \langle\langle Q | \tilde{\mathcal{G}}_{\mathrm{inv}} \tilde{\mathcal{G}}_m \cdots \widetilde{\mathcal{G}_1 \hat{\mathcal{G}}} | \rho \rangle\rangle, \tag{8.91}$$

for some input state ρ and output POVM $\{Q, \mathbb{1} - Q\}$ and $m \in \mathbb{M}$. This probability can be fitted to an exponential of the form

$$k_m^{\lambda'} =_{\text{fit}} A f_{\lambda'}^m + \varepsilon_m, \tag{8.92}$$

where A is a fitting parameter, $f_{\lambda'}$ is the largest eigenvalue of the operator $\mathbb{E}_{G\in\mathsf{G}}(\tilde{\mathcal{G}} \otimes \varphi_{\lambda'}(G))$ and $\varepsilon_m \leq \delta_1 \delta_2^m$ with

$$\delta_1 = |\hat{\varphi}| \max_{\hat{G}\in\hat{\mathsf{G}}} |\chi_{\hat{\varphi}}(\hat{G})| \max_{G\in\mathsf{G}} \|\Delta_G\|_\diamond , \tag{8.93}$$

$$\delta_2 = \mathbb{E}_{G\in\mathsf{G}} \|\Delta_G\|_\diamond , \tag{8.94}$$

where $\|\cdot\|_\diamond$ is the diamond norm on superoperators [25].

Proof. We begin by expanding $\widetilde{\mathcal{G}_1\hat{\mathcal{G}}} = \mathcal{L}\mathcal{G}_1\hat{\mathcal{G}}\mathcal{R} + \Delta_{G_1\hat{G}}$. This gives us

$$\begin{aligned} k_m^{\lambda'} = \mathop{\mathbb{E}}_{G_1,\ldots,G_m\in\mathsf{G}} \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} \chi_{\hat{\varphi}}(\hat{G})|\hat{\varphi}|\langle\!\langle Q|\tilde{\mathcal{G}}_{\text{inv}}\tilde{\mathcal{G}}_m \cdots \mathcal{L}\mathcal{G}_1\hat{\mathcal{G}}\mathcal{R}|\rho\rangle\!\rangle \\ + \chi_{\hat{\varphi}}(\hat{G})|\hat{\varphi}|\langle\!\langle Q|\tilde{\mathcal{G}}_{\text{inv}}\tilde{\mathcal{G}}_m \cdots \Delta_{G_1\hat{G}}|\rho\rangle\!\rangle. \end{aligned} \tag{8.95}$$

We now analyze the first term in eq. (8.95). Using the character projection formula, the fact that $\mathcal{G}_1 = (\mathcal{G}_{inv}\mathcal{G}_m \ldots \mathcal{G}_2)^\dagger$ and eq. (8.78) from theorem 8.1 we get

$$\mathop{\mathbb{E}}_{G_1,\ldots,G_m\in\mathsf{G}} \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} \chi_{\hat{\varphi}}(\hat{G})|\hat{\varphi}|\langle\!\langle Q|\tilde{\mathcal{G}}_{\text{inv}}\tilde{\mathcal{G}}_m \cdots \mathcal{L}\mathcal{G}_1\hat{\mathcal{G}}\mathcal{R}|\rho\rangle\!\rangle \tag{8.96}$$

$$= \mathop{\mathbb{E}}_{G_1,\ldots,G_m\in\mathsf{G}} \langle\!\langle Q|\tilde{\mathcal{G}}_{\text{inv}}\tilde{\mathcal{G}}_m \cdots \tilde{\mathcal{G}}_2\mathcal{L}\mathcal{G}_2^\dagger \ldots \mathcal{G}_{\text{inv}}^\dagger \mathcal{P}_{\hat{\varphi}}\mathcal{R}|\rho\rangle\!\rangle, \tag{8.97}$$

$$= \mathop{\mathbb{E}}_{G_3,\ldots,G_m\in\mathsf{G}} \langle\!\langle Q|\tilde{\mathcal{G}}_{\text{inv}}\tilde{\mathcal{G}}_m \cdots \tilde{\mathcal{G}}_3\mathcal{L}\mathcal{D}_{\mathsf{G}}\mathcal{G}_3^\dagger \ldots \mathcal{G}_{\text{inv}}^\dagger \mathcal{P}_{\hat{\varphi}}\mathcal{R}|\rho\rangle\!\rangle, \tag{8.98}$$

$$= \langle\!\langle Q|\mathcal{L}\mathcal{D}_{\mathsf{G}}^m\mathcal{P}_{\hat{\varphi}}\mathcal{R}|\rho\rangle\!\rangle, \tag{8.99}$$

$$= f_{\lambda'}^m \langle\!\langle Q|\mathcal{L}\mathcal{P}_{\hat{\varphi}}\mathcal{R}|\rho\rangle\!\rangle, \tag{8.100}$$

where we used that $\mathcal{D}_{\mathsf{G}}$ commutes with $\mathcal{G}$ for all $G \in \mathsf{G}$ and the fact that $\mathcal{D}_{\mathsf{G}}\mathcal{P}_{\hat{\varphi}} = f_{\lambda'}\mathcal{P}_{\hat{\varphi}}$. Next we consider the second term in eq. (8.95). For this we first need to prove a technical

statement. We make the following calculation for all $j \geq 2$ and $\hat{G} \in \hat{\mathsf{G}}$:

$$\mathop{\mathbb{E}}_{G_1,\ldots,G_m\in\mathsf{G}} \tilde{\mathcal{G}}_{\text{inv}}\tilde{\mathcal{G}}_m \cdots \tilde{\mathcal{G}}_{j+1}\mathcal{L}\mathcal{G}_j\mathcal{R}\Delta_{G_{j-1}}\ldots\Delta_{G_1\hat{G}} \tag{8.101}$$

$$= \mathop{\mathbb{E}}_{G_1,\ldots,G_m\in\mathsf{G}} \tilde{\mathcal{G}}_{\text{inv}}\tilde{\mathcal{G}}_m \cdots \tilde{\mathcal{G}}_{j+1}\mathcal{L}\mathcal{G}_{j+1}^\dagger \ldots \mathcal{G}_m^\dagger\mathcal{G}_{\text{inv}}\mathcal{G}_1^\dagger \tag{8.102}$$

$$\ldots\mathcal{G}_{j-1}^\dagger\mathcal{R}\Delta_{G_{j-1}}\ldots\Delta_{G_1\hat{G}} \tag{8.103}$$

$$= \mathop{\mathbb{E}}_{G_1,\ldots,G_m\in\mathsf{G}} \tilde{\mathcal{G}}_{\text{inv}}\tilde{\mathcal{G}}_m \cdots \tilde{\mathcal{G}}_{j+1}\mathcal{L}\mathcal{G}_{j+1}^\dagger \ldots \mathcal{G}_m^\dagger\mathcal{G}_{\text{inv}}\mathcal{G}_1^\dagger \tag{8.104}$$

$$\ldots\mathcal{G}_{j-1}^\dagger\mathcal{R}(\tilde{\mathcal{G}}_{j-1} - \mathcal{L}\mathcal{G}_{j-1}\mathcal{R})\Delta_{G_{j-2}}\ldots\Delta_{G_1\hat{G}} \tag{8.105}$$

$$= \mathop{\mathbb{E}}_{G_1,\ldots,G_{j-1},G_{j+1},\ldots G_m\in\mathsf{G}} \tilde{\mathcal{G}}_{\text{inv}}\tilde{\mathcal{G}}_m \cdots \tilde{\mathcal{G}}_{j+1}\mathcal{L}\mathcal{G}_{j+1}^\dagger \ldots \mathcal{G}_m^\dagger\mathcal{G}_{\text{inv}}\mathcal{G}_1^\dagger \tag{8.106}$$

$$\ldots\mathcal{G}_{j-2}^\dagger(\mathcal{D}_{\mathsf{G}} - \mathcal{D}_{\mathsf{G}})\mathcal{R}\Delta_{G_{j-2}}\ldots\Delta_{G_1\hat{G}} \tag{8.107}$$

$$= 0, \tag{8.108}$$

where we used the definition of $\Delta_{G_{j-1}}$, the fact that $G_{j-1} = (G_m \ldots G_{j+1})^\dagger G_{\text{inv}}(G_1 \ldots G_{j-1})^\dagger$ and eqs. (8.79) and (8.80). We can apply this calculation to the second term of eq. (8.95) to get

$$\mathop{\mathbb{E}}_{G_1,\ldots,G_m\in\mathsf{G}} \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} \chi_{\hat{\varphi}}(\hat{G})|\hat{\varphi}|\langle\langle Q|\tilde{\mathcal{G}}_{\text{inv}}\tilde{\mathcal{G}}_m\cdots\tilde{\mathcal{G}}_2\Delta_{G_1\hat{G}}|\rho\rangle\rangle \tag{8.109}$$

$$= \mathop{\mathbb{E}}_{G_1,\ldots,G_m\in\mathsf{G}} \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} \chi_{\hat{\varphi}}(\hat{G})|\hat{\varphi}|\langle\langle Q|\tilde{\mathcal{G}}_{\text{inv}}\tilde{\mathcal{G}}_m\cdots(\mathcal{L}\mathcal{G}_2\mathcal{R} + \Delta_{G_2})\Delta_{G_1\hat{G}}|\rho\rangle\rangle \tag{8.110}$$

$$= \mathop{\mathbb{E}}_{G_1,\ldots,G_m\in\mathsf{G}} \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} \chi_{\hat{\varphi}}(\hat{G})|\hat{\varphi}|\langle\langle Q|\tilde{\mathcal{G}}_{\text{inv}}\tilde{\mathcal{G}}_m\cdots\tilde{\mathcal{G}}_3\Delta_{G_2}\Delta_{G_1\hat{G}}|\rho\rangle\rangle \tag{8.111}$$

$$= \mathop{\mathbb{E}}_{G_1,\ldots,G_m\in\mathsf{G}} \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} \chi_{\hat{\varphi}}(\hat{G})|\hat{\varphi}|\langle\langle Q|\Delta_{G_{\text{inv}}}\Delta_{G_m}\ldots\Delta_{G_1\hat{G}}|\rho\rangle\rangle. \tag{8.112}$$

Hence we can write

$$k_m^\lambda = f_{\lambda'}^m\langle\langle Q|\mathcal{L}\mathcal{P}_{\hat{\varphi}}\mathcal{R}|\rho\rangle\rangle + \varepsilon_m \tag{8.113}$$

with

$$\varepsilon_m = \mathop{\mathbb{E}}_{G_1,\ldots,G_m\in\mathsf{G}} \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} \chi_{\hat{\varphi}}(\hat{G})|\hat{\varphi}|\langle\langle Q|\Delta_{G_{\text{inv}}}\Delta_{G_m}\ldots\Delta_{G_1\hat{G}}|\rho\rangle\rangle. \tag{8.114}$$

We can upper bound ε_m by

$$\mathop{\mathbb{E}}_{G_1,\ldots,G_m\in\mathsf{G}} \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} \chi_{\hat{\varphi}}(\hat{G})|\hat{\varphi}|\langle\langle Q|\Delta_{G_{\mathrm{inv}}}\Delta_{G_m}\ldots\Delta_{G_1\hat{G}}|\rho\rangle\rangle \tag{8.115}$$

$$\leq \mathop{\mathbb{E}}_{G_1,\ldots,G_m\in\mathsf{G}} \mathop{\mathbb{E}}_{\hat{G}\in\hat{\mathsf{G}}} |\chi_{\hat{\varphi}}(\hat{G})||\hat{\varphi}|\,\|\Delta_{G_{\mathrm{inv}}}\|_\diamond\,\|\Delta_{G_m}\|_\diamond\ldots\left\|\Delta_{G_1\hat{G}}\right\|_\diamond \tag{8.116}$$

$$\leq \max_{\hat{G}\in\hat{\mathsf{G}}}|\chi_{\hat{\varphi}}(\hat{G})||\hat{\varphi}| \max_{G\in\mathsf{G}}\|\Delta_G\|_\diamond \left(\mathbb{E}_{G\in\mathsf{G}}\|\Delta_G\|_\diamond\right)^m. \tag{8.117}$$

Setting

$$\delta_1 = |\hat{\varphi}|\left(\max_{\hat{G}\in\hat{\mathsf{G}}}|\chi_{\hat{\varphi}}(\hat{G})|\right)\left(\max_{G\in\mathsf{G}}\|\Delta_G\|_\diamond\right) \tag{8.118}$$

$$\delta_2 = \mathop{\mathbb{E}}_{G\in\mathsf{G}}\|\Delta_G\|_\diamond, \tag{8.119}$$

we complete the proof. ■

In [21] it was shown that δ_2 is small for realistic gate-dependent noise. This implies that for large enough m the outcome of a character randomized benchmarking experiment can be described by a single exponential decay (up to a small, exponentially decreasing factor). The rate of decay $f_{\lambda'}$ can be related to the largest eigenvalue of the operator $\mathbb{E}_{G\in\mathsf{G}}(\tilde{\mathcal{G}}\otimes\varphi_{\lambda'}(G))$. We can interpret this rate of decay following Wallman [21] by setting w.l.o.g. $\tilde{\mathcal{G}} = \mathcal{L}_G\mathcal{G}\mathcal{R}$ where $\mathcal{R}$ is defined as in theorem 8.1 and is invertible (we can always render $\mathcal{R}$ invertible by an arbitrary small perturbation). Now consider from $\tilde{\mathcal{G}} = \mathcal{L}_G\mathcal{G}\mathcal{R}$ and the invertibility of $\mathcal{R}$:

$$\mathop{\mathbb{E}}_{G\in\mathsf{G}}\mathrm{Tr}(\mathcal{G}^\dagger\mathcal{R}\tilde{\mathcal{G}}\mathcal{R}^{-1}) = \mathop{\mathbb{E}}_{G\in\mathsf{G}}\mathrm{Tr}(\mathcal{G}^\dagger\mathcal{R}\mathcal{L}_G\mathcal{G}\mathcal{R}\mathcal{R}^{-1}) \tag{8.120}$$

$$= \mathop{\mathbb{E}}_{G\in\mathsf{G}}\mathrm{Tr}(\mathcal{R}\mathcal{L}_G) \tag{8.121}$$

and moreover from eq. (8.79):

$$\mathop{\mathbb{E}}_{G\in\mathsf{G}}\mathrm{Tr}(\mathcal{G}^\dagger\mathcal{R}\tilde{\mathcal{G}}\mathcal{R}^{-1}) = \sum_{\lambda\in R_\mathsf{G}} f_\lambda\,\mathrm{Tr}(\mathcal{P}_\lambda). \tag{8.122}$$

From this we can consider the average fidelity of noise *between gates* (the map $\mathcal{R}\mathcal{L}_G$) averaged over all gates:

$$\mathop{\mathbb{E}}_{G\in\mathsf{G}} F_{\mathrm{avg}}(\mathcal{R}\mathcal{L}_G) = \mathop{\mathbb{E}}_{G\in\mathsf{G}}\frac{2^{-q}\,\mathrm{Tr}(\mathcal{R}\mathcal{L}_G)+1}{2^q+1} \tag{8.123}$$

$$= \frac{2^{-q}\sum_{\lambda\in R_\mathsf{G}} f_\lambda\,\mathrm{Tr}(\mathcal{P}_\lambda)+1}{2^q+1}. \tag{8.124}$$

Hence can interpret the quality parameters given by character randomized benchmarking as characterizing the average noise in between gates, extending the conclusion reached in [21] for standard randomized benchmarking to character randomized benchmarking. In [26] an alternative interpretation of the decay rate of randomized benchmarking in the presence of gate dependent noise is given in terms of Fourier transforms of matrix valued group functions. One could recast the above analysis for character randomized benchmarking in this language as well but we do not pursue this further here.

9

Experimental implementations of character benchmarking

In this chapter we report the results of a character randomized benchmarking experiment performed on a pair of Si\ SiGe quantum dots. In particular we will use the 2-for-1 interleaved randomized benchmarking protocol developed in chapter 8 to characterize the fidelity of a two-qubit CHPASE gate.

This chapter is adapted from sections of X. Xue, T. F. Watson, J. Helsen, et al., "*Benchmarking Gate Fidelities in a Si/SiGe Two-Qubit Device*", Physical Review X, 9, 021011 (2019)

9.1. Introduction

With steady progress towards practical quantum computers, it becomes increasingly important to efficiently characterize the relevant quantum gates. Quantum process tomography [1–3] provides a way to reconstruct a complete mathematical description of any quantum process, but has several drawbacks. The resources required increase exponentially with qubit number and the procedure cannot distinguish pure gate errors from state preparation and measurement (SPAM) errors, making it difficult to reliably extract small gate error rates. Randomized benchmarking (RB) was introduced as a convenient alternative [4–7]. It estimates the gate fidelity as a concise and relevant metric, requires fewer resources, is more robust against SPAM errors and works well even for low gate error rates.

Various randomized benchmarking methods have been investigated to extract fidelities and errors in different scenarios. In standard randomized benchmarking, sequences of increasing numbers of random Clifford operations are applied to one or more qubits [5, 6]. Then, loosely speaking, the average Clifford gate fidelity is extracted from how rapidly the final state diverges from the ideally expected state as a function of the number of random Clifford operations. In interleaved randomized benchmarking, the fidelity of a particular quantum gate is obtained by interleaving that gate in a reference sequence of random Clifford gates and studying how much faster the final state deviates from the ideal case [8]. Simultaneous randomized benchmarking uses simultaneously applied random Clifford operations to different qubits to characterize the degree of cross-talk [9].

A major drawback of traditional randomized benchmarking methods is that the number of native gates that needs to be executed in sequence to implement a Clifford operation, can rapidly increase with the qubit number. For example, it takes on average 1.5 controlled-phase (CPHASE) gates and 8.25 single-qubit gates to implement a two-qubit Clifford gate [10]. This in turns puts higher demands on the coherence time, which is still a challenge for near-term devices, and leads to rather loose bounds on the gate fidelity inferred from interleaved randomized benchmarking [8, 11]. Therefore, in early work characterizing two-qubit gate fidelities for superconducting qubits, the effect of the two-qubit gate projected in single-qubit space was reported instead of the actual two-qubit gate fidelity [12, 13]. For semiconductor spin qubits, even though two-qubit Bell states have been prepared [14–17] and simple quantum algorithms were implemented on two silicon spin qubits [15], the implementation issues of conventional randomized benchmarking have long stood in the way of quantifying the two-qubit gate fidelity. These limitations can be overcome either by using native gates that compile efficiently [17] or by using a new method called character randomized benchmarking (CRB) as discussed in chapter 8, which allows us to extract a two-qubit gate fidelity by interleaving the two-qubit gate in a reference sequence consisting of a small number of single-qubit gates only. As an additional benefit, CRB provides detailed information on separate decay channels and error correlations.

In this chapter we describe the use of character randomized benchmarking for the characterization of all relevant gate fidelities of two electron spin qubits in silicon quantum dots, including the single-qubit and two-qubit gate fidelity as well as the effect of cross-talk and correlated errors on single-qubit gate fidelities. This work is of strong interest

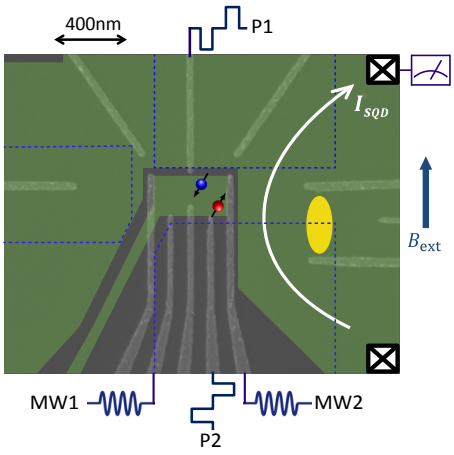


Figure 9.1: Device Schematic. A double quantum dot is formed in the Si/SiGe quantum well, where two spin qubits Q1 (blue spin) and Q2 (red spin) are defined. The green-shaded areas show the locations of the accumulation gates on top of the double dot and the reservoir. The blue dashed lines indicate the positions of three Co micro-magnets, which form a magnetic field gradient along the qubit region. MW1 and MW2 are connected to two vector microwave sources to perform EDSR for single-qubit gates. The yellow ellipse shows the position of a larger quantum dot which is used as a charge sensor for single-shot readout. Plunger gates P1 and P2 are used to pulse to different positions in the charge stability diagram as needed for initialization, manipulation, and readout, as well as for pulsing the detuning for controlling the two-qubit gate.

since silicon spin qubits are highly scalable, owing to their compact size ($<$ 100 nm pitch), coherence times up to tens of milliseconds and ability to leverage existing semiconductor technology [18, 19].

In section 9.2 we quickly describe the device under investigation before moving on to the results of the character randomized benchmarking experiment in section 9.3.

9.2. Device and qubit operation

Fig. 9.1 shows a schematic of the device, a double quantum dot defined electrostatically in a 12 nm thick Si/SiGe quantum well, 37 nm below the semiconductor surface. The device is cooled to $\sim$ 20 mK in a dilution refrigerator. By applying positive voltages on the accumulation gate, a two-dimensional electron gas is formed in the quantum well. Negative voltages are applied to the depletion gates in such a way that two single electrons are confined in a double well potential [15]. A 617 mT magnetic field is applied in the plane of the quantum well. Two qubits, Q1 and Q2, are encoded in the Zeeman split state of the two electrons.

Single-qubit rotations rely on electric dipole spin resonance (EDSR), making use of artificial spin-orbit coupling induced by the transverse magnetic field gradient from three cobalt micro magnets fabricated on top of the gate stack [20]. The longitudinal magnetic field gradient leads to well-separated spin resonance frequencies of 18.34 GHz and 19.72 GHz for Q1 and Q2 respectively. The rotation axis in the $\hat{x} - \hat{y}$ plane is set by the phase of the on-resonance microwave drive, while rotations around the $\hat{z}$ axis are implemented by changing the rotating reference frame in software [21].

We use the CPHASE gate as the native two-qubit gate. An exchange interaction $J(\varepsilon)$ is switched on by pulsing the detuning ε (electrochemical potential difference) between the two quantum dots, such that the respective electron wave functions overlap. Due to the large difference in qubit energy splittings, the flip-flop terms in the exchange Hamiltonian are ineffective and an Ising interaction remains [15, 16, 22, 23]. The resulting time

evolution operator in the standard $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ basis is given by

$$U_J(t) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{iJ(\epsilon)t/2\hbar} & 0 & 0 \\ 0 & 0 & e^{iJ(\epsilon)t/2\hbar} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{9.1}$$

Choosing $t = \pi\hbar/J(\epsilon)$ and adding single-qubit $\hat{z}$ rotations on both qubits, we obtain a CPHASE operator

$$Z_1\left(-\frac{\pi}{2}\right) Z_2(-\frac{\pi}{2}) U_J\left(\frac{\pi\hbar}{J(\epsilon)}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \tag{9.2}$$

with $Z_i(\theta)$ a $\hat{z}$ rotation of qubit i over an angle θ.
Spin initialization and single-shot readout of Q2 are realized by energy-selective tunneling [24]. Q1 is initialized to its ground spin state by fast spin relaxation at a hotspot [25]. For read-out, the state of Q1 is mapped onto Q2 using a conditional π rotation [15, 23], which enables extracting the state of Q1 by measuring Q2.

9.3. Results

In order to properly characterize the two-qubit CPHASE fidelity, we experimentally demonstrate a new approach to RB called character randomized benchmarking (CRB). As discussed in chapter 8, CRB is a powerful generic method that extends randomized benchmarking in a rigorous manner, making it possible to extract average fidelities from groups beyond the multi-qubit Clifford group while keeping the advantages of standard RB such as resistance to SPAM errors. The generality of CRB allows one to start from (a subset of) the natives gates of a particular device and then design an RB experiment tailored to that set. This can strongly reduce compilation overhead and gate dependent noise, a known nuisance factor in standard RB [26–28]. Moreover, since the accuracy of interleaved randomized benchmarking depends on the fidelity of the reference gates [8, 11], performing (through CRB) interleaved RB with a reference group generated by high fidelity gates can significantly improve the utility of interleaved RB.
Character randomized benchmarking requires us to average over two groups (the second one usually being a subgroup of the first). The first group is the “benchmark group" (denoted G in chapter 8). It is for the gates in this group that CRB yields the average fidelity. The second group is the “character group" (denoted $\hat{\mathsf{G}}$ in chapter 8). CRB works by performing standard randomized benchmarking using the benchmark group but augments this by adding a random gate from the character group before each RB gate sequence. By averaging over this extra random gate, but weighting the average by a special function known from representation theory as a character function, it guarantees that the average over random sequences can always be fitted to a single exponential decay, even when the benchmark group is not the multi-qubit Clifford group and even in the presence of SPAM errors.

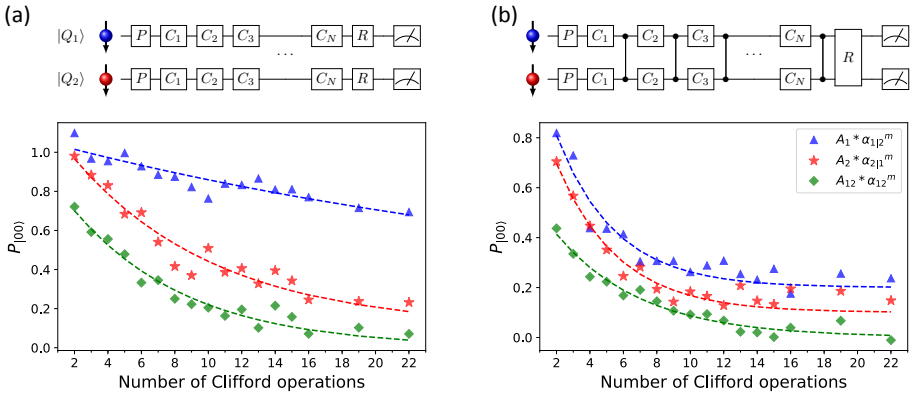


Figure 9.2: Character randomized benchmarking. (a) Reference CRB experiment. The probabilities P_1 (blue triangles), P_2 (red stars) and P_3 (green diamonds), obtained starting from the initial state $|00\rangle$ followed by a Pauli operation, as a function of the number of subsequent single-qubit Clifford operations simultaneously applied to both qubits (see the schematic of the pulse sequence). As the native gate set, we use $\{I, X(\pi), Z(\pm\pi), X(\pm\pi/2), Z(\pm\pi/2), \text{CPHASE}\}$. For each of the 16 Pauli operators, we apply 40 different random sequences, each with 20 repetitions. The dashed lines are fits to the data with a single exponential. Without SPAM errors, the data-points would decay from 1 to 0. (b) Interleaved CRB experiment. This experiment is performed in an analogous way to the reference CRB experiment, but with a two-qubit CPHASE gate interleaved after each Clifford pair, as seen in the schematic of the pulse sequence. The traces are offset by an increment of 0.1 for clarity.

Guided by the need for high reference fidelities, we choose for our implementation of CRB the benchmark group to be the parallel single-qubit Clifford group ($\mathrm{C}^{\otimes 2}$), the same as in standard simultaneous single-qubit RB) and the two-qubit Pauli group as the character group. This is the 2-for-1 interleaved character benchmarking protocol discussed in chapter 8. As already noted there, it is non-trivial that the $\mathrm{C}^{\otimes 2}$ group allows us to get information on two-qubit gates, since parallel single-qubit Clifford operations cannot fully depolarize the noise in the full two-qubit Hilbert space. In fact, for simultaneous single-qubit RB there are three depolarizing channels, each acting in a different subspace of the Hilbert space of density matrices, spanned by $I\otimes\sigma_i$, $\sigma_i\otimes I$, and $\sigma_i\otimes\sigma_i$, with I the identity operator and σ_i one of the Pauli operators. The three decay channels are reflected in the recovery probability for the final state, which is now described by (see section 8.3.2)

$$P_{C\otimes C} = A_1{\alpha_{1|2}}^m + A_2{\alpha_{2|1}}^m + A_{12}{\alpha_{12}}^m + B, \tag{9.3}$$

where $\alpha_{i|j}$ is again the depolarizing parameter for qubit i while simultaneously applying random Clifford operations to qubit j, and α_{12} is the depolarizing parameter for the two-qubit parity ($\{|00\rangle, |11\rangle\}$ versus $\{|01\rangle, |10\rangle\}$). We note that if the errors acting on both qubits are uncorrelated, then $\alpha_{12} = \alpha_{1|2}\alpha_{2|1}$ [9]. The question now is how to separate the three decays. Fitting the data using a sum of three exponentials will be very imprecise. Existing approaches combine the decay of specific combinations of the probabilities of obtaining $00, 01, 10$ and 11 upon measurement, but suffer from SPAM errors [9]. As discussed above, CRB offers a clean procedure for extracting the individual decay rates

that is immune to SPAM errors and does not incur additional overhead.
Concretely, CRB here proceeds as follows: (1) the two-qubit system is initialized to $|00\rangle$, then (2) one random Pauli operator on each qubit is applied to prepare the system in a state $|\phi_1\phi_2\rangle$ (one of $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$), followed by (3) a random sequence of simultaneously applied single-qubit Clifford operators. In practice, the random Pauli operator is absorbed in the first Clifford operation, making the Pauli gates effectively noise-free. A final Clifford operation is applied which ideally returns the system to the state $|\phi_1\phi_2\rangle$ and finally (4) both qubits are measured. Each random sequence is repeated to collect statistics on the probability $P_{\phi_1\phi_2}$ of obtaining measurement outcome 00 when starting from $|\phi_1\phi_2\rangle$ (note that each $P_{\phi_1\phi_2}$ averages over 4 Pauli operations). We combine these probabilities according to their character (section 8.3.2 for more details) to obtain three fitting parameters,

$$\begin{aligned} P_1 &= P_{00} - P_{01} + P_{10} - P_{11}, \\ P_2 &= P_{00} + P_{01} - P_{10} - P_{11}, \\ P_3 &= P_{00} - P_{01} - P_{10} + P_{11}. \end{aligned} \tag{9.4}$$

Each of these three fitting parameters is expected to decay as a single exponential, isolating one of the decay channels in Eq. 9.3:

$$\begin{aligned} P_1 &= A_1 {\alpha_{1|2}}^m, \\ P_2 &= A_2 {\alpha_{2|1}}^m, \\ P_3 &= A_{12} {\alpha_{12}}^m. \end{aligned} \tag{9.5}$$

Note that there is no constant offset B. This is also a feature of CRB. The three experimentally measured probabilities are shown in Fig. 9.2a. These contain a lot of useful information, including not only the separate depolarizing parameters but also the averaged CRB reference fidelity and information on error correlations. The blue (red) curve shows the decay in the subspace corresponding to Q1 (Q2), spanned by $\sigma_i \otimes I$ ($I \otimes \sigma_i$). The green curve shows the decay in the subspace spanned by $\sigma_i \otimes \sigma_j$. This decay can be interpreted as the parity decay. The fitted depolarizing parameters are $\alpha_{1|2} = 0.9738 \pm 0.0008$, $\alpha_{2|1} = 0.8902 \pm 0.0020$ and $\alpha_{12} = 0.8652 \pm 0.0022$.
The average CRB depolarizing parameter can be found from the separate depolarizing parameters as

$$P = \frac{3}{15}\alpha_{1|2} + \frac{3}{15}\alpha_{2|1} + \frac{9}{15}\alpha_{12}, \tag{9.6}$$

where the weights are proportional to the dimension of the corresponding subspaces of the 16-dimensional Hilbert space of two-qubit density matrices. We obtain a reference CRB fidelity of $91.9 \pm 0.1\%$, which represents the fidelity of two simultaneous single-qubit Clifford operators ($C \otimes C$) in the full two-qubit space.
Finally, from the three depolarizing parameters in Eq. 9.3, we can infer to what extent errors occur independently on each qubit or exhibit correlations between the two qubits. The fact that $\alpha_{12} - \alpha_{1|2}\alpha_{2|1} = -0.0017 \pm 0.0031$ indicates that the errors are essentially independent.
Next we perform the interleaved version of CRB, for which we insert a CPHASE gate after each single-qubit Clifford pair. Fig. 9.2b shows the three corresponding experimentally

measured decays. The fitting parameters we extract now reflect the combined errors from a single-qubit Clifford pair followed by a CPHASE gate. The fitted depolarizing parameters are $\alpha_{1|2} = 0.7522 \pm 0.0060, \alpha_{2|1} = 0.7623 \pm 0.0053$, and $\alpha_{12} = 0.8226 \pm 0.0030$. As can be expected, the three decays lie closer together than those for reference CRB: not only does the additional CPHASE gate contribute directly to all three decays, it also mixes the three subspaces. From the depolarizing parameters in interleaved and reference CRB measurement, we use eq. (8.26) to isolate the fidelity of the CPHASE gate, now in two-qubit space as desired, yielding $92.0 \pm 0.5\%$.
The dominant errors in the CPHASE gate arise from nuclear spin noise and charge noise. In natural silicon, the abundance of Si^{29} atoms is about 4.7%, and the Si^{29} nuclear spins dephase the electron spin states due to the hyperfine interaction [18]. Charge noise modulates the overlap of the two electron wave functions, and thus also the two-qubit coupling strength. In the present device, we could not access the symmetry point where the coupling strength is to first order insensitive to the detuning of the double dot potential [29, 30], hence charge noise directly (to first order) affects the two-qubit coupling strength.

9.4. CONCLUSION

Character randomized benchmarking provides a new method to effectively characterize multi-qubit behavior. It combines the advantages of simultaneous randomized benchmarking and interleaved randomized benchmarking, and gives tighter bounds on the fidelity number than standard interleaved randomized benchmarking due to its simpler compilation. CRB is useful in a wide variety of settings, far beyond the particular case studied here. The general approach to exploiting CRB is to start from a set of native gates that can be implemented easily and with high fidelity, and to construct a suitable reference sequence based on this set. The decay for the reference sequence contains any number of exponentials, which can be separated without suffering from SPAM errors and which provide relevant additional information, in the present case on the fidelity of simultaneously applied gates, cross-talk and on noise correlations. Comparison with interleaved CRB allows one to extract the fidelity of specific gates of interest.
We perform the first comprehensive study of the single-qubit, simultaneous single-qubit and two-qubit gate fidelities for semiconductor qubits, where the use of CRB, which allows for a compact reference sequence, was essential for extracting a reliable two-qubit gate fidelity. Summarizing, independent single-qubit gate fidelities are around 99%* in this system, these drop to 98.8% for qubit 1 and to 96.9% for qubit 2 when simultaneously twirling the other qubit, and the two-qubit CPHASE fidelity is around 92%. We expect that by working in an isotopically purified Si^{28}/SiGe substrate and performing the two-qubit gate at the symmetry point, a CPHASE gate fidelity above the fault-tolerant threshold ($> 99\%$) can be reached. A recent report on the fidelity of controlled rotations in Si/SiO_2 quantum dots already comes close to this threshold [17]. With further improvements in charge noise levels, two-qubit gate fidelities above 99.9% are in reach.

*In the interest of space we have not included the results of independent single qubit randomized benchmarking in this chapter. See however the paper this chapter is based on [X. Xue et al. Physical Review X, 9, 021011 (2019)] for details on these and other experiments

10

Quantum error correction in crossbar architectures

A central challenge for the scaling of quantum computing systems is the need to control all qubits in the system without a large overhead. A solution for this problem in classical computing comes in the form of so called crossbar architectures. Recently we made a proposal for a large scale quantum processor [Li et al. Science Advances 4 (7), eaar3960] to be implemented in silicon quantum dots. This system features a crossbar control architecture which limits parallel single qubit control, but allows the scheme to overcome control scaling issues that form a major hurdle to large scale quantum computing systems. In this work, we develop a language that makes it possible to easily map quantum circuits to crossbar systems, taking into account their architecture and control limitations. Using this language we show how to map well known quantum error correction codes such as the planar surface and color codes in this limited control setting with only a small overhead in time. We analyze the logical error behavior of this surface code mapping for estimated experimental parameters of the crossbar system and conclude that logical error suppression to a level useful for real quantum computation is feasible.

This chapter has been published, with minor changes, in J. Helsen, M. Steudtner, M. Veldhorst & S. Wehner, *Quantum error correction in crossbar architectures*, Quantum Sci. Technol. 3 035005 (2018)

10.1. Introduction

When attempting to build a large scale quantum computing system a central problem, both from experimental and theoretical perspectives, is what might be called the interconnect problem. This problem, which also exists in classical computing, arises when computational units (e.g. qubits in quantum computers, transistors in classical computers) are densely packed such that there is not enough room to accommodate individual control lines to every unit. A solution to this problem, which is commonplace in classical computing systems, is a so called 'crossbar architecture'. In this class of computing architecture we do not draw a control line to every qubit but rather organize computational units in a grid with control lines addressing full rows and columns of this grid. Control effects then happen at the intersection of column and row lines. In this way, using N control lines $O(N^2)$ computational units can be addressed. This makes it possible to scale the system to a large number of qubits. The price to pay for this is a reduced ability to perform operations on different units in the grid in parallel. For classical systems this is not a fundamental problem, but when the computational units are qubits, whose information decays over time, parallelism becomes absolutely essential. This introduces a formidable roadblock for the development of crossbar systems for quantum computing systems. Nevertheless various crossbar architectures for quantum computers have been proposed in the past [1–5]. Recently [4] we proposed a quantum computing platform based on spin qubits in silicon quantum dots featuring a crossbar architecture. This architecture features compatibility with modern silicon manufacturing techniques and in combination with recent advances in controlling quantum dot qubits and the inherent long coherence times of spin qubits in silicon we expect it to be a formidable step forwards in creating large scale quantum computing devices.

Any realistic quantum computing device, including the one we propose in [4], will suffer from noise processes that degrade quantum information. This noise can be combated by quantum error correction [6, 7], where quantum information is encoded redundantly in such a way that errors can be diagnosed and remedied as they happen without disturbing the encoded information. Many quantum error correction codes have been developed over the last two decades and several of them have desirable properties such as high noise tolerance, efficient decoders and reasonable implementation overhead. Of particular note are the planar surface [8] and color codes [9], which have the nice property that they can be implemented in quantum computing systems where only nearest-neighbor two-qubit gates are available.

However these codes, and all other quantum error correction codes, were developed under the (often implicit) assumption that all physical qubits participating in the code can be controlled individually and in parallel. For large (read: comprising many qubits) error correction codes this introduces a tension between the needs of the error correction code and the control limitations for large systems mentioned above. While practical large-scale quantum computers most likely pose control limitations, surprisingly little work has been done in this area [10]. Here we investigate the minimal amount of parallel control resources needed for quantum error correction and focus in particular on crossbar architectures. In figs. 10.1 and 10.5 we summarize the layout and control limitations of the

architecture in [4]. Overcoming these limitations motivates the current work.

10.1.1. Contributions

Analysis of the crossbar system

We analyze the crossbar architecture we propose in [4]. We give a full description of the layout and control characteristics of the architecture in a manner accessible to non-experts in quantum dots. We develop a language for describing operations in the crossbar system. Of particular interest here are the regular patterns (see e.g. section 10.3.4) that are implied by the crossbar structure. These configurations provide an abstraction on which we build mappings of quantum error correction codes (see below) This analysis is particular to the system in [4] but we believe many of the considerations to hold for more general crossbar architectures.

An efficient algorithms for control on crossbar architectures

We develop an algorithm for moving around qubits (shuttling) on crossbar architectures. We show that the task of shuttling qubits in parallel can be described using a matrix taking value in an idempotent monoid. The control algorithm then reduces to finding independent columns of this matrix, for a suitable notion of independence. This algorithm in principle allows the straightforward mapping of more complicated quantum algorithms which require long-range operations, with little operational overhead. We also expect this algorithm to be applicable to the control of more general crossbar architectures. We also sketch an algorithm for parallel two-qubit interactions in crossbar systems which produce *optimal* control sequences. This algorithm is based on computing the Schmidt-normal form of matrices with entries in the rings $\mathbb{Z}_2$ and $\mathbb{Z}_4$.

Mapping of surface and color codes

We map the planar surface code and the $6.6.6.$ (hexagonal) and $4.8.8.$ (square-octagonal) color codes [9] to the crossbar architecture, taking into account its limited ability to perform parallel quantum operations. The tools we develop for describing the mapping, in particular the configurations described in section 10.3.4, should be generalizable to other quantum error correction codes and general crossbar architectures.

Analysis of the surface code logical error

Due to experimental limitations the mappings mentioned above might not be attainable in near term devices. Therefore we adapt the above mappings to take into account practical limitations in the architecture [4]. In this version of the mapping the length of an error correction cycle scale with the distance of the mapped code. This means the mapping does not allow for arbitrary logical error rate suppression. Therefore we analyze the behavior of the logical error rate with respect to estimated experimental error parameters and find that the logical error rate can in principle be suppressed to below 10^{-20} (an error rate

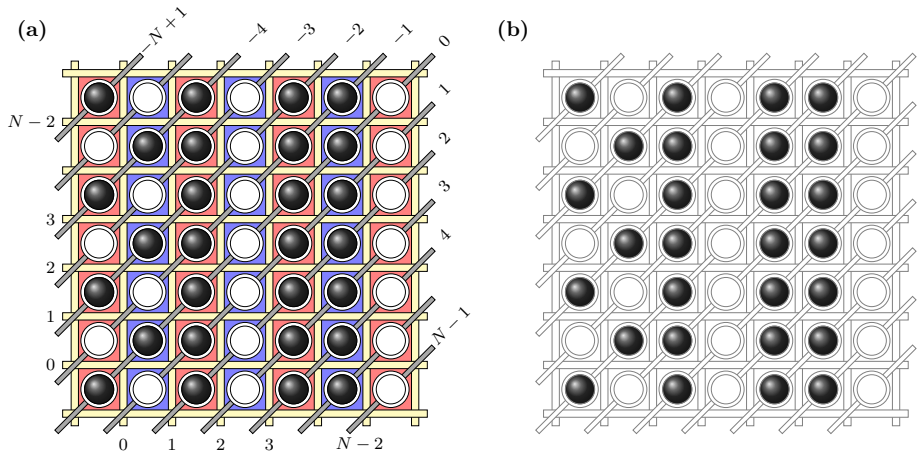


Figure 10.1: **(a)** A schematic of the Quantum Dot Processor (QDP) that we propose [4], see section 10.2.1 for details. The white circles correspond to quantum dots, with the black filling denoting the presence of electrons, whose spins are employed as qubits. All dots are embedded in either a red or a blue column. Single qubit gates can only be applied globally on either all qubits in all blue columns or all qubits in all red columns. The vertical, horizontal (both yellow) and diagonal lines (gray) are a feature of this crossbar scheme. The horizontal and vertical gate lines implement barriers that isolate the dots from each other. The diagonal lines simultaneously control the dot potentials of all dots coupled to one line. Quantum operations are effected by pulsing individuals lines. In order to perform two qubit operations on qubits in adjacent dots, one typically needs to lower the barrier that separates them, and change the dot potentials by operating the diagonal lines. Note that two-qubit gates applied to adjacent qubits in the same column are inherently different (by nature of the QDP design) from two-qubit gates between two adjacent qubits in the same row. With the control lines, we can also move qubits from dot to dot and measure them. However, since each control line influences $O(N)$ qubits, individual qubit control, as well as parallel operation on many qubits is limited. **(b)** Abstracted version of the QDP scheme representing the classical BOARDSTATE matrix. The BOARDSTATE holds no quantum information, but encodes where qubits are located on the QDP grid.

comparable to the error rate of classical computers [11]), allowing for practical quantum computation to take place.

Our work raises several interesting theoretical questions regarding the mapping of quantum algorithms to limited control settings, see section 10.6.

10.1.2. Outline

10

In section 10.2 we introduce the architecture we proposed in [4]. We forgo an explanation of the physics and focus on the abstract control aspects of the system (explaining them in a largely self-contained manner accessible to non experts in quantum dot physics). We introduce classical helper objects such as the BOARDSTATE which will aid later developments. We discuss one- and two-qubit operations, measurements, and qubit shuttling. In section section 10.3 we focus on parallel operations. We discuss difficulties inherent in parallel operation in a crossbar system and develop an algorithm for dealing with them efficiently. We also introduce several BOARDSTATE configurations which feature prominently in quantum error correction mappings and describe how to reach them efficiently

by parallel shuttling. In section 10.4 we give a quick introduction to quantum error correction with a particular focus on the planar surface code and the $4.8.8.$ and $6.6.6.$ color code. In section 10.4.4 we bring together all previous sections and devise a mapping of the planar surface code to the crossbar architecture. This we continue in section 10.4.5 for the $6.6.6.$ and $4.8.8.$ color codes. Finally in section 10.5 we analyze in detail the logical error probability of the surface code mapping as a function of the code distance and estimated error parameters of the crossbar system.

10.2. THE QUANTUM DOT PROCESSOR

In this section we will give an overview of the quantum dot processor (QDP) architecture as proposed in [4]. We will use this architecture as a concrete realization of the more general idea of quantum crossbar architectures. We will focus not so much on the details of the implementation but rather focus on abstract operational properties of the system as they are relevant for our purposes. The basic organization of the QDP is that for an $N \times N$ grid of qubits interspersed with control lines that effect operations on the qubits. The most notable feature of the QDP (and crossbar architectures in general) is the fact that any classical control signal sent to a control line will be applied simultaneously to all qubits adjacent to that control line. This means that every possible classical instruction applied to the QDP will affect $O(N)$ qubits (these qubits will not necessarily be physically close to each other). This has important consequences for the running of quantum algorithms on the QDP (or any crossbar architecture) that must be taken into account when compiling these algorithms to hardware level instructions. Notably it places strong restrictions on performing quantum operations in parallel on the QDP. To deal with these restrictions it is important to have a good understanding of how operations are performed on the QDP. It is for this reason that we begin our study of the QDP with an examination of its control structure at the hardware level. We describe the physical layout of the system and develop nomenclature for the fundamental control operations. This nomenclature might be called the 'machine code' of the QDP. From these basic instructions we go on to construct all elementary operations that can be applied to qubits in the QDP. These are quantum operations, such as single qubit gates, nearest-neighbor two-qubit gates and qubit measurements but also a non-quantum operation called coherent shuttling which does not affect the quantum state of the QDP qubits but changes their connectivity graph (i.e. which qubits can be entangled by two-qubit gates). All of these operations are restricted by the nature of the control architecture in a way that gives rise to interesting patterns (section 10.3.4) and which we will more fully examine in section 10.3.

10.2.1. LAYOUT

A schematic overview of the QDP architecture is given in fig. 10.1, where qubits (which are electrons, denoted by black balls) occupy an array of $N \times N$ quantum dots (hereafter often referred to as sites). The latter are denoted by white sites when empty, since they either are occupied by a qubit or not. We will label the dots by tuples containing row and column indices $(i, j) \in [0 : N-1]^{\times 2}$ (beginning from the *bottom left* corner), such that a single qubit state $|\psi\rangle$ living on the (i, j)'th site will be denoted by $|\psi\rangle_{(i,j)}$. We assume the qubits to be initialized in the state $|0\rangle$. For future reference we note that $|0\rangle$ corresponds

to the spin-up state and $|1\rangle$ to the spin-down state of the electron constituting the qubit.

Typically we will work in a situation where half the sites are occupied by a qubit and half the sites are empty (as seen in fig. 10.1 (a)). Because (as we discuss in section 10.2.3) the qubits can be moved around on the grid and the two-qubit gates depend on the filling of the grid, it is important to keep track of which sites contain qubits and which ones do not. This can be done efficiently in classical side-processing. To this end we introduce the BOARDSTATE object. BOARDSTATE consists of a binary $N \times N$ matrix with a 1 in the (i, j)'th place if the (i, j)'th site contains an electron (qubit) and a 0 otherwise. The BOARDSTATE does not contain information about the qubit state $|\psi\rangle_{(i,j)}$, only about the electron occupation of the grid. A particular BOARDSTATE is illustrated in the left panel of fig. 10.1.

We now turn to describing the control structures that are characteristic for this architecture. As a first feature, we would like to point out that each site is either located in a red or a blue region in fig. 10.1 (left panel). The blue (red) columns correspond to regions of high (low) magnetic fields, which plays a role in the addressing of qubits for single qubit gates. We will denote the set of qubits in blue columns (identified by their row and column indices) by $\mathcal{B}$ and the set of qubits in red columns by $\mathcal{R}$.

Much finer groups of sites can be addressed by the control lines that run through the grid. The crossbar architecture features control lines that are connected to $O(N)$ sites. At the intersections of these control lines individual sites and qubits can be addressed. This means that using $O(N)$ control lines $O(N^2)$ qubits can be controlled. As seen in fig. 10.1 the rows and columns of the QDP are interspersed with horizontal and vertical lines (yellow), as a means to control the tunnel coupling between adjacent sites. We refer to those lines as barrier gates, or barriers for short. Each line can be controlled individually, but a pulse has an effect on all $O(N)$ qubits adjacent to the line. Another layer of control lines is used to address the dots itself rather than the spaces in between them. The diagonal gate lines (gray), are used to regulate the dot potential. We label the horizontal and vertical lines by an integer running from 0 to $N - 2$ and the diagonal lines with integers running from $-(N - 2)$ to $N - 2$ where the $-(N - 2)$'th line is the top-left line and increments move towards the bottom right (see fig. 10.1(a)). Next we describe how these control lines can be used to effect operations on the qubits occupying the QDP grid.

10.2.2. CONTROL AND ADDRESSING

As described above, the QDP consists of quantum dots interspersed with barriers and connected by diagonal lines. For our purposes these can be thought of as abstract control knobs that apply certain operations to the qubits. In this section we will describe what type of gates operations are possible on the QDP. We will not concern ourselves with the details of parallel operation until section 10.3.

There are three fundamental operations on the QDP which we will call the “grid operations". These operations are “lower vertical barrier" (V), “lower horizontal barrier" (H) and “set diagonal line" (D). The first two operations are essentially binary (on-off) but the last

one (D) can be set to a value $t \in [0 : T]$ where T is a device parameter. (At the physical level this corresponds to how many clearly distinct voltages we can set the quantum dot plunger gates [4]). Although the actual pulses on those gates differ by amplitude and duration between the different gates and operations, this notation gives us a clear idea which lines are utilized. This can be done because realistically one will not interleave processes in which pulses have such different shapes. We can label the grid operations by mnemonics (which in a classical analogy we will call OPCODES) as seen in section 10.2.2. These OPCODES are indexed by an integer parameter that indicates which control line it applies to. We count horizontal and vertical lines starting at zero from the lower left corner of the grid (see fig. 10.1). Note that the lines at the boundary of the grid are never addressed in our model and are thus not counted.

We indicate parallel operation of a collection of OPCODES by ampersands, e.g. `D[1]&H[2]&D[5]`. We also define inherently parallel versions (in section 10.2.2) of the basic OPCODES that take as input a binary vector V of length N (for the diagonal line this is a T-valued vector of length N)

OPCODE	Effect
`V[i]`	Lower vertical barrier at index `i`
`H[i]`	Lower horizontal barrier at index `i`
`D[i][t]`	Set diagonal line at index `i` to value `t`

OPCODE	Effect
`V[V]`	Set vertical barrier to $V(i)$, $\forall i \in [0 : N-2]$
`H[V]`	Set horizontal barrier to $V(i)$, $\forall i \in [0 : N-2]$
`D[V]`	Set diagonal at height $V(i)$, $\forall i \in [-N+2 : N-2]$

These grid operations can be used to induce some elementary quantum gates and operations on the qubits in the QDP. Below we describe these operations.

10.2.3. ELEMENTARY OPERATIONS

Here we give a short overview of the elementary operations available in the QDP. We will describe basic single qubit gates, two-qubit gates, the ability to move qubits around by coherent shuttling [12] and a measurement process through Pauli Spin Blockade (PSB) [13]. All of these operations are implemented by a combination of the grid operations defined in section 10.2.2, and always have a dependence on the BOARDSTATE .

COHERENT QUBIT SHUTTLING

An elementary operation of the QDP is the coherent qubit shuttling [12, 14], of one qubit to an adjacent, empty site. That means that an electron (qubit) is physically moved to the other dot (site) utilizing at least one diagonal line and the barrier between the two sites. It thereby does not play a role whether the shuttling is in horizontal (from a red to a blue column or the other way around) or vertical direction (inside the same column). However,

the shuttling in between columns results in a Z rotation, that must be compensated by timing operations correctly, see [4] for details. This Z rotation can also by used as a local single qubit gate, see section 10.2.3. The operation is dependent on the BOARDSTATE by the prerequisite that the site adjacent to the qubit to must be empty. Collisions of qubits are to be avoided, as those will lead to a collapse of the quantum state (see however the measurement process in section 10.2.3). We now describe the coherent shuttling as the combination of grid operations.

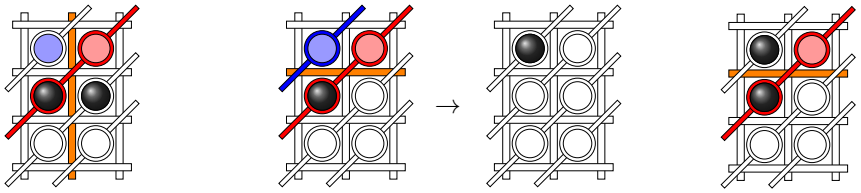
We lower the vertical (or horizontal) barrier in between the two sites and instigate a 'gradient' of the on-site potentials of the two dots. That is, the diagonal line of the site containing the qubit must be operated at $t \in [0 : T]$ while the line overhead the empty site must have the potential $\hat{t} \in [0 : T]$ with $\hat{t} = t - 1$. Note that this implies it might not be operated at all (if it is already at the right level). We will subsequently refer to the combination of a lowered barrier and such a gradient as a "flow". A flow will in general be into one of the four directions on the grid. We define the commands `VS[i,j,k]` (vertical shuttling) and `HS[i,j,k]` (horizontal shuttling). The command `VS[i,j,k]` shuttles a qubit at location (i, j) to $(i+1, j)$ for $k = 1$ (upward flow) and shuttles a qubit at location $(i+1, j)$ to (i, j) for $k = -1$ (downward flow). Similarly, the command `HS[i,j,k]` shuttles a qubit at location (i, j) to $(i, j + 1)$ for $k = 1$ (rightward flow) and shuttles a qubit at location $(i, j + 1)$ to (i, j) for $k = -1$ (leftward flow). See table 10.1 for a summary of these OPCODES.

Using only these control lines, we can individually select a single qubit to be shuttled. However, when attempting to shuttle in a parallel manner, we have to be carefully take into account the effect that the activation of several of those lines has on other locations. We will deal with this in more detail in section 10.3.1.

Measurement and readout

The QDP allows for local single qubit measurements in the computational basis $|0\rangle, |1\rangle$. We can measure a qubit by attempting to shuttle it to a horizontally adjacent site that is already occupied by an ancilla qubit and then detecting whether the shuttling was successful. This process is called Pauli Spin Blockade (PSB) measurement [4, 13]. However, the QDP's ability to perform this type of qubit measurements is limited by three factors.

Firstly, the measurement requires an ancilla qubit horizontally adjacent to the qubit to be measured. This ancilla qubit must be in a known computational basis state. Moreover, if the ancilla qubit is in the state $|0\rangle$ the ancilla qubit must be in the set $\mathcal{B}$ (blue columns in fig. 10.1) while the qubit to be measured must be in the set $\mathcal{R}$ (red columns in fig. 10.1). On the other hand, if the ancilla qubit is in the state $|1\rangle$ the ancilla qubit must be in the set $\mathcal{R}$ while the qubit to be measured is in the set $\mathcal{B}$. This means that when an qubit-ancilla pair is in the wrong configuration we must first shuttle both qubits one step to the left (or both the the right). Note that this takes two additional shuttling operations, which means it is important to keep track at all times where on the BOARDSTATE the qubit and its ancilla are or else incur a shuttling overhead (which might become significant when dealing with large systems and many simultaneous measurements). We will deal with this problem of



(a) CPHASE or Measurement **(b)** Coherent shuttling **(c)** $\sqrt{\text{SWAP}}$

Figure 10.2: Schematic representation of the use of control lines for the native operations in the QDP. Qubits are represented by black balls on the grid. Red or blue colored dots are empty, but their dot potentials change due to an operation of the diagonal line they are coupled to. Empty dots unaffected by grid operations are white. **(a)** Grid operations necessary to perform a measurement or a two-qubit effective CPHASE gate between the two qubits. The orange barrier between the two qubits is lowered, and the dot potentials along the red diagonal line is raised by pulsing the latter. Note that the empty, red colored dot is also effected by that action, and its barrier to the adjacent dot is lowered. If the two dots in the upper row were not empty, side effects would occur. See section 10.2.3 for more information on the nature of the two-qubit gates. Note also that the readout procedure of the measurement requires us to have the upper dot (light blue) empty, if the barrier gate between them is used for readout. **(b)** Vertical shuttling of a qubit (to the top dot) requires to lower the orange barrier. One can than either raise the dot potentials on the red diagonal line, or lower the potential on the blue dot by addressing the blue diagonal. **(c)** Schematic representation of the control lines used for performing two-qubit $\sqrt{\text{SWAP}}$ gate between the two qubits on that grid. The orange barrier is lowered and the red diagonal line is utilized to detune dot potentials.

qubit-ancilla pair placement in more detail in section 10.3.3.

Secondly, assuming that the qubit-ancilla pair is in the right configuration to perform the PSB process one still needs to perform a shuttling-like operation to actually perform the measurement. On the technical level, the operation is different from coherent shuttling, but the use of the lines is similar with the difference that after the readout, the shuttling-like operation is undone by the use of the same lines as before - which are not necessarily the lines one would use to reverse a coherent shuttling operation. However, scheduling measurement events on the QDP is at least as hard as the scheduling of shuttle operations discussed above. Depending on the state the qubit is in, it will now assume one of two possible states that can be distinguished by their charge distribution.

Thirdly, the readout process requires to have a barrier line that borders to the qubit pair, with an empty dot is across the spot of the qubit to be measured. This is a consequence of the readout procedure.
In table 10.1 we introduce the measurement OPCODE `M[i,j,k]` with $k \in \{-1, 1\}$ to denote a measurement of a qubit at location (i, j) with an ancilla located to the left $(k = -1)$ or to the right $(k = 1)$.

SINGLE-QUBIT ROTATIONS

There are two ways in which single qubit rotations can be performed on the QDP, both with drawbacks and advantages. The first method, which we call the semi-global qubit rotation, relies on electron-spin-resonance [15]. Its implementation in the QDP allows for any rotation in the single qubit special unitary group $SU(2)$ [16] to be performed but we

OPCODE	Control OPCODES	Effect
`HS[i,j,k]`	`V[i]&D[i-j][t-1/2-k/2]` `&D[i-j+1][t-1/2+k/2]`	$(k=1)$: Shuttle from (i,j) to $(i,j+1)$ $(k=-1)$: Shuttle from $(i,j+1)$ to (i,j)
`VS[i,j,k]`	`H[j]&D[i-j][t-1/2-k/2]` `&D[i-j-1][t-1/2+k/2]`	$(k=1)$: Shuttle from (i,j) to $(i+1,j)$ $(k=-1)$: Shuttle from $(i+1,j)$ to (i,j)
`M[i,j,k]`	`HS[i,j+1/2+k/2,-k]`	Measurement of qubit at (i,j) using the ancilla at $(i,j+k)$

Table 10.1: OPCODES for horizontal and vertical shuttling and measurement together with the control OPCODES required to implement these operations on the QDP.

do not have parallel control of individual qubits. The control architecture of the QDP is such that we can merely apply the same single qubit unitary rotation on all qubits in either $\mathcal{R}$ or $\mathcal{B}$ (even or odd numbered columns). Concretely we can perform in parallel the single qubit unitaries

$$U_{\mathcal{R}} = \bigotimes_{(i,j)\in\mathcal{R}} U_{i,j} \qquad U \in SU(2) \tag{10.1}$$

$$U_{\mathcal{B}} = \bigotimes_{(i,j)\in\mathcal{B}} U_{i,j} \qquad U \in SU(2), \tag{10.2}$$

where $U_{i,j}$ means applying the same unitary U to the state carried by the qubit at location (i,j). In general the only way to apply an arbitrary single qubit unitary on a single qubit in $\mathcal{B}$ (or $\mathcal{R}$) is by applying the unitary to all qubits in $\mathcal{B}$ ($\mathcal{R}$), moving the desired qubit into an adjacent column, i.e. from $\mathcal{B}$ to $\mathcal{R}$ ($\mathcal{R}$ to $\mathcal{B}$) and then applying the inverse of the target unitary to $\mathcal{R}$ ($\mathcal{B}$). This restores all qubits except for the target qubit to their original states and leaves the target qubit with the required unitary applied. The target qubit can then be shuttled to its original location. A graphical depiction of the BOARDSTATE associated with this manoeuvre can be found in fig. 10.3. This means applying a single unitary to a single qubit takes a constant amount of grid operations regardless of grid size.

The second method does allow for individual single qubit rotations but is limited to performing single qubit rotations of the form

$$U(\phi) = e^{i\phi Z}, \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \phi \in [0, 2\pi) \tag{10.3}$$

This operation can be performed on a given qubit $|\psi\rangle_{(i,j)}$ by shuttling it from (i,j) to $(i, j \pm 1)$. When the qubit leaves the column it was originally defined ($\mathcal{B}$ to $\mathcal{R}$ or vice versa) it will effectively start precessing about its Z axis [4]. This effect is always present but it can be mitigated by timing subsequent operations such that a full rotation happens between every operation (effectively performing the identity transformation, see section 10.2.3). By changing the timing between subsequent operations any rotation of the form eq. (10.3) can be effected. This technique will often be used to perform the Z gate (defined above) and the $S = \sqrt{Z}$ phase gate in error correction sequences.

TWO-QUBIT GATES

As the last elementary tool, we have the ability to apply entangling two-qubit gates on adjacent qubits. The QDP can perform two different types of two-qubit gates. Inside

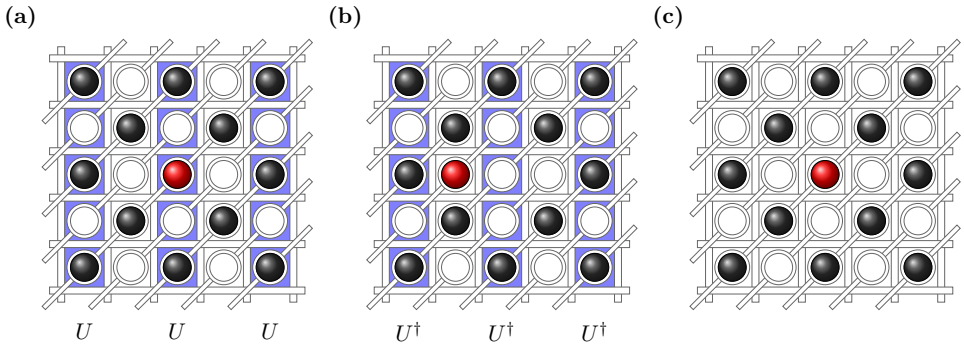


Figure 10.3: BOARDSTATE schematic for applying the unitary U to a single qubit (red). Time flows from left to right in the schematic. This process illustrates both, the possibility to retain single qubit control by using coherent shuttling, and the overhead that comes with it. **(a)** we firstly apply the unitary U (blue bars) to all qubits in $\mathcal{R}$ ($\mathcal{B}$). We then move the qubit to the adjacent column. Note that this takes two operations because we do not want any other qubits transitioning with it. In **(b)**, we apply the inverse unitary $U^\dagger$ to all qubits in $\mathcal{R}$ ($\mathcal{B}$). In the last step we move the red qubit back, such that it is in its original position in **(c)**.

one column, so between qubits at locations (i, j) and $(i \pm 1, j)$, a square-root of SWAP ($\sqrt{\text{SWAP}}$) can be realized [17]. This can be done by lowering the horizontal barrier between the two qubits and toggling the voltage on the diagonal lines overhead the two qubits. This situation is illustrated in fig. 10.2 (c). The $\sqrt{\text{SWAP}}$ gate is defined as

$$\sqrt{\text{SWAP}} = \begin{pmatrix} 1 & & & \\ & (1+i)/2 & (1-i)/2 & \\ & (1-i)/2 & (1+i)/2 & \\ & & & 1 \end{pmatrix}, \tag{10.4}$$

in the computational basis. Alternatively, between horizontally adjacent qubits, e.g. between $(i, j) \in \mathcal{R}$ and $(i, j \pm 1) \in \mathcal{B}$ the native two-qubit gate is an effective CPHASE gate which has matrix representation

$$\text{CPHASE} = \begin{pmatrix} 1 & & & \\ & e^{i\phi_1} & & \\ & & e^{i\phi_2} & \\ & & & 1 \end{pmatrix}, \tag{10.5}$$

in the computational basis and with the two angles $\phi_1 + \phi_2 \bmod 2\pi = \pi$ (demonstrated in [18–20]). This gate can be performed between horizontally adjacent qubits by lowering the vertical barrier between them and toggling the overhead diagonal lines. This is illustrated in fig. 10.2 (a). In practice we expect the $\sqrt{\text{SWAP}}$ gate to have significantly higher fidelity than the CPHASE gate [4] so in any application (e.g. error correction) the $\sqrt{\text{SWAP}}$ gate is the preferred native two-qubit gate on the QDP. In table 10.2 we define OPCODES for the horizontal interaction (CPHASE) and the vertical interaction ($\sqrt{\text{SWAP}}$).

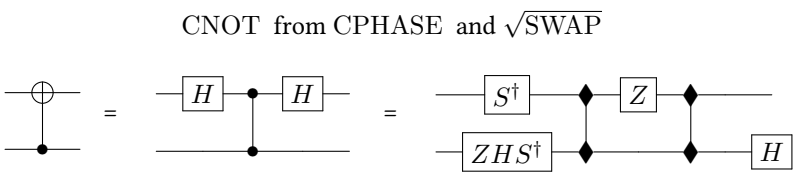


Figure 10.4: Construction of the CNOT gate out of the native CPHASE and $\sqrt{\text{SWAP}}$ gates. Note that one requires two $\sqrt{\text{SWAP}}$ gates to construct a CNOT gate [21]. When performing arbitrary algorithms it would be preferable to forgo this substitution and instead compile the algorithm directly into a gateset containing the $\sqrt{\text{SWAP}}$ gate.

OPCODE	Effect	Parameter
`HI[(i,j)]`	Perform CPHASE gate between sites `(i,j)` and `(i,j+1)`	$(i,j) \in [0:N-2]^{\times 2}$
`VI[(i,j)]`	perform $\sqrt{\text{SWAP}}$ gate between sites `(i,j)` and `(i+1,j)`	$(i,j) \in [0:N-2]^{\times 2}$
`HC[(i,j)]`	Perform CNOT (using CPHASE) between `(i,j)` and `(i,j+1)`	$(i,j) \in [0:N-2]^{\times 2}$
`VC[(i,j)]`	perform CNOT (using $\sqrt{\text{SWAP}}$) between `(i,j)` and `(i+1,j)`	$(i,j) \in [0:N-2]^{\times 2}$

Table 10.2: OPCODES for horizontal and vertical two-qubit operations on the QDP, respectively the CPHASE and $\sqrt{\text{SWAP}}$ gates. We also include OPCODES for the performing of CNOT gates composed of $\sqrt{\text{SWAP}}$ or CPHASE gates.

CNOT subroutine

Many quantum algorithms are conceived using the CNOT gate as the main two-qubit gate. However the QDP does not support the CNOT gate natively. It is easy to construct the CNOT gate from the CPHASE gate by dressing the CPHASE gate with single qubit Hadamard rotations as seen in fig. 10.4 (left). It is slightly more complicated to construct a CNOT gate using the $\sqrt{\text{SWAP}}$ but it can be done by performing two $\sqrt{\text{SWAP}}$ gates interspersed single qubit rotations [19–21] as seen in fig. 10.4 (right). If the control qubit is moved from an adjacent column on the QDP (as it is in most cases we will deal with) the Z and S gates can be performed by the Z-rotation-by-waiting technique described in the last section. For completeness we also define an OPCODE for the CNOT operation in table 10.2.

10.3. Parallel operation of a crossbar architecture

In this section we focus on performing operations in parallel on the QDP (or more general crossbar architectures). Because of the limitations imposed by the shared control lines of the crossbar architecture, achieving as much parallelism as possible is a non-trivial task. We will discuss parallel shuttle operations, parallel two qubit gates, parallel single qubit gates and parallel measurement. As part of the focus on parallel shuttling we also include some special cases relevant to quantum error correction where full parallelism is possible.

Before we start our investigation however, we would like to put three issues into focus that are likely to be encountered when attempting parallel operations. Firstly, it must be understood that an operation on one location on a crossbar system can cause unwanted side effects in other locations (that might be far away). As indicated in section 10.2 many elementary operations on the grid in particular take place at the crossing points of control

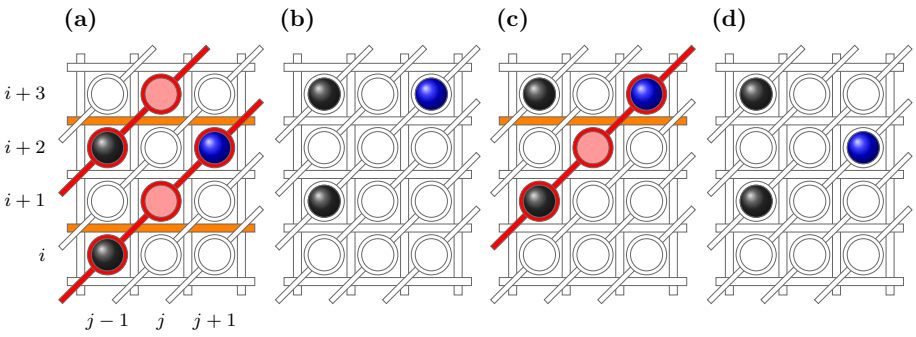


Figure 10.5: Spurious shuttle operations. Here we illustrate an example of unintended side effects that occur due to the limited control. We again denote qubits by colored balls, and color barriers and lines that are operated. Empty dots with changed potentials are colored as well, white dots are unaffected. **(a)** The black qubits are to be shuttled from $(i, j-1)$ to $(i+1, j-1)$ and from $(i+2, j)$ to $(i+3, j)$ respectively without moving the blue qubit. For that purpose, the (orange) barriers between the two dot pairs are lowered, as well as the (red) diagonal lines through $(i, j-1)$ and $(i+2, j)$ are pulsed, such that the dot potentials on those sites are raised. **(b)** The qubit on $(i+3, j+1)$ has unintentionally moved to $(i+2, j+1)$. **(c)** To remedy this situation, we lower the barrier number $i+2$ again (orange), and also raise the potential on $(i+3, j+1)$ and all other dots that are connected by the pulsed diagonal line (red). In **(d)**, the desired situation is achieved.

lines. This means that any parallel use of these grid operations must take into account "spurious crossings" which may have such unintended side effects. We can illustrate this with an example. Imagine we want to perform the vertical shuttling operations `VS[i,j-1,1]` and `VS[i+2,j-1,1]` in parallel (see fig. 10.5 for illustration). We can do this by lowering the horizontal barriers at rows i and $i+2$ (orange in illustration) and elevating the on-site potentials on the diagonal lines $i-j+1$ and $i+2-j+1$ (red in illustration). This will open upwards flows at locations $(i, j-1)$ and $(i+2, j-1)$. However it will also open an upward flow at the location $(i+2, j+1)$. This means, if a qubit is present at that location an unintended shuttling event will happen. To avoid this outcome we must either perform the operations `VS[i,j-1,1]` and `VS[i+2,j-1,1]` in sequence (taking two time-steps) or perform an operation `VS[i+2,j+1,-1]` to fix the mistake we made, again taking two time-steps. This is a general problem when considering parallel operations on the QDP.

Secondly, we would like to point out that in realistic setups, we expect a trade-off between parallelism (manifested in algorithmic depth) and operation fidelity (in particular this will be the case in the QDP system). In order to understand this, we have to be aware that most operations consist of applying the correct pulses for the right amount of time. These durations however can slightly vary from site to site (due to manufacturing imperfections), so we e.g. must be able to switch barriers back on again prematurely when accounting for a site with a shorter time required. If this is not possible (maybe because it would cause side effects) a loss in operation fidelity is a consequence of the resulting improperly timed operation. The most robust case is thus to schedule operations line-by-line. By this we mean that we attempt to perform $O(N)$ grid operations in a time-step while using every horizontal, diagonal or vertical line only once per individual grid operation. If we for instance schedule several vertical shuttle operations, we may choose to start by lowering

one of horizontal barrier first and then detune the dot potentials of all qubits adjacent to that barrier, by pulsing the corresponding diagonal lines. To account for the variations, we reset the diagonal lines at slightly different times. Line-by-line operations work with either line types for every two-dot operation (measurement, shuttling and two-qubit gates). Note however that for shuttling operations individual control over one line is sufficient, whereas for measurement and two-qubit gates we would ideally like to be able to control two lines per qubit pair individually, where one line should be the barrier separating the two paired qubits. Results presented in the following take into account these constraints for quantum error correction. The parallel operation nonetheless remains one of the greatest challenges of the crossbar scheme. In this section we will assume all operations to be perfect (even when performed in parallel) but in section 10.5 we perform a more detailed analysis of the behavior of the QDP when operational errors are taken into account.

Thirdly, from a performance perspective it is important to separate the operations that have to be done on the qubits on the crossbar grid from operations that can be done by classical side computation (which for our purposes is essentially free). We will deal with this by including classical side computation in the OPCODES for parallel operation. This way the complexity of dealing with spurious operations is abstracted away. We devise algorithms that take in an arbitrary list of shuttling or two-qubit gate locations and work out a sequence of shuttling or two-qubit gate steps that achieve that list. We begin with discussing parallel shuttle operations.

10.3.1. Parallel shuttle operations

We define parallel versions of the shuttling OPCODES `HS[i,j,k]` and `VS[i,j,k]` as

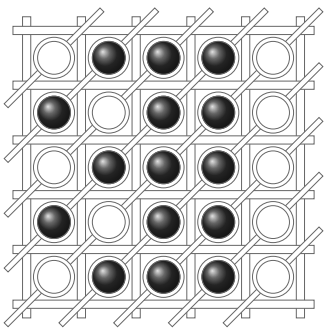
OPCODE	Effect
`HS[L]`	Perform `HS[i,j,k]` for all $(i,j,k) \in$ `L`
`VS[L]`	Perform `HS[i,j,k]` for all $(i,j,k) \in$ `L`

This code takes in a set (denoted as `L`) of tuples (i,j,k) which denote 'locations at which shuttling happens' (i,j) and 'shuttling direction' (k). From these codes it is not immediately clear how many of the shuttling operations can be performed in a single grid operation, i.e. setting the diagonal lines to some configuration and lowering several horizontal or vertical barrier. If multiple grid operations are needed (such as in the example fig. 10.5) we would like this sequence of grid operations to be as short as possible. However, given some initial BOARDSTATE and a parallel shuttling command `HS[L]` it is not clear what the sequence of parallel shuttling operations actualizing this command is. Below we analyze this problem of parallel shuttling in more detail and give a classical algorithm that produces, from an input `HS[L]` or `VS[L]` a sequence of parallel grid operations that performs this command. Ideally we would like this sequence to be as short as possible. This algorithm does not perform optimally in all circumstances (i.e. it does not produce the shortest possible sequence of parallel shuttling operations) but for many relevant cases it performs quite well. Note that this is a technical section and the details are not needed to understand the quantum error correction results in sections 10.4, 10.4.4 and 10.4.5. Readers

interested only in those may skip ahead to section 10.3.2

The flow matrix

We will only consider shuttling to the left and to the right but all mechanisms introduced work equally well for shuttling in the vertical directions. As will be seen in section 10.3.4 some BOARDSTATE configurations can be converted into each other in an amount of grid operations that is constant in the size of the grid. It can be seen that the problem of whether two shuttles can be performed in parallel is a problem with a matrix structure, as flows can only occur at the intersection open barriers and non-trivial diagonal line gradients. To capture this matrix intuition we construct, from the initial BOARDSTATE and the command HS[L] a matrix F which we call the flow matrix. This matrix will have entries corresponding to the crossing of the gradient line between two diagonal qubit lines and the vertical barrier lines. The flow matrix is defined with respect to a specific command HS[L] and its entries correspond to the locations on the grid where we want shuttling in certain directions to happen.



$$\texttt{List} = \{(1,0,1),(3,1,-1),(4,0,1),(3,3,1)\}$$

$$F = \begin{pmatrix} l & e & e & le \\ r & re & e & le \\ re & e & e & le \\ r & re & e & le \\ re & e & e & re \end{pmatrix}$$

Figure 10.6: Example of a BOARDSTATE , a parallel command HS[List] and the corresponding flow matrix F.

From a specific command HS[L] and a specific current BOARDSTATE we will define a flow matrix F. This matrix will have entries which take value in the set $\{r, l, e, re, le, *\}$. Each element of this set has a specific operational meaning. The elements r, l, e correspond to specific actions that can be taken on the qubit grid. They correspond specifically to 'shuttle to the right' (r), 'shuttle to the left' (l) and 'do nothing' (e). Note that these actions do not necessarily act on a fixed qubit. Rather they act on a specific location on the grid (where a qubit may or may not be present). The other three elements do not directly correspond to a shuttling action but rather signify that at this location we have a choice of different consistent actions. We will call these elements 'wildcards'. These wildcards signify the actions 'shuttle to the right or do nothing' (re), 'shuttle to the left or do nothing' (le), or 'any action is allowed' $(*)$.

We fill in the matrix entry F_{ij} with a symbol r for every $(i, j, 1)$ in L. This indicates that at some point in time we want to perform the operation HS[i,j,1] at that location. Similarly we fill in a symbol l on every matrix entry F_{ij} for every $(i, j+1, -1)$ in L. We

place the symbols re, le respectively on the matrix entries $F_{i(j-1)}$ and F_{ij} for every occupied site (i, j) in the BOARDSTATE that has no corresponding entry in L. This indicates that we would like for no shuttle operations to happen on these crossing points (since we want the qubit to stay put) but that we do not mind a `HS[i,j-1,1]` happening on the crossing point to the left of the qubit at (i, j) (since it will not affect the qubit) or mind a `HS[i,j,1]` happening to the right of the qubit at (i, j). Lastly we fill in the symbol e on every matrix entry F_{ij} where we want no shuttling operation to happen at any time to the right of the site (i, j) (for instance on the crossing point between two qubits that are in horizontally adjacent sites). In every other matrix entry F_{ij} we fill in the wildcard symbol $*$ indicating that we do not care if any operation happens at this crossing point. Let's summarize the above construction by

$$F_{ij} = \begin{cases} r & \text{if} \quad (i,j,1) \in \mathrm{L} \\ l & \text{if} \quad (i,j,1) \in \mathrm{L} \\ e & \text{if} \quad (\text{BOARDSTATE}\ (i,j) = 1 \wedge \text{BOARDSTATE}\ (i,j+1) = 1) \wedge ((i,j,k) \notin \mathrm{L}, k \in \{1,-1\}) \\ re & \text{if} \quad (\text{BOARDSTATE}\ (i,j) = 0 \wedge \text{BOARDSTATE}\ (i,j+1) = 1) \wedge ((i,j,k) \notin \mathrm{L}, k \in \{1,-1\}) \\ le & \text{if} \quad (\text{BOARDSTATE}\ (i,j) = 1 \wedge \text{BOARDSTATE}\ (i,j+1) = 0) \wedge ((i,j,k) \notin \mathrm{L}, k \in \{1,-1\}) \\ * & \text{if} \quad (\text{BOARDSTATE}\ (i,j) = 0 \wedge \text{BOARDSTATE}\ (i,j+1) = 0) \wedge ((i,j,k) \notin \mathrm{L}, k \in \{1,-1\}). \end{cases}$$

The flow matrix F takes values in the set $\{r, l, e, re, le, *\}$. In section 10.7 we discuss the mathematical structure of this set in more detail. The above construction gives us a matrix of operations we would like to apply to the initial BOARDSTATE . You can see an example of a BOARDSTATE and `HS[L]` command with corresponding flow matrix F in fig. 10.6.

AN ALGORITHM FOR PARALLEL SHUTTLING

The task is now to subdivide the flow matrix F into a sequence of shuttling operations that can be performed in parallel. Ideally we would like this sequence to be as short as possible. One simple way to generate a sequence of this form, as described in the beginning of the section, is to perform all operations one column at a time, i.e. lowering the first vertical barrier, setting the required gradients to shuttle every qubit adjacent to that vertical barrier and then move on to the second vertical barrier and so on. This yields a sequence of parallel shuttling operations of depth N. This solution is always possible for any flow matrix F. However, as can be seen in section 10.3.4 for some flow matrices this is far from an optimal solution. Below we set out in detail an algorithm that finds better (shorter sequences) solutions for many flow matrices. The algorithm is based on the idea that some columns of the flow matrix F can be 'dependent' on each other. For instance two columns could be composed of the exact same operations (up to a shift accounting for the fact that the diagonal lines do not run along the rows but diagonally). This means we can perform the shuttle operations in the two columns simultaneously by lowering barriers corresponding to these columns and setting the required gradient. More complicated forms of dependence are also possible. We can use dependence of columns to perform operations in parallel. For instance if a command `HS[L]` calls for exactly the same shuttling events to happen on two columns (up to a constant vertical shift proportional to the horizontal distance of the two columns) we can perform these shuttling operations in a

single time-step.

This notion of (in)dependence of columns is captured by a call to an 'independence subroutine'. We call these subroutines $\mathbf{CheckIndependence}(S, v)$ which takes in a set of columns S of the flow matrix F of and a column v of the flow matrix F and decides whether v is independent of the elements of S and $\mathbf{DependenceSet}(S, v)$ which takes in a set of columns S and a column v and returns a subset A of S containing all the columns on which v depends. We will discuss various versions of these subroutines leading to more or less refined notions of independence (and thus longer of shorter shuttling sequences) in section 10.7. We list all subroutines discussed in section 10.7 in table 10.3 together with their relative power and time complexity. Here we just treat the subroutines as a given and build the algorithm around it. This algorithm does not always yield optimal sequences of parallel shuttling operations, but it can be run using a polynomial amount of classical side-resources given that the subroutine can be constructed efficiently, (see theorem 10.1) while we expect an algorithm that always produces optimal shuttling sequences to require exponential computational resources. Below we give a pseudo-code version of the algorithm. Not that this algorithm only produces sequences of parallel shuttling operations where the ordering of the operations does not matter. See section 10.7 for more details on how this property is guaranteed.

Algorithm 3 Generate list of parallel shuttle operations

Input: Flow matrix F
Output: List of shuttle operations L

// We will consistently write columns of the flow matrix F as v_i where i indicates
// the column index of v_i in F.
Set S to an empty list
// Below we construct a set of independent columns S and sets of dependence A_i for the dependent columns v_i.
for $i \in [0 : N-2]$ **do**
    Set v_i to the i'th column of F
    //Check if the column v_i is independent of the columns already in the set S. This requires a
    // subroutine call to **CheckIndependence**. See Appendix for the construction of this subroutine.

    if **CheckIndependence**(v_i, S) is TRUE **then**
        // The function θ maps the symbols $*,\ re,\ le$ to e. We must do this since we want to make an operation
        // out of v_i later and the wildcard elements $*,\ re,\ le$ do not strictly correspond to operations. Other
        // choices are possible here but in keeping with the idea of doing a
        // minimal amount of operations, the mapping to e is a good choice.
        Add $\theta(v_i)$ to S
        Set A_i to $\{v_i\}$
    else
        Set A_i to **DependenceSet**(S, v_i)
    end if
end for
// Initialize an empty ordered set that will contain all HS [L] commands in sequence.
Set L to an empty ordered set
for $v_i \in S$ **do**
    // Initialize an empty set that will contain all tuples for a single HS [L] command.
    Set L to an empty set
    for $j \in [0 : N-2]$ **do**
        // Check if v_i is in the dependence set A_j.
        if $v_i \in A_j$ **then**
            // Loop over all components of v_i.
            for $k \in [0 : \text{length}(v_i) - 1]$ **do**
                // ϕ maps the r, l, e valued column v to an $1, -1, 0$ valued vector as $\phi(r) = 1, \phi(l) = -1, \phi(e) = 0$.
                if $\phi\big[(v_i)_k\big] \neq 0$ **then**
                    Add $\big(j, k-(i-j), \phi\big[(v_i)_k\big]\big)$ to L
                end if
            end for
        end if
    end for
    Add HS [L] to L
end for
return L

Name	Time Complexity	Relative power
Simple	$O(\textbf{CheckIndependence}(\,)) = O(NM)$ $O(\textbf{IndependenceSet}(\,)) = O(NM)$	Shorter sequences than line-by-line.
k-commutative	$O(\textbf{CheckIndependence}(\,)) = O(NMM^k k^4)$ $O(\textbf{IndependenceSet}(\,)) = O(NMM^k k^4)$	Shorter sequences than 'Simple'. Shorter sequences for increasing k.
Greedy commutative	$O(\textbf{CheckIndependence}(\,)) = O(NM^3)$ $O(\textbf{IndependenceSet}(\,)) = O(NM^3)$	Shorter sequences than 'Simple'. Relation to 'k-commutative' unknown.

Table 10.3: Table listing the time complexity and relative power of the **CheckIndependence**( ) and **IndependenceSet**( ) for three different classes of subroutine. The parameters N and M are the size of the QDP grid and the size of the input set S respectively. The subroutine classes 'simple' and 'greedy commutative' can be run in polynomial time while the class 'k-commutative' is fixed-parameter-tractable, with independent parameter k. This subroutine yields increasingly better results (shorter shuttling sequences) for increasing k but the time complexity grows rapidly with k. See section 10.7 for a detailed description of these subroutines. For an illustration of the advantages of these algorithms, one can consider the shuttle commands given in section 10.3.4. A naive line-by-line approach will take N time-steps while it is easy to see that the above algorithms find sequences of length one.

Theorem 10.1. The algorithm described in Algorithm 3 has a time complexity upper bounded by

$$\begin{aligned} O(N^4) + N\cdot O\big(\textbf{CheckIndependence}(S, v_i)\big) \\ + N\cdot O\big(\textbf{DependenceSet}(S, v_i)\big), \end{aligned} \tag{10.6}$$

where N is the number of columns in the input flow matrix F.
The subroutines **CheckIndependence**(S, v_i) and **DependenceSet**(S, v_i) both take in a set S of independent columns of the flow matrix F and a column v_i of the flow matrix F and respectively check whether v is independent of the set S or produce a subset A of S on which v depends. We list the time complexities pf various versions of these subroutines in table 10.3.

Proof. Begin by noting that the algorithm 3 consists of two independent **For**-loops. The first **For**-loop (lines 2-11) calls its body N times (ignoring constant factors). Calling the **For**-loop body (lines 3-10) in the worst case requires calling both **CheckIndependence**( ) and **DependenceSet**( ) plus some constant time instructions. This means the first **For** loop has a worst case complexity of $N\cdot O\big(\textbf{CheckIndependence}(\,)\big)+N\cdot O\big(\textbf{DependenceSet}(\,)\big)$.

The second **For**-loop (lines 13-25) consists of three nested **For** loops of length $O(N)$ with an **If**-clause inside the first two **For**-loops (line 16) constant time operation at the bottom (line 19). The first **For**-loop can be seen to be of order $O(N)$ by noting that the set of independent columns S can be no bigger than N in which case all columns are independent. The second **For**-loop (line 15) is $O(N)$ bounded by construction. Note that the **If** clause on line 16 can take time $O(N)$ to complete since for any dependency set A_j we can only say that $|A_j| \leq N$ (since A_j is a subset of the set of all columns of F). The third loop is also $O(N)$ bounded since $\text{length}(v_i) \leq N$ for all columns v_i of F. Tallying up all contributions we arrive at eq. (10.6), which completes the argument. ■

This concludes our discussion of parallel shuttling operations. Before we move on however, it is worth pointing out an interesting example where this shuttling can be used a subroutine to perform more complicated operations. This example will also be of use later

when discussing parallel measurement in section 10.3.3 and the mapping of quantum error correction codes in sections 10.4, 10.4.4 and 10.4.5.

Selective parallel single-qubit rotations

In this section we will discuss a particular example that illustrates the use of abstracting away the complexity of parallel shuttling. Imagine a QDP grid initialized in the so called *idle* configuration. This configuration can be seen in fig. 10.7. We will focus on the qubit in the odd columns (i.e. the set $\mathcal{B}$). Imagine a subset S of these qubits to be in the state $|1\rangle$ and the remainder of these qubits to be in the state $|0\rangle$. The qubits on in the set $\mathcal{R}$ can be in some arbitrary (and possibly entangled) multiqubit state $|\Psi\rangle$. We would like to change the states states of the qubits in the set S to $|0\rangle$ without changing the state of any other qubit. Due to the limited single qubit gates (see section 10.2.3) available in the QDP this is a non-trivial problem for some arbitrary set S. However using the power of parallel shuttling we can perform this task as follows. Begin by defining the set $\hat{S}$ to be the complement of S in $\mathcal{R}$. Now we begin by performing the parallel shuttling operation

$$\texttt{HS[L]}, \quad \texttt{L} = \{(i, j, 1) \parallel (i, j) \in \hat{S}\}. \tag{10.7}$$

Here we abuse notation a bit by referring to $\hat{S}$ as the set of locations of the qubits in $\hat{S}$. This operation in effect moves all qubits in $\hat{S}$ out of $\mathcal{R}$ (and into $\mathcal{B}$, note that the dots the qubits are being shuttled in are always empty because of the definition of the idle configuration). Now we can use a semi-global single qubit rotation (as discussed in section 10.2.3) to perform an X-rotation on all qubits in $\mathcal{R}$, which is now just all qubits in the set S. This flips changes the states of the qubits in S from $|1\rangle$ to $|0\rangle$ without changing the state of any other qubit. Following this we can restore the BOARDSTATE to its original configuration by applying the parallel shuttling command

$$\texttt{HS[L]}, \quad \texttt{L} = \{(i, j, -1) \parallel (i, j) \in \hat{S}\}. \tag{10.8}$$

Now we have applied the required operation. Note that at no point we had to reason about the structure of the set S itself. This complexity was taken care of by the classical subroutines embedded in `HS[L]`. Next we discuss performing parallel two-qubit gates.

10.3.2. Parallel two-qubit gates

Similar to parallel shuttling it is in general rather involved to perform parallel two-qubit operations in the QDP. We can again define parallel versions of the OPCODES for two-qubit operations and then analyze how to perform them as parallel as possible (again having access to classical side computation).

OPCODE	Effect
`HI[L]`	Perform `VI[(i,j)]` for $(i, j) \in$ `L`
`VI[L]`	Perform `HI[(i,j)]` for $(i, j) \in$ `L`

Given an BOARDSTATE and a `HI[L]` command one could use an algorithm similar to the algorithm presented for shuttling. We can again construct a matrix F such that $F_{ij} = 1$ is for all tuples (i, j) in `L` indicating the locations where we desire a two-qubit

operation to happen and $F_{ij} = 0$ everywhere else. Now we can use the algorithm presented above for shuttling to decompose the matrix F into a series of parallel `HI[L]` operations. However, since we have $\mathrm{CHPASE}^2 = \mathbb{1}$ the independence subroutine reduces to linear independence of the columns of F modulo 2. This means we can find an *optimal* decomposition into parallel operations by finding the Schmidt-normal [22, Chapter 14] form of the matrix F (Note that we do have to 'tilt' the matrix F to account for the fact that as posed the diagonal lines of the matrix F are its 'rows'). We can make the same argument given a BOARDSTATE and a `VI[L]` command but now the Schmidt-normal form must be found modulo 4 as $(\sqrt{\mathrm{SWAP}})^4 = \mathbb{1}$. As both addition modulo 2 ($\mathbb{Z}_2$) defines a number field, so finding the Schmidt-normal form here is easy suing Gaussian elimination. Addition modulo 4 ($\mathbb{Z}_4$) is not a field so finding the Schmidt-normal form is trickier but is can still be done efficiently (with high probability of success), see for instance [23]. The depth of the sequence of operations is now proportional to the rank of the matrix F over $\mathbb{Z}_2$ (CPHASE) or $\mathbb{Z}_4$ ($\sqrt{\mathrm{SWAP}}$). However, as mentioned before, the parallel operation of two-qubit gates in the QDP will mean taking a hit in operation fidelity vis-a-vis the more controllable line-by-line operation [4]. Since this operation fidelity is typically a much larger error source than the waiting-time-induced decoherence stemming from line-by line operation we will for the remainder of the paper assume line-by-line operation of the two-qubit gates. This will have an impact when performing quantum error correction on the QDP which we will discuss in more detail in section 10.5.

For the sake of completeness we also define a parallel version of the CNOT OPCODE. The same considerations of parallel operation hold for the parallel use of CNOT gates as they hold for the CPHASE and $\sqrt{\mathrm{SWAP}}$ gates. We continue the discussion of parallelism in the QDP by analyzing parallel measurements.

OPCODE	Effect
VC[L]	Perform VC[(i,j)] for every (i, j) in L

10.3.3. Parallel Measurements

Performing measurements on an arbitrary subset of qubits on the QDP is in general quite involved. Every qubit to be measured requires an ancilla qubit and this ancilla qubit must be in a known computational basis state, and an empty dot must be adjacent as a reference for the readout process. The qubits must then be shuttled such that they are horizontally adjacent to their respective ancilla qubits and must also be located in such a way such that they are in the right columns for the PSB process to take place (revisit section 10.2.3 for more information). This can be done using the algorithm for parallel shuttling presented above but in the worst case this will take a sequence of depth $O(N)$ parallel shuttle operations. On top of the required shuttling the PSB process itself (from a control perspective similar to shuttling) must be performed in a way that depends on the BOARDSTATE and the configuration of the qubit/ancilla pairs. In general this PSB process will be performed line-by-line (for the fidelity reasons mentioned in the beginning of the section) and hence requires a sequence of depth $O(N)$ parallel grid operations (plus the amount of shuttling operations needed to attain the right measurement configuration in the first place). Due

to this complexity we will not analyze parallel measurement in detail but rather focus on a particular case relevant to the mapping of the surface code. But first we define a parallel measurement OPCODE `M[L]` which takes in a list of tuples (i, j, k) denoting locations of qubits to be measured (i, j) and whether the ancilla qubit is to the left $(k = -1)$ or to the right $(k = 1)$ of the qubit to be measured

OPCODE	Effect
`M[L]`	Perform `M[`(i, j, k)`]` for every (i, j, k) in `L`

A specific parallel measurement example

Let us consider a specific example of a parallel measurement procedure that will be used in our discussion of error correction. We begin by imagining the BOARDSTATE to be in the *idle* configuration (fig. 10.7 top left). We next perform the shuttle operations needed to change the BOARDSTATE to the *measurement* configuration. This configuration (and how to reach it by shuttling operations from the idle configuration) will be discussed section 10.3.4 and can be seen in fig. 10.7 (c). Next take the qubits to be measured in the parallel measurement operation to be the red qubits in fig. 10.7. The qubits directly to the right or to the left of those qubits will be the required readout ancillas (blue in fig. 10.7). We will assume that the readout ancillas are in the $|0\rangle$ state. If some ancilla qubits are in the $|1\rangle$ state instead we can always perform the procedure given in section 10.2.3 to rotate them to $|0\rangle$ without changing the state of the other qubits on the grid. Note that all the ancilla qubits are in the set $\mathcal{B}$ whereas the qubits to be read out are in the set $\mathcal{R}$. This means that we can perform the PSB process by attempting to shuttle the qubit to be measured (red) into the sites occupied by the ancilla qubits (blue). In principle we could perform this operations in parallel by executing the operations

$$\texttt{VS[L]}, \quad \texttt{L} = \{(i, j, 1) \parallel i = 0 \bmod 2, \;\; j = 1 \bmod 2, \, i + j = 1 \bmod 4\} \tag{10.9}$$

to bring the qubits to be measured (red) horizontally adjacent to the ancilla qubits (blue) and then

$$\texttt{M[L]}, \texttt{L} = \{(i, j, 1) \parallel i = 1 \bmod 4, \, j = 1 \bmod 4\} \tag{10.10}$$

and

$$\texttt{M[L]}, \texttt{L} = \{(i, j, -1) \parallel i = 3 \bmod 4, \, j = 3 \bmod 4\}. \tag{10.11}$$

All of these operations can be performed in a single time-step. However for fidelity and control reasons laid out in the beginning we would prefer to perform these operations in a line-by-line manner. In particular we would like to perform these operations one row at a time since this gives us the ability to control both diagonal and vertical lines individually for each measurement. However we must take care to avoid spurious operations. For instance when performing measurements on the qubits at locations $(1, 1)$ and $(1, 5)$ we must avoid also performing a measurement on the qubit at location $(5, 5)$. To avoid this

situation we will bring only the bottom row of qubits to be measured horizontally adjacent to the ancilla qubits, perform the PSB process and readout on that row only and then shuttle the qubits to be measured back down again. This we repeat going up in rows until we reach the end of the grid. More formally we perform the following sequence of operations.

Algorithm 4 Loop over OPCODES to perform line-by-line measurements

for $i \in [0 : N-2]$ **do**
  if $i = 1 \bmod 4$ **then**
    $\texttt{VS[L]}, \quad \texttt{L} = \{(i-1, j, -1) \parallel j = 1 \bmod 4\}$
    $\texttt{M[L]}, \quad \texttt{L} = \{(i, j, 1) \parallel j = 1 \bmod 4\}$
    $\texttt{VS[L]}, \quad \texttt{L} = \{(i-1, j, 1) \parallel j = 1 \bmod 4\}$
  end if
  if $i = 3 \bmod 4$ **then**
    $\texttt{VS[L]}, \quad \texttt{L} = \{(i-1, j, -1) \parallel j = 3 \bmod 4\}$
    $\texttt{M[L]}, \quad \texttt{L} = \{(i, j, -1) \parallel j = 3 \bmod 4\}$
    $\texttt{VS[L]}, \quad \texttt{L} = \{(i-1, j, 1) \parallel j = 3 \bmod 4\}$
  end if
end for

We will use this particular procedure when performing the readout step in a surface code error correction cycle in section 10.4.4. This concludes our discussion of parallel operation on the QDP. We now move on to highlight some BOARDSTATE configurations that will feature prominently in the surface and color code mappings.

10.3.4. Some useful grid configurations

There are several configurations of the BOARDSTATE that show up frequently enough (for instance in the error correction codes in section 10.4.4) to merit some special attention. In this section we list these specific configurations and show how to construct them.

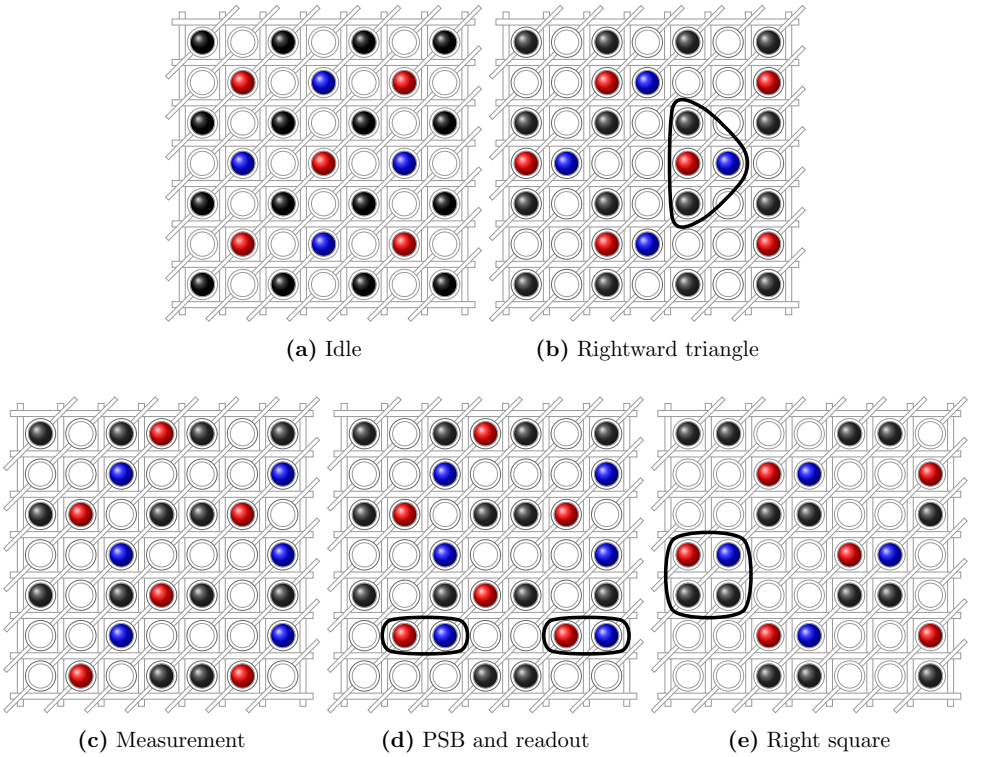
Idle configuration

The idle configuration is the configuration in which the QDP is initialized. As shown in fig. 10.7 it has a checkerboard pattern of filled and unfilled sites. In this configuration no two-qubit gates can be applied between any qubit pair but since it minimizes unwanted crosstalk between qubits [4], it is good practice to bring the system back to this configuration when not performing any operations. For this reason we consider the idle configuration to be the starting point for the construction of all other configurations.

Square configuration

As seen in fig. 10.7(e) the square configurations consist of alternating filled and unfilled 2×2 blocks of sites. The so-called right square configuration can be reached from the idle configuration by a shuttling operation $\texttt{HS[L]}$ with the set $\texttt{L}$ being

$$\begin{aligned} \texttt{L} = \{&(i, j, 1) \parallel i = 1 \bmod 2, j = 1 \bmod 2, \quad i + j = 2 \bmod 4\} \\ &\cup \{(i, j, -1) \parallel i = 0 \bmod 2, \quad j = 1 \bmod 2, i + j = 3 \bmod 4\}. \end{aligned} \tag{10.12}$$



(a) Idle **(b)** Rightward triangle

(c) Measurement **(d)** PSB and readout **(e)** Right square

Figure 10.7: Useful BOARDSTATE configurations. We denote data qubits with black color, X-measurement qubits by red and Z-measurement qubits by blue. Those will collect the parity of the data qubits in one error correction cycle, and one is the others reference at the PSB measurement. **(a)** The idle configuration is a starting point of all algorithms. All qubits are spread out and well separated. **(b)** The triangle configurations (here we have a rightward triangle, see the frame in the figure) is assumed when the proximity of measurement qubits to data qubits is required. This is the case for the parity measurements in error correction cycles. **(c)** The measurement configuration is formed to bring X- and Z-measurement qubits close to each other, such that a row can be selected in which the measurement is performed. **(d)** Certain measurement qubits are brought to adjacent dots in order to perform the PSB-based measurement and readout in a line-by-line fashion (encircled qubits). Since the rest of the grid is in the measurement configuration, individual control over the barrier lines and one potential is guaranteed without spurious measurements. **(e)** The (right) square configuration is a midway point between the idle and (right) triangle configuration. Going through the square configuration keeps the shuttling algorithm manageable, as not more that 2 different heights of the dot potentials are employed. One of the characteristic squares is framed in the figure.

Note that this operation only takes a single time-step, the square configuration is shown in fig. 10.7(e). The right square configuration is characterized by the red (Z-) ancilla being in the left corner of every square. Another flavor of this configuration is the left square configuration, where the red ancilla is in the upper right corner, and the blue one in the left. The left square configuration can be reached from the idle configuration by a shuttling operation HS [L] with the set L being .

$$\begin{aligned} \mathrm{L} = \{(i,j,1) \parallel i = 0 \bmod 2,\ \ j = 0 \bmod 2,\, i+j = 2 \bmod 4\} \\ \cup \{(i,j,-1) \parallel i = 1 \bmod 2,\ \ j = 0 \bmod 2,\, i+j = 1 \bmod 4\}. \end{aligned} \tag{10.13}$$

These configurations are used as an intermediate step for us to reach the triangle configurations.

Measurement Configuration

The measurement configuration can be reached from the idle configuration in three time-steps by the following sequence of parallel shuttling operations.

$$\begin{aligned} &\mathrm{HS}[\mathrm{A}], \quad \mathrm{A} = \{(i,j,-1),\ (i-1,j-1,1)\ \parallel i = 1 \bmod 4,\ j = 2 \bmod 4\}, \\ &\mathrm{HS}[\mathrm{B}], \quad \mathrm{B} = \{(i-1,j-1,1)\ \parallel i = 3 \bmod 4,\ j = 1 \bmod 4\}, \\ &\mathrm{VS}[\mathrm{C}], \quad \mathrm{C} = \{(i,j,-1) \parallel i = 0 \bmod 2,\ \ j = 1 \bmod 2,\ i+j = 1 \bmod 4\}. \end{aligned} \tag{10.14}$$

This configuration can be seen in fig. 10.7(d) and it is an intermediate state in the measurement process of the blue qubits using the red qubits as ancillas. How this measurement protocol works in detail is described in section 10.3.3.

Triangle configurations

In order to collect the parity of the data qubits in the error correction cycles, we need to align the ancilla qubits with the data qubits, according to the two-qubit gates used. This is reflected in the use of triangle configurations. There are two triangle configurations that can be reached in a single parallel shuttling step from the right square configuration. The first one, seen in fig. 10.7(b), is called the rightward triangle configuration. It can be reached from the square configuration by the grid operation HS [L] with the set L being

$$\mathrm{L} = \{(i,j,-1) \parallel 0 = 1 \bmod 2, \quad j = 1 \bmod 2, i+j = 3 \bmod 4\}, \tag{10.15}$$

which does as much as to shuttle the right data qubit of every square (framed squares in fig. 10.7(e)) to the empty dot on its right. In this configuration, we are able to perform high-fidelity two-qubit gates between the two data qubits and the ancilla in every triangle. In order to reach the neighboring pair of data qubits with the same ancilla, we start from the left square configuration and shuttle the left data qubit to the left. Operationally, we would do HS [L] with

$$\mathrm{L} = \{(i,j,1)\ \parallel\ i = 0 \bmod 2,\ \ j = 0 \bmod 2,\, i+j = 2 \bmod 4\}. \tag{10.16}$$

Note again that these parallel shuttling operations can be performed in a single time step. From these configurations the idle configuration can also be reached in a single time step. In the next section these configurations will feature prominently in the mapping of several quantum error correction codes to the QDP architecture.

10.4. Error correction codes

In this section we will apply the techniques we developed in the previous sections to map several quantum error correction codes to the QDP.

10.4.1. Introduction

First we recall some basic facts about quantum error correction codes and topological stabilizer codes in particular. The focus will be on practical application, for a more in depth treatment of quantum error correction and topological error correction codes we refer to [7]. Recall first the Pauli operators on a single qubit:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{10.17}$$

Given a system of n qubits we denote by P_i the Pauli operator $P \in \{X, Z\}$ acting on the i'th qubit. With this definition we can see write the n qubit Pauli group $\mathcal{P}_n$ as the group generated by the operators $\{X_i, Z_j \ : \ i, j \in [1 : n]\}$ under matrix multiplication. A stabilizer quantum error correction code acting on n physical qubits and encoding k logical qubits can then be defined as the joint positive eigenspace of an abelian subgroup $\mathcal{S}$ of $\mathcal{P}_n$ generated by $n - k$ independent commuting Pauli operators. Operationally, this code is then defined by measuring the generators of $\mathcal{S}$ and if necessary perform corrections to bring the state of the system back into the positive joint eigenspace of these generators. This is a very general definition and it is not guaranteed that a code defined this way

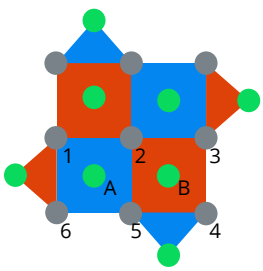


Figure 10.8: Schematic representation of a distance three rotated planar surface code [24]. The gray circles represent the data qubits supporting the code. The green circles represent ancilla qubits, which are used to perform the stabilizer measurements which define the code. These stabilizer measurements are represented by the red (Z-type stabilizers) and blue faces (X-type stabilizers). The ancilla qubit in the middle of a face will be used to perform a stabilizer measurement of the data qubits on the corners of that face. The actual quantum circuits used to perform these stabilizer measurements are shown in fig. 10.9.

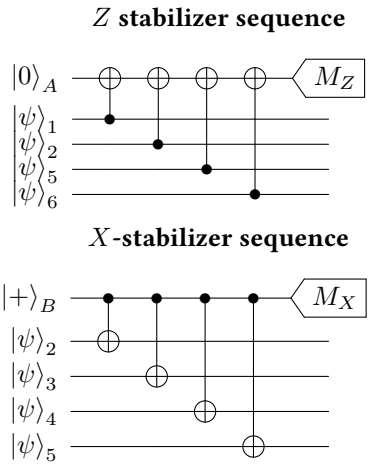


Figure 10.9: Quantum circuits for performing the X- and Z-stabilizer measurements of the planar surface code [6, 7, 25, 26]. The qubits A and B (see fig. 10.8) are ancilla qubits used to perform stabilizer measurements on the the data qubits on the corners of the faces defining the code. The data qubits associated to the face of qubit A are $\{1, 2, 5, 6\}$ and likewise $\{2, 3, 4, 5\}$ for qubit B.

yields any protection against errors happening. Below we will see some common examples of stabilizer error correction codes that do have good protection against errors. On top of that, these codes have the desirable property that their stabilizers are in some sense 'local'. That is they can be implemented on qubits lying on a lattice such that the stabilizer generators can be measured by entangling a patch of qubits that is small with respect to the total lattice size. The most well known example of a code of this type is the so-called planar surface code.

10.4.2. PLANAR SURFACE CODE

The planar surface code is probably the most well known practical quantum error correction code due to its high threshold [27], the availability of efficient decoding algorithms [28]. To construct the planar surface code (in particular we will use the so-called rotated planar surface code [24], as it uses less physical qubits per logical qubit) we will consider a regular $n \times n$ square lattice of degree four (every node has four connected neighbors) and we will place qubits on each node. We will define the generators of the abelian group $\mathcal{C}$ that defines the surface code by alternately placing X- and Z-quartets on the faces of the lattice (in fig. 10.8 the red faces correspond to X-stabilizer quarters while the green faces correspond to Z stabilizer quartets). This $X(Z)$ will indicate that we pick the generator $X^{\otimes 4}$ ($Z^{\otimes 4}$) on the four qubits on the corners of the X (Z) face. Note that this means that all of the generators commute with each other since they either act on disjoint sets of qubits or act on sets that have an overlap of exactly 2 qubits. Since $XZ = -ZX$ we have that $X^{\otimes 2}Z^{\otimes 2} = Z^{\otimes 2}X^{\otimes 2}$ which means that all generators commute. These generators (plus appropriate generators on the boundary of the lattice) define a stabilizer group which specifies a code space of dimension 2, i.e. a single logical qubit. We can locally measure these $X(Z)$ stabilizers by using the circuits [6, 7, 25, 26] illustrated in fig. 10.9. This construction calls for one ancilla qubit per lattice face.

10.4.3. 2D COLOR CODES

Another important class of planar topological codes are the 2D color codes [9]. These codes are defined on 3-colorable tilings of the Euclidean plane. Two popular tilings are the so called 6.6.6. and 4.8.8. tilings corresponding to hexagonal and square-octagonal

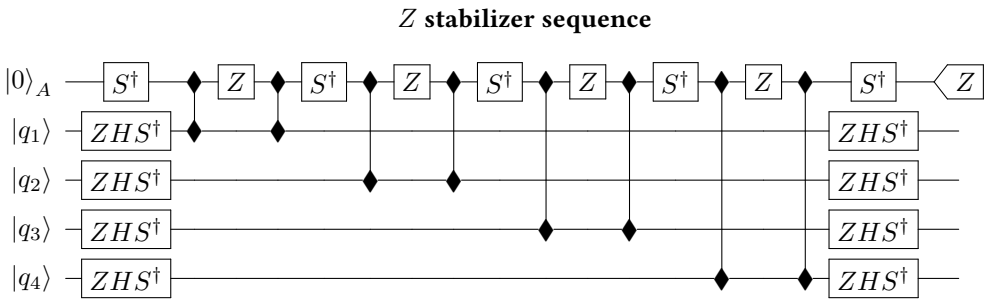


Figure 10.10: Z stabilizer measurement circuit using the $\sqrt{\text{SWAP}}$ as the main two-qubit gate. The Z- and S-rotations can be performed by the timing procedure described in section 10.2.3.

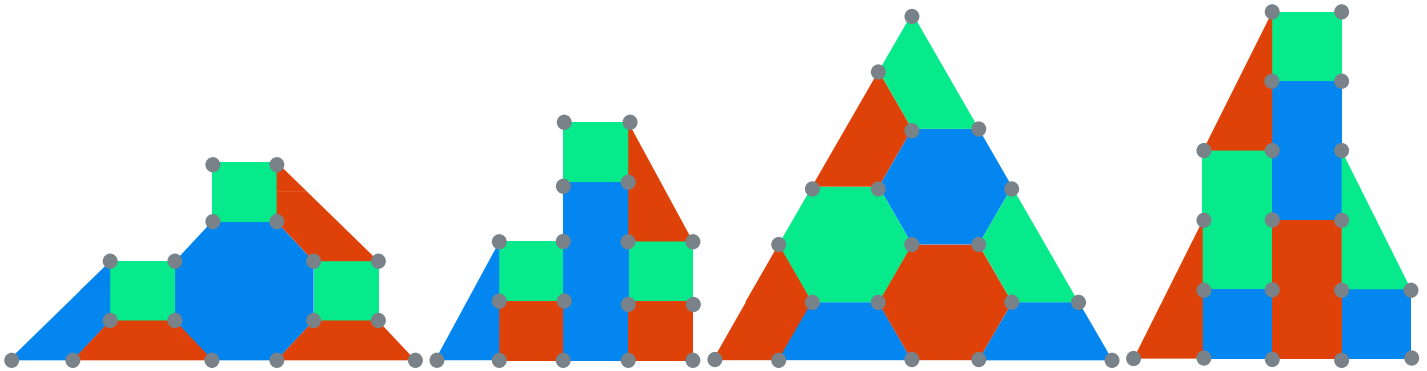


Figure 10.11: Distance 5 examples of the 4.8.8. (first from left) and 6.6.6. (third from left) color codes [9] and their deformed versions (second from left and fourth from left respectively). The vertices correspond to data qubits and every colored face corresponds to both an X- and a Z-stabilizer to be measured. These stabilizers can be measured by using weight 4, 6 and 8 versions of the circuits shown in fig. 10.9. The deformation of the codes does not change the code properties at all. They are a visual guide that facilitates the mapping the the crossbar grid in section 10.4.5.

tilings respectively. To construct the code qubits are places on all vertices of the tiling and X- and Z-stabilizers are associated to every tile by applying X (Z) to every qubit on the corner of the tile. With suitable boundary conditions this construction encodes a single logical qubit with a distance proportional to $\sqrt{n}$ with n the number of physical qubits. See fig. 10.11 for examples of the $6.6.6.$ and $4.8.8.$ color codes of distance five. Note that these pictures do not include ancilla qubits for measuring the stabilizers. The planar color codes have lower thresholds than the planar surface code but are more versatile when it comes to fault-tolerant gates. The planar color codes support the full Clifford group as a transversal set, making quantum computation on color codes more efficient than on the surface code. In the next section we will focus on mapping these codes to the QDP using the concepts introduced in section 10.3.

10.4.4. Surface code mapping

We now describe a protocol that maps the surface code on the architecture described in section 10.2. The surface code layout has a straightforward mapping that places the data qubits on the even numbered columns and the X- and Z-ancillas on the odd columns. This means we have single-qubit control over all data qubits and all ancilla qubits separately. There are two ways to perform the surface code cycle; we could use either the $\sqrt{\text{SWAP}}$ gate or the CPHASE gate as the main two-qubit gate. Since in practice the $\sqrt{\text{SWAP}}$
gate has higher fidelity [4] we will use this gate. We begin by changing the circuits performing the X- and Z-stabilizer measurements to work with $\sqrt{\text{SWAP}}$ rather than CNOT. We can emulate a CNOT gate by using two $\sqrt{\text{SWAP}}$ gates interspersed with a Z-gate on the control plus some single qubit gates. As described in section 10.2.3 the Z- and S-gates on the ancilla qubit can performed by waiting, which means they can be performed locally while the single qubit operations on the data qubits can be performed in parallel using the global unitary rotations described in section 10.2.3. The X- and Z-circuits using $\sqrt{\text{SWAP}}$ are shown in fig. 10.10.

We will split up the quantum error correction cycle by first performing all X-type stabilizers (the X-cycle) and then all Z-type stabilizers (Z-cycle). This means we can use the idle Z- (X-) ancilla to perform a measurement on the X- (Z-) ancilla at the end of the X

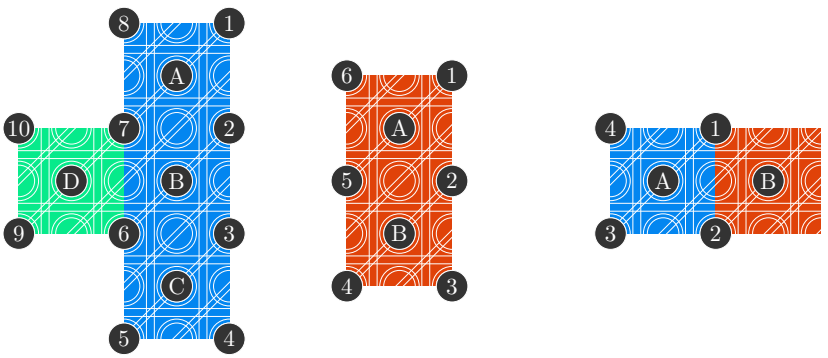


Figure 10.12: Unit cells of the deformed 4.8.8. and 6.6.6. codes (left and middle respectively) and the unit cell of the surface code (Z-cycle) with the gray circle corresponding to qubits. For the 4.8.8. unit cell the qubit labeled 'A' is the ancilla qubit for the octagon (now a rectangle) sub-cell while the qubit labeled 'D' is the ancilla for the square sub-cell. The qubit labeled 'B' is used to read out the qubit labeled 'D' and the qubit labeled 'C' is used to read out the ancilla qubit for the octagon cell directly below the square cell (not pictured). The qubits labeled by numbers are the data qubits. For the 6.6.6. unit cell the qubit labeled 'A' is the ancilla qubit used to perform the stabilizer measurement while the qubit labeled 'B' is used to read out the 'A' qubit for the unit cell directly to the bottom left (not pictured). The numbered qubits are again data qubits. For the surface code unit cell the qubit labeled 'A' is the ancilla used for the Z-cycle stabilizer measurement while the qubit labeled 'B' is the qubit used to read out the 'A' qubit. It is also the qubit used as the ancilla for the X-stabilizer cycle. The numbered qubits are again data qubits. Note that this unit cell mirrors when moving upwards. That is, the unit cell above the one pictured will have the ancilla qubit B to the right of qubit A instead of to the left as pictured.

(Z) cycle. For convenience we included a depiction of the surface code Z-cycle unit cell in fig. 10.12 (right). The qubit labeled 'A' is the ancilla used for the Z stabilizer circuit. The numbered qubits are data qubits and the qubit labeled 'B' is the qubit used for reading out the 'A' qubit. It is also the ancilla qubit for the X-cycle. We now describe the steps needed to perform the Z-cycle in parallel on the entire surface code sheet. For convenience we ignore the surface code boundary conditions but these can be easily included. The X-cycle is equivalent up to different single qubit gates ($XS^\dagger$ instead of $ZHS^\dagger$ on the data qubits, $HS^\dagger$ instead of $S^\dagger$ on the ancilla) and shifting every operation 2 steps up, e.g. setting i to $i+2$.

The surface code Z-cycle

1. Initialize in the idle configuration
2. Apply $ZHS^\dagger$ to all qubits in $\mathcal{R}$ (data) and $S^\dagger$ to qubits in $\mathcal{B}$ (ancilla)
3. Go to right square configuration
4. Go to rightward triangle configuration
5. Perform CNOT between qubits A and 1 by performing VC[L] with
$$\mathrm{L} = \{(i,j) \parallel i = 1 \bmod 2, j = 0 \bmod 2, i+j = 3 \bmod 4\}$$
6. Perform CNOT between qubits A and 2 by performing VC[L] with
$$\mathrm{L} = \{(i,j) \parallel i = 0 \bmod 2, j = 0 \bmod 2, i+j = 2 \bmod 4\}$$
7. Go to idle configuration
8. Go to left square configuration
9. Go to leftward triangle configuration
10. Perform CNOT between qubits A and 3 by performing VC[L] with
$$\mathrm{L} = \{(i,j) \parallel i = 1 \bmod 2, j = 0 \bmod 2, i+j = 1 \bmod 4\}$$
11. Perform CNOT between qubits A and 4 by performing VC[L] with
$$\mathrm{L} = \{(i,j) \parallel i = 0 \bmod 2, j = 0 \bmod 2, i+j = 0 \bmod 4\}$$
12. Go to idle configuration
13. Apply $ZHS^\dagger$ to all qubits in $\mathcal{R}$ (data) and $S^\dagger$ to qubits in $\mathcal{B}$ (ancilla)
14. Apply measurement ancilla correction step for qubit B as described in section 10.3.1
15. Go to measurement configuration
16. Perform Pauli Spin Blockade measurement process as described in section 10.3.3 using qubit B as ancilla to qubit A
17. Go to idle configuration

10.4.5. COLOR CODE MAPPING

The mapping of the color codes is largely analogous to that of the surface code. We begin with the 6.6.6. color code as it is easiest to map. We begin by deforming the tiling on which the color code is defined such that it is more amenable to the square grid struc-

ture of the QDP. This is fairly straightforward as can be seen from the $d = 5$ example in fig. 10.11. In the deformed tiling it is clear how to map the code to the crossbar grid layout. We once again place all data qubits in the even columns and all ancilla qubits in the odd columns. This places the unit 'hexagon' seen in the deformed code in a 3×5 tile on the QDP (see fig. 10.12 (right) for this unit tile). This places all data qubits in $\mathcal{R}$ and 2 extra qubits in $\mathcal{B}$, both of which could be used as an ancilla in the stabilizer circuit. We will always choose the top qubit (qubit 'A') of these two in the hexagon unit cell as the ancilla qubit for the error correction cycles. The extra (bottom) qubit (qubit 'B') in the unit cell will be used to perform the readout of the ancilla qubit of the unit hexagon to its direct left. This has the advantage of making the readout process independent of the measurement results of the previous cycles (as was the case in the surface code). Note also that the ancilla qubits are positioned along diagonal lines on the QDP grid. This makes the quantum error correction cycle very analogous to the surface code. We once again must split up the X- and Z-cycles (again due to the limited single qubit rotations possible). Below we present the steps needed to perform the Z-cycle (which now measures a weight 6 operator). The X-cycle is identical up to differing single qubit rotations on the data qubits.

The 6.6.6 color code Z-cycle

1. Perform **Steps 1 to 11** in the surface code Z-cycle to perform CNOT s between the ancilla (qubit A) and the data qubits $1, 2, 5, 6$ in the unit hexagon and end in the idle configuration
2. Go to idle configuration but with all even columns up and all odd columns down by performing VS [L] with
$$\begin{aligned} \mathrm{L} = \{(i,j,1) \parallel\ & i = 0 \bmod 2, j = 0 \bmod 2\} \\ & \cup \{(i,j,-1) \parallel\ i = 1 \bmod 2, j = 1 \bmod 2\} \end{aligned}$$
3. Go to right square configuration
4. Go to rightward triangle configuration
5. Perform CNOT between qubits A and 3 by performing VC [L] with
$$\mathrm{L} = \{(i,j) \parallel\ i = 1 \bmod 2, j = 0 \bmod 2, i+j = 1 \bmod 4\}$$
6. Go to idle configuration
7. Go to left square configuration
8. Go to leftward triangle
9. Perform CNOT by performing between qubits A and 4 VC [L] with
$$\mathrm{L} = \{(i,j) \parallel\ i = 0 \bmod 2, j = 0 \bmod 2, i+j = 2 \bmod 4\}$$
10. Go to idle configuration
11. Invert **Step 6** by performing VS [L] with
$$\begin{aligned} \mathrm{L} = \{(i,j,-1) \parallel\ & i = 0 \bmod 2, j = 0 \bmod 2\} \\ & \cup \{(i,j,1) \parallel\ i = 1 \bmod 2, j = 1 \bmod 2\} \end{aligned}$$
12. Apply $ZHS^\dagger$ to all qubits in $\mathcal{R}$ (data) and $S^\dagger$ to qubits in $\mathcal{B}$ (ancilla)
13. Go to measurement configuration
14. Perform Pauli Spin Blockade measurement process as described in section 10.3.3 using qubit B as ancilla to read out qubit A (unit cell to the right)
15. Go to idle configuration

Next up is the $4.8.8.$ color code. We deform the tiling on which the code is defined similarly to the $6.6.6.$ code. The deformed $4.8.8.$ code lattice can be seen in fig. 10.12 (left). We again place the data qubits in the set $\mathcal{R}$ the ancilla qubits in the set $\mathcal{B}$. See fig. 10.12 for a layout of the unit cell of the $4.8.8.$ code on the QDP. Note that there are two different types of tiles in this code. The square tile has one qubit (qubit 'D' in fig. 10.12) in $\mathcal{B}$, which we will use as ancilla qubit for that tile. The deformed octagon tile has three qubits in $\mathcal{B}$. We will use the topmost qubit (qubit 'A') as the ancilla qubit for the tile while the middle one (qubit 'B') serves as the readout qubit for the square tile ancilla directly to its left and the bottommost one (qubit 'C') will be used to perform the readout of the octagon directly below the square tile (not pictured). Because the structure of the $4.8.8.$ code is less amenable to direct mapping the stepping process is a little more complicated. We will again only write down the Z-cycle with the X-cycle being the same up to initial and final single qubit rotations on the data qubits.

The $4.8.8$ color code Z-cycle

1. Initialize in the idle configuration
2. Apply $ZHS^\dagger$ to all qubits in $\mathcal{R}$ (data) and $S^\dagger$ to qubits in $\mathcal{B}$ (ancilla)
3. Go to right square configuration
4. Go to rightward triangle configuration
5. Perform CNOT between qubits A and 1 and D and 7 by performing VC[L] with
$$\mathrm{L} = \{(i,j) \parallel i = 1 \bmod 2, j = 0 \bmod 2, [i+j = 3 \vee 7 \bmod 16]\}$$
6. Perform CNOT between qubits A and 2 and d and 6 by performing VC[L] with
$$\mathrm{L} = \{(i,j) \parallel i = 0 \bmod 2, j = 0 \bmod 2, [i+j = 2 \vee 6 \bmod 16]\}$$
7. Go to left square configuration
8. Go to left triangle configuration
9. Perform CNOT between qubits A and 8 and D and 9 by performing VC[L] with
$$\mathrm{L} = \{(i,j) \parallel i = 1 \bmod 2, j = 0 \bmod 2, [i+j = 1 \vee 5 \bmod 16]\}$$
10. Perform CNOT between qubits A and 7 and d and 10 by performing VC[L] with
$$\mathrm{L} = \{(i,j) \parallel i = 0 \bmod 2, j = 0 \bmod 2, [i+j = 0 \vee 4 \bmod 16]\}$$
11. Go to idle configuration
12. Go to idle configuration but with all even columns up and all odd columns down by performing VS[L] with
$$\mathrm{L} = \{(i,j,1) \parallel i = 0 \bmod 2, j = 0 \bmod 2\} \cup \{(i,j,-1) \parallel i = 1 \bmod 2, j = 1 \bmod 2\}$$
13. Go to right square configuration
14. Go to rightward triangle configuration
15. Perform CNOT between qubits A and 3 by performing VC[L] with
$$\mathrm{L} = \{(i,j) \parallel i = 1 \bmod 2, j = 0 \bmod 2, i+j = 3 \bmod 16\}$$
16. Perform CNOT between qubits A and 4 by performing VC[L] with
$$\mathrm{L} = \{(i,j) \parallel i = 0 \bmod 2, j = 0 \bmod 2, i+j = 2 \bmod 16\}$$
17. Go to idle configuration
18. Go to left square configuration
19. Go to leftward triangle configuration
20. Perform CNOT between qubits A and 6 by performing VC[L] with
$$\mathrm{L} = \{(i,j) \parallel i = 1 \bmod 2, j = 0 \bmod 2, i+j = 1 \bmod 16\}$$
21. Perform CNOT between qubits A and 5 by performing VC[L] with
$$\mathrm{L} = \{(i,j) \parallel i = 0 \bmod 2, j = 0 \bmod 2, i+j = 0 \bmod 16\}$$
22. Go to idle configuration
23. Invert **Step 6** by performing VS[L] with
$$\mathrm{L} = \{(i,j,-1) \parallel i = 0 \bmod 2, j = 0 \bmod 2\} \cup \{(i,j,1) \parallel i = 1 \bmod 2, j = 1 \bmod 2\}$$
24. Repeat **Steps 2-23** but setting i to $i+2$ and j to $j+1$
25. Apply $ZHS^\dagger$ to all qubits in $\mathcal{R}$ (data) and $S^\dagger$ to qubits in $\mathcal{B}$ (ancilla)
26. Go to measurement configuration
27. Perform Pauli Spin Blockade measurement process as described in section 10.3.3 using qubit B (unit cell to the right) as ancilla for qubit A and using qubit C as ancilla for qubit D
28. Go to idle configuration

10.5. Discussion

In this section we evaluate the mapping of the error corrections codes described above and argue numerically that it is possible to attain the error suppression needed for practical universal quantum computing. We will do this exercise for the planar surface code, as it is the most popular and best understood error correction code. The description given in section 10.4.4 assumes that all operations can be implemented perfectly in parallel. In practice though, for the reasons outlined in section 10.3 many operations that can in principle be done in parallel will be done in a line-by-line fashion. Note that for surface code in an array like this, the length of a quadratic grid scales linearly with the code distance as $N = 2d + 1$. This means that the time performing a surface code cycle and thus the number of errors affecting a logical qubit rises linearly with the code distance and hence this mapping of the surface code will not exhibit an error correction threshold. As a consequence the error probability of the encoded qubit (the logical error probability) cannot be made arbitrarily small but rather will exhibit a minimum for some particular code distance after which the logical error probability will start rising with increasing code distance. The code distance which minimizes the error will depend non-trivially on the error probability of the code qubits. This is not a very satisfactory situation from a theoretical point of view, but from the point of view of practical quantum computation we are not so much interested in asymptotic statements but rather if the logical error probability can be made small enough to allow for realistic computation [26]. As a target logical error probability we choose $P_L = 10^{-20}$ as at this point the computation is essentially error free (for comparison, a modern classical processor has an error probability around 10^{-19} [11]). We will use this number as a benchmark to assess if and for what error parameters the surface code mapping in the QDP yields a "practical" logical qubit. In order to assess this we must consider in more detail the sources of error afflicting the surface code operation on the QDP. We will begin by detailing how the surface code is likely to be implemented in practice on the QDP and afterward we will consider how this impacts the error behavior of the logical surface code qubit. We will distinguish two classes of error sources: operation induced errors and decoherence induced errors.

10.5.1. Practical implementation of the surface code

Here we present an mapping of the surface code based on the one presented in section 10.4.4 but differing in the amount of time-steps used to perform certain operations. In particular we choose to do all shuttle and two-qubit-gate operations in a line-by-line manner. This is a specific choice which we expect will work well but variations of this protocol are certainly possible. As mentioned above this will mean that the time an error correction cycler takes will scale will the code distance. This means it is important to keep careful track of the time needed to perform a cycle. We will do this while describing line-by-line operation of the surface code cycle in greater detail below.

In practice we will perform the protocol in section 10.4.4 in the following manner. We begin by performing step 1 and 2 for all qubits. Then we apply steps $3 - 7$ but only to the data and ancilla qubits in the columns 0 and 1. Note that after performing these steps on only the first two columns we are back in the *idle* configuration. Now we repeat the previous for columns 2 and 3 and so forth until we reach the end of the code surface. Hav-

Steps	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	X-Z cycle average	Full cycle total
$\sqrt{\text{SWAP}}$ gate					$2d$	$2d$				$2d$	$2d$							$8d$	$16d$
Z rotation					$2d$	$2d$				$2d$	d							$7d$	$14d$
Shuttling			d	d			d	d	d			d			$5d$	$2d$	$3d$	$16d$	$32d$
Global rotation		1											1		1			3	6
Measurement																d		d	$2d$

Table 10.4: Time-step count per step in terms of different types of possible gates for the line-by-line implementation of the surface code cycle described in section 10.4.4. The number of time-steps is quoted in terms of the code distance d. this table does not specify the exact order in which the operations happen, see section 10.5.1 for an explanation of the time flow. Note that the table shows the average of the time-step counts for the X- and Z-cycles. The actual time count for the individual X- and Z-cycles is slightly different due to the boundary conditions of the surface code. The exact count for the Z-cycle can be obtained by replacing d by $d-1$ in every entry (except for the last column) whereas the exact count for the X-cycle is obtained by replacing d with $d+1$ in every column bar the last one. Since $(d+1)+(d-1)=2d=d+d$ this makes no difference for the full cycle count. Table cells that are left empty signify zero entries.

ing done these operations we are at the end of step 7 (go to *idle* configuration) and the grid is the *idle* configuration. We now repeat the same process to perform steps $8-12$ of section 10.4.4. Next we perform step 13 which can be done globally. Hereafter we perform step 14 (ancilla correction) in standard line-by-line fashion. Note that even in an ideal implementation step 14 has to be done line-by-line in the worst case. After this we perform step 15 (go to *measurement* configuration) in a line-by line manner and similarly for steps 16 (PSB/readout procedure) and 17 (go to *idle* configuration).

Note that in this line-by-line implementation there is a slight asymmetry between the X- and Z-cycles. Due to the boundary conditions of the surface code the X-cycle will involve $d+1$ columns pairs whereas the Z-cycle will involve $d-1$ column pairs. However since $(d+1)+(d-1)=2d$ this is mathematically equivalent to saying that the average cycle involves d column pairs. With this understanding we quite in table 10.4 how many time-steps every step in section 10.4.4 takes (split up by gates involved in that step) in this particular implementation of the protocol. Note that in this table we do not specify the order in which the operations happen, only to which step they are associated. We also calculate the amount of time-steps (for different gate types) needed for the full surface code error correction cycle.

10.5.2. Decoherence induced errors

Decoherence induced errors are introduced into the computation by uncontrolled physical processes in the underlying system. The effect of these processes is called decoherence. Decoherence happens even if a qubit is not being operated upon and the amount of decoherence happening during a computation scales with the time that computation takes. Therefore, to account for decoherence induced errors during the error correction cycle we need to compute how long an error correction cycle takes. Generally any operation on the QDP takes a certain amount of time denoted by τ. We distinguish again five different operations: (1) two-qubit $\sqrt{\text{SWAP}}$ gates, (2) qubit shuttle operations, (3) single qubit Z gates by waiting, (4) global single qubit operations, (5) qubit measurements. The time they take we will denote by $\tau_{sw}, \tau_{sh}, \tau_z, \tau_{gl}$ and τ_m respectively. In table 10.4 we performed a count of the total time taken by the surface code error correction cycle using the mapping

described in sections 10.4.4 and 10.5.1. The table below summarizes the total number of time-steps for every gate type for a full surface code error correction cycle.

Symbol	Operation	time-steps per cycle
τ_{sw}	$\sqrt{\text{SWAP}}$ gate	$16d$
τ_{sh}	Shuttling	$32d$
τ_z	Z rotation by waiting	$14d$
τ_{gl}	Global qubit rotation	6
τ_m	Measurement	$2d$

We can now say the total time $\tau_{\text{total}}(d)$ as a function of the code distance d is given by

$$\begin{aligned}\tau_{\text{total}}(d) = 16d\tau_{sw} + 32d\tau_{sh} \\ + 14d\tau_z + 6\tau_{gl} + 2d\tau_m.\end{aligned} \tag{10.18}$$

This total time can be connected to an error probability by invoking the mean decoherence time of the qubits in the system, the so called T_2 time [16, 29] (We ignore the influence of T_1 in this calculation as it is typically much larger than T_2 in silicon spin qubits [4, 30]). We can find the decoherence induced error probability P_{dec} [16, Page 384] as

$$P_{dec}(d) = \frac{\tau_{\text{total}}(d)}{2T_2}. \tag{10.19}$$

Next we investigate operation induced errors. These will typically be larger than decoherence induced errors but will not scale with the distance of the code.

10.5.3. Operation induced errors

Operation induced errors are caused by imperfect application of quantum operations to the qubit states. There are five operations performed on qubits in the surface code cycle. These are: (1) two-qubit $\sqrt{\text{SWAP}}$ gates, (2) qubit shuttle operations, (3) single qubit Z gates by waiting, (4) global single qubit operations, (5) qubit measurements. We will denote the probability of an error afflicting these operations by $P_{sw}, P_{sh}, P_z, P_{gl}$ and P_m respectively. In table 10.5 we list the total number of gates of a given type a data qubit and an ancilla qubit participate in over the course of a surface code cycle. In section 10.8 we give a more detailed per-step overview of the operations performed on data qubits and ancilla qubits. For clarity we have chosen qubit 1 in fig. 10.12 (right) as a representative of the data qubits and qubit A in fig. 10.12 (right) as a representative of the ancilla qubits. Other qubits in the code might have a different ordering of operations but their gate counts will be the same, except for the qubits located at the boundary of the code which will have a strictly lower gate count (we can thus upper bound their operation induced errors by those of the representative qubits). For each gate we also calculate the average number of this gate data and ancilla qubits participate in. This average number will serve as our measure of operation induced error.

10.5.4. Surface code logical error probability

By tallying up the contributions from operational and decoherence induced errors we can construct a measure for the total error probability per QEC cycle experienced by all

	Data qubit			Z ancilla qubit			Average data/ancilla
	Z-cycle	X-cycle	Total	Z-cycle	X-cycle	Total	
$\sqrt{\text{SWAP}}$ gate	4	4	8	8	0	8	8
Z rotation	0	0	0	7	0	7	3.5
Shuttling	2	4	6	10	4	14	10
Global rotation	2	2	4	2	3	5	4.5
Measurement	0	0	0	1	1	2	1

Table 10.5: This table lists the total number of gates of a given type a data qubit and an ancilla qubit participate in over the course of a surface code cycle. In section 10.8 we give a more detailed per-step overview of the operations performed on data qubits and ancilla qubits. For clarity we have chosen qubit 1 in fig. 10.12 (right) as a representative of the data qubits and qubit A in fig. 10.12 (right) as a representative of the ancilla qubits. Other qubits in the code might have a different ordering of operations but there gate counts will be the same, except for the qubits located at the boundary of the code which will have a strictly lower gate count (we can thus upper bound their operation induced errors by those of the representative qubits).

physical qubits that make up the code. Note that this a rather crude model that disregards possible influences from inter-qubit correlated errors and time-like correlated errors. Nevertheless it serves as a useful first approximation to the performance of the surface code on the QDP. We define the average per qubit per cycle error probability P_{tot} as

$$\begin{aligned} P_{\text{tot}}(d) = 8P_{sw} + 3.5P_{sh} + 10P_z \\ + 4.5P_{gl} + P_m + P_{dec}(d). \end{aligned} \tag{10.20}$$

Note that this quantity depends linearly on the code distance d. We can plug this total per cycle error probability P_{tot} into an empirical equation for the logical error probability P_L derived in [26].

$$P_L = 0.03 \left(\frac{P_{tot}(d)}{8P_{th}} \right)^{\frac{d+1}{2}} \tag{10.21}$$

where P_{th} is the per-step fault-tolerance threshold of the surface code, which we take to be $P_{th} = 0.0057$ following the result in [26]. The factor of 8 is inserted to account for the fact that the empirical relation derived in [26] is between the physical *per-step* error rate and the logical *per cycle* error rate and the protocol analyzed in [26] requires 8 time-steps per surface code error correction cycle. This is an approximation but it will serve our purposes of getting a basic initial estimate of the logical error rate. The next step is to start plugging in experimental numbers into equation eq. (10.20). In the table below we quote error probabilities and operation times for all relevant parameters. These numbers are projections from [4] and references therein. To convert the operation times

Operation	Error probability	Time
two-qubit $\sqrt{\text{SWAP}}$ gate	$P_{sw} = 10^{-3}$	$\tau_{sw} = 20\text{ns}$
qubit shuttle	$P_{sh} = 10^{-3}$	$\tau_{sh} = 10\text{ns}$
Z rotation by waiting	$P_z = 10^{-3}$	$\tau_z = 100\text{ns}$
global qubit rotation	$P_{gl} = 10^{-3}$	$\tau_{gl} = 1000\text{ns}$
measurement	$P_m = 10^{-3}$	$\tau_m = 100\text{ns}$

into decoherence induced error we use the estimated T_2 time of quantum dot spin qubits in

^{28}Si quoted as $T_2 = 10^9$ns [4, 30] and eq. (10.19). Plugging these numbers into eq. (10.20) we get the following linear function of the code distance

$$P_{tot} = 2.7 \times 10^{-2} + 2.8d \times 10^{-5} \tag{10.22}$$

which we can plug into the empirical model eq. (10.21). In fig. 10.13 we plot the logical error probability P_L versus code distance. Note that for the experimental numbers provided the practical quantum computing benchmarking $\log(P_L) = -20$ is reached for a code distance of $d = 37$. The maximal code distance for the experimental parameters is $d = 155$ for which the log-logical error probability reaches $\log(P_L) = -41$, after which it starts increasing again. For completeness we have also plotted what would happen if we had the power to operate the QDP (with quoted device parameters) completely in parallel. We estimate the physical per cycle error rate of this situation by setting $d = 1$ in eq. (10.22). Note that the difference between parallel and crossbar style operation is not that big, the parallel version reaches $P_L = 10^{-20}$ for $d = 31$. This rough model provides some quantitative justification for the implementation of planar error correction codes in the QDP even in the absence of the ability to arbitrarily suppress logical error. Note also that, due to the long coherence times [4, 30] of the QDP spin qubits, the dominant terms in the expression for the total error probability P_{tot} are those associated with operation induced errors. This provides justification for the line-by-line application of two-qubit gates discussed in section 10.3.2, which takes a longer time to perform but improves gate quality. It also means that long coherence times and/or fast operation times are likely critical to the success of a crossbar based scheme. This concludes our discussion of the QDP mapping of the surface code. A similar exercise can be done for the 6.6.6. and 4.8.8. color codes but due to their lower thresholds [31], the results will likely be less positive for current experimental parameters.

10.6. CONCLUSION

We analyzed the architecture presented in [4], focusing on its crossbar control system. Building on this analysis we presented procedures for mapping the planar surface code and the 6.6.6. and 4.8.8. color codes. Because the line-by-line operation of the crossbar architecture means the noise in a single error correction cycle scales with the distance it is not possible to arbitrarily suppress the logical error rate by increasing the code distance. Instead there will be some “optimal" code distance for which the logical error rate is the lowest. Using numbers for [4] and an empirical model taken from [26] we analyzed the logical error behavior of the surface code mapping and found that, for current experimental numbers, it is at least in principle possible to achieve logical error probabilities below $P_{log} = 10^{-20}$, making practical quantum computation possible. However, we strongly stress that this is a rather crude estimate and a more detailed answer would have to take into account the details of the dominant error processes in quantum dot qubits. It must also take into account that while it is possible to achieve certain low noise gates and good coherence times in quantum dots qubits in isolation this does not necessarily mean they will be practically achievable in the current QDP design. A future research direction would be to perform much more detailed simulations of this crossbar system, perhaps with input from future experiments. In such a simulation the effect of correlated errors (which might

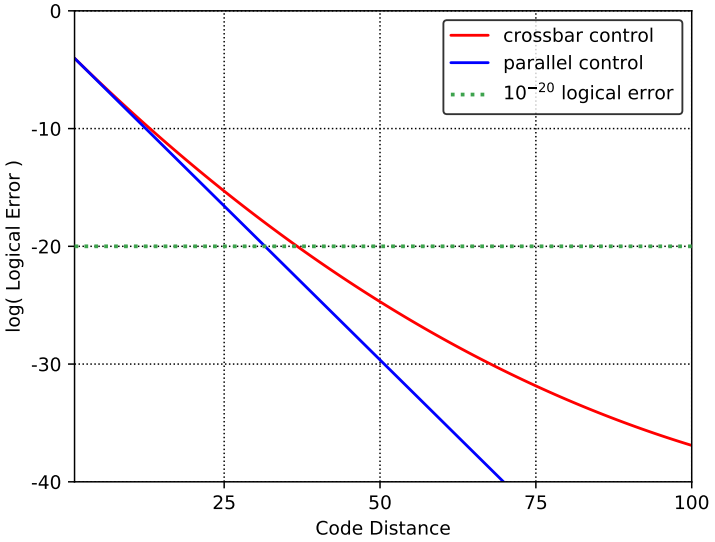


Figure 10.13: Plot of logical error probability versus code distance for the empirical model given in eq. (10.21) with experimental parameters given in section 10.5.4. Note that the logical error probability for crossbar operation goes below $P_L = 10^{-20}$ for $d = 37$. This is only slightly slower that parallel operation, which reaches $P_L = 10^{-20}$ for $d = 31$. Due to the scaling of crossbar operation with the code distance the logical error probability bottoms out at some point. This however does not happen until $d = 155$ (not shown) for a logical error rate of $P_L = 10^{-41}$, which is not practically relevant. This rough model gives good indication it is possible to create very low logical error surface code logical qubits in the QDP.

feasibly appear in a crossbar architecture) could be investigated.

Another possible research direction would be to use the currently developed machinery to map more exotic quantum error correction codes. A first step in this direction would be the implementation of variants of the surface code with more resistance to biased noise [32, 33]. Due to the possibility of qubit shuttling, also codes with long distance stabilizers could in principle be implemented. Codes such as the 3D gauge color codes might be prime candidates for this kind of treatment. However, barring some special cases, parallel shuttling is currently being performed in a line-by-line manner. A general classical algorithm for generating optimal (in time) shuttling-steps from an initial to a final BOARDSTATE would vastly simplify the task of mapping more exotic codes and also general quantum circuits. Such an algorithm would probably be useful for any future crossbar quantum architecture. In this work we constructed a non-optimal but classically efficient algorithm but finding an algorithm that generates optimal shuttling sequences and analyzing its resource use is still an open problem.

Lastly there are important aspects of quantum error correction that are not discussed in this paper. Two of these aspects are the ability to store multiple logical qubits simultaneously and the ability to perform quantum operations on the logical qubits. A popular way of performing these tasks is by encoding multiple logical qubits in a single surface code

sheet by introducing topological defects in to the surface code sheet [26]. This process involves not measuring stabilizers at certain points in the sheet, thus creating extra degrees of freedom which can store logical information. The code distance of the code is given by the physical distance (measured in number of physical qubits) between the defects. . Operations can then be performed on these logical qubits by moving the defects around each other, a process known as braiding. We think this approach is not natural to the constraints of the crossbar architecture for the following reasons

- Encoding qubits as defects would mean the size of the surface code sheet would scale as the number of encoded qubits. Hence also, in our implementation, the physical error probability per QEC cycle would scale with the number of qubits. This would put an upper limit on the number of qubits that can be implemented.
- Creating and moving around defects requires turning on and of measurements for certain stabilizers in a local manner. This locality runs counter to the design ideas of the crossbar architecture
- Given that the size of the surface code sheet would scale with the number of logical qubits one would likely face significant issues involving uniformity of control parameters of the entire sheet. This would be a significant issue even if the scaling of the physical error probability can be avoided by clever implementation

However we can envision a mode of computation that we speculate is more amenable to this architecture by thinking of an architecture composed of separate modules containing a single logical qubit. We refer to fig. 7 of [4] for a proposal of how this could be done. Inside each module our surface code protocol could be run with the ideal code distance given physical error parameters setting the size of these modules. We could then perform logical X and Z gates transversally within the modules and we could perform CNOT gates between adjacent modules via lattice surgery on the edges of the modules. Note that lattice surgery, which involves the turning on and off of stabilizer patches in regular patterns (see [24] for an introduction to lattice surgery), is very amenable to the constraints of the architecture, implying that a high degree of parallelization could be achieved when mapping lattice surgery techniques to the crossbar architecture.

10.7. Shuttling algorithm

In this technical section we go a little deeper into the shuttling algorithm presented in section 10.3.1. This algorithm takes as input a collection of desired shuttling operations in the form of a flow matrix F and outputs a sequence of parallel HS operations that, when applied sequentially, achieve this desired collection of operations. The algorithm, described in Algorithm 3 relies centrally on a notion of independence on the columns of this flow matrix F. This notion of independence is actualized by calling an 'independence subroutine' through the functions `CheckIndependence` and `DependenceSet`. Here we describe various independence subroutines and analyze their time complexity. Note that it is probably possible to optimize these subroutines and the time complexity estimates, hence the results in theorems 10.2 to 10.4 can best be seen as upper bounds on the worst case time complexity. Recall from section 10.3.1 that the flow matrix F has entries in the

set $e, r, l, re, le, *$ where e signifies doing nothing, r signifies a rightward shuttling operation, l signifies a leftward shuttling operation and re, le and $*$ are 'wildcard' symbols that indicate the operation at the point could be either r or e, or l or e, or r or l or e. We begin by analyzing the mathematical structure of the shuttle operations e, r, l.

10.7.1. THE LEFT-RIGHT MONOID

An idempotent monoid $\{M, \circ\}$ is a set M with a binary operation $\circ : M \times M \to M$ such that the following axioms hold

$$\begin{aligned} &\forall a, b, c \in M : && (a \circ b) \circ c = a \circ (b \circ c), && \text{(Associativity)} \\ &\exists e \in M, \forall a \in M : && e \circ a = a \circ e = a, && \text{(Identity element)} \\ &\forall a \in M : && a \circ a = a. && \text{(Idempotence)} \end{aligned}$$

Note that monoids are strict generalizations of groups (as the elements lack an inverse). We will argue that the set of shuttle operations r, l, e together with the binary operation 'composition of shuttle operations' is an idempotent monoid. Imagine for simplicity a BOARDSTATE with only one row and two columns. The shuttle operations that can be applied to this system are 'shuttle to the right' (r), shuttle to the left (l) and 'do nothing' (e). These shuttle operations can also be applied sequentially. We will denote the sequential application of operations a_1, a_2 as $a_1 \circ a_2$. This is read from the right, so we apply first a_2 and then a_1. For example shuttling to the right at a location and subsequently doing nothing at that location will be written as $e \circ r$. Note that this is equivalent to just shuttling to the right at that location so we have $e \circ r = r$. Other examples are more interesting.For instance shuttling to the right at a fixed location followed by shuttling to the left at that same location is equivalent to just shuttling to the left at that location. However shuttling to the left first and then shuttling to the right is equivalent to shuttling to the right. this means we have $r \circ l = r \neq l = l \circ r$ which means the operation $\circ$ is not commutative. Not also that shuttling to the right at a fixed location and then shuttling to the right again at that location is equivalent to shuttling to the right a single time at that location. Hence we have$r \circ r = r$. In general we have the following rules for the composition of the operations e, r, l

$$r \circ r = r, \qquad l \circ l = l, \tag{10.23}$$
$$r \circ l = r, \qquad l \circ r = r, \tag{10.24}$$
$$r \circ e = r, \qquad l \circ e = l, \tag{10.25}$$
$$e \circ r = r, \qquad e \circ l = e. \tag{10.26}$$

Note that these rules imply that the composition $\circ$ is also associative. This makes the set $S = \{e, r, l\}$ with the composition $\circ$ an idempotent monoid with e the identity element. Note also that neither r nor l is has an inverse. We call the idempotent monoid $\{S, \circ\}$ the left-right monoid. The non-commutativity of $\circ$ and the fact that neither r nor l have inverses makes finding a sequence of parallel shuttling operations that apply the shuttling operations encoded in the flow matrix F difficult in general as the order in which the operations are applied matters and no operation can be truly inverted.

10.7.2. COMPARING UP TO WILDCARDS

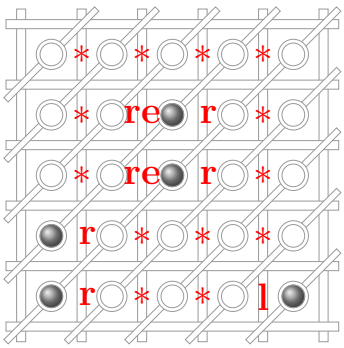
Recall that the entries of the flow matrix F take value in the set $\{r, l, e, re, le, *\}$. Of these elements only the first three ($\{r, l, e\}$) correspond to real actions. The other elements $\{re, le, *\}$ are called wildcard elements. If an entry in the flow matrix is wildcard valued it means we have some freedom which of the operations $\{r, l, e\}$ we apply there. When constructing the independence subroutines we need a way to compare elements of the flow matrix that takes this freedom into account. To this end we introduce the relation 'equality up to wildcards', signified by the symbol $=_w$. This relation formalizes the the intuitive notion that e.g. the elements re and r but also the elements re and e should be equal since if an entry of the flow matrix F takes the value re we can perform either the operation r or e without performing an illegal move. Below we list all elements that are equal up to wildcards (up to symmetry and reflexivity)

$$* =_w r \qquad re =_w r \qquad le =_w l \tag{10.27}$$
$$* =_w l \qquad re =_w e \qquad le =_w e \tag{10.28}$$
$$* =_w e \tag{10.29}$$
$$* =_w re \tag{10.30}$$
$$* =_w le. \tag{10.31}$$

10.7.3. COMPARING COLUMNS OF THE FLOW MATRIX

Checking an equality up to wildcards computationally takes at most five list comparisons. We would also like to be able to compare columns of the flow matrix F. This is so because repeated patterns in the flow matrix columns can be performed in a single parallel shuttle operation. To see why this is the case consider the following example flow matrix

$$F_{ex} = \begin{pmatrix} * & * & * & * \\ * & re & r & * \\ * & re & r & * \\ r & * & * & * \\ r & * & * & l \end{pmatrix} \tag{10.32}$$



For illustration we also included a BOARDSTATE configuration that could have led to this particular flow matrix. Here it is clear how all elements of the flow matrix correspond to a crossing point on the grid. Notice that columns 0 and 2 of F_{ex} are very similar. In fact they are the same up to a constant shift. Note also that we can apply all r operations encoded in the flow matrix by performing the operation HS[{(0,0,1),(1,0,1,),(2,2,1),(2,3,1)}]. This operation can be done in a single time-step by the parallel control OPCODE V[0]&V[1]&D[0][0] &D[1][1]&D[3][2]. This can be done because the two columns are equal up to a constant vertical shift. Because the diagonal lines run in a 45 degree angle over the QDP this shift is exactly equal to the

difference between the column indices of the two columns. In our example the shift is 2 since we are comparing the columns 0 and 2 but in general when checking if column i and j can be performed in parallel we must check if they are equal (up to wildcards, see above) up to a constant vertical shift of size $i-j$. The facilitate this process we define the padding function p:

$$p : \{r, l, e, re, le, *\}^{\times(N-1)} \times [1:N] \longrightarrow \{r, l, e, re, le, *\}^{\times(2N-2)} : \tag{10.33}$$

$$p(v, i) \longrightarrow (*, \ldots, \underset{i-1}{*}, v, \underset{N+i}{*}, \ldots, *). \tag{10.34}$$

Now because, up to wildcards, the $*$ symbol is equal to all symbols other we can check if two columns v_i, v_j of F are equal up to wildcards and a constant vertical shift by checking if the padded columns $p(v_i, i), p(v_j, j)$ are equal up to wildcards. This means checking $p(v_i, i)_t =_w p(v_j, j)_t$ for all $t \in [1 : 2N]$ where $p(v_i, i)_t$ is the t'th component of the padded column $p(v_i, i)$.

10.7.4. Composition of columns in the flow matrix

Sometimes it happens that columns in the flow matrix can be written as the composition of other columns. This is the case in the following example

$$F_{ex} = \begin{pmatrix} * & * & l \\ * & * & r \\ l & r & * \end{pmatrix}. \tag{10.35}$$

Note that we have $p(v_0, 0) \circ p(v_1, 1) =_w p(v_2, 2)$ where v_i is the i'th column of F_{ex} and the composition and equality up to wildcards are taken element-wise. This means that in a sense v_2 is 'dependent' on v_0 and v_1. This means we can apply the operations encoded by the flow matrix F in two sequential parallel shuttling steps by duplicating the operations encoded in columns 0 and 1 to also cover column 3 (up to the correct constant vertical shift). The shuttle operations that need to be performed to apply the shuttle operations encoded in the flow matrix are

```
HS[(0,0,-1), (2,2,-1)]   (10.36)
HS[(0,1,1), (1,2,1)].    (10.37)
```

Note that in this particular case we have $p(v_0, 0) \circ p(v_1, 1) =_w p(v_1, 1) \circ p(v_0, 0)$, i.e. $p(v_0, 0)$ and $p(v_1, 1)$ commute with respect to the composition '$\circ$', which means we can apply the operations above in any order. This does not need to be the case. Consider for instance in the following example:

$$F_{ex} = \begin{pmatrix} * & * & l \\ * & r & r \\ l & r & * \end{pmatrix}. \tag{10.38}$$

Note that we now have $p(v_0, 0) \circ p(v_1, 1) =_w p(v_2, 2)$ but not $p(v_1, 1) \circ p(v_0, 0) =_w p(v_2, 2)$! This means that in this case we can still apply all shuttling operation encoded in the flow matrix F_{ex} by extending the operations that apply the shuttle operations encoded in v_0 and v_1 to also include the operations encoded in v_2 but now the order in which we

perform the operations matters. We have to first apply the operation HS[(0,1,1),(1,1,1),(1,2,1),(2,2,1)] and then HS[(0,0,-1), (2,2,-1)] in order to apply the operations encoded in the flow matrix F_{ex}. The fact that the ordering of the operations matter is a difficulty we have not fully overcome. Therefore we will restrict ourselves only to compositions of columns that explicitly commute. Note that this means for two columns v_i, v_j and $k \in [1 : 2N-2]$ that if $p(v_i,i)_k =_w r$ we must have $p(v_j,j)_k =_w e$ or $p(v_j,j)_k =_w r$ or vice versa. A similar rule holds if $p(v_j,j)_k =_w l$. Using the above analysis we now present two simple subroutine algorithms that decide whether a column is dependent, i.e. can be written as a restricted composition of other columns in F (up to wildcards and vertical shifting).

10.7.5. SIMPLE SUBROUTINE

This subroutine simply detects whether the columns of a flow matrix F have duplicates up to wildcard symbols. This means the restriction of the previous section is basically the strictest possible, columns can only be written as compositions of something exactly equal. See eq. (10.32) for an example of a flow matrix with this property. We have two distinct subroutine functions. The first $\mathbf{CheckIndependence}(S, v_i)$ checks whether a column v_i of the flow matrix F is independent of the columns in the set S while the second, $\mathbf{DependenceSet}(S, v_i)$ returns a set of columns A_i on which v_i depends. For the [simple] subroutines this will be a set with a single entry, namely an exact copy (up to wildcards) of v_i in S. The two subroutine functions are given in algorithms 5 and 6 and their time complexity is analyzed in theorem 10.2.

Algorithm 5 CheckIndependence [simple]

Input: A column v_i, a set of columns S, the column index i of the column v_i
Output: Boolean a

```
// We will consistently write columns of the flow matrix F as v_i where i indicates the column index of v_i in F.

for all columns c_j ∈ S do
    if p(v_i, i) =_w p(c_j, j) then
        // Note that all the elements of C commute. This means that if p(v_i, i) =_w p(c_j, j) for a particular c_j it
        // must also hold that p(v_i, i) =_w p(c_k, k) for all c_k ∈ C.
        Set a to TRUE
    else
        Set a to FALSE
    end if
end for
return a
```

Algorithm 6 DependenceSet [simple]

Input: A column v_i, a set of columns S, the column index i of the column v_i
Output: Set of columns A_i that v_i depends on

1: // We will consistently write columns of the flow matrix F as v_i where i indicates the column index of v_i in F.
2: // We also tag the output set A_i with the subscript i to indicate it is connected to the i'th column v_i of the
3: // flow matrix F.
4: **Set** A_i to an empty set
5: **for** all columns $c_j \in S$ **do**
6: **if** $p(v_i, i) =_w p(c_j, j)$ **then**
7: Add c_j to A_i
8: **end if**
9: **end for**
10: **return** A_i

Theorem 10.2. The subroutines **CheckIndependence**( ) [simple] and **IndependenceSet**( ) [simple] have time complexity

$$O\big(\textbf{CheckIndependence}(\,)\ [\text{simple}]\big) = O(NM) \tag{10.39}$$

$$O\big(\textbf{IndependenceSet}(\,)\ [\text{simple}]\big) = O(NM) \tag{10.40}$$

where $M = |S|$ the size of the independent set S and N is the length of the input column v_i or equivalently the number of rows in the flow matrix F. Note that for our purposes we have $M \leq N$.

Proof. The complexity of the subroutine **CheckIndependence**( ) [simple] can be seen by straightforward counting. There is one **For**-loop (line 1) of length M and the **If**-clause on line 2 takes $O(N)$ time to evaluate since c_j is a column of length $O(N)$. This gives a total worst case complexity of $O(NM)$. Exactly the same argument holds for the subroutine **IndependenceSet**( ) [simple]. ■

10.7.6. K-COMMUTING SUBROUTINES

Next we present a class of subroutines collectively called the 'commuting' subroutines. These try to capture the intuition behind the example flow matrix in eq. (10.35), namely that some columns of the flow matrix can be written as a composition of a subset of pairwise commuting columns of the flow matrix. By a pairwise commuting subset C we mean concretely that for all columns v_i, v_j of F that are in C we have $p(v_i, i) \circ p(v_j, j) =_w p(v_j, j) \circ p(v_i, i)$. We restrict explicitly to pairwise commuting columns since this avoids the difficulty of time-ordering the resulting shuttle operations (as seen in eq. (10.38)). This subroutine relies on an initial construction of all maximal mutually commuting subsets of a set of columns S. Listing all these sets is in general hard. To see this we can construct the $M \times M$ matrix A that has entries $A_{pq} = 1$ whenever the p'th and q'th column in the set S commute and $A_{pq} = 0$ otherwise. If we now think of A as the adjacency matrix of a graph G with M vertices it is not hard to see that finding all maximal mutually commuting subsets of S is equivalent to finding listing all maximal cliques [34] in the graph G.

The best known algorithm for listing all maximal cliques in an arbitrary graph is called the Bron-Kerbosch algorithm [35] and has a worst-case complexity of $O(3^{M/3})$. There is no a priori way to restrict the number of possible maximal cliques generated by the commutation rules of the monoid-valued columns so currently any subroutine that searches over all possible maximal mutually commuting subsets of S will take at least $O(3^{M/3})$ time. We can however get out of this bind by restricting the size of the sets of mutually commuting columns to be less than a fixed parameter k. This will make our algorithm less effective but the problem of finding these reduces to finding all cliques of size less than k in the graph G which is a so called fixed-parameter-tractable problem. This problem has worst-case time complexity upper bounded by $O(M^k k^3)$ [36]. Using these sets we can construct a family of subroutines indexed by the parameter k. The two subroutine functions are given in algorithms 7 and 8 and their complexity is analyzed in theorem 10.3.

Algorithm 7 CheckIndependence [k-commuting]

Input: A column v_i, a set of columns S, the column index i of the column v_i and a fixed integer k

Output: Boolean a

1: // We will consistently write columns of the flow matrix F as v_i where i indicates the column index of v_i in F.
2: Construct All mutually commuting subsets C of S with $|C| \leq k$
3: Set a to FALSE
4: **for** all commuting subsets C **do**
5: **for** $t \in [0 : \text{length}(p(v_i, i)) - 1]$ **do**
6: **for** all columns $c_j \in C$ **do**
7: **if** $p(c_j, j)_t =_w p(v_i, i)_t$ **then**
8: // Note that all the elements of C commute. This means that if $p(v_i, i) =_w p(c_j, j)$ for a particular
9: // c_j it must also hold that $p(v_i, i) =_w p(c_k, k)$ for all $c_k \in C$.
10: **if** $t = \text{length}(p(v_i, i)) - 1$ **then**
11: Set a to TRUE
12: **end if**
13: **else**
14: Go to next commuting subset C in the loop at line 3
15: **end if**
16: **end for**
17: **end for**
18: **end for**
19: **return** a

Algorithm 8 DependenceSet [k - commuting]

Input: A column v_i, a set of columns S, the column index i of the column v_i
Output: Set of columns A_i that v_i depends on

```
1: // We will consistently write columns of the flow matrix F as v_i where i indicates the
   column index of v_i in F.
2: Construct All maximal mutually commuting subsets C of S with |C| ≤ k
3: for all commuting subsets C do
4:     for t ∈ [0 : length(p(v_i, i)) − 1] do
5:         for all columns c_j ∈ C do
6:             if p(c_j, j)_t =_w p(v_i, i)_t then
7:                 // Note that all the elements of C commute. This means that if
   p(v_i, i) =_w p(c_j, j) for a particular
8:                 // c_j it must also hold that p(v_i, i) =_w p(c_k, k) for all c_k ∈ C.
9:                 Add c_j to A_i
10:                if t = length(p(v_i, i)) − 1 then
       return A_i
11:                else
12:                    Set A_i to empty set
13:                    Go to next commuting subset C in the loop at line 2
14:                end if
15:            end if
16:        end for
17:    end for
18: end for
19: return A_i
```

Theorem 10.3. The subroutines **CheckIndependence**( ) [k − commuting] and **IndependenceSet**( ) [k − commuting] have time complexity

$$O\big(\textbf{CheckIndependence}(\,)\ [\text{k} - \text{commuting}]\big) = O(NM^{k+1}k^4) \tag{10.41}$$

$$O\big(\textbf{IndependenceSet}(\,)\ [\text{k} - \text{commuting}]\big) = O(NM^{k+1}k^4) \tag{10.42}$$

where $M = |S|$ the size of the independent set S, N is the length of the input column v_i or equivalently the number of rows in the flow matrix $F$and k is a fixed parameter indicating the maximal size of the mutually commuting subsets C. Note that for our purposes we have $M \leq N$.

Proof. We can again find the complexity of the subroutine **CheckIndependence**( ) [k − commuting] by a counting argument. We have already noted that the construction in line 1 takes $O(M^k k^3)$ time in the worst case. Apart from that we have a **For**-loop on line 3 that takes worst-case time $O(M^k k^3)$ to loop over and given that $|C| \leq k$ we can see that the **For**-loops on line 4 and 5 take respectively $O(N)$ and $O(k)$ time to complete. Finally the **If**-clause on line 6 takes $O(N)$ time to complete. Tallying this up we get a total worst-case time complexity of $O(NM^{k+1}k^4)$. We can make the same argument for the worst-case time complexity of **IndependenceSet**( ) [k − commuting]. ■

10.7.7. Greedy commuting subroutine

Building on the last subsection we introduce one last pair of subroutines dubbed greedy subroutines. As seen in the last section it seems hard to find all maximal mutually commuting subsets of a set of columns S. We made this problem fixed-parameter-tractable by restricting to mutually commuting subsets of size at most k with k some fixed parameter. Here we take a different approach based on the idea that while it is hard to find all mutually commuting subsets of S it is, given a column $v_i \in S$, tractable to find some maximal mutually commuting subset that contains v_i. Constructing this subset reduces to finding, given a graph G and some vertex v, some clique that contains v. Note that we do not get to choose which clique will be found, only that one will be found. This can be done in time $O(M)$ where M is the number of nodes in the graph G. Hence we can find, given a column v_i, a single maximal mutually commuting subset of S that contains v_i. We can use only this set to evaluate whether a given other column v_i can be written as the commuting composition of elements of A. The two subroutine functions are given in algorithms 9 and 10 and their complexity is analyzed in theorem 10.4.

Algorithm 9 CheckIndependence [greedy commuting]

Input: A column v_i, a set of columns S, the column index i of the column v_i
Output: Boolean a

```
1: // We will consistently write columns of the flow matrix F as v_i where i indicates the
   column index of v_i in F.
2:
3: Set a to FALSE
4: for columns w_j ∈ S do
5:     Construct maximal mutually commuting subset C of S containing w_j
6:     for t ∈ [0 : length(p(v_i, i)) − 1] do
7:         for all columns c_l ∈ C do
8:             if p(c_l, l)_t =_w p(v_i, i)_t then
9:                 if t = length(p(v_i, i)) − 1 then
10:                    // Note that all the elements of C commute. This means that if
   p(v_i, i) =_w p(c_j, j) for a
11:                    // particular c_j it must also hold that p(v_i, i) =_w p(c_k, k) for all
   c_k ∈ C.
12:                    Set a to TRUE
13:                end if
14:            else
15:                Go to next column w_j ∈ S at line 2
16:            end if
17:        end for
18:    end for
19: end for
20: return a
```

Algorithm 10 DependenceSet [greedy commuting]

Input: A column v_i, a set of columns S, the column index i of the column v_i
Output: Set of columns A_i that v_i depends on

1:
2: We will consistently write columns of the flow matrix F as v_i where i indicates the column index of v_i in F.
3: **for** columns $w_j \in S$ **do**
4: Construct a maximal mutually commuting subset C of S containing the column w_j
5: **for** $t \in [0 : \text{length}(p(v_i, i)) - 1]$ **do**
6: **for** all columns $c_l \in C$ **do**
7: **if** $p(c_l, l)_t =_w p(v_i, i)_t$ **then**
8: // Note that all the elements of C commute. This means that if $p(v_i, i) =_w p(c_j, j)$ for a particular c_j it
9: // must also hold that $p(v_i, i) =_w p(c_k, k)$ for all $c_k \in C$.
10: Add c_l to A_i
11: **if** $t = \text{length}(p(v_i, i)) - 1$ **then**
return A_i
12: **end if**
13: **else**
14: **Set** A_i to empty set
15: Go to next column $w_j \in S$ at line 1
16: **end if**
17: **end for**
18: **end for**
19: **end for**
20: **return** A_i

Theorem 10.4. The subroutines **CheckIndependence**( ) [greedy] and **IndependenceSet**( ) [greedy] have time complexity

$$O\big(\textbf{CheckIndependence}(\,)\ [\text{greedy}]\big) = O(NM^3) \tag{10.43}$$

$$O\big(\textbf{IndependenceSet}(\,)\ [\text{greedy}]\big) = O(NM^3) \tag{10.44}$$

where $M = |S|$ the size of the independent set S and N is the length of the input column v_i or equivalently the number of rows in the flow matrix F. Note that for our purposes we have $M \leq N$.

Proof. We again find the time complexity of **CheckIndependence**( ) [greedy] by a counting argument. The **For**-loop on line 2 takes $O(M)$ time to iterate over. We argued above that the greedy construction on line 3 can be done in $O(M)$ time, the **For**-loop on line 4 takes $O(N)$ time to iterate over, the **For**-loop on line 5 takes $O(M)$ time to iterate over and the **If**-clauses in the body can be evaluated in constant time. This means we get a total worst case time complexity of $O(NM^3)$. We can again make the same argument for **IndependenceSet**( ) [greedy]. ■

10.8. Surface code operation counts

Steps	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Z-cycle Total
$\sqrt{S}$ gate					2	2				2	2							8
Z rotation					2	2				2	1							7
Shuttling			1				1	1				1		2	1	2	1	10
Global rotation		1											1					2
Measurement																1		1

Table 10.6: Gate count per gate type and per step for the Z-ancilla during the Z-cycle of the surface code cycle described in section 10.4.4 and section 10.5.1. Specifically the Z-ancilla is taken to be qubit A in fig. 10.12 (right). Table cells that are left empty signify zero entries.

Steps	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	X-cycle Total
$\sqrt{S}$ gate																		0
Z rotation																		0
Shuttling														2	1		1	4
Global rotation		1											1		1			3
Measurement																1		1

Table 10.7: Gate count per gate type and per step for the Z-ancilla during the X-cycle of the surface code cycle described in section 10.4.4 and section 10.5.1. Specifically the Z-ancilla is taken to be qubit A in fig. 10.12 (right). Table cells that are left empty signify zero entries.

Steps	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Z-cycle Total
$\sqrt{S}$ gate					2						2							4
Z rotation																		0
Shuttling			1	1														2
Global rotation		1											1					2
Measurement																		0

Table 10.8: Gate count per gate type and per step for a data qubit during the Z-cycle of the surface code cycle described in section 10.4.4 and section 10.5.1. Specifically the data qubit is taken to be qubit 1 in fig. 10.12 (right) but other data qubits will have the same gate count up to a possible reordering of steps. Table cells that are left empty signify zero entries.

Steps	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	X-cycle Total
$\sqrt{S}$ gate						2						2						4
Z rotation																		0
Shuttling								1	1						1*		1*	4
Global rotation		1											1					2
Measurement																		0

Table 10.9: Gate count per gate type and per step for a data qubit during the X-cycle of the surface code cycle described in section 10.4.4 and section 10.5.1. Specifically the data qubit is taken to be qubit 1 in fig. 10.12 (right) but other data qubits will have the same gate count up to a possible reordering of steps. Table cells that are left empty signify zero entries.
*: Only half of the data qubits move during this step. In the total gate count this gate is counted towards all data qubits.

11

Conclusions and outlook

In this concluding chapter we summarize the research contributions described in this thesis, and attempt to look ahead into the future. This outlook proceeds along two paths: one the one hand we focus on immediate technical additions to the work performed in this thesis, while on the other hand we give a broader outlook on the fields in which the work presented in this thesis is situated.

11.1. Summary of results

The technical contributions of this thesis can be divided into two parts. The goal of the first part was to further understand the assumptions going into randomized benchmarking and related protocols and if possible justify or weaken them. Our contributions in this direction can be summarized as follows:

- We derived strong bounds on the number of random sequences needed for rigorous randomized benchmarking and unitarity randomized benchmarking, retroactively justifying the use of randomized benchmarking with common experimental choices for these parameters.
- We performed a detailed study of the representation theory of the Clifford group, in particular classifying all irreducible subrepresentations of the "2-copy" representations of the Clifford group for any qubit number. This formed a core structural element of the bounds discussed above.
- We studied on other statistical properties of randomized benchmarking and unitarity randomized benchmarking. Notably we worked out the correct estimators for the sequence purity in unitarity randomized benchmarking and discussed the impact of variance heterogeneity on the standard randomized benchmarking protocol.
- We introduced a new class of randomized benchmarking protocols we call character randomized benchmarking. These new methods allow for a principled extension of randomized benchmarking beyond the Clifford group while retaining the advantages of randomized benchmarking such as scalability and robustness to SPAM errors. It also enjoys similar statistical properties and moreover displays a similar robustness to gate-dependent noise fluctuations.
- In order to verify the character randomized benchmarking method we have implemented a particular instance, called 2-for-1 interleaved benchmarking, to characterize the average gate fidelity of a two-qubit CPHASE gate in a pair of Si\SiGe quantum dots.

The second research direction of this thesis concerns strategies for overcoming limited connectivity and control in future large scale quantum devices. Our contribution in this direction can be summarized as follows:

- We performed an in-depth study of the crossbar architecture proposed in [1]. Here we focused in particular on abstracting away enough physical details to allow us to focus on the issues of connectivity and parallel control. This study can function as a model for future investigations of crossbar-style architectures.
- We provided algorithms for the implementation of three types of planar topological error correction codes on the crossbar architecture and, for the planar surface code, provide a detailed operations count as a function of the code distance. We argued that, even though our implementations do not per se have a threshold, they still allow for sufficient error reduction for all practical purposes, especially when taking into account the savings in control requirements obtained by considering crossbar-control.

11.2. Future work

In this section we list a few questions that are close extensions of the work done in this thesis. They are mostly of a technical nature and probably have a nice answer, given some extra work. We think that answering these questions could lead to some intersitng developments.

Automated discovery of groups for character benchmarking: Character randomized benchmarking extends the randomized benchmarking protocol to gatesets defined by, in principle, any finite group. However, only a few families of finite have previously been discussed in the randomized benchmarking literature. This stands is sharp contrast with the amount of well characterized finite groups included in powerful algebra packages such as GAP [2] or MAGMA [3]. Therefore it would be interesting to take a systematic, computational approach to finding finite groups to benchmark. This will probably not find infinite families of groups that are defined on any number of qubits but rather be restricted to the few-qubit setting. On the other hand, this computational approach would allows us to extract all kinds of interesting and exotic quality parameters that would normally not be considered because the groups required are too complex to manipulate by hand.

Randomized benchmarking in a limited non-Markovian setting: A key assumption made when using randomized benchmarking, even its most general form, is that the noisy version of any given gate does not depend on when it is applied and what gates are applied before and after. This is an assumption that is known to be violated in real experiments so it would be very interesting to formally extend the validity of (an adapted version of) RB to at least some forms of non-Markovianity. The key idea here would be to source concrete forms of non-Markovianity and then attempt to formalize these. A recent example found in experimental practice [4] that might be tractable is where the implementation of a certain gate only depends on the gate that happened exactly before it.

Full analysis of the scalable unitarity RB implementation: In chapter 7 we discussed in detail the two different implementations of the unitarity randomized benchmarking protocol. Only one of these protocols is scalable in the number of qubits. However this implementation require $2q$ qubits to characterize the unitarity of q qubits. Moreover we must make the assumption that the there is no correlation between the noise affecting the first q qubits and the noise affecting the latter q. This is a strong assumption that will not be satisfied in experiment (especially given that an entangled state between the two registers must be established). This means it is interesting to see if this assumption can be weakened (by adapting the URB protocol), or to see what the interpretation of the number measured by URB is in this more general context.

11.3. Outlook

With this thesis we hope to have contributed in a small way to the development of quantum computers. However, much more still needs to be done before we can claim to have fully functioning quantum computers. The variety of research being performed to this

end is vast, and even the sub-fields that this thesis contributed to: diagnostics of quantum computers, and medium-scale quantum computer architecture design are rapidly growing and can very well be called research fields in their own right. Here we offer some concluding, personal thoughts on the directions these fields might move towards in the future.

Operational measures for concrete tasks Current diagnostic tools for quantum devices, such as randomized benchmarking or tomography, mostly focus on quality measures for the basic building blocks of quantum computers. However, as quantum computers move away from laboratory experiments and into a more industrial environment these measures will become less and less indicative of the actual performance of these devices. The prospective user of a medium size quantum device will care less about the fidelity of the average unitary the device can perform than they would care about the accuracy of simulation of specific classes of Hamiltonians. This latter factor is also influenced by many non-quantum parameters such as the implementation details of the simulation algorithm, the presence or absence of circuit optimization techniques or the use of algorithm specific error mitigation techniques. It is also likely that end-users will not have direct control over all these parameters. This means it will become increasingly important to develop benchmarks that test the ability of a quantum device to perform specific tasks in an integrated fashion. Moreover, these end-users will not all be experts in quantum computing so the interpretability usability criteria outlined in chapter 1 will become more important over time whereas the generality criterion will diminish in importance.

Single purpose logical units with minimal outside control In current devices all qubits are controlled directly by an external classical computer (that is often several meters and hundreds of Kelvins away from the quantum chip). However, as we already discussed in chapter 10, this control architecture runs into scalability issues. Therefore we suspect that any medium scale ($> 10^3$ qubits) quantum device must instead be organized in a hierarchical manner, with as much repetitive classical control as possible being integrated on the chips themselves. This meshes nicely with the need for quantum error correction in quantum computers. As we saw in chapter 10 QEC requires the repetition of a fixed sequence of instructions on a fixed subset of qubits (a logical unit) over and over again. Therefore, ideally, we would not need to send these instructions from an external source but rather include their source on the quantum chip itself, creating hybrid quantum-classical devices. There are many theoretical and even more engineering challenges to be overcome here, but a number of proposals in this direction have already been made [5, 6].

References for Chapter 2

[1] M. A. Nielsen and I. Chuang, *Quantum computation and quantum information,* (2002).

[2] J. Watrous, *Theory of quantum information,* CS 766/QIC 820 lecture notes, University of Waterloo (2011).

[3] R. A. Horn, R. A. Horn, and C. R. Johnson, *Matrix analysis* (Cambridge university press, 1990).

[4] R. Bhatia, *Graduate texts in mathematics: Matrix analysis,* (1997).

[5] J. F. Cornwell, *Group theory in physics: An introduction*, Vol. 1 (Academic press, 1997).

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

[7] J. J. Wallman and S. T. Flammia, *Randomized benchmarking with confidence,* New J. Phys. **16**, 103032 (2014).

[8] M. Wolf, *Quantum channels operations: Guided tour.* Lecture Notes (2012).

[9] M. A. Nielsen, *A simple formula for the average gate fidelity of a quantum dynamical operation,* Phys. Lett. A **303**, 249 (2002).

[10] J. J. Wallman, C. Granade, R. Harper, and S. T. Flammia, *Estimating the Coherence of Noise,* New J. Phys. **17**, 113020 (2015), arXiv:arXiv:1503.0786 .

References for Chapter 3

[1] W. Fulton and J. Harris, *Representation Theory* (Springer New York, 2004).

[2] R. Goodman and N. R. Wallach, *Symmetry, Representations, and Invariants* (Springer New York, 2009).

[3] D. Gross, K. Audenaert, and J. Eisert, *Evenly distributed unitaries: On the structure of unitary designs,* Journal of Mathematical Physics **48**, 052104 (2007).

[4] B. Dirkse, J. Helsen, and S. Wehner, *Efficient unitarity randomized benchmarking of few-qubit clifford gates,* arXiv preprint arXiv:1808.00850 (2018).

References for Chapter 4

[1] E. Magesan, J. M. Gambetta, and J. Emerson, *Scalable and Robust Randomized Benchmarking of Quantum Processes,* Phys. Rev. Lett. **106**, 180504 (2011).

[2] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D. G. Cory, and

R. Laflamme, *Symmetrized characterization of noisy quantum processes,* Science **317**, 1893 (2007).

[3] E. Magesan, J. M. Gambetta, and J. Emerson, *Characterizing quantum gates via randomized benchmarking,* Phys. Rev. A **85**, 042311 (2012).

[4] 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**, 012307 (2008).

[5] J. J. Wallman, C. Granade, R. Harper, and S. T. Flammia, *Estimating the Coherence of Noise,* New J. Phys. **17**, 113020 (2015), arXiv:arXiv:1503.0786 .

[6] J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, *Randomized benchmarking and process tomography for gate errors in a solid-state qubit,* Phys. Rev. Lett. **102**, 090502 (2009).

[7] 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**, 260503 (2012).

[8] L. Casparis, T. W. Larsen, M. S. Olsen, F. Kuemmeth, P. Krogstrup, J. Nygård, K. D. Petersson, and C. M. Marcus, *Gatemon benchmarking and two-qubit operations,* Phys. Rev. Lett. **116**, 150505 (2016).

[9] A. D. Córcoles, J. M. Gambetta, J. M. Chow, J. A. Smolin, M. Ware, J. Strand, B. L. T. Plourde, and M. Steffen, *Process verification of two-qubit quantum gates by randomized benchmarking,* Phys. Rev. A **87**, 030301 (2013).

[10] M. Veldhorst, C. H. Yang, J. C. C. Hwang, W. Huang, J. P. Dehollain, J. T. Muhonen, S. Simmons, A. Laucht, F. E. Hudson, K. M. Itoh, A. Morello, and A. S. Dzurak, *A two-qubit logic gate in silicon,* Nature **526**, 410 (2015).

[11] E. Kawakami, T. Jullien, P. Scarlino, D. R. Ward, D. E. Savage, M. G. Lagally, V. V. Dobrovitski, M. Friesen, S. N. Coppersmith, M. A. Eriksson, and L. M. K. Vandersypen, *Gate fidelity and coherence of an electron spin in an si/sige quantum dot with micromagnet,* Proceedings of the National Academy of Sciences **113**, 11738 (2016).

[12] D. Riste, J. Van Leeuwen, H.-S. Ku, K. W. Lehnert, and L. DiCarlo, *Initialization by measurement of a superconducting quantum bit circuit,* Physical review letters **109**, 050507 (2012).

[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**, 062119 (2013).

[14] D. Gross, K. Audenaert, and J. Eisert, *Evenly distributed unitaries: On the structure of unitary designs,* J. Math. Phys. **48**, 052104 (2007).

[15] J. Emerson, R. Alicki, and K. Życzkowski, *Scalable noise estimation with random unitary operators,* J. Opt. B **7**, S347 (2005).

[16] D. S. França and A.-L. Hashagen, *Approximate randomized benchmarking for finite groups,* arXiv preprint arXiv:1803.03621 (2018).

[17] J. Helsen, X. Xue, L. M. K. Vandersypen, and S. Wehner, *A new class of efficient randomized benchmarking protocols,* arXiv:1806.02048 (2018).

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

[19] D. Gottesman, *Theory of fault-tolerant quantum computation,* Phys. Rev. A **57**, 127 (1998).

[20] S. Aaronson and D. Gottesman, *Improved simulation of stabilizer circuits,* Physical Review A **70**, 052328 (2004).

[21] W. Hoeffding, *Probability inequalities for sums of bounded random variables,* Journal of the American Statistical Association **58**, 13 (1963).

[22] T. Proctor, K. Rudinger, K. Young, M. Sarovar, and R. Blume-Kohout, *What randomized benchmarking actually measures,* Physical review letters **119**, 130502 (2017).

[23] J. J. Wallman, *Randomized benchmarking with gate-dependent noise,* Quantum **2**, 47 (2018).

[24] M. A. Fogarty, M. Veldhorst, R. Harper, C. Yang, S. Bartlett, S. Flammia, and A. Dzurak, *Nonexponential fidelity decay in randomized benchmarking with low-frequency noise,* Physical Review A **92**, 022326 (2015).

[25] K. Rudinger, T. Proctor, D. Langharst, M. Sarovar, K. Young, and R. Blume-Kohout, *Probing context-dependent errors in quantum processors,* arXiv preprint arXiv:1810.05651 (2018).

[26] K. Takeda, J. Yoneda, T. Otsuka, T. Nakajima, M. Delbecq, G. Allison, Y. Hoshi, N. Usami, K. Itoh, S. Oda, *et al.*, *Optimized electrical control of a si/sige spin qubit in the presence of an induced frequency shift,* npj Quantum Information **4**, 54 (2018).

[27] T. F. Watson, S. G. J. Philips, E. Kawakami, D. R. Ward, P. Scarlino, M. Veldhorst, D. E. Savage, M. G. Lagally, M. Friesen, S. N. Coppersmith, M. A. Eriksson, and L. M. K. Vandersypen, *A programmable two-qubit quantum processor in silicon,* Nature **555**, 633 (2018).

[28] Z. Chen, J. Kelly, C. Quintana, R. Barends, B. Campbell, Y. Chen, B. Chiaro,

A. Dunsworth, A. Fowler, E. Lucero, *et al.*, *Measuring and suppressing quantum state leakage in a superconducting qubit,* Physical review letters **116**, 020501 (2016).

[29] T. H. Taminiau, J. Cramer, T. van der Sar, V. V. Dobrovitski, and R. Hanson, *Universal control and error correction in multi-qubit spin registers in diamond,* Nature nanotechnology **9**, 171 (2014).

[30] J. J. Wallman, M. Barnhill, and J. Emerson, *Robust characterization of leakage errors,* New J. Phys. **18**, 043021 (2016).

[31] C. J. Wood and J. M. Gambetta, *Quantification and characterization of leakage errors,* Phys. Rev. A **97**, 032306 (2018).

References for Chapter 5

[1] D. Gottesman, *An introduction to quantum error correction and fault-tolerant quantum computation,* , 13 (2010).

[2] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, *Quantum state tomography via compressed sensing,* Physical Review Letters **105** (2010), 10.1103/physrevlett.105.150401.

[3] D. P. DiVincenzo, D. W. Leung, and B. M. Terhal, *Quantum Data Hiding,* IEEE Trans. Inf. Theory **48**, 580 (2002), arXiv:0103098 [quant-ph] .

[4] S. Aaronson and D. Gottesman, *Improved simulation of stabilizer circuits,* Physical Review A **70** (2004), 10.1103/physreva.70.052328.

[5] 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**, 012304 (2009).

[6] E. Magesan, J. M. Gambetta, and J. Emerson, *Scalable and robust randomized benchmarking of quantum processes,* Physical Review Letters **106** (2011), 10.1103/physrevlett.106.180504.

[7] D. Gross, K. Audenaert, and J. Eisert, *Evenly distributed unitaries: On the structure of unitary designs,* Journal of Mathematical Physics **48**, 052104 (2007).

[8] R. Goodman and N. R. Wallach, *Symmetry, Representations, and Invariants* (Springer New York, 2009).

[9] H. Zhu, *Multiqubit clifford groups are unitary 3-designs,* Physical Review A **96** (2017), 10.1103/physreva.96.062336.

[10] Z. Webb, *The Clifford group forms a unitary 3-design,* ArXiv e-prints:1510.02769 (2015), arXiv:1510.02769 .

[11] J. J. Wallman and S. T. Flammia, *Randomized benchmarking with confidence,* New Journal of Physics **16**, 103032 (2014).

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

[13] W. Fulton and J. Harris, *Representation Theory* (Springer New York, 2004).

[14] R. A. Low, *Large deviation bounds for k-designs,* Proceedings of the Royal Society of London Series A **465**, 3289 (2009), arXiv:0903.5236 [quant-ph] .

[15] H. Zhu, R. Kueng, M. Grassl, and D. Gross, *The Clifford group fails gracefully to be a unitary 4-design,* ArXiv e-prints:1609.08172 (2016), arXiv:1609.08172 [quant-ph] .

[16] R. Kueng, H. Zhu, and D. Gross, *Distinguishing quantum states using Clifford orbits,* ArXiv e-prints:1609.08595 (2016), arXiv:1609.08595 [quant-ph] .

[17] R. Kueng, H. Zhu, and D. Gross, *Low rank matrix recovery from Clifford orbits,* ArXiv e-prints:1610.08070 (2016), arXiv:1610.08070 [cs.IT] .

[18] 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,* Physical Review A **77** (2008), 10.1103/physreva.77.012307.

[19] J. J. Wallman, C. Granade, R. Harper, and S. T. Flammia, *Estimating the Coherence of Noise,* New J. Phys. **17**, 113020 (2015), arXiv:arXiv:1503.0786 .

References for Chapter 6

[1] E. Magesan, J. M. Gambetta, and J. Emerson, *Characterizing quantum gates via randomized benchmarking,* Phys. Rev. A **85**, 042311 (2012).

[2] J. J. Wallman and S. T. Flammia, *Randomized benchmarking with confidence,* New J. Phys. **16**, 103032 (2014).

[3] J. M. Epstein, A. W. Cross, E. Magesan, and J. M. Gambetta, *Investigating the limits of randomized benchmarking protocols,* Phys. Rev. A **89**, 062321 (2014), arXiv:1308.2928 [quant-ph] .

[4] C. Granade, C. Ferrie, and D. G. Cory, *Accelerated Randomized Benchmarking,* New J. Phys. **17**, 013042 (2014), arXiv:1404.5275v1 .

[5] H. Zhu, R. Kueng, M. Grassl, and D. Gross, *The Clifford group fails gracefully to be a unitary 4-design,* (2016), arXiv:1609.08172 .

[6] 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**, 012307 (2008).

[7] J. T. Muhonen, A. Laucht, S. Simmons, J. P. Dehollain, R. Kalra, F. E. Hudson, S. Freer, K. M. Itoh, D. N. Jamieson, J. C. McCallum, A. S. Dzurak, and A. Morello, *Quantifying the quantum gate fidelity of single-atom spin qubits in silicon by randomized benchmarking,* Journal of Physics Condensed Matter **27**, 154205 (2015), arXiv:1410.2338 [quant-ph] .

[8] M. A. Fogarty, M. Veldhorst, R. Harper, C. H. Yang, S. D. Bartlett, S. T. Flammia, and A. S. Dzurak, *Nonexponential fidelity decay in randomized benchmarking with low-frequency noise,* Phys. Rev. A **92**, 022326 (2015), arXiv:1502.05119 .

[9] J. V. Beck and K. J. Arnold, *Parameter estimation in engineering and science,* AIChE Journal **24**, 367 (1978).

[10] W. Hoeffding, *Probability inequalities for sums of bounded random variables,* Journal of the American Statistical Association **58**, 13 (1963).

[11] E. Magesan, J. M. Gambetta, and J. Emerson, *Scalable and Robust Randomized Benchmarking of Quantum Processes,* Phys. Rev. Lett. **106**, 180504 (2011).

[12] G. A. F. Seber and C. J. Wild, *Nonlinear Regression* (John Wiley & Sons, Inc., 1989).

[13] J. J. Wallman, C. Granade, R. Harper, and S. T. Flammia, *Estimating the Coherence of Noise,* New J. Phys. **17**, 113020 (2015), arXiv:arXiv:1503.0786 .

[14] 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**, 080505 (2012), arXiv:1203.4550 .

[15] A. C. Dugas, J. J. Wallman, and J. Emerson, *Characterizing Universal Gate Sets via Dihedral Benchmarking,* (2015), arXiv:1508.06312 .

[16] A. W. Cross, E. Magesan, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, *Scalable randomized benchmarking of non-clifford gates,* Npj Quantum Information **2**, 16012 (2016), arXiv:1510.02720 .

[17] 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**, 012304 (2009).

[18] P. S. Turner and D. Markham, *Derandomizing quantum circuits with measurement-based unitary designs,* Phys. Rev. Lett. **116**, 200501 (2016).

[19] A. K. Hashagen, S. T. Flammia, D. Gross, and J. J. Wallman, *Real Randomized Benchmarking,* (2018), arXiv:1801.06121 .

[20] R. Harper and S. Flammia, *Fault tolerance in the IBM Q Experience,* (2018),

arXiv:1806.02359 .

[21] D. Gross, K. Audenaert, and J. Eisert, *Evenly distributed unitaries: On the structure of unitary designs,* J. Math. Phys. **48**, 052104 (2007).

[22] V. V. Fedorov, *Theory of optimal experiments* (Elsevier, 2013).

[23] T. Proctor, K. Rudinger, K. Young, M. Sarovar, and R. Blume-Kohout, *What randomized benchmarking actually measures,* Physical review letters **119**, 130502 (2017).

[24] J. J. Wallman, *Randomized benchmarking with gate-dependent noise,* Quantum **2**, 47 (2018).

[25] M. Wolf, *Quantum channels operations: Guided tour.* Lecture Notes (2012).

[26] J. J. Wallman and J. Emerson, *Bounding experimental quantum error rates relative to fault-tolerant thresholds,* (2015), arXiv:1512.01098 .

[27] D. P. DiVincenzo, D. W. Leung, and B. M. Terhal, *Quantum Data Hiding,* IEEE Trans. Inf. Theory **48**, 580 (2001), arXiv:0103098 [quant-ph] .

[28] M. A. Nielsen, *A simple formula for the average gate fidelity of a quantum dynamical operation,* Phys. Lett. A **303**, 249 (2002).

[29] E. Magesan, R. Blume-Kohout, and J. Emerson, *Gate fidelity fluctuations and quantum process invariants,* Phys. Rev. A **84**, 012309 (2011).

[30] R. Sagastizabal, X. Bonet-Monroig, T. O'Brien, M. Singh, M. Rol, C. Bultink, X. Fu, N. Muthusubramanian, N. Bruno, and L. DiCarlo, *Error mitigation by symmetry verification on a variational quantum eigensolver,* In Preparation (2018).

[31] D. Riste, J. Van Leeuwen, H.-S. Ku, K. W. Lehnert, and L. DiCarlo, *Initialization by measurement of a superconducting quantum bit circuit,* Physical review letters **109**, 050507 (2012).

[32] J. Emerson, E. Livine, and S. Lloyd, *Convergence conditions for random quantum circuits,* Physical Review A **72**, 060302 (2005).

[33] I. L. Chuang and M. A. Nielsen, *Prescription for experimental determination of the dynamics of a quantum black box,* J. Mod. Opt. **44**, 2455 (1997).

[34] J. J. Wallman, *Bounding experimental quantum error rates relative to fault-tolerant thresholds,* (2015), arXiv:1511.00727 .

[35] J. Helsen, J. J. Wallman, and S. Wehner, *The two copy representation of the multiqubit clifford group,* (2016), arXiv:1609.08188 .

[36] H. Zhu, *Multiqubit Clifford groups are unitary 3-designs,* , 1 (2015), arXiv:1510.02619

.

[37] D. Gottesman, *Theory of fault-tolerant quantum computation,* Phys. Rev. A **57**, 127 (1998).

[38] A. S. Holevo, *Additivity conjecture and covariant channels,* International Journal of Quantum Information **03**, 41 (2005), http://www.worldscientific.com/doi/pdf/10.1142/S0219749905000530 .

[39] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, New York, NY, USA, 2004).

References for Chapter 7

[1] E. Magesan, J. M. Gambetta, and J. Emerson, *Characterizing quantum gates via randomized benchmarking,* Physical Review A **85**, 042311 (2012), arXiv:1109.6887 .

[2] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, *Estimating the coherence of noise,* New Journal of Physics **17**, 113020 (2015), arXiv:1503.0786 .

[3] D. Gross, K. Audenaert, and J. Eisert, *Evenly distributed unitaries: On the structure of unitary designs,* Journal of Mathematical Physics **48**, 052104 (2007), arXiv:0611002 [quant-ph] .

[4] J. Helsen, J. J. Wallman, S. T. Flammia, and S. Wehner, *Multi-qubit Randomized Benchmarking Using Few Samples,* (2017), arXiv:1701.04299 .

[5] C. Granade, C. Ferrie, and D. G. Cory, *Accelerated randomized benchmarking,* New Journal of Physics **17**, 013042 (2015), arXiv:1404.5275 .

[6] 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,* Physical Review A **77**, 012307 (2008), arXiv:0707.0963 .

[7] W. Hoeffding, *Probability Inequalities for Sums of Bounded Random Variables,* Journal of the American Statistical Association **58**, 301 (1963).

[8] N. Johnston, *QETLAB: A MATLAB toolbox for quantum entanglement, version 0.9,* (2016).

[9] J. M. Epstein, A. W. Cross, E. Magesan, and J. M. Gambetta, *Investigating the limits of randomized benchmarking protocols,* Physical Review A **89**, 062321 (2014), arXiv:1308.2928 .

[10] I. Hincks, J. J. Wallman, C. Ferrie, C. Granade, and D. G. Cory, *Bayesian Inference for Randomized Benchmarking Protocols,* arXiv (2018), arXiv:1802.00401 .

[11] C. A. Fuchs and J. van de Graaf, *Cryptographic Distinguishability Measures for Quantum Mechanical States,* IEEE Transactions on Information Theory **45**, 1216 (1999), arXiv:9712042 [quant-ph] .

[12] N. A. Weiss, *A Course in Probability* (Addison-Wesley, 2005) pp. 380–383.

[13] H. Zhu, R. Kueng, M. Grassl, and D. Gross, *The Clifford group fails gracefully to be a unitary 4-design,* arXiv (2016), arXiv:1609.08172 .

[14] H. Zhu, *Multiqubit Clifford groups are unitary 3-designs,* Physical Review A **96**, 062336 (2017), arXiv:1510.02619 .

[15] J. Helsen, J. J. Wallman, and S. Wehner, *Representations of the multi-qubit Clifford group,* Journal of Mathematical Physics **59**, 072201 (2018), arXiv:1609.08188 .

[16] D. Pérez-García, M. M. Wolf, D. Petz, and M. B. Ruskai, *Contractivity of positive and trace-preserving maps under LPnorms,* Journal of Mathematical Physics **47**, 083506 (2006), arXiv:0601063 [math-ph] .

[17] T. Chasseur and F. K. Wilhelm, *Complete randomized benchmarking protocol accounting for leakage errors,* Physical Review A **92**, 042333 (2015), arXiv:1505.00580 .

[18] T. Proctor, K. Rudinger, K. Young, M. Sarovar, and R. Blume-Kohout, *What Randomized Benchmarking Actually Measures,* Physical Review Letters **119**, 130502 (2017), arXiv:1702.01853 .

[19] J. J. Wallman, *Randomized benchmarking with gate-dependent noise,* Quantum **2**, 47 (2018), arXiv:1703.09835 .

[20] J. J. Wallman and S. T. Flammia, *Randomized benchmarking with confidence,* New Journal of Physics **16**, 103032 (2014), arXiv:1404.6025 .

[21] C. Dankert, R. Cleve, J. Emerson, and E. Livine, *Exact and Approximate Unitary 2-Designs: Constructions and Applications,* Physical Review A **80**, 012304 (2009), arXiv:0606161 [quant-ph] .

[22] A. Carignan-Dugas, J. J. Wallman, and J. Emerson, *Characterizing universal gate sets via dihedral benchmarking,* Physical Review A **92**, 060302 (2015), arXiv:1508.06312 .

[23] A. W. Cross, E. Magesan, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, *Scalable randomized benchmarking of non-Clifford gates,* npj Quantum Information **2**, 16012 (2016), arXiv:1510.02720 .

[24] D. S. França and A.-L. Hashagen, *Approximate Randomized Benchmarking for Finite Groups,* arXiv (2018), arXiv:1803.03621 .

[25] A. K. Hashagen, S. T. Flammia, D. Gross, and J. J. Wallman, *Real Randomized Benchmarking,* arXiv (2018), arXiv:1801.06121 .

[26] W. G. Brown and B. Eastin, *Randomized benchmarking with restricted gate sets,* Physical Review A **97**, 062323 (2018), arXiv:1801.04042 .

[27] J. Helsen, X. Xue, L. M. K. Vandersypen, and S. Wehner, *A new class of efficient randomized benchmarking protocols,* (2018), arXiv:1806.02048 .

[28] M. A. Nielsen, *A simple formula for the average gate fidelity of a quantum dynamical operation,* Physics Letters A **303**, 249 (2002), arXiv:0205035 [quant-ph] .

[29] J. Watrous, *The Theory of Quantum Information* (Cambridge Academic Press, 2018).

[30] D. Braun, O. Giraud, I. Nechita, C. Pellegrini, and M. Žnidarič, *A universal set of qubit quantum channels,* Journal of Physics A **47**, 135302 (2014), arXiv:1306.0495 .

[31] R. A. Horn and C. R. Johnson, *Matrix analysis*, 2nd ed. (Cambridge University Press, 2013).

[32] C. D. Aliprantis and O. Burkinshaw, *Principles of Real Analysis*, 3rd ed. (Academic Press, 1998).

[33] W. Fulton and J. Harris, *Representation Theory* (Springer, 2004).

References for Chapter 8

[1] E. Magesan, J. M. Gambetta, and J. Emerson, *Characterizing quantum gates via randomized benchmarking,* Phys. Rev. A **85**, 042311 (2012).

[2] A. W. Cross, E. Magesan, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, *Scalable randomised benchmarking of non-clifford gates,* npj Quantum Information **2** (2016).

[3] W. G. Brown and B. Eastin, *Randomized benchmarking with restricted gate sets,* arXiv preprint arXiv:1801.04042 (2018).

[4] A. Hashagen, S. Flammia, D. Gross, and J. Wallman, *Real randomized benchmarking,* arXiv preprint arXiv:1801.06121 (2018).

[5] D. S. França and A.-L. Hashagen, *Approximate randomized benchmarking for finite groups,* arXiv preprint arXiv:1803.03621 (2018).

[6] A. Carignan-Dugas, J. J. Wallman, and J. Emerson, *Characterizing universal gate sets via dihedral benchmarking,* Physical Review A **92**, 060302 (2015).

[7] 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).

[8] S. T. Flammia and Y.-K. Liu, *Direct Fidelity Estimation from Few Pauli Measurements,* Phys. Rev. Lett. **106**, 230501 (2011).

[9] 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**, 012307 (2008).

[10] J. T. Muhonen, A. Laucht, S. Simmons, J. P. Dehollain, R. Kalra, F. E. Hudson, S. Freer, K. M. Itoh, D. N. Jamieson, J. C. McCallum, A. S. Dzurak, and A. Morello, *Quantifying the quantum gate fidelity of single-atom spin qubits in silicon by randomized benchmarking,* Journal of Physics Condensed Matter **27**, 154205 (2015), arXiv:1410.2338 [quant-ph] .

[11] J. Helsen, J. J. Wallman, S. T. Flammia, and S. Wehner, *Multi-qubit randomized benchmarking using few samples,* arXiv preprint arXiv:1701.04299 (2017).

[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).

[13] 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**, 012304 (2009).

[14] A. C. Dugas, J. J. Wallman, and J. Emerson, *Efficiently characterizing the total error in quantum circuits,* arXiv preprint arXiv:1610.05296 (2016).

[15] A. D. Córcoles, J. M. Gambetta, J. M. Chow, J. A. Smolin, M. Ware, J. Strand, B. L. Plourde, and M. Steffen, *Process verification of two-qubit quantum gates by randomized benchmarking,* Physical Review A **87**, 030301 (2013).

[16] T. Watson, S. Philips, E. Kawakami, D. Ward, P. Scarlino, M. Veldhorst, D. Savage, M. Lagally, M. Friesen, S. Coppersmith, *et al.*, *A programmable two-qubit quantum processor in silicon,* Nature **555**, 633 (2018).

[17] W. Fulton and J. Harris, *Representation Theory: A First Course*, Readings in Mathematics (Springer-Verlag New York, 2004).

[18] W. Hoeffding, *Probability inequalities for sums of bounded random variables,* Journ. Am. Stat. Assoc. **58**, 13 (1963).

[19] J. J. Wallman and S. T. Flammia, *Randomized benchmarking with confidence,* New J. Phys. **16**, 103032 (2014).

[20] D. M. Titterington, A. F. Smith, and U. E. Makov, *Statistical analysis of finite mixture distributions* (Wiley„ 1985).

[21] J. J. Wallman, *Randomized benchmarking with gate-dependent noise,* Quantum **2**, 47 (2018).

[22] J. Helsen, J. J. Wallman, and S. Wehner, *The two copy representation of the multiqubit clifford group,* arXiv:1609.08188, to appear in J. Math. Phys. (2016).

[23] C. R. MacCluer, *The many proofs and applications of perron's theorem,* Siam Review **42**, 487 (2000).

[24] J. J. Sakurai, J. Napolitano, *et al.*, *Modern quantum mechanics*, Vol. 261 (Pearson, 2014).

[25] J. Watrous, *Theory of quantum information,* CS 766/QIC 820 lecture notes, University of Waterloo (2011).

[26] S. T. Merkel, E. J. Pritchett, and B. H. Fong, *Randomized benchmarking as convolution: Fourier analysis of gate dependent errors,* arXiv preprint arXiv:1804.05951 (2018).

References for Chapter 9

[1] I. L. Chuang and M. A. Nielsen, *Prescription for experimental determination of the dynamics of a quantum black box,* Journal of Modern Optics **44**, 2455 (1997).

[2] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White, *Quantum process tomography of a controlled-not gate,* Phys. Rev. Lett. **93**, 080502 (2004).

[3] 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**, 062119 (2013).

[4] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D. G. Cory, and R. Laflamme, *Symmetrized characterization of noisy quantum processes,* Science **317**, 1893 (2007).

[5] 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**, 012307 (2008).

[6] 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**, 260503 (2012).

[7] J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, *Randomized benchmarking and process tomography for gate errors in a solid-state qubit,* Phys. Rev. Lett. **102**, 090502 (2009).

[8] 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**, 080505 (2012).

[9] 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**, 240504 (2012).

[10] A. D. Córcoles, J. M. Gambetta, J. M. Chow, J. A. Smolin, M. Ware, J. Strand, B. L. T. Plourde, and M. Steffen, *Process verification of two-qubit quantum gates by randomized benchmarking*, Phys. Rev. A **87**, 030301 (2013).

[11] A. C. Dugas, J. J. Wallman, and J. Emerson, *Efficiently characterizing the total error in quantum circuits,* arXiv:1610.05296 (2016).

[12] 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**, 220502 (2014).

[13] L. Casparis, T. W. Larsen, M. S. Olsen, F. Kuemmeth, P. Krogstrup, J. Nygård, K. D. Petersson, and C. M. Marcus, *Gatemon benchmarking and two-qubit operations,* Phys. Rev. Lett. **116**, 150505 (2016).

[14] M. D. Shulman, O. E. Dial, S. P. Harvey, H. Bluhm, V. Umansky, and A. Yacoby, *Demonstration of entanglement of electrostatically coupled singlet-triplet qubits,* Science **336**, 202 (2012).

[15] T. F. Watson, S. G. J. Philips, E. Kawakami, D. R. Ward, P. Scarlino, M. Veldhorst, D. E. Savage, M. G. Lagally, M. Friesen, S. N. Coppersmith, M. A. Eriksson, and L. M. K. Vandersypen, *A programmable two-qubit quantum processor in silicon,* Nature **555**, 633 (2018).

[16] D. M. Zajac, A. J. Sigillito, M. Russ, F. Borjans, J. M. Taylor, G. Burkard, and J. R. Petta, *Resonantly driven cnot gate for electron spins,* Science **359**, 439 (2017).

[17] W. Huang, C. H. Yang, K. W. Chan, T. Tanttu, B. Hensen, R. C. C. Leon, M. A. Fogarty, J. C. C. Hwang, F. E. Hudson, K. M. Itoh, A. Morello, A. Laucht, and A. S. Dzurak, *Fidelity benchmarks for two-qubit gates in silicon,* arXiv:1805.05027 (2018).

[18] F. A. Zwanenburg, A. S. Dzurak, A. Morello, M. Y. Simmons, L. C. L. Hollenberg, G. Klimeck, S. Rogge, S. N. Coppersmith, and M. A. Eriksson, *Silicon quantum electronics,* Rev. Mod. Phys. **85**, 961 (2013).

[19] L. M. K. Vandersypen, H. Bluhm, J. S. Clarke, A. S. Dzurak, R. Ishihara, A. Morello, D. J. Reilly, L. R. Schreiber, and M. Veldhorst, *Interfacing spin qubits in quantum dots and donors–hot, dense, and coherent,* npj Quantum Information **3**, 34 (2017).

[20] M. Pioro-Ladrière, T. Obata, Y. Tokura, Y.-S. Shin, T. Kubo, K. Yoshida, T. Taniyama, and S. Tarucha, *Electrically driven single-electron spin resonance in a slanting zeeman field,* Nature Physics **4**, 776 (2008).

[21] L. M. K. Vandersypen and I. L. Chuang, *Nmr techniques for quantum control and computation,* Rev. Mod. Phys. **76**, 1037 (2005).

[22] T. Meunier, V. E. Calado, and L. M. K. Vandersypen, *Efficient controlled-phase gate for single-spin qubits in quantum dots,* Phys. Rev. B **83**, 121403 (2011).

[23] M. Veldhorst, C. H. Yang, J. C. C. Hwang, W. Huang, J. P. Dehollain, J. T. Muhonen, S. Simmons, A. Laucht, F. E. Hudson, K. M. Itoh, A. Morello, and A. S. Dzurak, *A two-qubit logic gate in silicon,* Nature **526**, 410 (2015).

[24] J. M. Elzerman, R. Hanson, L. H. Willems van Beveren, B. Witkamp, L. M. K. Vandersypen, and L. P. Kouwenhoven, *Single-shot read-out of an individual electron spin in a quantum dot,* Nature **430**, 431 (2004).

[25] V. Srinivasa, K. C. Nowack, M. Shafiei, L. M. K. Vandersypen, and J. M. Taylor, *Simultaneous spin-charge relaxation in double quantum dots,* Phys. Rev. Lett. **110**, 196803 (2013).

[26] J. J. Wallman, *Randomized benchmarking with gate-dependent noise,* Quantum **2**, 47 (2018).

[27] 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 Journal of Physics **20**, 092001 (2018).

[28] T. Proctor, K. Rudinger, K. Young, M. Sarovar, and R. Blume-Kohout, *What randomized benchmarking actually measures,* Phys. Rev. Lett. **119**, 130502 (2017).

[29] F. Martins, F. K. Malinowski, P. D. Nissen, E. Barnes, S. Fallahi, G. C. Gardner, M. J. Manfra, C. M. Marcus, and F. Kuemmeth, *Noise suppression using symmetric exchange gates in spin qubits,* Phys. Rev. Lett. **116**, 116801 (2016).

[30] M. D. Reed, B. M. Maune, R. W. Andrews, M. G. Borselli, K. Eng, M. P. Jura, A. A. Kiselev, T. D. Ladd, S. T. Merkel, I. Milosavljevic, E. J. Pritchett, M. T. Rakher, R. S. Ross, A. E. Schmitz, A. Smith, J. A. Wright, M. F. Gyure, and A. T. Hunter, *Reduced sensitivity to charge noise in semiconductor spin qubits via symmetric operation,* Phys. Rev. Lett. **116**, 110402 (2016).

References for Chapter 10

[1] J. Colless, *Control and readout of scaled-up quantum dot systems,* PhD Thesis, University of Sydney (2014).

[2] L. M. K. Vandersypen, H. Bluhm, J. S. Clarke, A. S. Dzurak, R. Ishihara, A. Morello, D. J. Reilly, L. R. Schreiber, and M. Veldhorst, *Interfacing spin qubits in quantum dots and donors—hot, dense, and coherent,* npj Quantum Information **3** (2017), 10.1038/s41534-017-0038-y.

[3] C. D. Hill, E. Peretz, S. J. Hile, M. G. House, M. Fuechsle, S. Rogge, M. Y. Simmons, and L. C. Hollenberg, *A surface code quantum computer in silicon,* Science advances **1**, e1500707 (2015).

[4] R.Li, L. Petit, D. Franke, J. Dehollain, J. Helsen, M. Steudtner, N. Thomes, Z. Yoscovits, K. Singh, S. Wehner, L. Vandersypen, J. Clarke, and M. Veldhorst, *A crossbar network for silicon quantum dot qubits,* arXiv preprint arXiv:1711.03807 (2017).

[5] M. Veldhorst, H. G. J. Eenink, C. H. Yang, and A. S. Dzurak, *Silicon CMOS architecture for a spin-based quantum computer,* Nature Communications **8** (2017), 10.1038/s41467-017-01905-6.

[6] D. Gottesman, *Theory of fault-tolerant quantum computation,* Physical Review A **57**, 127 (1998).

[7] D. A. Lidar, T. A. Brun, and T. Brun, eds., *Quantum Error Correction* (Cambridge University Press, 2009).

[8] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, *Topological quantum memory,* Journal of Mathematical Physics **43**, 4452 (2002).

[9] H. Bombin and M. A. Martin-Delgado, *Topological quantum distillation,* Physical Review Letters **97** (2006), 10.1103/physrevlett.97.180501.

[10] R. Versluis, S. Poletto, N. Khammassi, B. Tarasinski, N. Haider, D. Michalak, A. Bruno, K. Bertels, and L. DiCarlo, *Scalable quantum circuit and control for a superconducting surface code,* Physical Review Applied **8** (2017), 10.1103/physrevapplied.8.034021.

[11] T. Heijmen, *Soft errors from space to ground: Historical overview, empirical evidence, and future trends, Soft Errors in Modern Electronic Systems*, , 1 (2010).

[12] T. Fujita, T. A. Baart, C. Reichl, W. Wegscheider, and L. M. K. Vandersypen, *Coherent shuttle of electron-spin states,* npj Quantum Information **3** (2017), 10.1038/s41534-017-0024-4.

[13] R. Hanson, L. P. Kouwenhoven, J. R. Petta, S. Tarucha, and L. M. K. Vandersypen, *Spins in few-electron quantum dots,* Reviews of Modern Physics **79**, 1217 (2007).

[14] J. M. Taylor, H.-A. Engel, W. Dür, A. Yacoby, C. M. Marcus, P. Zoller, and M. D. Lukin, *Fault-tolerant architecture for quantum computation using electrically controlled semiconductor spins,* Nature Physics **1**, 177 (2005).

[15] M. Veldhorst, J. C. C. Hwang, C. H. Yang, A. W. Leenstra, B. de Ronde, J. P. Dehollain, J. T. Muhonen, F. E. Hudson, K. M. Itoh, A. Morello, and A. S. Dzurak, *An addressable quantum dot qubit with fault-tolerant control-fidelity,* Nature Nanotechnology **9**, 981 (2014).

[16] M. A. Nielsen and I. L. Chuang, *Quantum circuits,* .

[17] J. R. Petta, *Coherent manipulation of coupled electron spins in semiconductor quantum dots,* Science **309**, 2180 (2005).

[18] T. Meunier, V. E. Calado, and L. M. K. Vandersypen, *Efficient controlled-phase gate for single-spin qubits in quantum dots,* Physical Review B **83** (2011), 10.1103/physrevb.83.121403.

[19] T. F. Watson, S. G. J. Philips, E. Kawakami, D. R. Ward, P. Scarlino, M. Veldhorst, D. E. Savage, M. G. Lagally, M. Friesen, S. N. Coppersmith, M. A. Eriksson, and L. M. K. Vandersypen, *A programmable two-qubit quantum processor in silicon,* Nature **555**, 633 (2018).

[20] M. Veldhorst, C. H. Yang, J. C. C. Hwang, W. Huang, J. P. Dehollain, J. T. Muhonen, S. Simmons, A. Laucht, F. E. Hudson, K. M. Itoh, A. Morello, and A. S. Dzurak, *A two-qubit logic gate in silicon,* Nature **526**, 410 (2015).

[21] N. Schuch and J. Siewert, *Natural two-qubit gate for quantum computation using the xy interaction,* Physical Review A **67**, 032301 (2003).

[22] A. L. Gorodentsev, *Algebra I: Textbook for Students of Mathematics* (Springer, 2016).

[23] M. Elsheikh, M. Giesbrecht, A. Novocin, and B. D. Saunders, *Fast computation of smith forms of sparse matrices over local rings,* in *Proceedings of the 37th International Symposium on Symbolic and Algebraic Computation* (ACM, 2012) pp. 146–153.

[24] C. Horsman, A. G. Fowler, S. Devitt, and R. V. Meter, *Surface code quantum computing by lattice surgery,* New Journal of Physics **14**, 123011 (2012).

[25] B. M. Terhal, *Quantum error correction for quantum memories,* Reviews of Modern Physics **87**, 307 (2015).

[26] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, *Surface codes: Towards practical large-scale quantum computation,* Physical Review A **86** (2012), 10.1103/physreva.86.032324.

[27] D. S. Wang, A. G. Fowler, and L. C. L. Hollenberg, *Surface code quantum computing*

with error rates over 1%, Physical Review A **83** (2011), 10.1103/physreva.83.020302.

[28] A. G. Fowler, A. M. Stephens, and P. Groszkowski, *High-threshold universal quantum computation on the surface code,* Physical Review A **80** (2009), 10.1103/physreva.80.052312.

[29] Y. Tomita and K. M. Svore, *Low-distance surface codes under realistic quantum noise,* Physical Review A **90** (2014), 10.1103/physreva.90.062320.

[30] A. M. Tyryshkin, S. Tojo, J. J. L. Morton, H. Riemann, N. V. Abrosimov, P. Becker, H.-J. Pohl, T. Schenkel, M. L. W. Thewalt, K. M. Itoh, and S. A. Lyon, *Electron spin coherence exceeding seconds in high-purity silicon,* Nature Materials **11**, 143 (2011).

[31] A. J. Landahl, J. T. Anderson, and P. R. Rice, *Fault-tolerant quantum computing with color codes,* arXiv preprint arXiv:1108.5738 (2011).

[32] J. R. Wootton, A. Peter, J. R. Winkler, and D. Loss, *Proposal for a minimal surface code experiment,* Physical Review A **96** (2017), 10.1103/physreva.96.032338.

[33] D. K. Tuckett, S. D. Bartlett, and S. T. Flammia, *Ultrahigh error threshold for surface codes with biased noise,* Physical Review Letters **120** (2018), 10.1103/physrevlett.120.050505.

[34] D. B. West *et al.*, *Introduction to graph theory*, Vol. 2 (Prentice hall Upper Saddle River, 2001).

[35] C. Bron and J. Kerbosch, *Algorithm 457: finding all cliques of an undirected graph,* Communications of the ACM **16**, 575 (1973).

[36] R. G. Downey and M. R. Fellows, *Fixed-parameter tractability and completeness II: On completeness for w[1],* Theoretical Computer Science **141**, 109 (1995).

References for Chapter 11

[1] R. Li, L. Petit, D. P. Franke, J. P. Dehollain, J. Helsen, M. Steudtner, N. K. Thomas, Z. R. Yoscovits, K. J. Singh, S. Wehner, *et al.*, *A crossbar network for silicon quantum dot qubits,* Science advances **4**, eaar3960 (2018).

[2] GAP, *GAP – Groups, Algorithms, and Programming, Version 4.10.0*, The GAP Group (2018).

[3] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language,* J. Symbolic Comput. **24**, 235 (1997), computational algebra and number theory (London, 1993).

[4] M. A. Rol, F. Malinkowski, B. Francesco, B. Tarasinski, B. Terhal, and L. DiCarlo, *Atomic flux pulses for a superconducting quantum processor,* in preparation (2018).

[5] L. Vandersypen, H. Bluhm, J. Clarke, A. Dzurak, R. Ishihara, A. Morello, D. Reilly, L. Schreiber, and M. Veldhorst, *Interfacing spin qubits in quantum dots and donors—hot, dense, and coherent,* npj Quantum Information **3**, 34 (2017).

[6] R. Versluis, S. Poletto, N. Khammassi, B. Tarasinski, N. Haider, D. Michalak, A. Bruno, K. Bertels, and L. DiCarlo, *Scalable quantum circuit and control for a superconducting surface code,* Physical Review Applied **8**, 034021 (2017).