Characterizing Quantum Devices Using the Principles of Quantum Information

by

Anthony M. Polloreno

B.A., University of California, Berkeley, 2016

M.S., University of Colorado at Boulder, 2022

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Physics

2023

Committee Members:

Graeme Smith, Chair

Joshua Combes

Oliver DeWolfe

Murray Holland

Daniel Slichter

James Thompson

Polloreno, Anthony M. (Ph.D., Physics)

Characterizing Quantum Devices Using the Principles of Quantum Information

Thesis directed by Prof. Graeme Smith

The title of this thesis is "Characterizing Quantum Devices Using the Principles of Quantum Information", named after the NSTGRO fellowship of the same name. As with any work, the focus has become more clear after the fact. The structure is roughly as follows:

(1) Quantum computation.

(2) Quantum metrology.

(3) Quantum characterization.

This thesis outlines work pertaining to the development and characterization of quantum devices. The suggested methodologies borrow ideas and rely on techniques from quantum information theory. Importantly, this thesis is meant to be a thesis on theoretical physics, rather than mathematics or computer science, and so we hope that the reader is able to glean physical intuition from this work.

Arguably the most important question in recent years is "Is this a quantum computer?" This answer, in all of its various realizations, is the prerogative and impetus of the field of quantum characterization, verification and validation (QCVV). We will generalize this question to "Is this a quantum device," provide theoretical motivation for a number of interesting experiments, and develop theories that we hope provide new insights into the field and have the potential to lead to further advancements in quantum technology.

Specifically, seven projects are discussed: developing single-site rotations in a Penning trap as a scalable mode of quantum computation, enabling large-scale quantum optimization in a neutral atom trap, benchmarking quantum computers with a quantum error detecting code, bounding the integrated quantum Fisher information for metrological protocols, developing the theory of

reservoir computing with stochastic reservoirs, rigorously deriving a theory of direct randomized benchmarking and understanding the impact of Markovian errors in random circuits.

The first two projects are experimental collaborations working towards quantum computation with hundreds of qubits. In recent years quantum devices appear poised to break the hundred qubit mark, with varying levels of control and fidelity. Surely there is much work to be done in developing these technologies if we want to perform genuine quantum computation, but the demonstration of control over a large number of quantum degrees of freedom is, in itself, a marvel of engineering and demonstrates a growth of our mastery of physics. In these two projects, we propose a new method of single-qubit control in an ion trap which routinely has hundreds of qubits, and investigate the impact of measurement rate of neutral atom traps on our ability to execute quantum algorithms.

The next three projects work towards developing techniques to demonstrate genuine quantum information theoretic phenomena. The first work is simple, but was done at a time when many results were coming out suggesting the demonstration of fault tolerance, when they were instead implementing fault tolerant circuits, above threshold. We propose a simple Bacon-Shor code, over a family of circuits, that should be easy to implement on existing hardware, with current error rates, that should give a clear demonstration of actual fault-tolerance - a break-even of the logical error rate and physical error rate. While quantum computation has stolen the scene, quantum metrology has already been demonstrated to provide significant improvement over classical methods, if only in the context of scientific experiments. In particular, the probing of quantum gravitational effects and dark matter are only enabled by considering quantum devices. We first explore an application of sensing that relates to both dark matter detection and quantum computation, and then consider a theoretical investigation of a computational technique that may be able to leverage these metrological advantages.

Finally, the last two chapters work towards the more general goal of characterizing quantum circuits. Despite significant research into the characterization of quantum states and gates, it is relatively poorly understood how errors impact the performance of structured circuits such as those used in error correction. By using techniques from the theory of randomized benchmarking we first

establish an extremely general result about the decay rate of a large class of random circuits, and then use this result to understand how certain physical error rates impact the success probabilities of those circuits.

This work was done in collaboration with others. With the exception of the last chapter, all chapters have included with them a reference to the published article or preprint. The sections of these papers that were done primarily by another author have been elided from this document. The interested reader should refer to the referenced works to see the elided results.

Finally, this work was supported by a variety of funding agencies that I would like to acknowledge support from. This work was supported by a NASA Space Technology Graduate Research Opportunity award, the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator (QSA), the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR001120C0068, the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Quantum Testbed Pathfinder, the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), the NSF JILA PFC grant 1734006 and NSF award DMR-1747426.

Dedication

This thesis is dedicated to my family.

Acknowledgements

My thesis is the culmination of four years of work, but of course, I myself am currently the culmination of 28 years of work, and that is, emphatically, a collaborative piece!

Naturally, the first thanks go to my family - Mom, Dad, Jana, Nana, and Pa - thank you all for supporting me and loving me. Briefly:

Dad - you frequently told me, "follow your passion, discover your bliss," and looking back, this has been apt advice and has paid dividends.

Jana - thank you for growing with me, being supportive, speaking your mind, being true to yourself, and being a great friend and sister.

Pa - besides imparting on me your sense of humor, you've always instilled in me a passion for science. There is a lot more I could say, but it suffices to say that a passion for science has been a large part of my life.

Nana - thank you for always being there for me, looking out for me, and helping me in every and any way you can.

Mom - thanks for being my daily hype squad, support group, team captain, and coach. It looks like I "got that Ph.D." Especially - especially - thank you for being a grounded source of stability nearly every day. You have had an effect on my life that is hard to put into words. Surely, for the five of you, there is much more to say, but perhaps in brevity we can find abundance.

Besides my family, I've had the enormous blessing of a large number of friends and mentors who have significantly shaped not only my career, but also who I am.

Alexander Papageorge - naturally, I must begin with you! How different my life would have

been if I hadn't met you at Rigetti! Your advice, friendship, support, and patience, among other things, have been essential to my development. I wouldn't be the scientist I am today, nor the man, without having met you - and both of those are of equal importance. Your influence on me as a young physicist at Rigetti was substantial. Your insistence on precision, intuition, and general bias towards pedantic pedagogy grounded me in a very important way. I look forward to our future ventures together, and I am happy beyond words that I have had the opportunity to both be taught by you and be your friend.

Nikolas Tezak - while working with you was where our friendship began, the past few years in Boulder have been where it has blossomed. Your advice, both technical and personal, has profoundly influenced me. You are a truly fantastic and modern scientist, and your style of problem solving has molded me into a more sensible and committed physicist. Fittingly, it was from you that I heard the suggestion to choose who you work with, rather than what you work on. Following this advice, I am extremely happy we were able to work on the quantum reservoir computing work together. Most importantly for this thesis, I have had the opportunity to be exposed to your physical intuition, understanding, artistry, analytic skills, coding abilities, pragmatism, and company.

Eric Peterson - you're not only a great and patient teacher of mathematics, but also a truly wonderful friend. I am incredibly happy that we have spoken as regularly as we have, both because it means I get to learn about algebraic topology, but also because it means that we have had a chance to further our friendship. To the first point, you have maintained in me a feeling of mathematical competency and eagerness that I hope to capitalize on. To the second point, my life has been substantially richer due to our conversations. I say with confidence that you have had a profound influence on me, both as a mentor and a friend.

Prasahnt Sivarajah - you've been a constant friend since we met at Rigetti. You are surely a mentor, but most importantly a colleague. Thank you for reading with me, thank you for always expressing interest in my personal and professional life and success, and thank you for being a constant source of scientific camaraderie. While it is unrelated to my thesis, thank you also for being a very accommodating neighbor at Rigetti. It was surely on the red team that my career as

a scientist began to blossom.

Reuben Wang - in addition to being a lovely person, thank you for being "that-other-student" in the program that I got to do science with. You made Boulder feel more like a home and community to me.

Graeme Smith - naturally my advisor should find a place on this list! Thank you for advocating for me, for providing interesting problems, and helping me achieve my goals. I look forward to crossing paths with you again in the future. I'm not sure I could have had a better advisor than you. There is surely more I could say, but, as with my family above, perhaps in the brevity of my thanks, we can find depth in the space around it.

Ana Maria Rey - thank you for working very much as a second local advisor to me. You taught me a lot about being a diligent scientist, and I feel extremely fortunate for having had the opportunity to work with you.

Robin Blume-Kohout - you've provided something that is quintessential but hard to communicate. A long time ago, I tried to burn off any kind of mystical romanticism about doing science that I had. You rekindled that in me and helped me double down on a project. Thank you for listening, investing, and contributing. As a child, I found a lot of joy in believing in magic and reading fantasy. You've convinced me that it's tenable to continue finding that magic in science. Furthermore, you have pushed me to develop my scientific communication, something that I am grateful for.

Tim Proctor - I think you are the first example of an early-late stage career scientist I've had the pleasure of working with. You already do, and will continue to, serve as an excellent archetype of how I should push through the next few years of my life. Thank you for always being jovial and for taking me under your wing. You are wonderful to work with, and I am grateful for the learning opportunities, friendship, and projects we've been able to share.

Kevin Young - thank you for working with me in 2015! That's really when this all started for me. I'm glad we finally got our paper published. But most importantly, you've been a great mentor and friend, and I'm glad to have you in my life.

Eleanor Rieffel - thank you for supporting my research goals, providing helpful advice, and furthering my interest in the mathematics of quantum computing. I look forward to our future collaborations.

To you all, a lesson I learned at Rigetti, and relearned in graduate school, was that working with people you like is extremely important. I will be forever grateful that I have managed to find you all.

It takes a village, and I am so happy to have found one.

And beyond the matter at hand, there is the fact that growing through your 20s and living through a global pandemic is hard. I've leaned heavily on my friends and they must be called out.

Chadwick Casper - I'm glad that we've remained in touch, kept skiing, and stayed in each other's lives. I'm looking forward to seeing where life takes both of us in the coming years.

André Melo and Genya Crossman - a special shout-out to the two of you. While you both have developed as scientists alongside and with me, the more substantial and crucial contribution to my life during these years has been through our lofi chat. One might say that matters of the heart are separate from the trials and tribulations of graduate school, but I think anyone who experienced the pandemic can attest to the fact that productivity and progress in work are strongly dependent on mental state. You two, despite being across the world, have had one of the most profound impacts on these past four years of my life. Thank you for being a fantastic support network and fantastic people.

Bryan Berggren - thank you for keeping me sane! Despite not having the honor of working with you, I did have the honor of having you in my life as a grounded, personable, cheery scientist who was always willing to push me to my best climbing, go on crag adventures with me and play guitar with me. You kept me happy and healthy in times that I otherwise surely would have been less than that. Thanks for being my friend, neighbor, colleague and accountability buddy. I look forward to staying in touch, and seeing where you go.

The Worts - thank you for being a daily community to share life with. You all have played an

important role in my life and time in graduate school, and are exceptional friends and colleagues.

Chloe Rossin - you've provided a bright light to the end of the graduate school tunnel, a warm companionship, and I am excited to start our life together in San Francisco.

I truly love you all dearly.

And of course, I must thank all of my collaborators - every piece of work in this thesis has been collaborative, in the literal sense. So, I would like to thank Jacob L. Beckey, Robin Blume-Kohout, John Bollinger, Arnaud Carignan-Dugas, Michael Foss-Feig, David Hayes, Jordan Hines, Joshua Levin, Timothy Proctor, Ana Maria Rey, Ariel Shlosberg, Graeme Smith, Nikolas A. Tezak, James Thompson, Reuben R.W. Wang, and Kevin Young.

And finally, I must thank Michael Perlin for his help with compiling this thesis and for being yet another great colleague.

Contents

Chapter

Figures

Figure

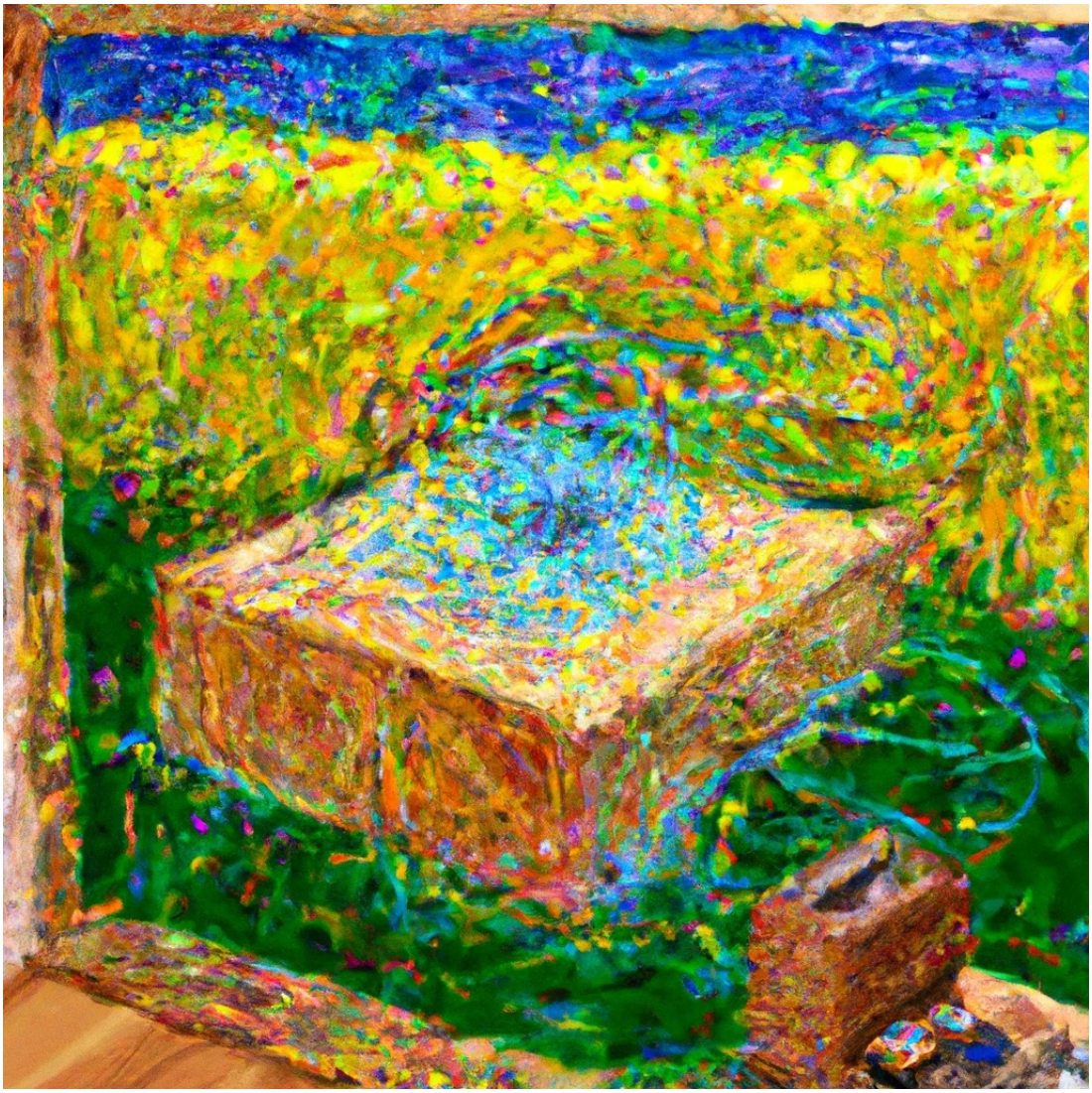


Figure 1: “A Quantum Computer in the Style of Van Gogh” - DALL·E 2

Chapter 1

Individual qubit addressing of rotating ion crystals in a Penning trap [1]

1.1 Abstract

Trapped ions boast long coherence times and excellent gate fidelities, making them a useful platform for quantum information processing. Penning traps offer the possibility of trapping large two-dimensional crystals of several hundred ions. The ions are confined by controlling the rotation of the ion crystal in the presence of a strong magnetic field. However, the rotation of the ion crystal makes single ion addressability a significant challenge. We propose a protocol that takes advantage of a deformable mirror to introduce AC Stark shift patterns that are static in the rotating frame of the crystal. Through numerical simulations we validate the potential of this protocol to perform high-fidelity single-ion gates in crystalline arrays of hundreds of ions.

1.2 Introduction

Ions confined in RF traps are one of the leading platforms for quantum information processing [2–5]. They hold the record for the highest fidelity entangling gates [6–10], and have exceptional coherence times [11]. However, scaling to larger numbers of ion qubits, potentially solved by photonic interconnects between Paul traps [12], or by shuttling ions in the quantum charge-coupled device (QCCD) architecture [2], remains a key challenge. In this work, we consider Penning ion traps which offer the possibility of quantum information processing with samples of as many as 500 trapped ions self-assembled in a large two-dimensional (2D) Coulomb crystal. They use a set of cylindrical electrodes and static voltages to generate axial confinement. The radial confinement is

provided by the Lorentz force experienced by the ions as they undergo a controlled rotation about the trap symmetry axis in the presence of a strong axial magnetic field, typically generated by a superconducting magnet [see Fig. 1.1a]. Following the standard approach of encoding a qubit in two internal levels of the ions, Penning traps have the potential to perform quantum information processing with hundreds of qubits. These qubits typically have transition frequencies from 10's to 100 GHz, making microwaves suitable for global addressing [13]. We note that other ion trapping technologies also allow for the generation of 2D crystals of ions. For instance, it is possible to use the radial degree of a freedom in a Paul trap, or to introduce arrays of closely spaced individual traps [14–17]. To date the number of ions used for quantum information processing in these experiments has been modest (< 20).

Two-qubit gates for entangling the ions in Penning traps have been engineered via spin-dependent optical dipole forces [18]. These forces are generated by interfering two lasers with a difference frequency (sometimes called the beatnote) adjusted to excite phonons in the crystal [see Fig. 1.1a]. Virtual excitation of the center-of-mass mode, for example, generates collective spin-spin interactions across the ion array.

Single-site rotations along with global rotations and a global entangling operation form a universal set of operations for quantum computation—that is, they can be used to implement every unitary operation [19]. Thus, an important step in enabling general quantum information processing in a Penning trap is the ability to perform individual-qubit rotations.

A well-known technique for introducing single-site qubit rotations with one-dimensional (1D) ion strings is through AC Stark shifts with off-resonant, focused laser beams [20]. AC Stark shifts produce σ^Z rotations, which can be turned into more general rotations through the application of global rotations.

The analogous implementation of variable AC Stark shifts in a Penning trap is feasible, but requires introducing focused co-rotating laser beams. The fast rotation frequency used in typical experiments [21–25], ranging from tens to a few hundred kHz in recent NIST experiments, makes this task challenging. Variable AC Stark shifts can also be implemented with spatially fixed, focused

beams directed at the correct radius so that an ion experiences a time-varying AC Stark shift as it rotates through the off-resonance laser beam. Similarly a fixed, focused pair of laser beams in a stimulated Raman configuration can be used to generate a spin rotation as an ion qubit passes through the laser beam waists. However, such approaches would in general require some sequential addressing of the ions, which is inherently slower than parallel addressing.

Here we propose another path for introducing variable AC Stark shifts that are static in the rotating frame of the ion crystal by using the same optical dipole force that is used for implementing a global entangling operation [18, 21]. This can be done by introducing distortions (or, more precisely, spatially dependent phase offsets) to the wavefront of the optical dipole force while setting the beatnote frequency to be a multiple of the rotation frequency [see Figs. 1.1b and 1.1c]. As we will see, a nice feature of this technique is that the rotations of the ions' spins can be conducted in parallel.

Spatially dependent wave front deformations can be implemented with a deformable mirror (DM). DM technology has been under development for many decades, motivated by the desire of adaptive wavefront control in astronomy [26, 27]. For the application discussed in this article, DMs with good wavefront control that employ many physically small actuators with fast switching speeds are desirable. Devices which employ more than 4000 actuators that are 300-micrometer in size with a mechanical response time of order 50 microseconds are commercially available. Wavefront control of better than 10 nm has been documented [26]. The surface of the DM can be coated with aluminum providing good reflectivity down to near ultraviolet wavelengths (< 300 nm). The technology continues to rapidly develop with anticipated improvements in the actuator size and speed.

Any wavefront distortion on the unit disk can be decomposed into the basis of Zernike polynomials [28]. Such functions are generally expressible as

$$Z_n^m(\rho,\phi) = \begin{cases} R_n^{|m|}(\rho)\cos(m\phi), & \text{for } m \geq 0, \\ R_n^{|m|}(\rho)\sin(m\phi), & \text{for } m < 0, \end{cases} . \tag{1.1}$$

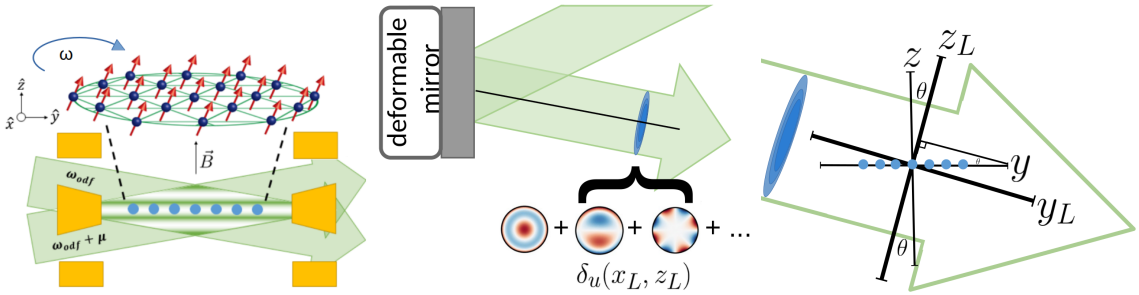
Here $n \geq |m|$, $R_n^{|m|}(\rho)$ are radial polynomials defined on the unit disk. For example, the first few Zernike polynomials are

$$Z_0^0(\rho,\phi) = 1 \tag{1.2}$$

$$Z_1^{-1}(\rho,\phi) = 2\rho\sin(\phi) \tag{1.3}$$

$$Z_1^1(\rho,\phi) = 2\rho\cos(\phi). \tag{1.4}$$

Through the decomposition of wavefront distortions in the basis of Zernike polynomials, we motivate two protocols in Sec. 1.3 for imprinting an AC Stark shift pattern across the crystal of ions. A numerical simulation of the protocols is outlined in Sec. 1.4, followed by a discussion of three primary sources of error in Sec. 1.5. These errors are from ignoring off-resonant terms, considering a finite number of Zernike polynomials, and applying distortions that are too large. However, in Sec. 3.5 we show with numerics that the errors can be controlled to have maximum infidelities as small as 10^{-3}. In particular, we demonstrate the faithful reconstruction of an annulus, an elliptical Gaussian, and a displaced Gaussian using parameters that are representative of typical Penning trap conditions [18, 21, 22]. The annulus and elliptical Gaussian patterns produce initial states that are interesting for quantum simulation. For example, both patterns are 2D analogs of a domain wall, and might allow us to study non-trivial spin transport phenomena [29]. The displaced Gaussian is chosen to rotate a single qubit in the crystal. The high-fidelity reconstructions suggest that our protocol provides a path forward for implementing high-fidelity single- and multi-site qubit rotations in a Penning trap.



(a) (b) (c)

Figure 1.1: (a) Ions in a Penning trap. The ions are axially confined by a set of cylindrical electrodes and rotate with frequency ω. In the presence of a strong magnetic field $\vec{B}$ along the $\hat{z}$ direction, the induced Lorentz force radially confines the ions. A moving 1D optical lattice formed by interfering two beams with angular frequencies ω_{odf} and $\omega_{\text{odf}} + \mu$ respectively, generates an optical dipole force. (b) Imprinting a phase pattern with a DM. By reflecting one of the laser beams off of a deformable mirror (grey) surface, we can imprint a phase pattern on the wavefront (blue) of $\delta_u(x_L, z_L)$, where x_L, y_L and z_L are the beam-centric coordinates. The phase pattern $\delta_u(x_L, z_L)$ can be decomposed into a basis of Zernike polynomials. (c) To generate a distortion $\delta(x, y)$ at the ion crystal, one must generate a distortion in the upper beam given by $\delta_u(x_L, z_L) \equiv \delta[x, z_L/\sin\theta]$ with θ the angle between the beam propagation direction and the $y-$axis. The ions (blue) are at $z = 0$.

1.3 Experimental Set-up and Protocols

In a Penning trap, a system of N ions is confined axially by voltages applied to a set of cylindrically symmetric electrodes. Radial confinement is implemented by the Lorentz force experienced by the ions as they rotate around the trap symmetry axis in the presence of a static $B_0\hat{z}$ field [see Fig. 1.1a]. The ion crystal rotation frequency is precisely controlled by a rotating electric field [30]. In work at NIST [18, 21, 22], the $^2S_{1/2}$ ground-state valence electron spin $|\uparrow\rangle \equiv |m_J = +1/2\rangle$ ($|\downarrow\rangle \equiv |m_J = -1/2\rangle$) in each trapped $^9Be^+$ ion encodes a spin-1/2 degree of freedom, which can be globally controlled by external microwaves resonant with the 124 GHz frequency splitting of the electronic spin states in the $B_0 = 4.5$ T magnetic field of the trap.

In a frame rotating with the ion crystal, the motion of the ions can be decomposed into in-plane modes, which describe ion motion in the plane of the ion crystal, and axial or drumhead modes, which describe ion motion transverse to the plane. For performing quantum simulations [18, 21, 22], the drumhead modes are coupled to the spin degree of freedom by a spin-dependent optical dipole force (ODF) produced by a pair of off-resonant laser beams far detuned from the nearest optical transitions. The beams generate a 1D traveling-wave lattice potential at a frequency μ [see Fig. 1.1a]. The system can be well described by the Hamiltonian

$$\hat{H}_{\mathrm{ODF}}/\hbar = U \sum_i \cos\left(\delta k \hat{z}_i - \mu t + \psi\right)\hat{\sigma}_i^Z. \tag{1.5}$$

Here U is the zero-to-peak AC Stark shift, δk the wave vector of the moving lattice potential, and ψ is an initial phase. For clarity, we will use capital letters to denote directions in spin space, e.g., $\hat{\sigma}^Z$, and lowercase letters to denote directions in real space, e.g., $\hat{z}$.

In what follows we show how we can modify the ODF interaction in Eq. (1.5) to introduce an AC Stark shift pattern that is static in the rotating frame of the crystal through the introduction of wave front deformations. We assume Eq. (1.5) does not excite spin-dependent motion, which is reasonable if μ is far off-resonant with any modes.

It is convenient to assume that without a DM the two ODF beams have flat wavefronts. Then position-dependent phase offsets (or wavefront distortions) $\delta(x, y)$ can be implemented in Eq. (1.5)

with a single DM that is inserted into one of the ODF laser beam lines, say the upper beam. The DM generates a position-dependent phase offset $\delta_u(x_L, z_L)$ where x_L, z_L are the distances from the center of the beam in a coordinate system perpendicular to the k-vector of the laser beam [beam-centric coordinates, see Figs. 1.1b and 1.1c]. We can choose the x_L coordinate in the upper beam-centric coordinate system to be the same as the laboratory frame x coordinate ($x_L = x$) at the single-plane crystal. Because the ODF beam crosses the ion crystal with an angle θ, an ion located at the laboratory frame coordinates (x, y) samples the DM generated phase offset $\delta_u(x, y\sin\theta)$. Therefore, to generate a wavefront offset pattern $\delta(x, y)$ at the ion crystal requires generating a wavefront pattern in the upper beam given by $\delta_u(x_L, z_L) \equiv \delta(x_L, z_L/\sin(\theta))$. In general, the DM will be located at some distance from the ion crystal and an optical imaging set-up is required to image the DM surface to the ion crystal with different demagnification ratios in the z_L and x_L directions.

Suppose, as sketched in the previous paragraph, we introduce waveform deformations to the 1D optical-dipole lattice potential [Eq. (1.5)] so that the interaction is described by

$$\hat{H}_{\mathrm{ODF}}/\hbar = U \sum_i \cos\left(\delta k \hat{z}_i - \mu t + \psi + \delta(x_i, y_i)\right)\hat{\sigma}_i^Z. \tag{1.6}$$

Here $x_i = \rho_i \cos[\phi_i^{\mathrm{lab}}(t)]$ and $y_i = \rho_i \sin[\phi_i^{\mathrm{lab}}(t)]$ are the laboratory frame coordinates of ion i in the $z = 0$ plane, and $\delta(x, y)$ is a local phase shift that describes the distortions of the optical dipole force wavefronts. The azimuthal angle in the laboratory frame is related to the azimuthal angle in the rotating frame by $\phi_i^{\mathrm{lab}} = \phi_i - \omega t$ where ϕ_i is independent of time for a stable crystal and ω is the ion crystal rotation frequency.

Our goal is to employ wave front deformations $\delta(x, y)$ described in Eq. (1.6) to generate an effective Hamiltonian given by

$$\hat{H}_{EFF}/\hbar = U \sum_i F(\rho_i, \phi_i)\hat{\sigma}_i^Z. \tag{1.7}$$

Here F describes an AC Stark shift pattern that is static in the rotating frame of the crystal. To do this, we will consider two different protocols with slightly different choices of $\delta(x, y)$. Before

introducing these choices, we motivate a convenient decomposition of $F(\rho,\phi)$ by expanding in terms of Zernike polynomials, as described in Eq. (1.1). This gives a decomposition of the form

$$\begin{aligned} F(\rho,\phi) &= \sum_{n=0}^{\infty}\sum_{m=-\infty}^{\infty} A c_n^m Z_n^m(\rho,\phi) \\ &= \sum_{n,m=0}^{\infty} A R_n^m(\rho)\Big(c_n^m\cos(m\phi) + c_n^{-m}\sin(m\phi)\Big) \\ &= \sum_{m=0}^{\infty} A P^m(\rho)\cos(m\phi) + A Q^m(\rho)\sin(m\phi), \end{aligned} \tag{1.8}$$

where we included A as an overall amplitude scaling of the pattern $F(\rho,\phi)$, and set $\max[F(\rho,\phi)]/A = 1.0$, which provides a normalization condition for the c_n^m. The radial functions $P^m(\rho)$ and $Q^m(\rho)$ are weighted sums of the appropriate $R_n^{|m|}(\rho)$ for $n \geq |m|$. Specifically, in expanding $F(\rho,\phi)$ in this way, we gathered all terms with the same $\cos(m\phi)$ or $\sin(m\phi)$ azimuthal dependence.

We explore using the deformable mirror to generate each term in the sum of Eq. (1.8). Because F in Eq. (1.7) can, in general, have arbitrarily high spatial frequency components (and the protocols we consider necessarily generate AC Stark shifts up to some maximum order $m_{\max}$) truncation error will be introduced by considering a finite number of terms in Eq. (1.8). This error will be considered in Secs. 1.5 and 3.5 where we carry out a numerical analysis.

We first consider generating an AC Stark shift pattern proportional to the term

$$A P^m(\rho)\cos(m\phi) \tag{1.9}$$

in Eq. (1.8). Suppose the DM is set to generate a distortion in Eq. (1.6) of the form

$$\tilde{\delta}(\rho,\phi^{\text{lab}}) = \delta_e^m(\rho)\cos(m\phi^{\text{lab}}). \tag{1.10}$$

We now set $\hat{z}_i = 0$ (the ions are located in the $z=0$ plane) and substitute $\phi_i^{\text{lab}} = \phi_i - \omega t$. Then Eq. (1.6) can be written

$$\hat{H}_{\text{ODF}}/\hbar = \frac{U}{2}\sum_i \left\{e^{i[-\mu t+\psi+\tilde{\delta}(\rho,\phi_i-\omega t)]} + \text{c.c}\right\}\hat{\sigma}_i^Z. \tag{1.11}$$

The phase modulation term can be expanded in terms of Bessel functions, using the Jacobi–Anger expansion [$e^{iz\cos\theta} = \sum_{\ell=-\infty}^{\infty} i^\ell J_\ell(z) e^{i\ell\theta}$],

$$\hat{H}_{\mathrm{ODF}}/\hbar = \frac{U}{2}\sum_i \left\{ e^{i[-\mu t+\psi]} \sum_{\ell=-\infty}^{\infty} i^\ell J_\ell[\delta_e^m(\rho_i)] e^{i\ell(m\phi_i - m\omega t)} + \text{c.c.} \right\} \hat{\sigma}_i^Z. \tag{1.12}$$

By setting $\mu = m\omega$ only the $\ell = -1$ term is static in the rotating frame. All other terms are rapidly oscillating and can be ignored. It is also possible to get static terms by choosing μ to be a higher integer multiple of $m\omega$. However, these terms will be scaled by a higher-order Bessel function, and therefore produce a smaller static AC Stark shift, assuming sufficiently small arguments $\delta_e^m(\rho_i)$. Furthermore, we will show in Sec. 1.5.1 and Appendix A that the contribution of the fast rotating terms is exactly zero if we choose to apply H_{ODF} for a duration T satisfying $\omega T = 2\pi r$, with r an integer value.

After some algebra we obtain,

$$\hat{H}_{\mathrm{ODF}}/\hbar \approx U \sum_i J_1[\delta_e^m(\rho_i)] \sin(m\phi_i - \psi)\hat{\sigma}_i^Z. \tag{1.13}$$

Therefore by choosing a distortion of the form

$$\delta_e^m(\rho) = J_1^{-1}(\frac{A}{2} P^m(\rho)), \tag{1.14}$$

Eq. (1.6), under the approximations discussed above, reduces to

$$\hat{H}_{\mathrm{ODF}}/\hbar \approx \frac{U}{2}\sum_i A P^m(\rho_i) \sin(m\phi_i - \psi)\hat{\sigma}_i^Z. \tag{1.15}$$

With $\psi = -\pi/2$ this is exactly the targeted AC Stark shift pattern of Eq. (1.9) (up to a factor of $1/2$ which we introduce for convenience as we will discuss later).

In Eq. (1.14), we denote the choice of the wave front pattern described by $J_1^{-1}[\frac{A}{2}P^m(\rho)]$ rather than $\frac{A}{2}P^m(\rho)$ as precompensation. Precompensation is always possible as long as $|\frac{A}{2}P^m(\rho)| \lesssim 0.58$, where 0.58 is the approximate maximum value that the J_1 Bessel function can take. This condition is always possible to satisfy by choosing A small enough in Eq. (1.14). Note that this limits the phase offset $\delta_e^m(\rho)$ to be $|\delta_e^m(\rho)| \leq 1.84$.

An identical derivation for odd Zernike polynomials reveals that a deformation of the form,

$$\delta(\rho, \phi^{\text{lab}}) = J_1^{-1}(\frac{A}{2}Q^m(\rho)) \sin(m\phi^{\text{lab}}) \tag{1.16}$$

gives,

$$\hat{H}_{\text{ODF}}/\hbar \approx -\frac{U}{2}\sum_i AQ^m(\rho_i)\cos(m\phi_i - \psi)\hat{\sigma}_i^Z. \tag{1.17}$$

Additionally, for $m = 0$ corresponding to a circularly symmetric pattern, Eq. (1.11) reduces to

$$\hat{H}_{\text{ODF}}/\hbar = U\sum_i \cos(\delta_e^0(\rho_i) + \psi)\hat{\sigma}_i^Z, \tag{1.18}$$

so that choosing $\delta_e^0(\rho) = \cos^{-1}[AP^0(\rho)] - \psi$ for the $m = 0$ terms reproduces the desired phase pattern Eq. (1.9). By sequentially setting the DM to generate each even $[AP^m(\rho)\cos(m\phi)]$ and odd $[AQ^m(\rho)\sin(m\phi)]$ term in Eq. (1.8) for each m, the above derivation shows that one can apply any AC Stark shift pattern $F(\rho, \phi)$ [see Eq. (1.8)]. However, sequential application can take a long time if there are many terms and the reset time of the DM is slow. Thus, it would be good to have a technique for applying all azimuthal phase patterns in parallel.

Applying in parallel means applying even and odd orders at the same time and applying different beat note frequencies at the same time. First, we show that we can apply both even and odd terms simultaneously, by considering a distortion of the form

$$\delta(\rho, \phi^{\text{lab}}) = \delta_e^m(\rho)\cos(m\phi^{\text{lab}}) + \delta_o^m(\rho)\sin(m\phi^{\text{lab}})\,. \tag{1.19}$$

Using the Jacobi-Anger expansion, and setting $\mu = m\omega$, we find

$$\hat{H}_{\text{ODF}}/\hbar = \frac{U}{2}\sum_i \left(e^{i(-\mu t+\psi)}\exp\{(i[\delta_e^m(\rho_i)\cos(m\phi_i - m\omega t) + \delta_o^m(\rho_i)\sin(m\phi_i - m\omega t)])\} + \text{c.c.}\right)\hat{\sigma}_i^Z, \tag{1.20}$$

$$\hat{H}_{\text{ODF}}/\hbar = \frac{U}{2}\sum_i \left(e^{i(-\mu t+\psi)}\sum_{a=-\infty}^{\infty}\sum_{b=-\infty}^{\infty} i^a J_a(\delta_e^m(\rho_i))J_b[\delta_o^m(\rho_i)]e^{i[(a+b)(m\phi_i - m\omega t)]} + \text{c.c.}\right)\hat{\sigma}_i^Z, \tag{1.21}$$

Neglecting the fast rotating terms or operating with application times T where $\omega T = 2\pi r$ with r an integer value (a condition at which the contribution of all non-static terms vanish), we obtain

$$\hat{H}_{\rm ODF}/\hbar \approx U \sum_i \Big(\sum_{\substack{a+b=-1 \\ a,b\in\mathbb{Z}}} J_a[\delta_e^m(\rho_i)] J_b[\delta_o^m(\rho_i)] \cos\left(a\frac{\pi}{2} - m\phi_i + \psi\right)\Big)\hat{\sigma}_i^Z. \tag{1.22}$$

For $a = -1$, and $\delta_o^m(\rho) = 0$, this agrees with the expression in Eq. (1.13) ,

$$\begin{aligned}\hat{H}_{\rm ODF}/\hbar &\approx U \sum_i J_{-1}(\delta_e^m(\rho_i)) \cos\left(-\frac{\pi}{2} - m\phi_i + \psi\right)\hat{\sigma}_i^Z \\ &= U \sum_i J_1(\delta_e^m(\rho_i)) \sin\left(m\phi_i - \psi\right)\hat{\sigma}_i^Z.\end{aligned} \tag{1.23}$$

However, for both $\delta_e^m(\rho)$ and $\delta_o^m(\rho)$ non-zero, there are now terms given by higher-order Bessel functions that are static and non-zero.

$$\begin{aligned}\hat{H}_{\rm ODF}/\hbar \approx\ & U \sum_i \Big(J_1(\delta_e^m(\rho_i)) J_0(\delta_o^m(\rho_i)) \sin(m\phi_i - \psi) && (1.24)\\ & - J_1(\delta_o^m(\rho_i)) J_0(\delta_e^m(\rho_i)) \cos\left(m\phi_i - \psi\right) + && (1.25)\\ & J_1(\delta_e^m(\rho_i)) J_2(\delta_o^m(\rho_i)) \sin\left(m\phi_i - \psi\right) && (1.26)\\ & - J_2(\delta_e^m(\rho_i)) J_1(\delta_o^m(\rho_i)) \cos\left(m\phi_i - \psi\right) - \ldots \Big)\hat{\sigma}_i^Z.\end{aligned}$$

Fortunately, for small arguments, $J_\ell(x) \approx \frac{1}{\ell!}(\frac{x}{2})^n$ and $J_0(x) \approx 1 - (\frac{x}{2})^2$, so that if we can choose $\delta_e^m(\rho_i) = AP^m(\rho_i)$ and $\delta_o^m(\rho_i) = AQ^m(\rho_i)$ to be small (by choosing A to be small), the first two terms will reduce to the desired results [Eqs. (1.15) and (1.17)]. Note that we introduced the factor of $1/2$ in the precompensation step [Eqs. (1.14) and (1.16)] to make the outcome of the serial and parallel protocols the same. The remaining terms will also be made small. Explicitly,

$$\hat{H}_{\rm ODF}/\hbar \approx \frac{U}{2} \sum_i \Big(AP^m(\rho_i) \sin(m\phi_i - \psi) - AQ^m(\rho_i) \cos\left(m\phi_i - \psi\right)) + O(A^2)\Big)\hat{\sigma}_i^Z. \tag{1.27}$$

To first order in the arguments of the Bessel function, we see the even ($AP^m(\rho)\cos(m\phi)$) and odd ($AQ^m(\rho)\sin(m\phi)$) terms in Eq. (1.8) can be treated additively, and the Hamiltonian considered in Eq. (1.20) can be used to apply both the even and odd m components in parallel.

A similar analysis can be applied to show that it is possible to apply all different orders m at the same time. For instance, consider the simplified case of two different nonzero even orders, m_1 and m_2, and suppose we set the beatnote frequency to $\mu = m_1\omega$. Then our wavefront deformation is given by

$$\delta(\rho, \phi^{\text{lab}}) = AP^{m_1}(\rho)\cos\left(m_1\phi^{\text{lab}}\right) + AP^{m_2}(\rho)\cos\left(m_2\phi^{\text{lab}}\right) \tag{1.28}$$

and our Hamiltonian is

$$\hat{H}_{\text{ODF}}/\hbar = \tag{1.29}$$

$$\frac{U}{2}\sum_i \Big(e^{i(-\mu t+\psi)}\exp\{(i[AP^{m_1}(\rho_i)\cos(m_1\phi_i - m_1\omega t) + AP^{m_2}(\rho_i)\cos(m_2\phi_i - m_2\omega t)\}]) + c.c.\Big)\hat{\sigma}_i^Z,$$

$$\hat{H}_{\text{ODF}}/\hbar = \tag{1.30}$$

$$\frac{U}{2}\sum_i \Big(e^{i(-\mu t+\psi)}\sum_{a=-\infty}^{\infty}\sum_{b=-\infty}^{\infty} i^{a+b}J_a(AP^{m_1}(\rho_i))J_b(AP^{m_2}(\rho_i))e^{i(a(m_1\phi_i - m_1\omega t)+b(m_2\phi_i - m_2\omega t))} + c.c.\Big)\hat{\sigma}_i^Z$$

As before, neglecting the fast rotating terms, or operating with application times T where ωT is a positive integer multiple of 2π (a condition at which all non-static terms vanish) we get

$$\hat{H}_{\text{ODF}}/\hbar \approx U\sum_i \Big(\sum_{\substack{am_1+bm_2=-m_1\\ a,b\in\mathbb{Z}}} J_a(AP^{m_1}(\rho_i))J_b(AP^{m_2}(\rho_i))\cos\left((a+b)\frac{\pi}{2} - m_1\phi_i + \psi\right)\Big)\hat{\sigma}_i^Z. \tag{1.31}$$

The lowest order terms occur when $a = -1$ and $b = 0$, resulting in

$$\hat{H}_{\text{ODF}}/\hbar \approx \frac{U}{2}\sum_i \Big(AP^{m_1}(\rho_i)\sin\left(m_1\phi_i - \psi\right) + O(A^2)\ldots\Big)\hat{\sigma}_i^Z. \tag{1.32}$$

For small A this is approximately the desired AC Stark shift pattern. If we set the ODF beatnote $\mu = m_2\omega$ we select an AC Stark shift pattern described by the second term in Eq. (1.28).

When considering terms with $m = 0$ and setting $\psi = -\pi/2$, the leading order contribution is instead $UAP^0(\rho_i)$, so that there is an additional factor of 2 multiplying the radial polynomial. Note that the precompensation for $m = 0$ [see Eq. (1.18)] was chosen to also make the outcome of the serial and parallel protocols the same.

The above analyses support two experimental procedures for generating an AC Stark shift pattern $F(\rho, \phi)$ that is static in the rotating frame of the crystal. The first is sequential: for a phase pattern with terms of at most order $m_{\max}$, we sequentially set the DM to $2m_{\max} + 1$ different azimuthal phase patterns, applying the appropriate beatnote frequency at each step and the corresponding precompensation in the applied waveform. This has the advantage of allowing for larger amplitudes A and higher accuracy.

The second procedure is a parallel application: we set the DM once to a phase pattern proportional to $F(\rho, \phi)$ at the ion crystal, and simultaneously (or in rapid succession) apply all beatnote frequencies $\mu_m = m\omega$, for $0 \leq m \leq m_{\max}$. The beatnote at μ_m will imprint an AC Stark shift in the rotating frame of the ions proportional to $P^m(\rho)\cos(m\phi) + Q^m(\rho)\sin(m\phi)$, rotating the ion's spins according to the mth-order component of $F(\rho, \phi)$. This has the benefit of being faster if the pattern has a large number of frequency components, but at the cost of lower accuracy and requiring smaller amplitudes. We note that small amplitudes [i.e., A in Eq. (1.8) or δ in Eq. (1.6)] can be offset through the use of large U or long application durations T.

1.4 Numerical Simulation

In this section we outline a numerical study whose results are presented in Secs. 1.5 and 3.5 for preparing arbitrary qubit rotation profiles, $F(\rho, \phi)$, across the crystal. In the work that follows, we set $U = 2\pi \times 10$ kHz and $\omega = 2\pi \times 180$ kHz, which are typical experimental parameters [21, 22]. Our goal will be to prepare the ions in the state

$$|\psi(T)\rangle = \bigotimes_i e^{-iUF(\rho_i,\phi_i)T\hat{\sigma}_Z^i} |+\rangle_i \,, \tag{1.33}$$

where T is the gate duration.

First, we will prepare all of the ions in the $|+\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle)$ state, which can be easily done by preparing all ions in $|\downarrow\rangle$ and then applying a global rotation around the Y axis. We will then determine a maximum m and n, based on the desired fidelity of the state preparation, such that we approximately reconstruct F as $\tilde{F}$, using only Z_m^n for all $|m| \leq m_{\max}$ and $n \leq n_{\max}$. Writing α_n^m for the coefficients of F in the Zernike basis, we have

$$F = A \sum_{\substack{-n\leq m\leq n \\ 0\leq n\leq\infty}} \alpha_n^m Z_n^m \approx A \sum_{\substack{-m_{\max}\leq m\leq m_{\max} \\ 0\leq n\leq n_{\max}}} \alpha_n^m Z_n^m = \tilde{F}. \tag{1.34}$$

Experimentally, $n_{\max}$ could be constrained by the available resolution of the deformable mirror—since an nth order polynomial is determined by $n+1$ points, a mirror with N actuators in a dimension can only hope to parametrize a family of polynomials of degree $N-1$. In our analysis we will assume that the DM has a sufficiently large number of actuators and prioritize minimizing $m_{\max}$, which sets the number of terms in the decomposition of F [see Eq. (1.8)] that will be included in the reconstruction $\tilde{F}$. Larger $m_{\max}$ in general requires a longer gate duration or higher laser power.

As discussed in Sec. 1.3, there are two ways to apply the full phase pattern—sequentially, and in parallel. In the analysis that follows we will consider both of these approaches. Although in principle one could consider using these techniques to apply arbitrary qubit rotations, for the purposes of assessing the performance of our protocols we consider the experimentally useful example of π rotations. That is, in the examples considered in Secs. 1.5 and 3.5 the time evolution will be set for a time T such that in the final state $|\psi(T)\rangle$ the ion located at the maximum of the phase-pattern F will be rotated by π radians in the XY-plane.

1.5 Sources of Error

We will now discuss three sources of error that can occur in the protocols described in Sec. 1.3. To quantify the error and to analyze the performance of our protocol we will use the single-spin infidelity

$$I_j = 1 - |\langle \mathbb{P}^j \tilde{\psi}(T) | \mathbb{P}^j \psi(T) \rangle|^2, \tag{1.35}$$

where $\mathbb{P}^j$ traces out all but the jth ion. The quantity I_j is the infidelity of spin j in the simulated state $|\tilde{\psi}(T)\rangle$ with respect to the target state $|\psi(T)\rangle$. State-of-the-art quantum information processing platforms often have single-qubit gate infidelities of 10^{-3} – 10^{-2}[31, 32] or less, and so this will be the standard of comparison in our analysis. While trapped ion computers often have substantially better single-qubit infidelities [6, 7, 33], our goal here is to demonstrate that Penning traps are a compelling candidate for quantum information processing, for which it is sufficient to study these larger infidelities.

1.5.1 Rotating Wave Approximation

The first type of error comes from the rotating wave approximation, where we ignore the rapidly oscillating terms in Eq. (1.12). As we will show below these errors nevertheless can be avoided if one chooses the gate duration to be commensurate with the ion crystal rotation frequency: $\omega T = 2\pi r$, with r an integer.

As a simple example, we consider a phase pattern with a single angular order m and with an additional amplitude parameter A that will be chosen sufficiently small so that precompensation is not necessary. Thus our phase function in Eq. (1.6) is given as

$$\delta(\rho, \phi) = AP^m(\rho)\cos(m\phi). \tag{1.36}$$

For our simple Hamiltonian, we can analytically express the expectation value $\langle\sigma_X\rangle$ of the spin after evolving for time T. Writing $f(\tau)$ as the time-dependent coefficient in Eq. (1.12), i.e., $\hat{H}_{\mathrm{ODF}}/\hbar = \sum_j f_j(t)\hat{\sigma}_j^Z$, and setting $\mu = m\omega$, we find (for $m \neq 0$)

$$\langle\tilde{\psi}(T)|\sigma_X^j|\tilde{\psi}(T)\rangle = \cos\Big(2\int_0^T f_j(\tau)d\tau\Big). \tag{1.37}$$

We can write

$$\int_0^T d\tau f_j(\tau) = \sum_{\ell=-\infty}^{\infty} s_j(\ell), \tag{1.38}$$

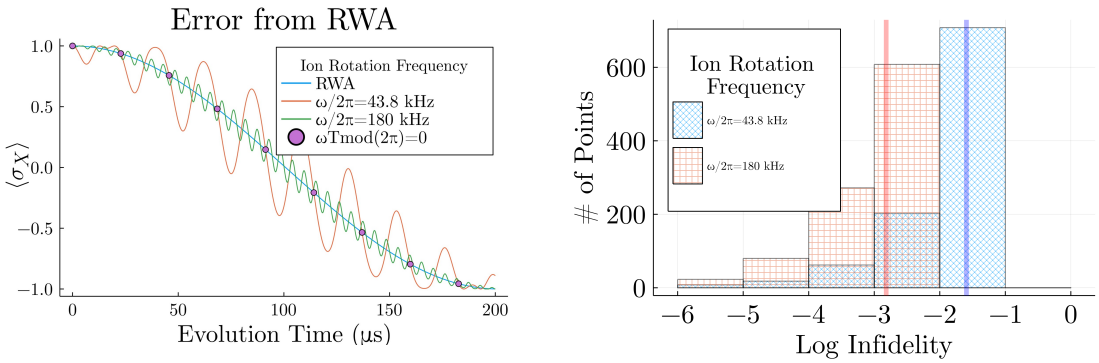


Figure 1.2: (a) **Error from the Rotating Wave Approximation.** Starting in the state $|\psi(0)\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle)$, we plot $\langle\sigma_X\rangle$ as a function of time under evolution by Eq. (1.37) with $\mu = \omega$ and $\delta(\rho, \phi) = A\rho\cos(\phi)$ for $A = 0.25$, and different values of ω at $(\rho, \phi) = (1, 0)$. These expectation values are compared with those coming from the rotating wave approximation given by Eq. (1.13) with $m = 1$. We see that by increasing the ion rotation frequency from $\omega = 2\pi \times 43.8$ kHz to $\omega = 2\pi \times 180$ kHz, the correction terms in Eq. (1.40) get suppressed so that the maximum infidelity is no larger that 10^{-2}. (See Fig. 1.2b). The pink dots mark times that are positive integer multiples of $2\pi/43.8$kHz where the RWA error vanishes for the slower trap rotation frequency. (b) Distribution of errors from RWA. Panel (a) samples $\langle\sigma_X\rangle$ at 1000 different points in time. Panel (b) shows the histogram of the log infidelities between the exact evolution and evolution under Eq. (1.13) evaluated at the same points in time. Choosing $\omega = 2\pi \times 43.8$ kHz results in most of the infidelities being larger than 10^{-2}. Increasing the rotation frequency to $\omega = 2\pi \times 180$ kHz, the maximum infidelity is decreased to smaller than 3×10^{-3}(orange). By sampling many different values of evolution time, we can be confident that the infidelity contributed by the RWA for arbitrary angles of rotation are sufficiently small for high-fidelity generation of $|\psi(T)\rangle$ [Eq.(1.33)]. Red and blue lines have been added to the red and blue distributions, respectively, to highlight the mean and median, which are indistinguishable at this scale.

where

$$s_j(\ell) = \begin{cases} UJ_1[AP^m(\rho_j)]\sin[m\phi_j - \psi]T, & \text{if } \ell = -1, \\ \frac{4UJ_\ell[AP^m(\rho_j)]\sin\left(\frac{1}{2}m(\ell+1)T\omega\right)\cos\left(m\ell\phi+\frac{1}{2}[\pi\ell-m(\ell+1)T\omega]+\psi\right)}{m(\ell+1)\omega} & \text{if } \ell \neq -1. \end{cases} \tag{1.39}$$

Therefore, if $UJ_\ell[AP^m(\rho_j)]/(m\omega) \ll 1$ then,

$$\langle\tilde{\psi}(T)|\sigma_X^j|\tilde{\psi}(T)\rangle \approx \cos(\Theta) - \sin(\Theta)\sum_{\ell=-\infty,\ell\neq-1}^{\infty} a_\ell[T(\ell+1)m\omega]\frac{U}{(\ell+1)m\omega}, \tag{1.40}$$

$$\Theta = 2UJ_1[AP^m(\rho_j)]\sin[m\phi_j - \psi]T, \tag{1.41}$$

for some function a_n. For $m = 0$ we have simply that

$$\int_0^T f_j(\tau)d\tau = U\cos[AP^0(\rho_j) + \psi]T, \tag{1.42}$$

which has no time-dependent corrections.

As an example, choosing $m = 1$ and $P^1(\rho) = \rho$, we will evaluate Eq. (1.37) at the point that will have the largest infidelity, $\rho = 1$ and $\phi = 0$, with $A = 0.25$. This choice of A allows us to focus primarily on effects from the RWA, and ignore the other sources of error discussed below. This gives superimposed oscillations around a cosinuisoidal evolution, with corrections proportional to $\frac{U}{m\omega}$. Thus, by increasing ω for fixed U, the RWA becomes more accurate. This is shown in Fig. 1.2a, where the evolution is sampled at 1000 points in time. As ω is increased from $2\pi \times 43.8$ kHz (blue) to $2\pi \times 180$ kHz (orange), the oscillations become smaller. We see that the ion completes a full π rotation in nearly 200 μs, as expected from Eq. (1.13). Additionally, we see that at evolution times that are positive integer multiples of $2\pi/\omega$ the difference between the exact evolution and the RWA is zero [see pink points in Fig. (1.2)]. This can be seen in Eq. (1.39), as the terms with $n \neq -1$ are zero at these points. See also Appendix A for a more generic case. In Sec. 3.5 we choose evolution times that take advantage of this fact. Note that for this simple case the fast-rotating terms also vanish at $T = \pi/\omega$.

Figure 1.2b shows a histogram of the log infidelity obtained at the different evolution times sampled in Fig. 1.2a. By increasing the rotation frequency from $2\pi \times 43.8$ kHz (blue) to $2\pi \times 180$ kHz (orange), the maximum single-spin infidelity is decreased from approximately 5×10^{-2} to 3×10^{-3}.

This implies that setting $\omega = 2\pi \times 180$ kHz is sufficient for obtaining infidelities of 3×10^{-3} or smaller.

1.5.2 Truncation

The second source of error comes from the fact that in practice we apply a finite number of beatnote frequencies, truncating the basis expansion at finite order. This will produce imperfect reconstructions of the desired phase pattern. While choosing finite $n_{\max}$ is also a possible source of error, as discussed earlier we choose $n_{\max}$ large enough that it is not the limiting factor—effectively assuming that the DM has enough actuators to give good resolution.

The Zernike polynomials form an orthogonal set of polynomials on the disk, D, and therefore arbitrary functions can be decomposed into these polynomials. The inner product on the space of functions on the disk is given by:

$$\langle F, G\rangle = \int_D \rho d\rho d\theta F \cdot G \tag{1.43}$$

We can write the coefficients of F from Sec. 1.4 as

$$\alpha_n^m = \frac{2n+2}{\epsilon_m \pi} \langle \frac{F}{A}, Z_n^m \rangle, \tag{1.44}$$

where ϵ_m is 2 if $m = 0$, and 1 otherwise. The prefactor in Eq. (1.44) is due to the fact that the polynomials are not normalized,

$$\langle Z_n^m(\rho,\phi) Z_{n'}^{m'}(\rho,\phi)\rangle = \frac{\epsilon_m \pi}{2n+2} \delta_{n,n'} \delta_{m,m'}. \tag{1.45}$$

Truncating the number of terms we include will give us a different phase function, $\tilde{F}$, from which we can define the error from truncation as

$$\mathcal{E} = \max_D(|F - \tilde{F}|/A). \tag{1.46}$$

Note that $\max[F(\rho,\phi)] = A$ [see discussion after Eq. (1.8)], so $\mathcal{E}$ is the truncation error normalized to the maximum value of F.

The amount that this truncation contributes to the infidelity will vary depending on the particular phase function F being considered, as we will see in Sec. 3.5. Here, to provide a rough estimate, we consider the maximum single-spin infidelity ε across the crystal, defined as

$$\begin{aligned}
\varepsilon &= \max_j(I_j) \\
&= \max_j(1 - |\,\langle \mathbb{P}^j\tilde{\psi}(T)|\mathbb{P}^j\psi(T)\rangle\,|^2) \\
&\approx \max_j\left(1 - \left|1 - \frac{1}{2}\Big(UAT[F(\rho_j,\phi_j) - \tilde{F}(\rho_j,\phi_j)]/A\Big)^2\right|^2\right) \\
&\approx \mathcal{E}UAT)^2, \qquad (1.47)
\end{aligned}$$

where $\mathbb{P}^j$ traces out all but the jth ion.

We consider rotations where the ion located at the maximum of F is rotated by π radians, corresponding to $UAT = \pi/2$. An infidelity requirement of ε, therefore, in general necessitates a truncation error $\mathcal{E} \lessapprox \frac{2}{\pi}\sqrt{\varepsilon}$. For an infidelity requirement of 10^{-2} (10^{-3}) the maximum truncation error should be less than 0.064 (0.02).

1.5.3 Linear Approximation

The final source of infidelity, which is only relevant for the parallel application discussed in Sec. 1.3, is in assuming that A in Eqs. (1.27) and (1.32) is small enough so that ignoring higher-order terms is justified. By increasing the product UT and decreasing the amplitude A, the linear approximation can be made arbitrarily good. However, increasing UT will also increase decoherence due to off-resonant light scattering from the ODF beams **[Uys2010]** during the qubit rotations.

To estimate the contribution of higher-order terms in A on the infidelity for the target case of a π-rotation on a single ion, we consider the parallel application of two P^{m_1}, P^{m_2} terms [see (Eq. 1.28)]. It can be shown that leading corrections, of order $O(A^2)$, arise when $2m_1 = m_2$. In this case the leading-order static terms generate a Hamiltonian of the form

$$\hat{H}_{\mathrm{ODF}}/\hbar \approx \frac{U}{2}\Big(\sum_i AP^{m_1}(\rho_i)\sin(m_1\phi_i - \psi) - \frac{A^2}{2}P^{m_1}(\rho_i)P^{2m_1}(\rho_i)\cos(m_1\phi_i - \psi) + O(A^3)\Big) \quad (1.48)$$

We now compute the infidelity between the state $|\phi\rangle = e^{-i\hat{H}_{\mathrm{ODF}}T/\hbar}|+\rangle$ and $|\tilde{\phi}\rangle = e^{-i\frac{UATP^{m_1}(\rho_0)}{2}\hat{\sigma}_Z}|+\rangle$ obtained by considering only the desired first order term. We choose $\psi = -\pi/2$ in the rest of this discussion. By Taylor expanding, we find the infidelity to be

$$I(\tilde{\phi},\phi) \approx \left|\frac{UT}{4}A^2P^{m_1}(\rho_i)P^{2m_1}(\rho_i)\sin(m_1\phi_i)\right|^2 \tag{1.49}$$

Thus, for $P^{m_1}(\rho_i)P^{2m_1}(\rho_i)\sin(m_1\phi_i) < 1$, and $UTA/2 \sim \pi$, the infidelity reduces to

$$I(\tilde{\phi},\phi) \lesssim (\frac{\pi}{2}A)^2. \tag{1.50}$$

We see that if $A \leq 0.02$ the infidelity can be constrained to be less than 10^{-3}. If we relax our infidelity requirements to 10^{-2} we can choose A as large as $A = 0.06$. While these amplitude requirements may seem strict, we note that this estimate is pessimistic — we considered the worst-case situation when $2m_1 = m_2$, which gives leading order error contributions of size $O(A^2)$. As we will see in Sec. 3.5, the amplitude can often be made larger. In fact, as we will see in Sec. 1.6.3 where we look at the case of flipping the spin of a single ion, the amplitude can be taken more than an order of magnitude larger while achieving the same infidelity goals.

1.6 Numerical Results

Following the discussion of Secs. 1.4 and 1.5, we now numerically demonstrate a few interesting examples of implementing different AC Stark shift patterns across a circular crystal with our protocols from Sec. 1.3. The preparation of initial states with targeted spatial profiles can be of great utility for investigating propagation of quantum information and entanglement. With that purpose in mind, here we consider a range of geometries including an annulus, an elliptical Gaussian, and a displaced Gaussian. For these patterns results for $\langle\sigma_X\rangle$ as well as the log infidelity across a crystal of 91 ions are shown. This number was chosen to have inter-ion spacings of 0.1 of the crystal diameter. For the phase patterns considered, Figs. 1.3, 1.6, and 1.10 show the truncation error,

$|\mathcal{E}|$, from considering a finite number of Zernike polynomials. Next, we study the error generated during the dynamical evolution. First we apply the protocol in series, evolving under each even $[AP^m(\rho)\cos(m\phi)]$ and odd $[AQ^m(\rho)\sin(m\phi)]$ term in Eq. (1.8) that constitutes $\tilde{F}(\rho,\phi)$ one-by-one, for two different choices of target maximum infidelity. The infidelities of the final state are shown in Figs. 1.5, 1.8, and 1.12. We discuss contributions to the infidelity arising from the RWA and truncation errors (see Sec. 1.5). Next, we apply the protocol in parallel for all m such that $0 \leq m \leq m_{\max}$ in Eq. (1.34). In Figs. 1.9 and 1.13 we show the corresponding infidelities for the elliptical and displaced Gaussians. (The annulus only requires implementing a single $m = 0$ term.)

1.6.1 Annulus

As a first example, we consider preparing ions in an annulus. To make the problem of reconstruction in a basis of continuous functions easier, we will smooth the edges with sigmoid functions, giving

$$g(\rho) = \frac{1}{1+e^{-\kappa(\rho-r_1)}} - \frac{1}{1+e^{-\kappa(\rho-r_2)}}. \tag{1.51}$$

Scaling this function to be one at its maximum, we have a targeted normalized AC Stark shift pattern given by:

$$F(\rho,\phi) = Ag(\rho)/g(r_1 + \frac{1}{2}(r_2 - r_1)), \tag{1.52}$$

which corresponds to the phase function

$$\delta(\rho,\phi) = \cos^{-1}\left[F(\rho,\phi)\right] \tag{1.53}$$

in Eq. (1.6). For our numerical experiment we will set $r_1 = 0.45$, $r_2 = 0.55$, and $\kappa = 10$. This value of κ was chosen to avoid sharp rising and falling edges for the annulus.

1.6.1.1 Reconstruction

Because the phase pattern is azimuthally symmetric, the only nonzero coefficients have $m = 0$ and only the application of a single beatnote with frequency $\mu = m\omega = 0$ is required. We see that all terms are static, and therefore, incur no error from the RWA. We choose $n_{\max}$ to be sufficiently

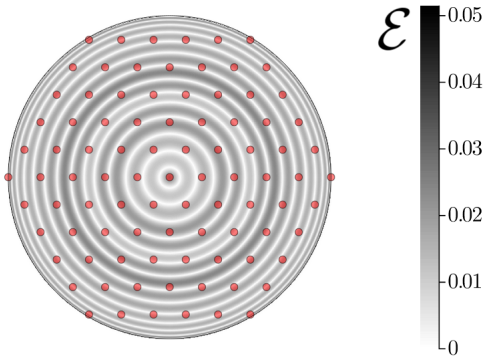


Figure 1.3: Absolute value of the error $\mathcal{E} = |F - \tilde{F}|/A$ in reconstructing an annulus. With $n_{\max} = 24$ and $m = 0$ we can reconstruct an annulus with an error no larger than 0.05. This is sufficiently low error to reproduce a high fidelity state. (See Fig. 1.5.) The red dots represent the ion positions and are shown for reference.

large ($n_{\max} = 24$) so that the reconstruction error $\mathcal{E}$, presented in Fig. 1.3, is less than 0.064. From the discussion in Sec. 1.5 this should enable a single-spin infidelity of less than 10^{-2}. The next contributing error term is radially symmetric, since the pattern itself is radially symmetric, which can be seen clearly in Fig. 1.3.

1.6.1.2 Evolution

Given that all the terms in the Zernike expansion have the same value of m, they can be applied simultaneously using the precompensation technique [Eq. (1.53)] above, incurring no errors from the linear approximation. Setting $A = 1.0$, we obtain a gate duration of 25 μs for $U = 2\pi \times 10$ kHz, which is significantly faster than typical decoherence times in trapped ions.

The expectation values $\langle \sigma_X \rangle$ after performing the precompensation protocol is shown in Fig. 1.4, and the infidelity to the target state is shown in Fig. 1.5. Here the infidelity is due to finite n truncation. Figure 1.5a shows an infidelity better than 10^{-2} for $n_{\max} = 24$. An infidelity of 10^{-3} can be obtained with $n_{\max} = 54$ as shown in Fig. 1.5b. A histogram of the infidelities for the two different $n_{\max}$ values is shown in Fig. 1.5c. The presented analysis shows that the protocol can produce a faithful reconstruction of the annulus pattern with a small state infidelity ($< 10^{-3}$).

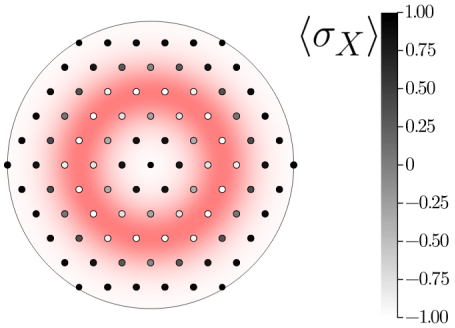
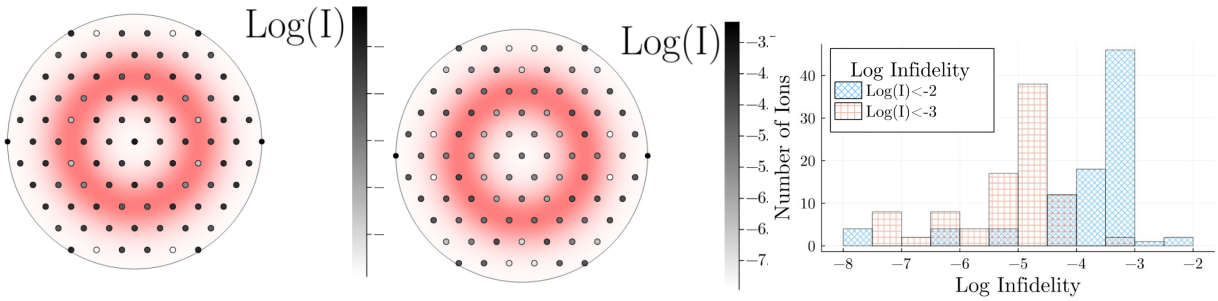


Figure 1.4: $\langle \sigma_X \rangle$ for an annulus after following the protocol in Sec. 1.4. Using $n_{\max} = 24$ and $m = 0$, we see that $\langle \sigma_X \rangle$ is -1 on the annulus and 1 outside the annulus, as desired.The pink pattern illustrates the targeted AC Stark shift pattern F.



(a) (b) (c)

Figure 1.5: Infidelity I [see (Eq. 1.35)] for preparing an annulus following the protocol of Sec. 1.6.1. (a) For $n \leq 24$ and $m = 0$, the maximum infidelity is smaller than 10^{-2}. (b) For $n \leq 54$ and $m = 0$, the maximum infidelity is smaller than 10^{-3}. (c) A histogram over the ions' infidelities, comparing (a) and (b).

1.6.2 Elliptical Gaussian

The next example we consider is an elliptical Gaussian,

$$F(\rho,\phi) = \frac{A}{2}\exp\{(-(\rho\cos(\phi))^2/(2\eta_x^2) - (\rho\sin(\phi))^2/(2\eta_y^2)\} \tag{1.54}$$

We choose $\eta_x = \sqrt{2}/10$ and $\eta_y = \sqrt{2}$ to generate an elliptical pattern with a “narrow" and “wide" distribution covering the 91-ion crystal. Because there is no longer azimuthal symmetry, we must consider $m > 0$.

1.6.2.1 Reconstruction

We use a numerical integrator to compute the Zernike basis coefficients in Eq. (1.44). The reconstruction error for $m_{\max} = 10$ and $n_{\max} = 26$ is shown in Fig. 1.6. Interestingly, there are radial bands, set by the value of $m_{\max}$. For $m_{\max} = 10$, the next higher-order omitted term has 12 full periods of oscillation corresponding to the 12 pairs of dark fringes in Fig. 1.6. Since the error is minimized at the bright fringes, we can, in principle, achieve a higher fidelity reconstruction by adjusting $m_{\max}$ so that ions near the edge of the crystal in the y-direction are positioned near the minima of the reconstruction error. At any ion in the crystal, the reconstruction error $\mathcal{E}$ shown in Fig. 1.6 is less than 0.035, which should enable an implementation of the elliptical Gaussian AC Stark shift pattern with an infidelity less than 10^{-2}.

The error in the reconstruction is seen to be smaller near the center of the disk, since including lower-degree radial polynomials can match the behavior there. To capture the phase pattern far from the center of the disk requires higher-order radial polynomials. In fact the error is observed to be maximum at the vertical wings of the distribution since to capture the decay of the Gaussian closer to the $\rho = 1$ boundary requires higher-order terms.

1.6.2.2 Series Application

Figure. 1.7 shows $\langle\sigma_X\rangle$ at the end of a series evolution with $m_{\max} = 10$, $n_{\max} = 26$, and $A = 0.5$. Note that because of the elliptical symmetry, m is restricted to non-negative even values,

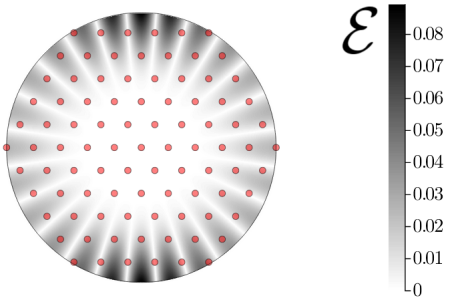


Figure 1.6: Absolute value of the error $\mathcal{E} = |F - \tilde{F}|/A$ in reconstructing an elliptical Gaussian. By using $n_{\max} = 26$ and $m_{\max} = 10$ with m always positive we can reconstruct an elliptical Gaussian pattern with a maximum error of less than 0.09. Nevertheless the maximum error at an ion is approximately 0.035 since there are no ions sitting at the maximum of the phase pattern. The periodicity in the error pattern is mainly set by the $m_{\max}$ value considered.

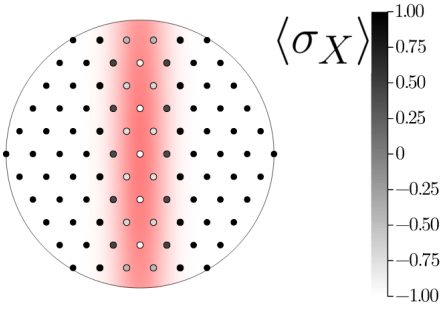


Figure 1.7: $\langle\sigma_X\rangle$ for an elliptical Gaussian. We use $n \leq 26$, $0 \leq m \leq 10$, and $A = 0.5$. Ions along the center of the elliptical phase pattern are rotated towards $|-\rangle$, while those far away from the center remain in $|+\rangle$.

corresponding to six terms in total. We choose a gate duration of $T = 18 \times (2\pi/\omega) = 100\,\mu s$ and hence we remove the errors from the rotating wave approximation. The protocol generates spin rotations along the three vertical columns of ions in the center of the disk, while very nearly leaving all other ions in the $|+\rangle$ state. Figure 1.8a shows a maximum single-spin infidelity of 10^{-2}. With $m_{\max} = 12$, corresponding to seven terms, and making $n_{\max} = 32$, Fig. 1.8c shows clearly that the maximum infidelity is reduced to 10^{-3}.

Each application of one even $[AP^m(\rho)\cos(m\phi)]$ term in Eq. (1.8) takes 100 μs. When done in series, the six different values of m employed to obtain an infidelity of 10^{-2} sets a gate duration of 600 μs, neglecting the reset time of the DM. A gate duration of approximately 700 μs is required for the seven terms that enable an infidelity of 10^{-3}. The reset time of the DM can significantly increase the gate duration. For example, a DM reset time of 50 μs increases the gate duration by 50%.

1.6.2.3 Parallel Application

The protocol where all terms are applied in parallel has the advantage that the DM is only set once. This can possibly lead to shorter gate durations. The targeted AC Stark shift [Eq. (1.54)] and choices of parameters for the parallel application are the same as those considered in the previous paragraphs with the exception of the choice of A and therefore T. The parameter A is chosen

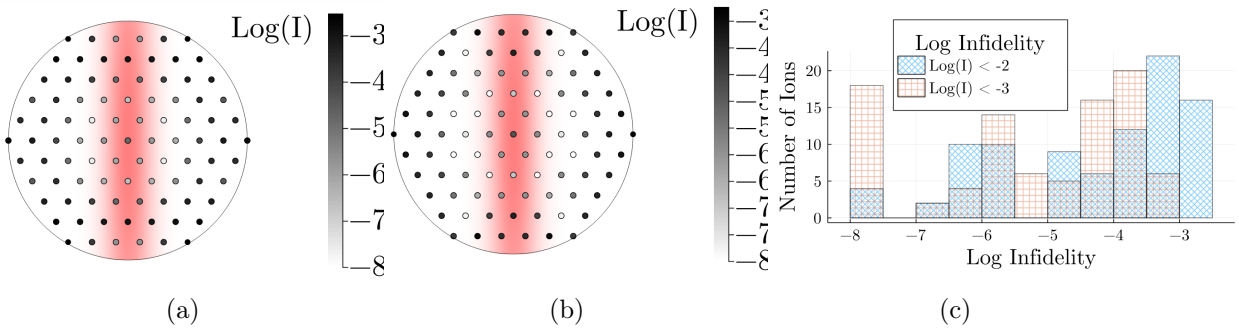


Figure 1.8: Infidelity for preparing an elliptical Gaussian with the serial protocol. (a) For $n \leq 26$, $0 \leq m \leq 10$, and $A = 0.5$, the maximum infidelity is smaller than 10^{-2}. (b) For $n \leq 32$, $0 \leq m \leq 12$, and $A = 0.5$, the maximum infidelity is smaller than 10^{-3}. (c) A histogram over the ions' infidelities, comparing (a) and (b).

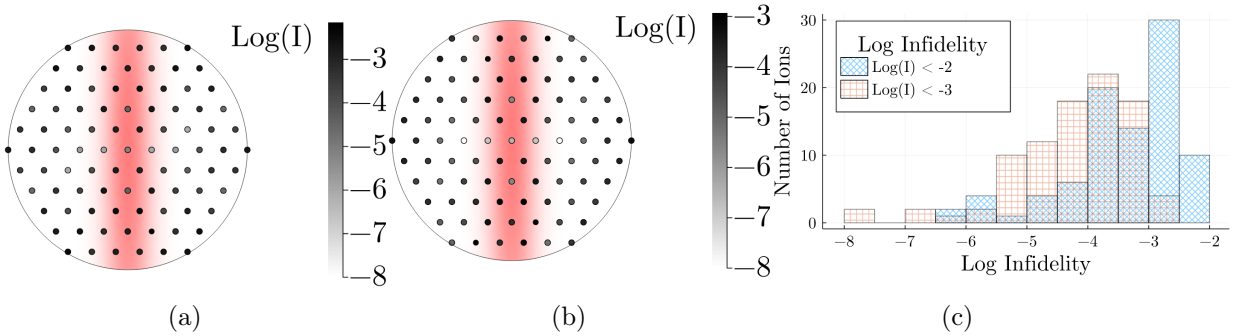


Figure 1.9: Infidelity for preparing an elliptical Gaussian with the parallel protocol. (a) For $n \leq 26$ and $0 \leq m \leq 10$. By applying all polynomials and beatnotes simultaneously, with $A = 0.4$, the resulting maximum infidelity is less than 10^{-2}. (b) With $n \leq 32$, $0 \leq m \leq 12$, but and with $A = 0.2$, the maximum infidelity is no more than 3×10^{-3} with all polynomials and beatnotes applied simultaneously. (c) A histogram over the ions' infidelities, comparing (a) and (b).

sufficiently small so that the linear approximation discussed in Sec. 1.3 holds, and T is picked so that the spins at the maxima of the phase pattern experience a full π rotation, as discussed in Sec. 1.4. By applying the different orders in parallel, we incur all of the errors from the previous section and additional errors from the linear approximation discussed in Sec. 1.5.3. Because the n and m chosen in the previous section were minimally large to meet our fidelity requirements, we keep them the same. For a maximum infidelity of 10^{-2}, Fig. 1.9a demonstrates that choosing $A = 0.4$ is sufficient. This changes the gate duration to 250 μs, obtained with $T = 45 \times (2\pi/\omega)$. If we increase our infidelity requirements to 3×10^{-3}, we see in Figs. 1.9b and 1.9c that we can choose $A = 0.2$. For the gate duration to be commensurate with the crystal rotation frequency we find $T = (90 \times 2\pi/\omega)$ giving a longer gate duration of 500 μs. We note that we chose 3×10^{-3} as the target infidelity rather than 10^{-3} as in the other examples that we show.

1.6.3 Displaced Gaussian

Finally, we consider an AC Stark shift pattern described by a displaced Gaussian,

$$F(\rho, \phi) = \frac{A}{2} \exp\big(-\{[\rho \cos(\phi) - \delta_x]^2 + [\rho \sin(\phi) - \delta_y]^2\}/(2\eta^2)\big) \tag{1.55}$$

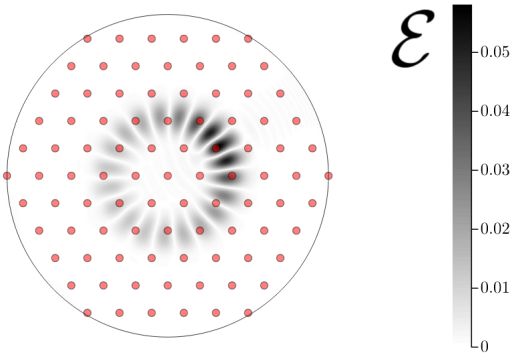


Figure 1.10: Absolute value of the error $\mathcal{E} = |F - \tilde{F}|/A$ in reconstructing a displaced Gaussian. With $n_{\max} = 40$ and $|m_{\max}| = 9$ we get a maximum error smaller than 0.06. The error is worse around the targeted ion since this is the sharpest feature trying to be reconstructed with 20 dark fringes corresponding to the extrema of the next omitted $m = 10$ term.

with standard deviation $\eta = 0.1/\sqrt{2}$, displaced by $\delta_x = 0.3$ in x and $\delta_y = 0.1\sqrt{3}$ in y. The displacement was chosen to coincide with an ion in the crystal, and the width was chosen to achieve a single-spin rotation. In particular, this choice of η corresponds to a Gaussian profile that decays by a factor of $1/e$ at a diameter of the inter-particle spacing. In this case, we no longer have azimuthal symmetry, so we must include $m \geq 0$ and $m < 0$ as well.

1.6.3.1 Reconstruction

In Fig. 1.10, we included up to $m = \pm 9$ and $n_{\max} = 40$ terms. We see that the reconstruction has a maximal error in a region surrounding the targeted ion. The maximum truncation error is approximately 0.06, which should enable a single-spin infidelity of 10^{-2}. We note that choosing an ion closer to the boundary of the crystal will require a large m value to reach similar truncation error.

1.6.3.2 Series Application

In this example we consider $A = 3.0$, motivated by our infidelity goal of 10^{-2}. We picked this value of A to be as large as possible while still requiring that each term being applied in the expansion [Eq. (1.8)] can be inverted as discussed in Sec. 1.3 for precompensation. Additionally,

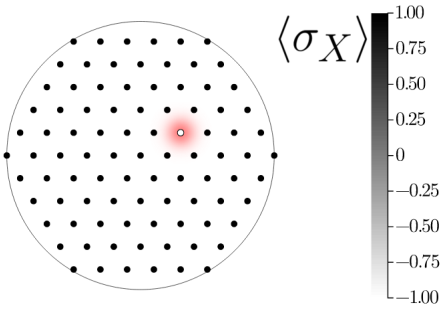
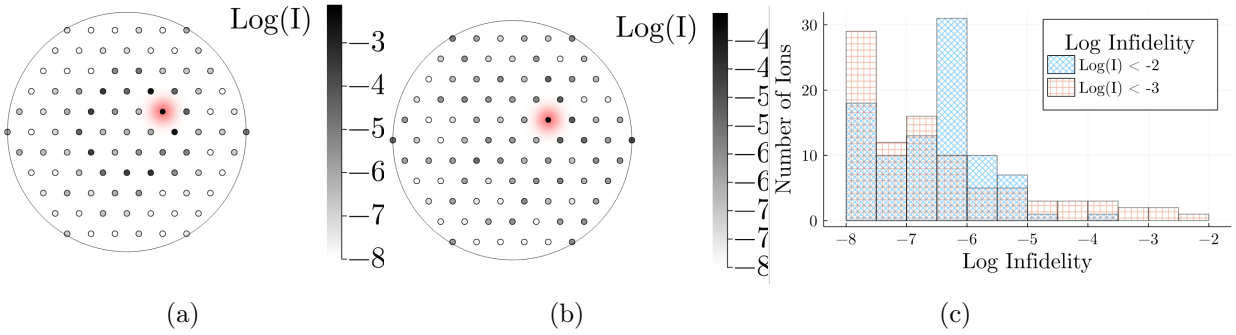


Figure 1.11: $\langle\sigma_X\rangle$ for a displaced Gaussian after following the protocol in Sec. 1.4. Using $n_{\max} = 40$ and $|m_{\max}| = 9$, we see that a single ion spin is rotated to a very good approximation.

we chose $T = 3 \times (2\pi/\omega)$, to remove the error from the RWA. For $A = 3.0$, the application time required for each beatnote is approximately 16.66 μs. When the polynomials are applied in series for 19 different values of m, we get a gate duration of approximately 316.66 μs, again assuming a DM with zero reset time. If we increase our infidelity goals to 10^{-3}, we can choose $|m| \leq 20$ and $A = 3.0$, giving a gate duration of 683.33 μs. As in the previous two cases, we see excellent agreement with the desired phase pattern in the evolution of $\langle\sigma_X\rangle$ shown in Fig. 1.11. The protocol very nearly rotates a single spin, as all of the ions surrounding the desired ion are very nearly in the $|+\rangle$ state. This behavior leads us to conclude that single ion addressability is feasible with this technique, requiring about a factor of 2 more terms than in the other cases, due to the additional odd $(AQ^m \sin(m\phi))$ terms. The infidelities in Figs. 1.12a and 1.12b are worse in a radial band of ions containing the ion being flipped. This is intuitive — suppressing that error requires the introduction of high-order angular terms, while we used only up to $m = 9$.

1.6.3.3 Parallel Application

The parameters in this section are the same as those considered in the previous paragraphs, with the exception of the choice of A and therefore T. Figures 1.13a and 1.13b demonstrate that choosing $A = 0.3$ is sufficient for the 10^{-2} and 10^{-3} infidelity requirements. Consequently, the gate duration in both cases is approximately 333.33 μs, which is given as $T = 60 \times (2\pi/\omega)$ and thus the error incurred from the rotating wave approximation is zero. This reduces the gate duration for the



(a) (b) (c)

Figure 1.12: Infidelity for preparing a displaced Gaussian with the serial protocol. (a) For $n \leq 40$, $-9 \leq m \leq 9$, and $A = 3.0$ the maximum infidelity is smaller than 10^{-2}. (b) For $n \leq 40$, $-20 \leq m \leq 20$, and $A = 3.0$ the maximum infidelity is smaller than 10^{-3}. (c) A histogram over the ions' infidelities, comparing (a) and (b).

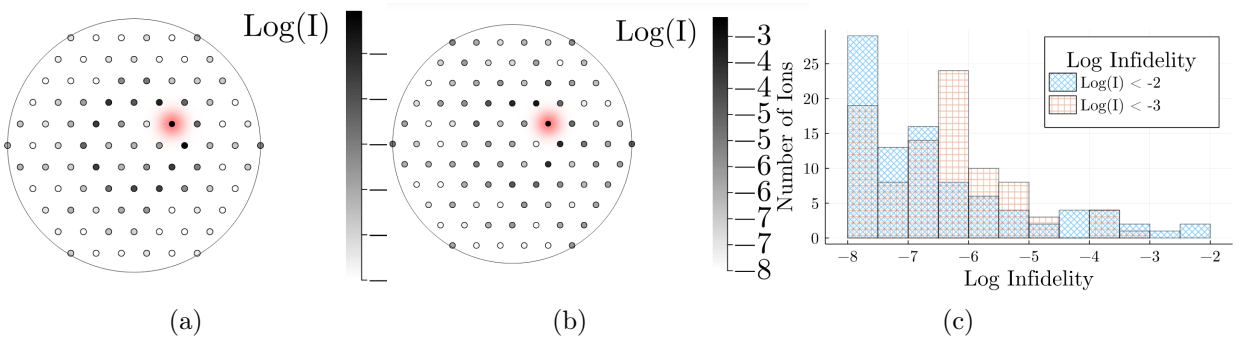


Figure 1.13: Infidelity for preparing a displaced Gaussian with the parallel protocol. (a) For $n \leq 40$ and $-9 \leq m \leq 9$. By applying all polynomials and beatnotes simultaneously, with $A = 0.3$, the maximum infidelity is smaller than 10^{-2}. (b) For $n \leq 40$, $-20 \leq m \leq 20$, and still with $A = 0.3$, the maximum infidelity is less than 10^{-3}. (c) A histogram over the ions' infidelities, comparing (a) and (b).

parallel application compared to the serial application only for the case of a targeted infidelity of 10^{-3}. However, this neglects the reset time of the DM. For a targeted infidelity of 10^{-2}, the serial application needed $m_{\max} = 9$, which would require the DM to be set 18 times. With a reset time as high as 50 μs, this incurs an overhead of 900 μs. This is substantially larger than the gate duration itself, and highlights a potential reason to instead consider the parallel protocol.

1.7 Conclusion

Penning traps are promising candidates as platforms for quantum information processing, due to their ability to control hundreds of qubits, and perform non-local entangling operations. However, existing experiments lack the ability to address individual ions, and therefore, fail to meet the criteria for universal quantum information processing. In this paper we discussed a method for implementing programmable Z-rotations in a Penning trap, thus providing a path forward for more complex quantum simulations and general large-scale quantum information processing.

By employing a DM in the path of one of the laser beams that creates the optical dipole force in Refs. [18, 21], we showed how wave front deformations introduced by the DM can be used to generate AC Stark shift patterns that are static in the rotating frame of the crystal. A pattern

of azimuthal order m $[P^m(\rho)\cos(m\phi), Q^m(\rho)\sin(m\phi)]$ is introduced by setting the frequency μ of the optical dipole force to the mth harmonic of the rotation frequency, $\mu = m\omega$. General AC Stark shift patterns are obtained through the introduction of different azimuthal orders. We analytically and numerically demonstrated the feasibility of this approach for generating single-site rotations. Choosing a Gaussian phase pattern with a $1/e$ diameter equal to the interparticle spacing, we demonstrated that we can resolve single ions in a crystal of about 100 ions under typical experimental conditions [18, 21]. Moreover, by applying the required beatnote frequencies $m\omega$ in parallel, one can obtain 99% fidelity single-qubit gate durations of 333.33 μs, which is faster than typical single particle decoherence times, ~ 10 ms, in current Penning trap experiments.

In this paper we assume a perfectly performing DM and analyze the sources of infidelity for two different protocols where patterns of different azimuthal order are introduced serially or in parallel. In particular, we assume that the number of DM actuators is large compared to the number of ions and that the surface of the DM can be set with arbitrary precision. Deformable mirrors with greater than 4000 actuators, surface figures of less than 10 nm, and mechanical response times of less than 50 μs are available commercially. The desired AC Stark shift pattern as well as the performance of the DM will impact whether the serial or parallel protocol should be employed. In general, the serial protocol enables the implementation of larger amplitudes and therefore higher accuracy phase patterns. However, for patterns that require introducing many azimuthal orders m, the reset time of the DM can add significant overhead in the time required to implement the desired AC Stark shift pattern. The parallel protocol removes any overhead due to the reset time of the DM, but the restriction on the amplitude of the phase pattern can impact the accuracy with which the desired phase pattern can be implemented.

1.8 Appendix

1.8.1 The RWA is Exact for Integer Multiples of the Trap Rotation Frequency

As discussed in Sec. 1.5.1, choosing the evolution time T such that $\omega T = 2\pi r$ with r a positive integer can remove all error from the rotating wave approximation. In Sec. 1.5.1 we showed that this was the case when we applied different P_m patterns. We now argue that this can be generalized to arbitrary order. The Hamiltonian we consider will have both even $[AP^{m_i}\cos(m_i\theta)]$ and odd $[AQ^{m_i}\sin(m_i\theta)]$ terms, up to m_{N_f}, so that $i \leq N_f$. From the Jacobi-Anger expansion each of these terms will introduce a new sum, as in Eq. (1.21), with an index a_i or b_i, respectively. In general, we then get phase factors in the sum, $f(t)$, of the form

$$f(t) = \exp\left\{(-i\mu t - i\phi\sum_i (a_i + b_i)m_i - i\sum_i (a_i + b_i)m_i\omega t)\right\}. \tag{1.56}$$

Because this is the only time dependence, we can integrate over time from 0 to $2\pi r/\omega$ to get

$$\int_0^{2\pi k/\omega} dt f(t) = \frac{\exp\{[-i\mu t - i\phi\sum_i (a_i + b_i)m_i - i\sum_i (a_i + b_i)m_i\omega t)]\}}{-i[\mu + \sum_i (a_i + b_i)m_i\omega]}\Bigg|_0^{2\pi r/\omega} = 0. \tag{1.57}$$

This is true except when the denominator vanishes, which is also the case where the term in Eq. (1.56) is static. If we have a drive $\mu = m_j\omega$ the denominator vanishes when

$$m_j + \sum_{i=1}^{N_f} (a_i + b_i)m_i = 0, \tag{1.58}$$

as we have seen previously in the case of two terms. The case in Sec. 1.5.1 was special in that it repeats every π/ω. The general argument is given above, and shows that in general a multiple of $2\pi/\omega$ is needed.

Chapter 2

The QAOA with Slow Measurements [34]

2.1 Abstract

The Quantum Approximate Optimization Algorithm (QAOA) was originally developed to solve combinatorial optimization problems, but has become a standard for assessing the performance of quantum computers. Fully descriptive benchmarking techniques are often prohibitively expensive for large numbers of qubits ($n \gtrsim 10$), so the QAOA often serves in practice as a computational benchmark. The QAOA involves a classical optimization subroutine that attempts to find optimal parameters for a quantum subroutine. Unfortunately, many optimizers used for the QAOA require many shots ($N \gtrsim 1000$) per point in parameter space to get a reliable estimate of the energy being minimized. However, some experimental quantum computing platforms such as neutral atom quantum computers have slow repetition rates, placing unique requirements on the classical optimization subroutine used in the QAOA in these systems. In this paper we investigate the performance of a gradient free classical optimizer for the QAOA - dual annealing - and demonstrate that optimization is possible even with $N = 1$ and $n = 16$.

2.2 Introduction

The Quantum Approximate Optimization Algorithm (QAOA) is a hybrid quantum-classical algorithm that uses a subroutine run on a quantum computer together with a classical optimizer to find approximate solutions to combinatorial optimization problems [35, 36]. While many results have demonstrated that the QAOA faces substantial challenges if it is to outperform classical algorithms

[37–45], it has nevertheless become a standard technique for benchmarking quantum computers [46–48], serving as a holistic test.

The classical optimization routine in the QAOA is used to select a collection of angles that parameterize a quantum circuit. While [35] discusses the feasibility of an open-loop grid search for the QAOA, in practice classical optimizers are closed-loop and require accurate estimates of the energy expectation value for any choice of angles. Previous work [46, 47, 49, 50] has often chosen classical optimizers that require a large number of measurement repetition rates [35, 46, 51–54]. Furthermore, theoretical works also often assume that the optimizer has access to the expectation value at each point in parameter space [55, 56].

However, this assumption is not always justified - some quantum computing platforms, such as neutral atoms, have substantially slower circuit repetition rates. In the case of neutral atoms, this is due in part to atom reloading [57, 58]. Modern neutral atom quantum computers use lossy measurements which remove atoms from the trap, requiring atoms to be reloaded between measurements. Moreover modern systems such as [48], can have measurement durations as long as 30ms. These facts together can give total shot-to-shot measurement rates of around 5Hz [59].

This places substantial difficulties on the classical optimizer used in the QAOA. While noise is commonly modeled as coming from decoherence and miscalibration, it can also be a result of shot noise in sampling from the quantum computer a limited number of times. This noise can cause unreliable estimates of quantities such as the gradient and the energy.

In this paper we explore a classical optimization routine for circumventing these difficulties – dual annealing. Dual annealing is a form of simulated annealing, which explores the search space while annealing a "temperature" that parameterizes how likely the algorithm is to jump to higher energy configurations. This can be useful for avoiding local minima. We first discuss modifications that can be made to the algorithm to accommodate the error from taking a limited number of shots. Next, we find empirically that this optimizer performs well even without modification. In particular, we find that dual annealing can find global maxima with a single shot at each point in parameter space that it evaluates.

To assess the efficacy of this optimization technique, we propose a family of graphs and use the QAOA to find the maximum cut of a member of this family. We provide complexity theoretic arguments that suggest these graphs have maximum cuts that may be hard for classical computers to approximate. While several families of graphs are known to be poor choices for showing quantum advantage [35, 37, 40, 60], it is an open problem to find families of graphs on which the QAOA can be shown to outperform classical algorithms.

2.3 Theory

The quantum circuit used by the QAOA is described by $2p$ parameters, in a layered ansatz. Specifically, the unitary evolution of the quantum computer has input angles $\vec{\gamma}, \vec{\beta}$ and is given as

$$U(\vec{\gamma}, \vec{\beta}) = \prod_{j=1}^{p} e^{(-i\beta_j \hat{X}^{\otimes n})} e^{(-i\gamma_j H)}, \tag{2.1}$$

where H is called the driver and $\sigma_x^{\otimes n}$ is called the mixer. An example of this circuit is shown in Fig. 2.1 and the details of H for the problem considered in this paper are given in Eq. 2.2. H is a Hamiltonian whose ground state encodes the solution an optimization problem. In this paper we consider the MAX-CUT problem, with associated Hamiltonian

$$H = \sum_{(i,j)\in E} \frac{1}{2}\omega_{ij}(1 + \hat{Z}_i\hat{Z}_j). \tag{2.2}$$

Here E is the set of edges in a graph, and ω_{ij} are the weights for the associated edges.

The goal of the QAOA is to find optimal values of $\vec{\gamma}$ and $\vec{\beta}$ that minimize

$$C(\vec{\gamma}, \vec{\beta}) = \left\langle +|U(\vec{\gamma}, \vec{\beta})^\dagger H U(\vec{\gamma}, \vec{\beta})|+ \right\rangle. \tag{2.3}$$

In this paper, instead of minimizing $C(\vec{\gamma}, \vec{\beta})$ we equivalently maximize $-C(\vec{\gamma}, \vec{\beta})$. We will only consider the $p = 1$ QAOA, and therefore only have γ_1 and β_1. Because there is only one of each angle in the examples we consider, we will refer to them as γ and β for the rest of the paper. Numerous descriptions of the QAOA and MAX-CUT exist in the literature [35, 47], and further discussion is omitted here. MAX-CUT is a common problem to consider for demonstrations of QAOA due to

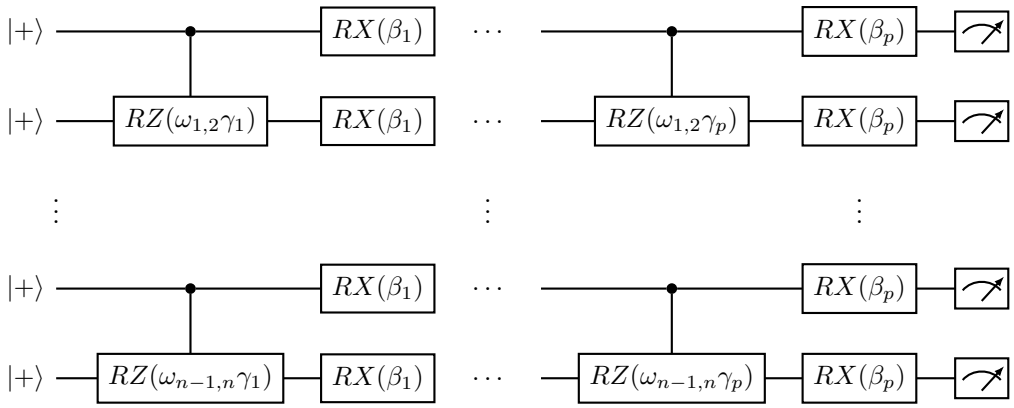


Figure 2.1: An example of a circuit implementing the QAOA from Eq. 2.1, using the MAX-CUT Hamiltonian in Eq. 2.2. In this example, we've assumed that there are only edges between nearest neighbors.

the locality of the Hamiltonian, requiring only two-body terms. This makes it attractive for early demonstrations of hardware such as that in [48].

To find the optimal angles, γ and β, we consider simulated annealing algorithms. Simulated annealing is traditionally introduced in analogy to the metallurgical process, wherein the optimizer has a temperature which controls random fluctuations and an energy function that is used for preferentially exploring certain configurations. The system has "thermal kinetic energy", with a visiting temperature which allows it to probabilistically explore the search space while avoiding local minima. The transition between any two points has an associated hopping probability that characterizes how likely a transition between the two states is. As the temperature is lowered, exploration becomes more difficult in the presence of energy barriers.

Simulated annealing over a discrete set (such as that achieved by discretizing γ and β) can be shown to converge to the set of global minima [61]. For each temperature, the hopping probabilities form a Markov chain, whose stationary distribution is given by

$$\pi_T(i) = \frac{1}{Z_T} \exp\left\{(-\frac{J(i)}{T})\right\} \tag{2.4}$$

[61], with Z_T a normalizing constant. In the limit that $T \to 0$, we see that this distribution concentrates around the collection of minima of J.

We further note that if the energy estimate is noisy, this convergence is not guaranteed. This can easily be seen by considering the space with two elements, i_1 and i_2, and function K such that $K(i_1) = 0$ and

$$K(i_2) = \begin{cases} 2, & \text{with probability } 1/2 \\ -1, & \text{otherwise} \end{cases}. \tag{2.5}$$

Taking $J = \langle K \rangle$ to be the energy function, we see that minimum is i_1, however the stationary point of the Markov chain will be the uniform distribution over i_1, i_2.

The algorithm can, however, be modified to maintain its convergence properties. We can model finite shot noise as a random variable for the k^{th} step in the annealing chain that depends

on the state being measured,

$$X_k = (\frac{1}{N}\sum_{\ell=1}^{N} f_\ell(x_j) - J(x_j)) - (\frac{1}{N}\sum_{\ell=1}^{N} f_\ell(x_i) - J(x_i)), \tag{2.6}$$

where $J(m) = \langle m|C|m\rangle$ and f_ℓ is the ℓ^{th} evaluation on the quantum computer - i.e. the ℓ^{th} shot. One of the conditions for convergence given in [62] is that $\sigma_k \in o(T_k)$, where σ_k is the standard deviation of X_k, T_k is the annealing temperature and $o(\cdot)$ is the standard asymptotic notation for a strict upper bound. Likewise in what follows, $O(\cdot)$ denotes an asymptotic upper bound with possible equality, $\omega(\cdot)$ denotes a strict asymptotic lower bound, $\Omega(\cdot)$ denotes an asymptotic lower bound with possible equality and $\Theta(\cdot)$ means both $O(\cdot)$ and $\Omega(\cdot)$.

From [35] we know that the values of the sampled cost function are concentrated about their mean $C(\gamma, \beta)$, with variance, σ^2 upper bounded by

$$\sigma^2 \leq 2e\frac{(v-1)^{(2p+2)} - 1}{(v-1) - 1}, \tag{2.7}$$

for a graph with e edges and v vertices. Thus

$$\sigma^2_{\hat{\mu}} = \sigma^2/N \in O(e/N), \tag{2.8}$$

where $\sigma^2_{\hat{\mu}}$ is the variance of sample mean. This shows that we should expect the number of shots required to be at worst proportional to the number of edges in the graph.

From [62], we expect the modified annealing algorithm to work if $\sqrt{N/e} \in \omega(1/T_k)$. We will show, however, in Sec. 3.5 that even without modifying dual annealing, it can find the local maxima with a single shot per point in parameter space.

The implementation details of the dual annealing algorithm can be found in [63] and theoretical details can be found in [64]. We reproduce some relevant details here. In particular, the algorithm uses a distorted Cauchy-Lorentz distribution for the "visiting temperature"

$$g_{q_v}(\Delta x(t)) \propto \frac{[T_{q_v}(t)]^{-\frac{D}{3-q_v}}}{[1 + (q_v - 1)\frac{(\Delta x(t))^2}{[T_{q_v(t)}]^{\frac{2}{3-q_v}}}]^{\frac{1}{q_v-1}+\frac{D-1}{2}}}, \tag{2.9}$$

where $x(t)$ is a jumping distance, and D is the dimension of the optimization problem [64]. For the $p = 1$ QAOA $D = 2$, since there are only two values to optimize, γ and β. This temperature is annealed according to

$$T_{q_v}(t) = T_{q_v}(1)\frac{2^{q_v-1} - 1}{(1+t)^{q_v-1} - 1}, \tag{2.10}$$

and new points are accepted with probability

$$p_{q_a} = \min\{1, [(1 - q_a)\Delta E/T_{q_v}]^{\frac{1}{1-q_a}}\}. \tag{2.11}$$

q_a and q_v are called the acceptance and visiting parameters, and are hyperparameters used to control how exploratory the annealing algorithm is. This annealing routine is accompanied by local search at the end of each annealing attempt and a number of restart attempts. Simulated annealing, and hence dual annealing, benefits from a wealth of analysis and convergence guarantees [65].

2.4 Choice of Graph

While the primary purpose of this paper is to discuss the choice of classical optimizer in the QAOA, we need to choose a problem instance. So as to avoid solving a problem that is known to be trivial, we give a heuristic argument for our choice of graph. The particular problem we choose due to its simple-to-implement Hamiltonian and abundance in the quantum computing community is MAX-CUT. The difficulty of MAX-CUT depends on the family of graphs being considered, and the general problem has been well-studied classically - the Goemans-Williamson (GW) algorithm gives an optimal approximation ratio assuming the unique games conjecture [66–68]. Therefore one would like to identify a collection of graphs that classical computers perform worst on.

In this work we choose random graphs $G(v, e)$ on v vertices with e edges chosen at random from all possible edges. MAX-CUT on these graphs has a phase transition when $e/v = 1/2$, in the following sense: below $1/2$ there are efficient classical algorithms for solving the problem, however above $1/2$ there seem to be hard instances [69, 70]. Furthermore, if we consider signed graphs, then if the cover number of the positively signed edges is near $\sqrt{v}$, the problem becomes strongly

NP hard [71]. A problem is strongly NP hard if the problem remains NP hard when its numerical parameters are bounded by a polynomial in the input size. This means that hard instances are, roughly speaking, easy to generate, and so in particular randomly generating instances should result in graphs that may be more challenging for classical algorithms.

The cover number of the positively signed edges is the minimum size of a subset V of the vertices, such that every edge with positive weight has at least one endpoint in V. In particular, [72] shows this family is strongly NP hard for graphs with cover number $\Omega(v^{-k})$, with k a positive integer. We have already specified $e \in \Theta(v)$ so that $k = 2$ is the smallest k that is guaranteed to be compatible. This suggests that algorithms like the GW algorithm may perform less favorably on these graphs. An example of this graph is shown in Fig. 2.2. Similar graphs have been considered in [56].

2.5 Numerics

We simulate the $p = 1$ QAOA on a graph sampled from the family of graphs discussed in Sec. 2.4 using quimb [73]. We emphasize that the purpose of this paper is to demonstrate the performance of the optimizers with low shot rates. The $p = 1$ QAOA has been well-studied, and its performance is well-understood analytically [35, 74, 75]. Thus, our emphasis is that these techniques allow for the demonstration and benchmarking of a quantum computer - for the case of $p = 1$ optimization, one should use the known-optimal choice of angles if the goal is produce a best approximate solution to the problem being studied. In certain problem instances of MAX-CUT, we can reduce the search space for both γ and β to the range $[0, \pi]$ [51]. In general we see this is not possible, since scaling the edge weight by a factor of $1/c$ moves the maximum in γ by a factor of c. Nevertheless, we restrict our search space, since we are interested primarily showing that optimization is possible. For our example, we choose a graph on $v = 20$ vertices, which has $n = 16$ qubits that interact with at least one other qubit. We note that $n = 16$ is a non-trivial example for initial demonstrations of new quantum computing architectures [48].

The particular graph considered in this section is shown in Fig. 2.2, and is sampled from

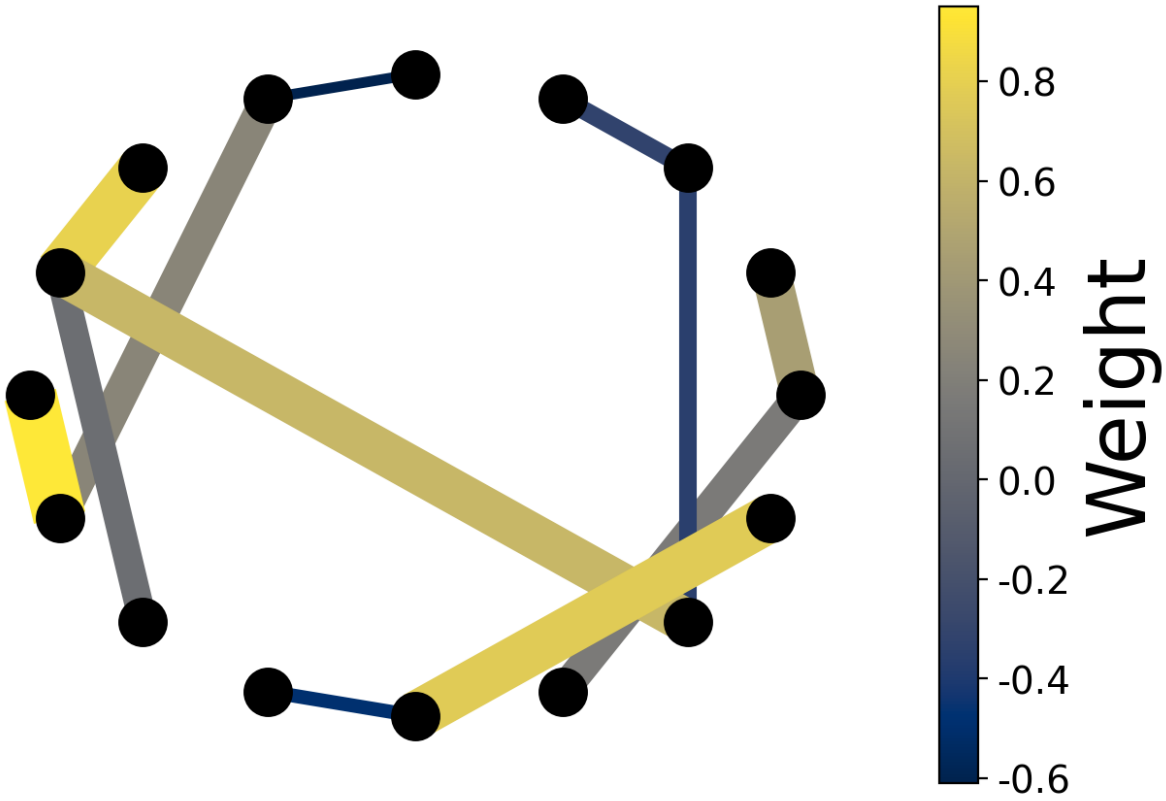


Figure 2.2: **Example Graph from Section 2.4.** Above we see a graph on $v = 20$ vertices, generated according to the distribution described in Sec. 2.4. Edgeless vertices are omitted from the figure. We sample $e = 3v/5$ edges at random, and edge weights are chosen uniformly at random from $[-1, 1]$.

the distribution described in Sec. 2.4. The graph is chosen from a distribution with $e = \frac{3v}{5}$ edges, with $v = 20$, and therefore only contains 16 vertices with edges. Such a graph would be useful for demonstrating the QAOA on an architecture of a 4×4 grid of atoms. As mentioned in Sec. 3.3, rather than minimizing Eq. 2.3, we maximize $-C(\gamma, \beta)$.

In Fig. 2.3 we see the performance of dual annealing on the graph in Fig. 2.2. We use one shot per point in parameter space ($N = 1$) and allow the dual annealing algorithm ten opportunities to restart and re-anneal. In total the optimizer sampled 326 times. For a clock rate of 5Hz this would take just over a minute to run. Fig. 2.3 also provides two kernel density estimates of the distribution of sampled points which highlights that there are two distinct attractors over β, and one over γ, corresponding to the optima of the landscape. Kernel density estimates use Gaussian kernels to produce a non-parametric estimate of the sampling distribution that is given by

$$\hat{f}_h(x) = \frac{1}{nh} \sum \exp\{(-(x - x_i)^2/h^2)\}, \tag{2.12}$$

where the bandwidth h given by

$$h = n^{-1/(d+4)}, \tag{2.13}$$

and n is the number of samples (326) and d is the dimension (1) as suggested in [76]. This demonstrates that the dual annealing optimizer not only can find a good cut, but also is taking advantage of the structure of the optimization landscape.

Another interesting feature of single-shot optimization, which we leave to future work, is the behavior of the single-shot performance statistics. While many optimizers attempt to estimate and maximize the QAOA energy, the goal of the QAOA is to optimize the single best observed bitstring. The shot to shot energy is plotted against the cumulative best energy in Fig.2.4. While the optimizer quickly finds a good cut and Fig. 2.3 shows that the optimizer samples preferentially from the basins of attraction of the optima, it can be seen that the single shot data varies substantially. The maximum value in the grid of values produced in Fig. 2.3 is 3.78, however the optimum reported by the optimizer is 4.13. This is in fact the sum of the positive weights, and hence the maximum cut.

We see that dual annealing is able to perform with just a single shot per choice of angle pair γ

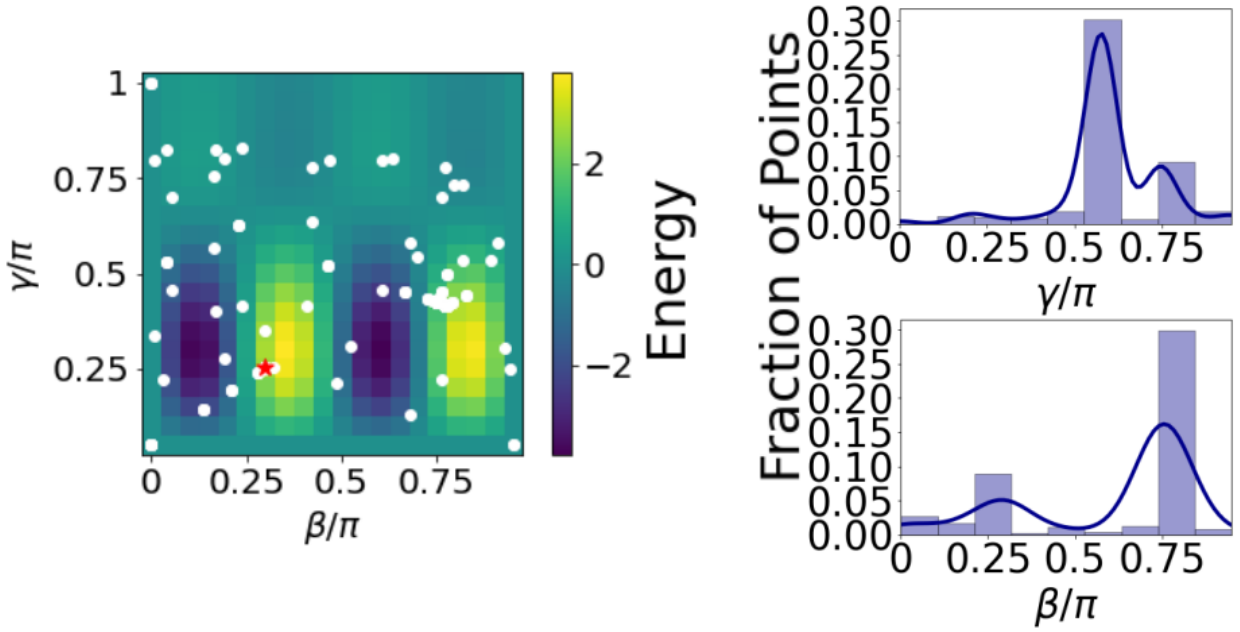


Figure 2.3: **Optimization using dual annealing.** Each white point corresponds to a single shot measured by dual annealing. The kernel density estimates show peaks corresponding to the local maxima, demonstrating that optimizer preferentially samples from higher energies. Moreover, the red star shows the final sampled point which lies near a global maximum.

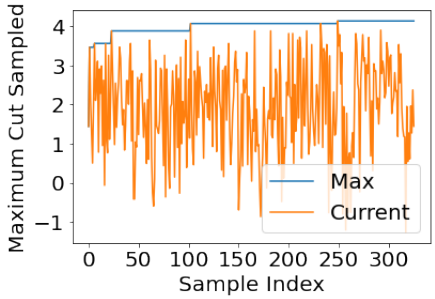


Figure 2.4: **Cumulative maximum statistics.** The orange line plots all 326 points sampled by the dual annealing optimizer, while the blue is the maximum cut seen so far. While Fig. 2.3 shows that the optimizer preferentially samples from higher energies, this figure shows that the single shot data can vary substantially. Despite this, the optimizer quickly finds a good cut.

and β. For quantum computers with low shot rates this provides a tractable method for performing closed-loop quantum-classical optimization, with noisy estimates of the energy at each point.

2.6 Conclusion

Due to its easy implementation and interpretation, the QAOA is an appealing holistic benchmark for quantum computers. However, different quantum computing architectures place different requirements on the classical optimization subroutine of the QAOA. In particular, neutral atom quantum computers often have substantially slower measurement rates than conventional superconducting quantum computers. To use the QAOA as a benchmark for these platforms thus requires choosing a classical optimizer that can use noisy estimates of the QAOA energy.

In this paper, we have demonstrated an optimization technique that is well-suited for this problem, dual annealing. Dual annealing searches the parameter space in a thermal way, accepting transitions if the cost is lower or if the annealing temperature is sufficiently high to allow jumps to worse points. We have shown that this algorithm can use a single shot per point in parameter space while still finding good cut values for the MAX-CUT problem on as many as $n = 16$ qubits.

Additionally, we have suggested a new family of graphs which have a phase transition in their known classical hardness. There are several straightforward analyses to extend this work - including realistic noise models, scaling to larger numbers of qubits and extending the algorithm considered here to model the shot uncertainty explicitly [62], as described in Sec. 3.3. Finally, while the intention of this work was to discuss the optimizer choice for demonstrating the performance of NISQ hardware, and thus focused on $p = 1$, exploring the increased dimensionality of the search space for higher p is an interesting direction for future research.

Chapter 3

Towards Demonstrating Fault Tolerance in Small Circuits Using Bacon-Shor Codes [77]

3.1 Abstract

Quantum error correction is necessary to perform large-scale quantum computations in the presence of noise and decoherence. As a result, several aspects of quantum error correction have already been explored. These have been primarily studies of quantum memory[78, 79], an important first step towards quantum computation, where the objective is to increase the lifetime of the encoded quantum information. Additionally, several works have explored the implementation of logical gates[80–82]. In this work we study a next step - fault-tolerantly implementing quantum circuits. We choose the $[[4, 1, 2]]$ Bacon-Shor subsystem code, which has a particularly simple error-detection circuit. Through both numerics and site-counting arguments, we compute pseudo-thresholds for the Pauli error rate p in a depolarizing noise model, below which the encoded circuits outperform the unencoded circuits. These pseudo-threshold values are shown to be as high as $p = 3\%$ for short circuits, and $p = 0.6\%$ for circuits of moderate depth. Additionally, we see that multiple rounds of stabilizer measurements give an improvement over performing a single round at the end. This provides a concrete suggestion for a small-scale fault-tolerant demonstration of a quantum algorithm that could be accessible with existing hardware.

3.2 Introduction

In order to perform useful quantum computational tasks that require many qubits or deep circuits, it is necessary to use quantum error-correcting codes that have the ability to handle faulty physical hardware[83]. Historically, fidelities of quantum devices have been too far below the necessary thresholds to warrant any attempts to demonstrate fault-tolerant quantum computation. However, slow and steady progress has pushed both single and two-qubit gate infidelities as low as 10^{-3}[6, 7, 32]. It has therefore become an important theoretical challenge to develop initial small-scale quantum error correction (QEC) experiments to validate the theory.

The aim of error correction is to make logical qubits and gates that have lower effective error rates than bare physical systems. Over the past decade, there have been multiple non-fault-tolerant QEC code demonstrations using platforms such as superconducting qubits, trapped-ion systems, and microwave modes[84–87]. While these have all been important building blocks towards large-scale quantum computation, we are more generally interested in fault-tolerant schemes, which effectively mitigate the errors that occur throughout the operation of a QEC scheme. In addition to fault-tolerant quantum circuits, other hardware-efficient fault-tolerant schemes have been proposed, such as performing stabilizer extraction through coupling to nonlinear-oscillators[88]. In this paper, however, we will restrict our attention to the quantum circuit model. In recent years, experiments have studied the near-threshold performance of quantum error-correcting and error-detecting schemes. This has included exploring multiple rounds of error-detection/correction in small surface codes[79] and a large repetition code[78]. Recent trapped ion experiments have additionally yielded promising results. For example, Egan et. al implemented circuits designed to be fault-tolerant and studied the implementation of logical gates in a $[[9, 1, 3]]$ Bacon-Shor code[80], although the logical gates were still outperformed by the physical gates. In another experiment, Ryan-Anderson et. al prepared codewords of the [[7, 1, 3]] color code and demonstrated several rounds of error correction, also with circuits designed to be fault-tolerant[89].

These experiments have been primarily quantum memory experiments, which aim to preserve

quantum information, rather than process it. However, the ultimate goal of QEC codes is to perform computations on faulty physical hardware in addition to preserving quantum memory. Therefore, we must consider QEC schemes that can handle errors that occur in the logical manipulation of quantum data.

In this paper, inspired by Gottesman[90], we explore a small quantum error-detecting scheme to demonstrate a limited form of fault-tolerant quantum computation. The advantage of considering quantum error-detection, as opposed to error-correction, is that the physical circuit we take consists of only 5 physical qubits, greatly simplifying the experimental procedure. In particular, the $[[4, 2, 2]]$ code requires only one additional ancilla to encode two logical qubits. By treating one of those logical qubits as a gauge qubit, we can construct a Bacon-Shor subsystem code, which has fault-tolerant error-detection circuits. These error-detection circuits can be repeated intermittently throughout a computation to detect errors as they occur and to significantly reduce the likelihood of multiple undetectable errors.

Gottesman defines fault tolerance in small experiments as being when any encoded circuit, from a given family of circuits $\mathcal{C}$, outperforms its unencoded counterpart[90]. That is to say that every encoded circuit in $\mathcal{C}$ must have a lower error rate than the error rate of the equivalent bare circuit. Given the computational difficulty of testing every element of $\mathcal{C}$, which is exponentially large for any given circuit depth, we loosen the notion of fault tolerance to only requiring that the average encoded circuit from a uniform random sample collected from $\mathcal{C}$ have increased performance. For our family we have chosen circuits which consist of X, Z and H gates, that are transversally implementable in the code we consider, and thus fault tolerant. An example of such a circuit is shown in Fig. 3.1, consisting of a single X and Z gate.

In what follows, we use threshold to mean the pseudo-threshold for the particular code being considered. We take two approaches to estimate the fault tolerance threshold for the four-qubit Bacon-Shor code. First, we study depolarizing noise numerically. By simulating circuits of varying depth, we find a threshold for the depolarizing parameter as high as 0.6% for deep circuits. Next, we consider a site-counting argument which uses a stochastic noise model, where errors occur at

each site within a circuit with some probability[91]. This site-counting argument is more pessimistic than the depolarizing model - it fails to account for benign sets of errors, and assumes that any error that occurs is the worst possible one. Even so, this analysis yields a threshold of 0.2%. These analyses show that the threshold for seeing increased performance over the unencoded circuit could be achievable using existing hardware.

3.3 Bacon-Shor Subsystem Codes

The fundamental idea of QEC is to encode the state of a given set of qubits, with Hilbert space $\mathcal{H}_S$, into a larger physical system. Generically this entails mapping states on $\mathcal{H}_S$ into the codespace, $\mathcal{H}_{\rm c}$, of a higher-dimensional Hilbert space that can be decomposed as $\mathcal{H} = \mathcal{H}_{\rm c} \oplus \mathcal{H}_{{\rm c}_\perp}$. Physical errors that move the state of the qubits out of the codespace, and into $\mathcal{H}_{{\rm c}_\perp}$, can be detected and in some cases corrected. A subsystem code breaks apart the codespace into a product of subsystems($\mathcal{H}_{\rm c} = \mathcal{H}_L \otimes \mathcal{H}_G$)[92, 93]. One of the subsystems, $\mathcal{H}_L$, is taken as the logical subsystem and the remaining subsystem, $\mathcal{H}_G$, contains gauge degrees of freedom. In this construction, the state on $\mathcal{H}_G$ does not affect the encoded logical state.

A large class of QEC codes are the stabilizer codes, where logical qubits are encoded as codewords that are in the shared +1-eigenspace of a set of commuting Pauli operators[94]. The two-dimensional Bacon-Shor code is a subsystem code with a codespace defined by an associated stabilizer group on a square lattice[93]. The stabilizer framework provides a method of detecting whether errors occur in the process of computation, and the division of the codespace into logical and gauge subsystems allows for a simplified scheme of error syndrome extraction. In the case of the Bacon-Shor code, stabilizer measurements can be performed by measuring a set of spatially-local, two-qubit gauge operators, $\mathcal{S}$, whose products contain the stabilizers. The gauge operators do not affect the logical state, but may act non-trivially on the gauge subsystem.

In this work, we consider the four-qubit Bacon-Shor subsystem code, in which there are two elements in the tensor factorization, giving a single logical qubit and a single gauge qubit. We choose the set of gauge operators to be $\mathcal{S} = IIXX, XXII, ZIZI, IZIZ$. The logical space of the code

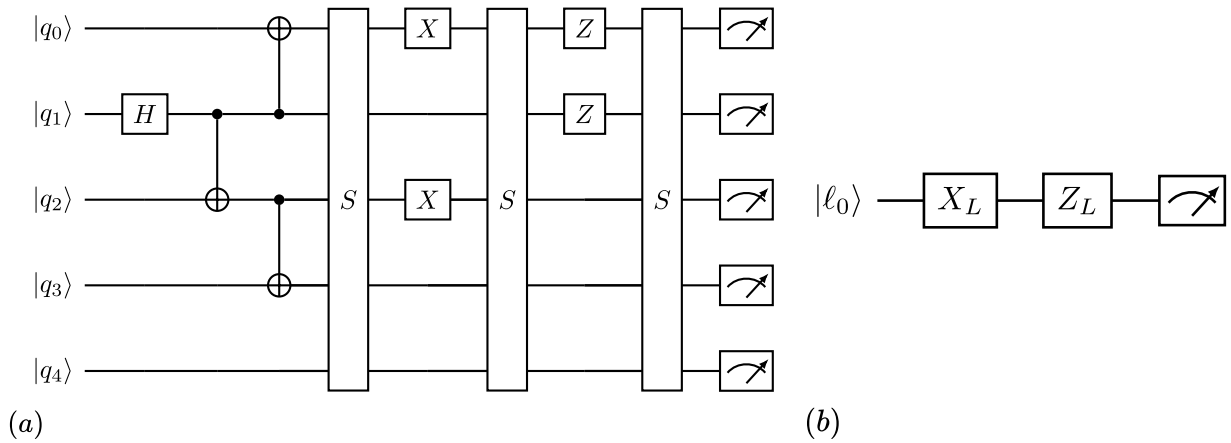


Figure 3.1: Comparison of the encoded fault-tolerant circuit to the bare circuit. a) A physical, fault-tolerant circuit encoding the logical circuit shown in Fig. b. Following state preparation and each logical gate, a syndrome measurement is applied, denoted by S. This corresponds to the case of one gate per stabilizer measurement round. Final measurement outcomes are post-selected on obtaining trivial error syndromes. The logical operators X_L and Z_L in Fig. b are implemented via the physical operators $XIXI$ and $ZZII$. b) Qubit ℓ_0 is acted on by logical gates X_L and Z_L, and then measured. We say an experiment demonstrates fault tolerance if a family of logical circuits (one member of which is shown in b)) has a higher success probability on average when implemented with encoded circuits (one example shown in a)), rather than with the corresponding unencoded circuits (a single X followed by a single Z, in this example).

is given as the simultaneous +1-eigenspace of the center of $\mathcal{S}$, forming the stabilizer group of the code - in this case $IIII, XXXX, YYYY$, and $ZZZZ$. The remaining operators will in general act non-trivially on the gauge qubit, meaning that each logical state has in general a family of physical states associated with it.

To see how this works, we will identify the logical operations and states of the code being considered. A logical zero state is given by $|0\rangle_L = \frac{1}{\sqrt{2}}(|0000\rangle + |1111\rangle)$. However as we are considering a subsystem code, then there is another orthogonal state that can be associated with logical zero, given by $|\tilde{0}\rangle_L = \frac{1}{\sqrt{2}}(|0011\rangle + |1100\rangle)$. Any state in the subspace spanned by $|0\rangle_L$ and $|\tilde{0}\rangle_L$ corresponds to the same logical zero state, and each of these states only differ by the state of the gauge qubit. The associated $[[4, 2, 2]]$[95, 96] code has the same stabilizers as the Bacon-Shor code discussed in this paper. The codewords of the $[[4, 2, 2]]$ code are

$$\begin{aligned} |00\rangle &\to \frac{1}{\sqrt{2}}(|0000\rangle + |1111\rangle) \\ |01\rangle &\to \frac{1}{\sqrt{2}}(|1100\rangle + |0011\rangle) \\ |10\rangle &\to \frac{1}{\sqrt{2}}(|1010\rangle + |0101\rangle) \\ |11\rangle &\to \frac{1}{\sqrt{2}}(|0110\rangle + |1001\rangle), \end{aligned} \tag{3.1}$$

and so we see that $|\tilde{0}\rangle_L$ is generated by acting on $|0\rangle_L$ with the gauge operator $XXII \in \mathcal{S}$. The operator $ZIZI$ similarly performs a Z rotation on the second logical qubit in the $[[4, 2, 2]]$ code (which is taken as the gauge qubit), but does not act on the logical codespace. This is because gauge operators are logical operators on only one of the qubits in the $[[4, 2, 2]]$ code, and therefore don't affect the logical state of the qubit encoded in the $[[4, 1, 2]]$ Bacon-Shor code. In addition to gauge operators, there are the logical operators $X_L = XIXI$ and $Z_L = ZZII$, which transition the state between distinct logical codewords ($|0\rangle_L$ to $|1\rangle_L$ in the case of $XIXI$).

While the number of logical qubits is fewer than that associated with the $[[4, 2, 2]]$ code, we get several good properties from using a Bacon-Shor code. First, a convenience of working with this code is that the stabilizer operators are all weight-two, which are generally easier to measure fault-tolerantly. As shown in Fig. **??**, a measurement of all the error syndromes only requires eight

two-qubit gates. While these gates may require additional SWAPs in a superconducting architecture, trapped-ion architectures may incur no overhead, as they typically have all-to-all connectivity. This suggests that the required threshold for being able to fault-tolerantly measure the stabilizers will be lower for trapped-ion systems.

Finally, the preparation and stabilizer circuits in Fig. 3.2 and Fig. **??** can be implemented fault-tolerantly. In fact, while the state preparation circuit for the $[4, 2, 2]$ code requires an ancilla, Fig. 3.2 shows that the Bacon-Shor code does not, since a single error can at worst propagate to errors on the first and second qubit, or the third and fourth qubit. In both cases, these correspond to performing a gauge operation that changes the state of the gauge qubit, which is irrelevant for this code. The logical codeword $|0_L\rangle$ can thus be prepared through the initialization circuit shown in Fig. 3.2.

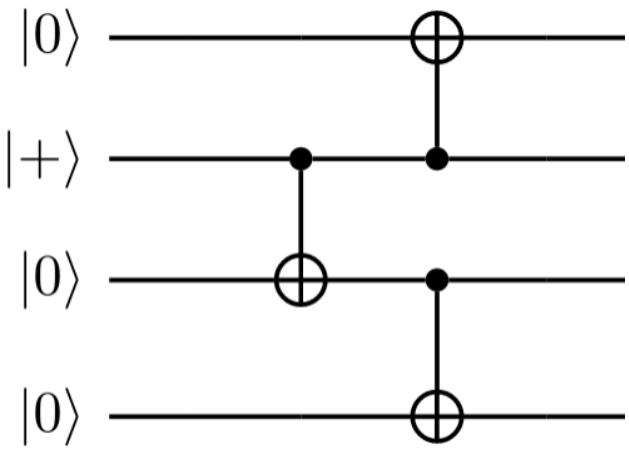


Figure 3.2: Logical zero ($|0_L\rangle$) preparation circuit for the $[[4, 1, 2]]$ Bacon-Shor code.

While a universal gate set is not transversally implementable on the Bacon-Shor subsystem code (meaning that not all single-qubit logical gates can be decomposed entirely into combinations of single-qubit phyical gates), it is possible to transversally implement the logical gates $X_L = X_1X_3$, $Z_L = Z_1Z_2$, and $H_L = \text{SWAP}_{23}H_1H_2H_3H_4$. SWAP_{23}, for our purposes, can be implemented just by updating the labels of the qubits.

3.4 Site-Counting

For another estimate of the threshold of this code, we will follow the site-counting argument in [90, 91] closely. We will derive a bound on the logical error rate for an analytic prediction of when we can expect an encoded algorithm to outperform its unencoded counterpart. As discussed in both Sec. 3.2 and Sec. 3.5, we consider circuits consisting uniformly of X_L, Z_L and H_L gates. These require two, two and four physical gates (as noted previously, SWAP gates can be implemented by relabeling qubits) respectively, so a circuit with T logical gates, will on average contain $(2+2+4)T/3 = 8T/3$ physical sites where an error can occur.

We assume perfect initialization of both the $|0\rangle$ and $|+\rangle$ states, perfect measurements and no idle noise. While in principle these will contribute more infidelity to the logical circuit, the contribution will be small. Using the initialization circuit in Fig. 3.2 with 3 CNOTs, we get $N_A = 6$ locations in the logical state preparation circuit where errors can occur.

The stabilizer measurements require measuring $IIXX$ and $XXII$, which both require 2 CNOTs and 1 Hadamard, contributing 10 locations for error, and measuring $IIZZ$ and $ZZII$ requires 2 CNOTs, giving 8 locations for errors. Thus we have $N_B = 18$ locations per stabilizer measurement. A logical error only arises if more than one error occurs between measurements. Using site-counting we find that the probability for a logical error, p_l, in a circuit with M stabilizer measurements, and $N = \frac{8}{3}T/(M+1)$ physical gates between measurements is bounded as

$$1 - p_l \geq (1 - \binom{N_A+N}{2}p^2)(1 - \binom{N_B+N}{2}p^2)^M, \quad (3.2)$$

where p is the physical error rate. Additionally, we can bound the post-selection probability. For each block of the circuit between stabilizers, a bound on the post-selection probability is given by $1 - Sp$ where S is the number of sites per block, giving

$$p_{ps} \geq (1 - (N_A+N)p)(1 - (N_B+N)p)^M. \quad (3.3)$$

For the unencoded circuit, we have that $p_u \approx Tp$. Finally, to compare the two, we will consider the conditional probability of success. The resulting threshold for when the encoded circuit begins to

outperform the unencoded circuit is shown in Fig. 3.3a. We can additionally use site-counting to estimate the rejection probabilities as shown in Fig. 3.3b. In both cases, the result is qualitatively similar - by repeatedly measuring stabilizers, the post-selection probability decreases exponentially, however by measuring the stabilizers fewer times the post-selection probability increases, due to the potential for errors to collude and become undetectable. In both cases, site-counting gives a more conservative threshold than the numerics. This is due to the looseness of the union bound and because our argument fails to account for benign errors. We could improve our bound by considering benign errors, as in [91].

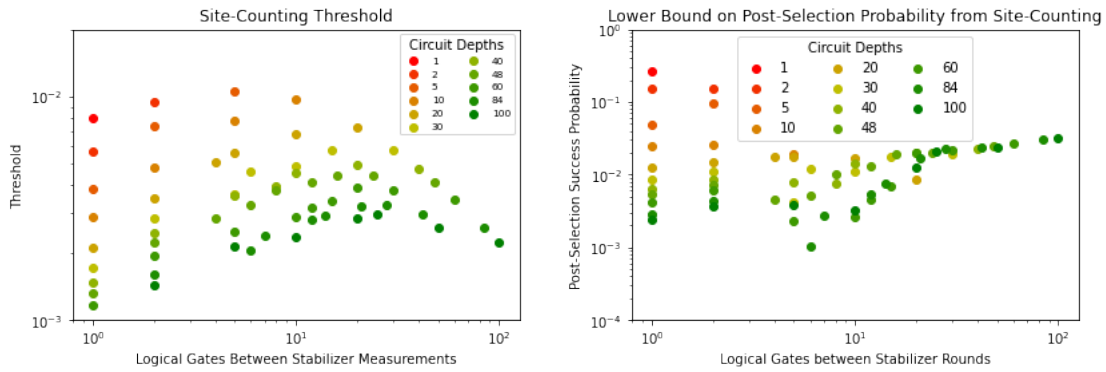
As in Sec. 3.5, we can estimate the optimal number of stabilizer measurements to perform for each circuit depth. Taking the maximum of each curve in Fig. 3.3a we get (1, 2, 5, 10, 20, 15, 20, 16, 20, 28, 25) gates between stabilizer measurements for depths (1, 2, 5, 10, 20, 30, 40, 48, 60, 84, 100) respectively. The site-counting argument is overly pessimistic about the errors contributed by the stabilizer measurements, and so is more sparing in their use, which can be seen by comparing to the 15 logical gates between stabilizer measurements suggested by the numerics in Sec. 3.5.

3.5 Numerics

Elided.

3.6 Conclusion

Quantum error correction is a necessary step towards the development of large-scale quantum computers. Quantum fault tolerance is the demonstration of a quantum error correction scheme that is robust to errors in the error correction schemes themselves. Experimental platforms have recently achieved impressive fidelities that put us near the range that we should consider small quantum codes for both quantum memories and small demonstrations of fault tolerance. Indeed, we have yet to see an unambiguous demonstration of improvement from fault tolerance. In this paper, we have proposed a simple demonstration of fault tolerance that can be used to perform encoded circuits.



(a) (b)

Figure 3.3: (a) Error probability thresholds computed through site-counting. While the typical value is an order of magnitude lower than in the case of the numerics, the shape of the bound is qualitatively similar, showing an improved threshold with a nontrivial rate of intermediate measurements. (b) Post-selection probabilities computed through site-counting, evaluated at the threshold. We see that the results approach a fixed value when only one measurement is performed at the end.

Through both numerics and a site-counting argument, we have shown that Pauli error rates for short circuits (with a depth of five) as high as $p = 0.03$ are capable of demonstrating an improvement of the encoded circuit over the unencoded circuit. Moreover, for deeper circuits (with a depth of 100), we find that performing more than one round of gauge measurements (one round per 15 logical gates) provides a threshold error rate as high as $p = 0.005$, making experimental observation of the threshold potentially realizable on superconducting and trapped ion platforms.

In our analysis, we have assumed that single- and two-qubit gates have similar error rates, modeling two-qubit gates as being subjected to two single-qubit errors. However, state-of-the-art quantum computing platforms have single-qubit error rates that are often smaller by at least an order of magnitude. Nevertheless, the code considered in this paper encodes a single logical qubit. Thus the natural quantity to compare against in a discussion of fault tolerance, and the one we chose to compare against, is another single qubit physical error rate.

There are at least two natural extensions to this work. The first is considering two logical qubits, each encoded in their own four-qubit Bacon-Shor code. In this case, the natural quantity to compare against is the two-qubit gate infidelity, and in the analysis of the logical circuit one could consider single-qubit gates as free. Then one would only count the two-qubit gates used in the state initialization circuit, stabilizer measurements, and to perform entangling operations between the two encoded qubits. This model makes the more realistic assumption of different error rates for gates involving different numbers of qubits. This assumption is particularly relevant in trapped ion systems where single-qubit gate infidelities are exquisitely small[6, 7].

The other straightforward extension of this work is to consider the $[[9, 1, 3]]$ Bacon-Shor code. This code also encodes a single logical qubit, but is distance three and therefore can correct one error. Thus, rather than simply rejecting circuits as discussed in this paper, we can consider implementing error-correcting circuits which should improve the threshold for deeper circuits.

While this work focused on an uncorrelated depolarizing noise model, an additional direction of research is to consider more realistic noise models for trapped-ion and superconducting architectures. However, we are finally reaching an exciting point in hardware where quantum codes can be

implemented and have the ability to outperform bare physical qubit schemes, and are therefore on the brink of scalability.

Chapter 4

Opportunities and Limitations in Broadband Sensing [97]

Abstract

Detecting a signal at an unknown frequency is a common task, arising in settings from dark matter detection to magnetometry. For any detection protocol, the precision achieved depends on the signal's frequency and can be quantified by the quantum Fisher information (QFI). To study limitations in broadband sensing, we introduce the integrated quantum Fisher information and derive inequality bounds that embody fundamental trade-offs in any sensing protocol. Our inequalities show that sensitivity in one frequency range must come at the cost of reduced sensitivity elsewhere. For many protocols, including those with small phase accumulation and those consisting of π-pulses, we find the integrated quantum Fisher information scales linearly with T. We also find protocols with substantial phase accumulation that can have integrated QFI that grows quadratically with T, and prove that this scaling is asymptotically optimal. These protocols may allow the very rapid detection of a signal with unknown frequency over a very wide bandwidth. We discuss the implications of these results for a wide variety of contexts, including dark matter searches and dynamical decoupling. Thus we establish fundamental limitations on the broadband detection of signals and highlight their consequences.

4.1 Introduction

Quantum systems, from atoms to SQUIDs, can be excellent sensors. Indeed detecting weak signals requires the consideration of quantum effects. Even better, entangled detectors are well-

known to be more sensitive than their unentangled counterparts [98]. In practice, quantum sensors have been used for dark matter searches, entanglement-enhanced magnetometry, microwave clocks, and matterwave interferometers. [99–108].

The quantum Fisher information (QFI) captures the performance of a parameter estimation protocol [109]. For a pure state parameterized by θ, $\rho_\theta := |\psi_\theta\rangle \langle\psi_\theta|$, the QFI is

$$J(\theta) = 2\mathrm{Tr}\left[(\partial_\theta \rho_\theta)^2\right], \tag{4.1}$$

[110]. The QFI tells us, via the Cramer-Rao bound [111], how well an unbiased estimator of θ can approximate its true value. In particular, given m copies of $|\psi_\theta\rangle$ the variance of any unbiased estimator $\hat{\theta}$ must satisfy $\mathrm{Var}(\hat{\theta}) \geqslant \frac{1}{mJ(\theta)}$. This bound can be saturated, so the more quantum Fisher information a protocol has, the better one can estimate θ [111].

We consider a coupling between the qubit and the signal as

$$H(t) = \mu B \cos(\omega t + \varphi) Z, \tag{4.2}$$

where μ is the magnetic moment of the qubit. For instance, we may wish to estimate the strength of an AC magnetic field [112–120]. While we have included φ in the analysis in Eq. (4.7), we have omitted it for convenience elsewhere.

To gather information about this Hamiltonian, we need to establish a protocol, which generically consists of preparing the sensing qubit in an initial state, applying a time-dependent control sequence, and finally performing a measurement. The performance of a protocol will depend on the frequency of the signal ω. For example, preparing $|+\rangle$ followed by free evolution for time T and measurement in the $|\pm\rangle$ basis is optimal for $\omega = 0$ but performs poorly for $\omega \gg 1/T$. In fact, we will show that trade-offs in sensitivity at different frequencies are inevitable. We make this quantitative by considering the integrated QFI (IQFI).

While this may be different than the integrated sensitivity of a single protocol, it provides an analytically tractable method of analysis which is tight in many cases of interest, namely weak fields and protocols consisting only of $\pi-$pulses. In general, a choice of protocol includes a choice

of measurement basis, as defined above. In the majority of examples we consider in this paper, the optimal measurement basis is frequency independent, so that the the optimization over the final measurement in the definition of QFI can be ignored. By bounding this integral we formalize the idea that there is a fundamental tension between having sensitivity in different frequency bands.

The longer we observe a signal, the more we can expect to learn about it. Thus, it is no surprise that the IQFI will typically grow with the duration T of a protocol. In fact, we find a number of constraints on how IQFI grows with T. First, we find that any protocol starting on the equator of the Bloch sphere that involves only π-pulses has an IQFI of $2\pi\zeta^2 T$, where $\zeta \equiv \mu/\hbar$ is the inverse gyromagnetic ratio of the system being used as a sensor [121]. Second, for an arbitrary protocol with $\zeta BT \ll 1$, the IQFI is close to $2\pi\zeta^2 T$. Then we study a particular protocol that significantly exceeds $2\pi\zeta^2 T$ — by continuously driving a spin with a transverse gX term, our protocol has a peak sensitivity around $2g$, with IQFI scaling quadratically with time. We further show that the IQFI can not exceed quadratic scaling with time, so that this protocol is in a sense optimal. However, the practical restriction to the small signal regime $\zeta BT \ll 1$ is not uncommon, due in part to phase ambiguities that may arise in the accrued phase if one begins to leave this regime.

Beyond signal estimation and detection, our results can be applied to better understand the performance of dynamical decoupling [122–125]. Here we find that for many dynamical decoupling protocols, the average (over the initial state) IQFI is at least linearly proportional to T. As a result, dynamical decoupling can at best move the noise sensitivity of a qubit around in frequency space, rather than eliminating sensitivity at all frequencies. This is reminiscent of the filter functions discussed in [126] which considers CPMG sequences [127–129], a subset of the protocols considered in this work.

We note that the setting under consideration is different from the waveform estimation studied in [130]. That work studied how to simultaneously estimate a large number of parameters representing the full time series of a waveform. We consider the sensing problem of estimating a single Fourier amplitude, i.e. the systems we consider couple to a monochromatic signal $B\cos(\omega t)$. The relative simplicity of this setting admits a global analysis of the performance of an arbitrary

protocol at different frequencies.

4.2 Preliminaries

We consider Hamiltonians of the form Eq. (4.2), motivated by a spin-$\frac{1}{2}$ particle in a magnetic field. We first consider estimation protocols composed of instantaneous, arbitrary unitary rotations P_i followed by periods of free evolution. The choice of measurement at the end of the protocol is assumed to be optimal in the sense that it maximizes the Fisher information of the resulting classical probability distribution. For a state starting in the $+1$ X eigenstate, we have the final state

$$|\psi(T,\omega,B)\rangle = U_{N-1}(t_N,t_{N-1})P_{N-1}...U_0(t_1,t_0)P_0\,|+\rangle\,, \tag{4.3}$$

where $t_N = T$ and $U(t_{i+1},t_i)$ is the time evolution operator under the Hamiltonian in Eq. (4.2) between times t_i and t_{i+1}.

Given Eq. (4.3), the QFI tells us how well we can estimate B. We write $J(B|\omega)$ to indicate that Fisher information with respect to B will in fact depend on the signal frequency ω. Writing $|\phi\rangle := \partial_B\,|\psi(T,\omega,B)\rangle$ and $|\psi\rangle := |\psi(T,\omega,B)\rangle$ the QFI can be expressed as [110, 131]

$$J(B|\omega) = 4(\langle\phi\rangle\,\phi + \mathrm{Re}\{\langle\phi\rangle\,\psi^2\}). \tag{4.4}$$

To understand the total sensitivity of a protocol across all frequencies we define the integrated QFI (IQFI) for a protocol with total evolution time T as [a]

$$K(T) = \int_0^\infty d\omega J(B|\omega). \tag{4.5}$$

Statistically, one way of understanding the IQFI is as an unnormalized expected quantum Fisher information (EQFI) for estimating a magnetic field amplitude at an unknown frequency with a uniform prior. While there is no uniform distribution over the positive real line, because the QFI vanishes at high frequencies, it is possible, in principle, to establish a cutoff at which

[a] Note that the integral in Eq. (4.5) always exists since the QFI is bounded and asymptotically we have $J(\rho_\omega) \sim (\sum_{i=0}^{N-1}\Theta_i)^2 \sim 1/\omega^2$

the IQFI is approximately an unnormalized expected quantum Fisher information. As we show in Appendix 4.A.4, this cutoff is independent of time, and so can be varied independently of protocol duration. While the EQFI is an interesting statistical quantity, because this equivalence is only approximate, and because we are interested in a measure of how the total (unnormalized) bandwidth sensitivity of a protocol changes with time, we consider the IQFI rather than the EQFI. One way of understanding this distinction is that the former has units of time, while the latter has units of time squared, and so they measure different physical quantities.

Nevertheless, it is possible to give an interpretation of our results in the Bayesian setting. For the choice of a delta function prior, it is known that EQFI grows at most quadratically with time [131]. In this work, however, we are interested in understanding the optimal performance of a protocol given an arbitrary prior. We will show that the optimal performance for the uniform prior is also quadratic in time, and thus for any prior quadratic scaling is also optimal.

To see this, consider the following integral where μ is the uniform prior, ξ is any other prior, and Ω is some cutoff frequency beyond which the QFI and ξ are both negligible,

$$\int_0^\Omega d\xi(\omega) J(B|\omega) \leqslant M \int_0^\Omega d\mu(\omega) J(B|\omega). \tag{4.6}$$

In this equation $M = \max(\xi(\omega))/\mu_0$ where μ_0 is the value of the uniform prior. Any error in this expression is independent of time and can be suppressed by increasing Ω, and so we see that the asymptotic behavior of the EQFI with a uniform prior provides an upper bound on the EQFI with any prior.

4.3 Ramsey and π-pulse protocols

Within the family of control sequences consisting of instantaneous rotations interleaved with free-evolution, we now consider Ramsey spectroscopy, where a qubit is prepared on the equator of the Bloch sphere and allowed to freely precess. Defining $\Theta(t_{i+1}, t_i) = [\sin(\omega t_{i+1} - \sin(\omega t_i]/(\hbar\omega)$, we find $|\phi\rangle = -i\zeta\Theta(T,0)|\psi\rangle$ and Eq. (4.4) gives $J(B|\omega) = 4\zeta^2\Theta^2(T,0)$. Defining $\zeta \equiv \mu/\hbar$, the IQFI

follows as

$$K(T) = 4\zeta^2 \int_0^\infty d\omega \frac{(\sin(\omega T + \varphi) - \sin(\varphi))^2}{\omega^2}, \tag{4.7}$$

$$= 2\zeta^2 T(\pi - \ln(4)\sin(2\varphi)). \tag{4.8}$$

If φ is unknown and therefore random in each experiment, averaging over φ gives $K(T) = 2\pi\zeta^2 T$, but if φ is known and we wish to maximize the IQFI, we would lock the experimental sequence to $\varphi = 3\pi/4$ to get $K(T) = 2\zeta^2 T(\pi + \ln 4)$. For convenience, in what follows we assume $\varphi = 0$.

Now consider a protocol applying π-pulses at times $t_1, \ldots, t_N = T$. At time t_i we apply either X, Y or Z. Additionally, we can apply any unitary that leaves the expectation value of Z invariant [b]. Then we have

$$\begin{aligned} K(T) &= 4\zeta^2 \int_0^\infty d\omega \Big(\sum_{i=0}^{N-1} \Theta(t_{i+1}, t_i)\Big)^2 \\ &= 2\pi\zeta^2 T. \end{aligned} \tag{4.9}$$

$K(T)$ was derived assuming that the system was initialized in the $|+\rangle$ state. If we instead integrate over all initial states we find that the average IQFI is given as $\frac{4\pi\zeta^2 T}{3}$. This average has important implications for dynamical decoupling protocols based on π-pulses. It shows that for such protocols a qubit must maintain a sensitivity to environmental noise over a substantial frequency range. Indeed, in the presence of white noise any π-pulse protocol will leave the qubit equally degraded by the noise. Specifically, we imagine that a qubit is subjected to a background power spectrum of magnetic field noise fluctuations whose noise spectrum is flat. A lower bound on the IQFI implies a lower bound on how much of this noise spectrum the qubit will experience, and therefore on its decoherence. For noise with more structure, these protocols do not allow sensitivity to noise to be eliminated, but can simply move that sensitivity to a frequency range where the environmental noise is fairly low.

---

[b] Consider a $\pi-$rotation about the axis halfway between the X- and Y- axes, given by $\exp\Big\{(-i\frac{\pi}{2\sqrt{2}}(X+Y)\Big\})$

4.4 B=0 Bound

We now present an argument to bound $K(T)$ at $B = 0$, and approximately bound $K(T)$ for short times and weak magnetic fields. Consider protocols with initial state $|\psi_0\rangle$ and a control Hamiltonian $H_0(t)$ in addition to the signal Hamiltonian $H(t)$. We can then write $|\psi(T, \omega, B)\rangle = U\mathcal{U}|\psi_0\rangle$, where U is the time evolution due to H_0 and $\mathcal{U}$ is the interaction picture time-evolution, given by

$$\mathcal{U} = 1 - i\zeta B \int_0^T \cos(\omega\tau) Z_I(\tau) d\tau + O(B^2), \tag{4.10}$$

where we have used $Z_I(\tau)$ to express Z in the interaction picture. Then we have

$$\partial_B |\psi(T, \omega, B)\rangle \Big|_{B=0} = -i\zeta U \int_0^T \cos(\omega\tau) Z_I(\tau) d\tau |\psi_0\rangle . \tag{4.11}$$

Substituting into Eq. (4.4) and integrating, we find that the IQFI is at most $2\pi\zeta^2 T$. For small B and T, the next term on dimensional grounds should be $\mathcal{O}(B^2T^3)$, since we can show the term linear in B is zero. This dimensional analysis assumes that there are no other dimensionful quantities - for instance, if the interpulse spacings are not functions of T, then there may be other terms. Thus we find

$$K(T) \leqslant 2\pi\zeta^2 T + O(\zeta^4 B^2 T^3) \tag{4.12}$$

The full proof is provided in Appendix 4.A.2. This shows that for small magnetic fields and short times we should expect a roughly linear scaling of the IQFI.

4.5 Entangled probe advantage

From standard results [132, 133] we expect that entangled inputs can outperform this bound. Indeed an $n-$qubit GHZ state

$$|GHZ\rangle_n = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n}). \tag{4.13}$$

accumulates phase n times more rapidly, so that an analogous argument gives an IQFI at $B = 0$ of

$$K(T) = 2\pi n^2 \zeta^2 T. \tag{4.14}$$

Conversely, product input states can be reduced to the single qubit example, since $J(\rho^{\otimes n}) = nJ(\rho)$, so that for n qubits starting in a product state, again with $B = 0$, we have

$$K(T) \leqslant 2\pi n\zeta^2 T. \tag{4.15}$$

So, while entanglement allows us to increase the coefficient in front of T, IQFI still increases linearly with time.

4.6 Quadratic scaling of IQFI

We have seen that π-pulse protocols and protocols with $BT\zeta \ll 1$ have IQFI that scales no faster than $2\pi\zeta^2 T$. Even an entanglement-enhanced protocol gives linear scaling of IQFI with T, albeit with an improved coefficient. We now study a simple protocol with IQFI scaling quadratically in T.

We consider a continuous-time protocol which applies a transverse field, $\hbar gX$, to the sensing qubit. This gives a full Hamiltonian of

$$H(T) = \hbar gX + \mu B\cos(\omega T)Z. \tag{4.16}$$

Assuming $\omega \sim 2g$, $BT\zeta \gg 1$ and using the rotating wave approximation [134], we find

$$J(B|\omega) \sim \frac{(\zeta T)^2}{(1 + (\frac{\omega - 2g}{B\zeta})^2)^2} \tag{4.17}$$

which we can integrate from $\omega = g$ to $\omega = 3g$ to get a lower bound of

$$K(T) \gtrsim \zeta^2 T^2 \left(\frac{g}{1 + \frac{g^2}{\zeta^2 B^2}} + \zeta B \tan^{-1}\left(\frac{g}{\zeta B}\right)\right). \tag{4.18}$$

A detailed derivation is given in Appendix 4.A.3. Representative dynamics in this regime are shown in Fig. 3 and 4 of [135].

In Fig. 4.2 we numerically study this protocol. We first approximate it by a discrete pulse-based protocol, described by Eq. (4.3) and given by Trotterization, that intersperses instantaneous rotations around the X-axis by periods of free evolution under the magnetic field. For time T, we

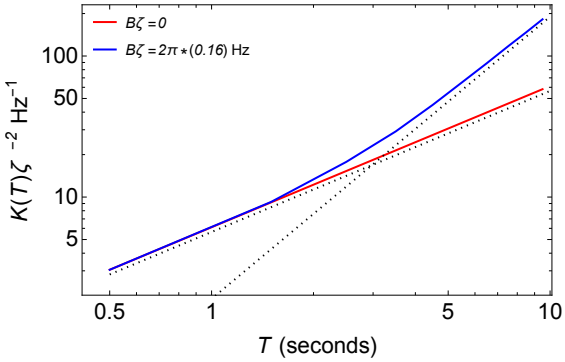


Figure 4.1: IQFI ($K(T)$) for using $X_{\pi/2}$ pulses. Depending on the magnetic field strength we observe a more rapid accumulation of IQFI. The dashed lines are $\sim T$ and $\sim T^2$ scalings to guide the eye. For non-zero magnetic field, we see that a crossover from quadratic to linear scaling occurs when when $BT\zeta \ll 1$ (where the perturbative results are valid).

can approximate evolution under Eq. (4.16) by m periods of free evolution of duration $\frac{T}{m}$ separated by rotations of angle $\frac{T\pi}{m}$ about the X-axis. Indeed, in Fig. 4.1, we see quadratic scaling choosing $m = 2T$. In Fig. 4.2 we compare the QFI of the $m = T$ and $m = 2T$ cases with the gX protocol and a Ramsey protocol, where it is evident that both the $m = 2T$ protocol and the gX protocol accumulate IQFI more rapidly than the other two protocols. These discrete protocols with quadratic scaling of IQFI have the property that the number of pulses scales with the total time of the protocol. In fact, this is necessary, as we will now see.

Consider a protocol with N pulses P_i applied between periods of free evolution U_i. Because $\langle\psi\rangle\,\psi = 1$, $\partial_B \langle\psi\rangle\,\psi = 0$, we have $\langle\phi\rangle\,\psi = -\langle\psi\rangle\,\phi$. This means that $Re(\langle\phi\rangle\,\psi) = 0$, and so $\langle\psi\rangle\,\phi^2$ has real part that is non-positive. Thus from Eq. (4.4) we see

$$J(B|\omega) \leqslant 4\,\langle\phi\rangle\,\phi. \tag{4.19}$$

We see that $\langle\phi\rangle\,\phi = \zeta^2 \sum_{i,j}^{N} \Theta_i \Theta_j V_{ij} = \zeta^2 \mathbf{\Theta}^T V \mathbf{\Theta}$, where $\mathbf{\Theta}$ is a vector whose i^{th} entry is $\Theta_i \equiv \Theta(t_{i+1}, t_i)$. V is an $N \times N$ complex matrix with entries of norm at most 1, so its eigenvalues have norm at most N. We thus find

$$K(T) = \int_0^\infty d\omega J(\omega) \leqslant 4 \int_0^\infty d\omega\,\langle\phi\rangle\,\phi \tag{4.20}$$

$$= 4\zeta^2 \int_0^\infty d\omega \mathbf{\Theta}^T V \mathbf{\Theta} \leqslant 4\zeta^2 \int_0^\infty d\omega N |\mathbf{\Theta}|^2 = 2\pi N \zeta^2 T.$$

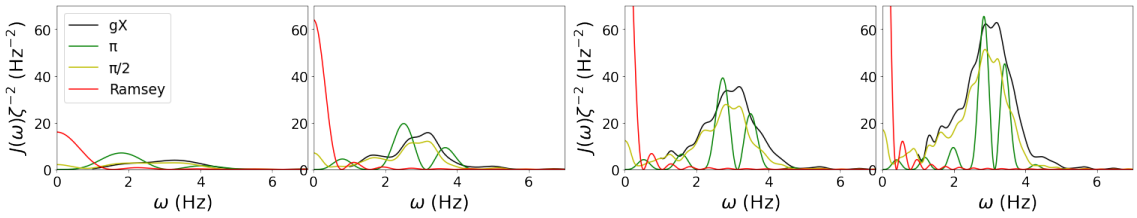


Figure 4.2: QFI as a function of frequency for some protocols considered in this paper, for different protocol times T [(a) $T = 2$, (b) $T = 4$, (c) $T = 6$, and (d) $T = 8$ seconds], with $B\zeta = 1$ Hz. Ramsey has a large DC QFI, as expected, and very little ability to detect any AC signal. π-pulses, on the other hand, can be used to measure signals at higher frequencies, as might be expected from spectroscopy techniques such as CPMG. The gX protocol, with $g = 2\pi \times \frac{1}{4}$Hz in this example, is seen to be sensitive near $2g$. Moreover, it is seen to be sensitive over a broad bandwidth. The π-protocol shown here consists of π rotations about the X axis, at each integer value of time. The $\pi/2$-protocol consists of $\pi/2$ rotations about the X axis, every half second. Similar to the π-protocol, the $\pi/2$-protocol has AC sensitivity, as well as more broadband sensitivity.

We will now see that the bound Eq. (4.20) can be used to show that the IQFI can scale at most quadratically with time, so that up to a constant the IQFI of the gX protocol grows as rapidly as possible in time.

A continuous-time protocol involves a control Hamiltonian G that gives a total Hamiltonian of the system

$$H(t) = \hbar G(t) + \mu B \cos(\omega t) Z. \tag{4.21}$$

We can Trotterize this evolution into a discrete sequence [136] like those considered in proving Eq. (4.20). Then, by linearity of the derivative we can constrain the derivative of the Trotterized evolution to be close to the actual derivative. Using $U(t)$ to refer to the continuous time protocol, and $\hat{\epsilon}''(t) = \partial_B U(t) - \partial_B U'(t)$, we see from Eq. (4.19),

$$J(\omega) \leqslant J_d(\omega) + 8Re(\Big\langle + |\hat{\epsilon}''(T)^\dagger (\partial_B U'(T) + 4\hat{\epsilon}''(T))| + \Big\rangle),$$

where we have used $J_d(\omega)$ to denote the QFI of the protocol given by $U'(t)$. As we show in Appendix **??**, this gives a bound of

$$K(T) \leqslant \frac{2\pi\zeta^2 T^2}{\delta t} \quad + \quad \alpha(\delta t, \Omega, \mu B, \hbar||G||) T^2,$$

where $||G||$ is maximum spectral norm of the control, $\max_{t \leqslant T} ||G(t)||$ and α is a function that controls the error in approximating the IQFI of the continuous protocol by the IQFI of the Trotterization. Crucially it does not depend on T. This proves that $K(T)$ scales at most quadratically in time. Thus, our examples with quadratic scaling are asymptotically optimal in the amount of IQFI they accumulate. This bound can be extended to consider both entangled probes and separable probes as shown in Appendix 4.A.4, and thus establishes a fundamental limit on the broadband sensitivity of detectors.

4.7 Conclusions

The QFI provides an ultimate bound on how well a quantity can be estimated, in our case the amplitude of a sinusoid with fixed frequency. Integrating the QFI over all frequencies, we found

fundamental limits on the broadband performance of quantum sensors. For tasks such as axion detection [137], this implies that spectral sensitivity is a scarce resource that needs to be carefully considered when designing metrological protocols. While conventional spectroscopy protocols such as Ramsey interferometry and CPMG [127, 128] consist only of π-pulses, and therefore linearly accumulate IQFI, we found both continuous and discrete protocols that quadratically increase this accumulation. Moreover, we have shown that this is asymptotically the largest scaling one can achieve.

We have seen that there are protocols with IQFIs that scale as both $\mathcal{O}(\zeta^2 T)$ and $\mathcal{O}(\zeta^3 B T^2)$, but which is better? It depends—if the goal is sensitivity to a wide range of frequencies, $\mathcal{O}(\zeta^3 B T^2)$ may allow the protocol to work over a wider frequency band. If the goal is sensitivity to a very narrow frequency range, $\mathcal{O}(\zeta^2 T)$ protocols may have support over a small band. Thus, we may see enhancements when searching for a weak signal over a wide frequency range. In such a setting, long integration times could give a quadratic enhancement of the accumulated QFI compared to the $\mathcal{O}(\zeta^2 T)$ protocols.

Additionally if T_d is a characteristic decoherence time of the system, then we are practically constrained to $T < T_d$. Thus for $B \ll 1/(\zeta T_d)$ and fixed T_d, we also find $\zeta B T \ll 1$. In these contexts, our results show that sensing should be expected to be limited by the bound of $2\pi\zeta^2 T$. Physically, this corresponds to when the peak angular excursion of the Bloch vector is much less than π. For weak fields, if we can sense for long times so that $\zeta B T \ll 1$, then we can accumulate quadratically more IQFI.

The gX protocol is sensitive to frequencies around $2g$, making it an excellent candidate for broadband detection around a particular frequency. It is an open question how to design optimal metrological protocols with sensitivity spread evenly over wide bands. Techniques like GRAPE [138] may be useful for this task [139].

Many dynamical decoupling protocols consist solely of π-pulses (e.g., [140]). Such techniques may be described by Eq. (4.3). Consequently, our results show these decoupling strategies are fundamentally limited - while they can move noise sensitivity, they cannot remove it. We leave

open whether such bounds apply to arbitrary protocols.

Our key conceptual contribution is the idea that IQFI can be used to understand the trade-offs inherent in broadband sensing. In some cases, this metric provides a conservation law that can be summarized by the slogan "no free QFI". In particular, in the case where the interaction picture operator being sensed ($Z_I(t)$) commutes with itself at all times and in the small signal limit ($\zeta BT \ll 1$) we showed that QFI at one frequency ultimately comes at the cost of less QFI at another frequency. This is also true for sequences consisting only of π pulses, when the sensor state begins on the equator. Moreover we have shown that for *any* protocol there is a limited amount of IQFI that can be accumulated. This demonstrates that while broadband sensing is possible, there is an upper limit on how wide the bandwidth of a given protocol can be if one desires a certain sensitivity. We do not currently know if other classes of control protocols yield strict conservation laws, and we leave this to future work.

The bounds on IQFI that we have found on single qubit initial states can be extended, via Eq. (4.15), to arbitrary separable states due to the additivity and convexity of the QFI [141, 142]. As Eq. (4.15) applies only to separable probe states, we can think of it as a kind of standard quantum limit that cannot be exceeded without entanglement. Indeed, we see that an n-qubit cat state can significantly exceed the $2\pi n\zeta^2 T$ performance of unentangled π-pulse-based protocols. This points to the possibility of using IQFI as a form of entanglement witness, so that the quantum Fisher information at any particular frequency may be consistent with a separable state but the breadth of such sensitivity can only be explained by an entangled state. In general, however, we find that the quadratic-in-time asymptotic upperbound of the IQFI holds for both separable and entangled probes. Finally, another interesting open question is if other transformations of QFI spectra might generate new insights into broadband sensing limitations.

4.A Appendix

4.A.1 Derivation of Average IQFI for Instantaneous π-Pulses

Using the identity

$$[A, UBU^\dagger] = U[U^\dagger AU, B]U^\dagger \tag{4.22}$$

we find the N-pulse QFI is given by

$$J(\omega) = -2\,\mathrm{Tr}\left(\left(\left[U_0^\dagger P_0^\dagger U_1^\dagger P_1^\dagger ... U_{N-1}^\dagger P_{N-1}^\dagger Z P_{N-1} U_{N-1} ... P_0 U_0, \frac{X}{2}\right]\Theta_{N-1} + ... + \left[Z, \frac{X}{2}\right]\Theta_0\right)^2\right)$$

Integrating over all frequency and using the orthogonality condition

$$\int d\omega \Theta_i \Theta_j = 2\pi(t_{i+1} - t_i)\delta_{ij}, \tag{4.23}$$

we see that when this square is integrated this will sum to be $2\pi T$, so that for a protocol starting in the $|+\rangle$ state and using only π-pulses we find

$$K(T) = 2\pi T. \tag{4.24}$$

We can also average over the input state, denoted with angle brackets $\langle\cdot\rangle$, to find

$$\langle K(T)\rangle = \frac{1}{4\pi}\int_0^{2\pi} d\beta \int_0^{\pi} d\alpha \sin(\alpha) 2\pi\zeta^2 T \sin^2\alpha, \tag{4.25}$$

where one copy of sin comes from the integration measure, and two copies come from the off-diagonal element of the density matrix. This gives the result from the main text,

$$\langle K(T)\rangle = \frac{4\zeta^2\pi}{3}T. \tag{4.26}$$

4.A.2 Perturbative Expansion to $\mathcal{O}(B^2)$

Elided.

4.A.3 Derivation of the IQFI under the Rotating Wave Approximation

Elided.

4.A.4 Bound on IQFI of Continuous Protocols

As in the main text, consider a sensing protocol defined by a time-dependent control $G(T)$. In particular, the full Hamiltonian we are considering is

$$H(T) = \hbar G(T) + \mu B \cos(\omega T) Z. \tag{4.27}$$

We can Trotterize the evolution into a discrete sequence [136], which will look like those considered in the proof above, with some step size δt. The evolution will be

$$U'(T) = \mathcal{T} e^{i \int_{(N-1)\delta t}^{N\delta t} dt G(t)} e^{i\zeta B \Theta_N Z} \dots \mathcal{T} e^{i \int_0^{\delta t} dt G(t)} e^{i\zeta B \Theta_1 Z} + O(\delta t^2) \tag{4.28}$$

with $\Theta_k = [\sin(\omega k \delta t) - \sin(\omega(k-1)\delta t)]/(\hbar\omega)$, $N\delta t = T$, and where $\mathcal{T}$ denotes the time-ordering operator, which is necessary because in general $G(t)$ will not commute with itself at all times. The number of pulses, N, in the discrete protocol described in Eq. (4.28) gives, to zeroth order,

$$\int d\omega J_c(\omega) \leqslant \frac{2\pi\zeta^2 T^2}{\delta t}, \tag{4.29}$$

for all δt. There are however error terms from the Trotter expansion that we need to propagate through the IQFI - this is what we will do now. By linearity of the derivative operator we can also constrain the derivative of the Trotterized evolution to be close to the derivative of the actual evolution. In particular, we have that

$$U(T) = U'(T) + \hat{\epsilon}'(T) \tag{4.30}$$

$$\partial_B U(T) = \partial_B U'(T) + \hat{\epsilon}''(T) \tag{4.31}$$

where $\hat{\epsilon}''(T) = \partial_B \hat{\epsilon}'(T)$ is the B derivative of the error term in the Trotter expansion. Then we see from Eq. (4.19),

$$J_c(\omega) \leqslant J_d(\omega) + 4(\left\langle +|\hat{\epsilon}''^{\dagger}(T)\partial_B U'(T) + \partial_B U'^{\dagger}(T)\hat{\epsilon}''(T)|+\right\rangle + \left\langle +|\hat{\epsilon}''(T)^{\dagger}\hat{\epsilon}''(T)|+\right\rangle) \tag{4.32}$$

$$= J_d(\omega) + 8Re(\left\langle +|\hat{\epsilon}''(T)^{\dagger}\partial_B U'(T)|+\right\rangle) + 4\left\langle +|\hat{\epsilon}''(T)^{\dagger}|\hat{\epsilon}''(T)|+\right\rangle). \tag{4.33}$$

We will analyze the integrals of the error terms piecewise, first up to a frequency $\Omega > 0$. The derivative of the first order Trotterized evolution defined above is

$$\partial_B U'(T) = -i\zeta \sum_{j=0}^{N-1} \Theta_j P_0 U_0 ... P_j Z U_j ... P_{N-1} U_{N-1}. \tag{4.34}$$

Noting that $|\Theta_k| \leqslant 2T/N$ for a uniformly spaced pulse sequence, this gives

$$|| \int_0^{\Omega} \hat{\epsilon}''(T)(-\frac{i\mu}{\hbar}) \sum_{j=0}^{N-1} \Theta_j P_0 U_0 ... P_j U_j ... P_{n-1} U_{n-1}|| \tag{4.35}$$

$$\leqslant \frac{\mu}{\hbar} \text{max}_{\omega<\Omega} ||\hat{\epsilon}''(T)|| \int_0^{\Omega} \sum_{j=0}^{N-1} |\Theta_j| ||P_0 U_0 ... P_j U_j ... P_{n-1} U_{n-1}|| \tag{4.36}$$

$$\leqslant \frac{\mu}{\hbar} \text{max}_{\omega<\Omega} ||\hat{\epsilon}''(T)|| \int_0^{\Omega} \sum_{j=0}^{N-1} 2T/N \tag{4.37}$$

$$= \frac{2\mu}{\hbar} \text{max}_{\omega<\Omega} ||\hat{\epsilon}''(T)|| \Omega T, \tag{4.38}$$

Similarly,

$$4 \int_0^{\Omega} ||\hat{\epsilon}''(T)^{\dagger}\hat{\epsilon}''(T)|| \leqslant 4 \text{max}_{\omega<\Omega} ||\hat{\epsilon}''(T)||^2 \Omega \tag{4.39}$$

where the maximum is over ω on which the error term implicitly depends. Then we have, for the last two terms in Eq. (4.33)

$$8|| \int_0^{\infty} Re(\hat{\epsilon}''^{\dagger}(T)\partial_B U'(T))|| \leqslant \frac{16\mu}{\hbar} \Omega T \text{max}_{\omega<\Omega} ||\hat{\epsilon}''(T)|| + || \int_{\Omega}^{\infty} 8Re(\hat{\epsilon}''^{\dagger}(T)\partial_B U'(T))||, \tag{4.40}$$

and

$$4 \int_0^{\infty} ||\hat{\epsilon}''(T)^{\dagger}\hat{\epsilon}''(T)|| \leqslant 4 \text{max}_{\omega<\Omega} ||\hat{\epsilon}''(T)||^2 \Omega + 4 \int_{\Omega}^{\infty} ||\hat{\epsilon}''(T)^{\dagger}\hat{\epsilon}''(T)||. \tag{4.41}$$

$\hat{\epsilon}''(T)$ is at worst proportional to T, since the error in a time step δt is independent of T (it's proportional to $||[\hbar G(t), \mu B\cos(\omega t)Z]||\delta t^2$), and there are $N = T/\delta t$ times steps. Furthermore

$$||\partial_B U'(T)|| = ||\frac{-i\mu}{\hbar}\sum_j \Theta_j P_1 U_1 ... P_j U_j ... P_n U_n|| \leqslant \mu N/(\hbar\omega) = \mu T/(\delta t\hbar\omega), \tag{4.42}$$

so we need only understand $||\hat{\epsilon}''(T)||$ at high frequency, where $||\cdot||$ is the spectral norm. To this end, consider the error in time step δt given in [143] as $U'(t_k, t_{k-1})F(t_k, t_{k-1})$, where $U'(a,b)$ is the Trotterized evolution from time a to time b and

$$F(t_k, t_{k-1}) = \int_{t_{k-1}}^{t_k} C(v, t_{k-1})F(v, t_{k-1})dv \tag{4.43}$$

$$C(t_k, t_{k-1}) = \exp\left\{(i\int_{t_{k-1}}^{t_k} G(s)ds)\right\}^\dagger \int_{t_{k-1}}^{t_k} du\ \exp\left\{(i\int_{t_{k-1}}^{u} A(s)ds)\right\}^\dagger \tag{4.44}$$

$$\times[A(u), \hbar G(t_k)]\exp\left\{(-i\int_{t_{k-1}}^{u} A(s)ds)\right\}\exp\left\{(-i\int_{t_{k-1}}^{t_k} G(s)ds)\right\}, \tag{4.45}$$

where $A(s) = \frac{\mu B}{\hbar}\cos(\omega s)Z$. Thus, to lowest order in $1/\omega$,

$$\begin{aligned}\exp\left\{(-i\int_{t_{k-1}}^{u} A(s)ds)\right\} &= \exp\{(iBZ(\cos(\omega u) - \cos(\omega t_{k-1}))/\omega)\} \\ &\approx I - i\frac{\mu B}{\hbar}Z(\cos(\omega u) - \cos(\omega t_{k-1}))/\omega\end{aligned} \tag{4.46}$$

so that the relevant integrals over u are

$$\int_{t_{k-1}}^{t_k} du\cos(\omega u) = (\sin(\omega t_k) - \sin(\omega t_{k-1}))/\omega \tag{4.47}$$

$$\int_{t_{k-1}}^{t_k} du\cos(\omega u)(\cos(\omega u) - \cos(\omega t_{k-1})) \tag{4.48}$$

$$= (2\delta t\omega + \sin(2t_{k-1}\omega) - 4\cos(t_{k-1}\omega)\sin((t_{k-1}+\delta t)\omega) + \sin(2(t_{k-1}+\delta t)\omega))/(4\omega) \tag{4.49}$$

Plugging these in we find

$$C(t_k, t_{k-1}) = \frac{\mu B}{\hbar}((\sin(\omega t_k) \tag{4.50}$$

$$-\sin(\omega t_{k-1}))/\omega)(\exp\left\{(i\int_{t_{k-1}}^{t_k}\hbar G(s)ds)\right\}^\dagger [Z, \hbar G(t_k)]\exp\left\{(-i\int_{t_{k-1}}^{t_k}\hbar G(s)ds)\right\}) \tag{4.51}$$

$$+\frac{i2\zeta^2 B^2}{4\omega^2}((2\delta t\omega + \sin(2t_{k-1}\omega) - 4\cos(t_{k-1}\omega)\sin((t_{k-1}+\delta t)\omega) + \sin(2(t_{k-1}+\delta t)\omega)) \tag{4.52}$$

$$\times(\exp\left\{(i\int_{t_{k-1}}^{t_k}\hbar G(s)ds)\right\}^\dagger Z[Z, \hbar G(t_k)]\exp\left\{(-i\int_{t_{k-1}}^{t_k}\hbar G(s)ds)\right\}) \tag{4.53}$$

$$+(\exp\left\{(i\int_{t_{k-1}}^{t_k}\hbar G(s)ds)\right\}^\dagger [Z, \hbar G(t_k)]Z\exp\left\{(-i\int_{t_{k-1}}^{t_k}\hbar G(s)ds)\right\})))), \tag{4.54}$$

where we have only kept the lowest order correction terms. By inspection, we see that there is no way for a T dependence to enter for higher order terms, and moreover we see that for ω large enough ($\Omega >> \mu B/\hbar$) the terms are both integrable over ω and arbitrarily small. We are interested in N times the B-derivative of this error, where the extra factor of N is because the total error accumulates at worst as N times the step-wise error. But now we are done, since this shows that for all $\epsilon > 0$ there is an $\Omega(||G||, B, \delta t)$ such that in total we have a bound of

$$\int d\omega J_c(\omega) \leqslant \frac{2\pi\zeta^2 T^2}{\delta t} + (c(\hbar||G||, \mu B, \delta t) + \epsilon(\hbar||G||, \mu B, \delta t))T^2 = \frac{2\pi\zeta^2 T^2}{\delta t} + \alpha(\hbar||G||, \mu B, \delta t)T^2, \tag{4.55}$$

where we've denoted two parts in the coefficient of the second term - one coming from the lowest order contribution to the right tail of the QFI, and one coming from the inaccuracy in this approximation. So in total we see that $\int d\omega K(T) \in \mathcal{O}(T^2)$. We see furthermore that this argument generalizes to arbitrary entangled probes. The $B-$derivative of the time evolution is maximized on entangled inputs and contributes a factor of n, the number of qubits. To see this, note Eq. (4.34) for n qubits is instead

$$\partial_B U'(T) = -i\zeta \sum_{j=0}^{N-1}\sum_{k=0}^{k=n-1} \Theta_j P_0 U_0 ... P_j Z^{(k)} U_j ... P_{N-1} U_{N-1}, \tag{4.56}$$

where the new index (k) denotes the Z operation on the k^{th} qubit, and the other operators are generalized to their natural multi-qubit counterparts. Similarly, the Trotter error has a factor of n

because it is proportional to the sum of $Z^{(k)}$, and so the $B-$derivative of the Trotter error also is proportional to n. Because both terms in the QFI are quadratic in $|\phi\rangle$ we find a factor of n^2 as in the GHZ state example, giving a total bound of $\int d\omega K(T) \in \mathcal{O}(n^2T^2)$ for n qubits.

Chapter 5

Noisy Reservoir Computation [144]

5.1 Abstract

In this note we extend the definition of the Information Processing Capacity (IPC) by Dambre et al [145] to include the effects of stochastic reservoir dynamics. We quantify the degradation of the IPC in the presence of this noise.

[1] Dambre et al. **Scientific Reports** 2, 514, (2012).

5.2 Reservoir Computation

In the traditional framework of a classical reservoir computer, [146] one considers a dynamical system observed at discrete time-steps $t = 0, 1, 2, \ldots$, with internal states $\boldsymbol{s}(t) \in \mathbb{R}^n$, driven by inputs $\boldsymbol{U}(t) \in \mathbb{R}^d$. The reservoir dynamics are then encompassed by the dynamical update rule

$$\boldsymbol{s}(t) = \boldsymbol{F}\big(\boldsymbol{s}(t-1), \boldsymbol{U}(t)\big), \tag{5.1}$$

from which m outputs $x_k(t) \in \mathbb{R}$ with $k \leqslant m \leqslant n$, can be constructed from the reservoir as

$$x_k(t) = \boldsymbol{G}\big(\boldsymbol{s}(t), \boldsymbol{U}(t)\big). \tag{5.2}$$

These outputs can then be utilized to approximate a target function $y(t) \in \mathbb{R}$, via a supervised learning scheme.

Specifically, we first define the concatenated h-step sequence of recent inputs $\boldsymbol{U}^{-h}(t) = [\boldsymbol{U}(t-h+1), \boldsymbol{U}(t-h+2), \ldots, \boldsymbol{U}(t)]$. While we may use the reservoir to learn a function of time, the

reservoir outputs themselves can be approximated by maps $x_k^h : \boldsymbol{U}^{-h}(t) \mapsto \mathbb{R}$. In particular, this is because we require that the reservoir satisfies the fading memory property [145]. A dynamical system has fading memory if, for all $\epsilon > 0$, there exists a positive integer $h_0 \in \mathbb{N}$, such that for all $h > h_0$, for all initial conditions, and for all sufficiently long initialization times $T' > h$, the $x_k(t)$ at any time $t \geqslant 0$ are well-approximated by functions x_k^h:

$$\mathbb{E}(x_k(t) - x_k^h[\boldsymbol{U}^{-h}(t)])^2 < \epsilon \tag{5.3}$$

where the expectation is taken over the $t + T'$ previous inputs.

The signals $x_k(t)$ can used to construct estimators (denoted with hats) of $y(t)$

$$\hat{y}_{\boldsymbol{w}}(t) = \boldsymbol{w}^T \boldsymbol{X}(t), \tag{5.4}$$

where $\boldsymbol{w}$ is a vector of weights, and $\boldsymbol{X}(t)$ is the vector whose entries are $x_k(t)$.

Characterizing the learning capabilities of a dynamical system requires analyzing it in a learning-task independent manner. We do so by utilizing a quantity known as the information processing capacity (IPC) [145]. We first define the *capacity* to reconstruct some arbitrary signal y as

$$C_T[y] = 1 - \min_{\boldsymbol{w}} \frac{\langle (\hat{y}_{\boldsymbol{w}} - y)^2 \rangle_T}{\langle y^2 \rangle_T}, \tag{5.5}$$

where $\langle \chi \rangle_T = \frac{1}{T} \sum_{t=1}^{T} \chi(t)$ denotes a time average. Then for a complete and countably infinite set of basis functions $\{y_1, y_2, \ldots\}$ for the Hilbert space of functions with fading memory [145], the IPC of a dynamical system is given as

$$\text{IPC} = \lim_{D \to \infty} \lim_{T \to \infty} \sum_{\ell}^{D} C_T[y_\ell] \leqslant n. \tag{5.6}$$

Intuitively, the IPC can be thought of as a normalized measure of the size of the subspace of functions learnable by the dynamical system. Here the basis set should be ordered in some sense by the complexity (e.g., the polynomial degree and the memory time h) of the basis functions. The numerical estimate of the IPC requires defining a probability measure over all possible sequences of inputs (thus promoting $\boldsymbol{U}(t)$ to a random process with measure ξ) and truncating the basis set at some sufficiently high, but finite cutoff.

5.3 Noisy Reservoir Computation

In this section we extend the ideas from Sec. 5.2 and the formula for the IPC in [145] to describe the performance of a reservoir in the presence of noise. The internal state and readouts $(\boldsymbol{s}(t), \boldsymbol{X}(t))$, thus evolve stochastically, even for a fixed input sequence. Formally, what this means is that in the following, $\boldsymbol{X}(t)$ and $\boldsymbol{U}(t)$ are random variables with probability measures η and ξ. Furthermore, because the state of the reservoir is a function of the input to the reservoir, we have that in general $\boldsymbol{X}(t)$ is conditioned on all past $\boldsymbol{U}(t)$.

In what follows, we use the overline $\overline{\boldsymbol{A}}$ to denote the expectation value of some variable $\boldsymbol{A}$ over the input signal distribution η. We use $\langle \boldsymbol{A} \rangle$ to denote expectation values of $\boldsymbol{A}$ over the reservoir noise distribution ξ. We have otherwise suppressed the notational dependence on $\boldsymbol{U}(t)$ in this analysis.

We find that the IPC is

$$\mathrm{IPC} \leq \mathrm{Tr}\Big((\mathbf{I} + \tilde{\mathbf{Q}}_\xi)^{-1} \Big) \tag{5.7}$$

$$= \sum_{k=1}^{n} \frac{1}{1+\tilde{\sigma}_k^2} \leq n, \tag{5.8}$$

where $\mathbf{I}$ is the identity matrix and $\tilde{\mathbf{Q}}_\xi$ measures the normalized reservoir output noise covariance. This is by design a positive (semi-)definite matrix whose eigenvalues we denote as $\{\tilde{\sigma}_k^2 : k = 1, \ldots, n\}$ and which are the generalized reservoir noise variances. Thus IPC $= n$ only occurs when $\tilde{\mathbf{Q}}_\xi = 0$ is the trivial matrix, demonstrating that noise strictly decreases the achievable IPC of a reservoir computer of given (output) size n. In the following derivation of the inequality (5.7) we assume that the average reservoir outputs $\langle \boldsymbol{X} \rangle$ (i.e, averaged over the reservoir noise) satisfy

$$\mathbf{Q}_\eta \coloneqq \overline{\langle \boldsymbol{X} \rangle \langle \boldsymbol{X} \rangle^T} > 0. \tag{5.9}$$

We will designate this assumption as the full signal rank condition. We note, however, that this only simplifies the analysis, and that the result is correct even when the condition does not hold. We revisit this case at the very end of our discussion.

5.3.1 Proof

We limit our proof to the finite case of a truncated basis set (of size D). A common quantity to consider in any regression analysis is the expected squared error, J, between the predictions made by the model and the data. Thus we start by defining the expected squared reconstruction error

$$\begin{aligned} J &= \overline{\left\langle \sum_{\ell=1}^{D}\sum_{k=1}^{n}(y_\ell(t) - x_k(t)w_{k\ell})^2 \right\rangle} \\ &= \overline{\left\langle \sum_{\ell=1}^{D}\sum_{k=1}^{n}(y_\ell(t) - x_k(t)w_{k\ell})(y_\ell(t) - w_{k\ell}x_k(t)) \right\rangle} \\ &= \overline{\langle (\boldsymbol{Y}^T - \boldsymbol{X}^T\mathbf{W})(\boldsymbol{Y} - \mathbf{W}^T\boldsymbol{X})\rangle} \\ &= \mathrm{Tr}\,\overline{\langle (\boldsymbol{Y} - \mathbf{W}^T\boldsymbol{X})(\boldsymbol{Y}^T - \boldsymbol{X}^T\mathbf{W})\rangle}, \end{aligned} \tag{5.10}$$

where here vectors are column vectors, and transposes are row vectors. In particular, $\boldsymbol{Y} \in \mathbb{R}^{D\times 1}$, $\mathbf{W} \in \mathbb{R}^{m\times D}$ are the vector of estimated signals and weight matrix, respectively, and $\boldsymbol{X} \in \mathbb{R}^{m\times 1}$ is defined as before. We have chosen the convention for $\mathbf{W}^T$ so that $\mathbf{W}^T\boldsymbol{X}$ is an estimator of the target functions.

In this notation we have suppressed and will continue to suppress the dependence on time, so that $\boldsymbol{Y} \coloneqq \boldsymbol{Y}(\boldsymbol{U}^{-h}(t))$ are target functions of the finite-sequence inputs $\boldsymbol{U}^{-h}(t)$, and will be approximated in this scheme by linear functions of the measured outputs $\boldsymbol{X} \coloneqq \boldsymbol{X}(t)$. That is, the goal is to find a weight matrix $\mathbf{W}$ such that $\mathbf{W}^T\boldsymbol{X} \approx \boldsymbol{Y}$. For computing the IPC, we will take $\boldsymbol{Y}(\boldsymbol{U}^{-h}(t))$ to constitute a complete basis of functions on the finite input sequences, as in Eq. (5.6). Differentiating Eq. (5.10) with respect to $\mathbf{W}$ then gives

$$\begin{aligned} \nabla_{\mathbf{W}} J &= \nabla_{\mathbf{W}}\overline{\langle \boldsymbol{X}^T\mathbf{W}\mathbf{W}^T\boldsymbol{X} - \boldsymbol{Y}^T\mathbf{W}^T\boldsymbol{X} - \boldsymbol{X}^T\mathbf{W}\boldsymbol{Y}\rangle} \\ &= \nabla_{\mathbf{W}}\,\mathrm{Tr}\Big(\overline{\langle \boldsymbol{X}^T\mathbf{W}\mathbf{W}^T\boldsymbol{X} - 2\boldsymbol{X}\boldsymbol{Y}^T\mathbf{W}^T\rangle}\Big) \\ &= 2\overline{\langle \boldsymbol{X}\boldsymbol{X}^T\rangle}\mathbf{W} - 2\overline{\langle \boldsymbol{X}\boldsymbol{Y}^T\rangle}, \end{aligned} \tag{5.11}$$

allowing us to solve for $\mathbf{W}$ via the first order optimality condition $\nabla_{\mathbf{W}} J = 0$ as

$$\mathbf{W}_* = \langle \overline{\boldsymbol{X}\boldsymbol{X}^T} \rangle^+ \langle \overline{\boldsymbol{X}\boldsymbol{Y}^T} \rangle \tag{5.12}$$

where A^+ denotes the Moore-Penrose pseudo-inverse of A. Under the full signal rank condition (5.9) this is a proper inverse.

Defining the reservoir output noise as

$$\Delta \boldsymbol{X} \coloneqq \boldsymbol{X} - \langle \boldsymbol{X} \rangle, \tag{5.13}$$

allows decomposing the output into its deterministic and its noise part $\boldsymbol{X} = \langle \boldsymbol{X} \rangle + \Delta \boldsymbol{X}$. With this, we find

$$\begin{aligned} \overline{\langle \boldsymbol{X}\boldsymbol{X}^T \rangle} &= \underbrace{\overline{\langle \boldsymbol{X} \rangle \langle \boldsymbol{X} \rangle^T}}_{\mathbf{Q}_\eta} + \underbrace{\overline{\langle \Delta \boldsymbol{X} \Delta \boldsymbol{X}^T \rangle}}_{\mathbf{Q}_\xi} \\ &\quad + \overline{\langle \boldsymbol{X} \rangle \underbrace{\langle \Delta \boldsymbol{X}^T \rangle}_{0}} + \overline{\underbrace{\langle \Delta \boldsymbol{X} \rangle}_{0} \langle \boldsymbol{X} \rangle^T} \end{aligned} \tag{5.14}$$

$$= \mathbf{Q}_\eta + \mathbf{Q}_\xi, \tag{5.15}$$

i.e., we can decompose this second moment matrix into the second moment matrix of the deterministic output signal $\mathbf{Q}_\eta$ and the noise covariance $\mathbf{Q}_\xi$.

Under the full signal rank condition (5.9), we can perform a spectral decomposition $\overline{\langle \boldsymbol{X} \rangle \langle \boldsymbol{X} \rangle^T} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^T$ with positive definite, diagonal matrix $\boldsymbol{D}$ and an orthogonal matrix $\boldsymbol{V}$ and factor the second moment matrix as

$$\overline{\langle \boldsymbol{X}\boldsymbol{X}^T \rangle} = \boldsymbol{V}\boldsymbol{D}^{\frac{1}{2}} \left(\mathbf{I} + \tilde{\mathbf{Q}}_\xi \right) \boldsymbol{D}^{\frac{1}{2}} \boldsymbol{V}^T \tag{5.16}$$

where we have defined

$$\tilde{\mathbf{Q}}_\xi \coloneqq \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{V}^T \mathbf{Q}_\xi \boldsymbol{V} \boldsymbol{D}^{-\frac{1}{2}} = \overline{\langle \Delta \tilde{\boldsymbol{X}} \Delta \tilde{\boldsymbol{X}}^T \rangle}. \tag{5.17}$$

$\tilde{\boldsymbol{X}}$ in turn is related to the above diagonalization of the deterministic second moment matrix $\overline{\langle \boldsymbol{X} \rangle \langle \boldsymbol{X} \rangle^T}$, corresponding to a basis change for the outputs

$$\tilde{\boldsymbol{X}} = \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{V}^T \boldsymbol{X} \Leftrightarrow \boldsymbol{X} = \boldsymbol{V}\boldsymbol{D}^{\frac{1}{2}} \tilde{\boldsymbol{X}}. \tag{5.18}$$

We can use this same transform to re-express the overlap matrix

$$\begin{aligned}\mathbf{C}_{XY} &= \overline{\langle \boldsymbol{X}\boldsymbol{Y}^T \rangle} = \overline{\langle \boldsymbol{X} \rangle \boldsymbol{Y}^T} \\ &= \boldsymbol{V}\boldsymbol{D}^{\frac{1}{2}} \overline{\langle \tilde{\boldsymbol{X}} \rangle \boldsymbol{Y}^T} = \boldsymbol{V}\boldsymbol{D}^{\frac{1}{2}} \mathbf{C}_{\tilde{X}Y}.\end{aligned} \tag{5.19}$$

Substituting $\mathbf{W}_*$ back into J and transforming to the normalized outputs gives

$$\begin{aligned}J(\mathbf{W}_*) &= \operatorname{Tr} \overline{\langle \boldsymbol{Y}\boldsymbol{Y}^T - 2\boldsymbol{Y}\boldsymbol{X}^T\mathbf{W}_* + \mathbf{W}_*^T\boldsymbol{X}\boldsymbol{X}^T\mathbf{W}_* \rangle} \\ &= \operatorname{Tr}\left(\overline{\langle \boldsymbol{Y}\boldsymbol{Y}^T \rangle} - \mathbf{C}_{\tilde{X}Y}^T \left(\mathbf{I} + \tilde{\mathbf{Q}}_\xi \right)^{-1} \mathbf{C}_{\tilde{X}Y} \right) \\ &= D - \operatorname{Tr}\left(\left(\mathbf{I} + \tilde{\mathbf{Q}}_\xi \right)^{-1} \underbrace{\mathbf{C}_{\tilde{X}Y}\mathbf{C}_{\tilde{X}Y}^T}_{\leq \mathbf{I}} \right) \geq D - n\end{aligned} \tag{5.20}$$

We note that the trace in the last line is over the reservoir output indices $k = 1, 2, \ldots, n$ whereas the previous lines have traces over the basis set $l = 1, 2, \ldots, D$. Finally, our claim follows from the Cauchy-Schwarz inequality and the fact that $\operatorname{Tr} AB \geq 0$ for positive semi-definite $A, B \geq 0$. Also note that under the full signal rank condition we have $\lim_{D,T\to\infty} \mathbf{C}_{\tilde{X}Y}\mathbf{C}_{\tilde{X}Y}^T \to \mathbf{I}$, where in typical physicist fashion we ignore the subtleties around how to carefully take the infinite limits.

For a simple example, consider a deterministic reservoir with orthonormal outputs, i.e. $\mathbf{Q}_\eta = \mathbf{I}$. By introducing Gaussian noise on the outputs with covariance matrix $\boldsymbol{\Sigma}$, the IPC is bounded exactly as in Eq. (5.7), giving

$$\mathrm{IPC} \leqslant \sum_{k=1}^{n} \frac{1}{1+\sigma_k^2}, \tag{5.21}$$

where σ_k^2 are the eigenvalues of $\boldsymbol{\Sigma}$, since $\mathbf{Q}_\eta = \mathbf{I}$ implies $\tilde{\mathbf{Q}}_\xi = \mathbf{Q}_\xi = \boldsymbol{\Sigma}$. While we have specified that the noise is on the outputs of the reservoir, our derivation shows that the result is the same independent of whether the noise is added to the outputs or is internal to the reservoir.

Finally, note that this proof can easily be extended to the case without the full signal rank condition (5.9). E.g., if the second signal moment has only rank $\tilde{n} < n$, then it requires replacing $\boldsymbol{D}^{-\frac{1}{2}}$ with its pseudo-inverse version $(\boldsymbol{D}^{\frac{1}{2}})^+$ and some special care needs to be taken when canceling $(\boldsymbol{D}^{\frac{1}{2}})^+\boldsymbol{D}^{\frac{1}{2}} = \mathbf{I}_{\tilde{\mathbf{n}}}$ where $\mathbf{I}_{\tilde{\mathbf{n}}}$ is a diagonal matrix with $\tilde{n}$ ones on its diagonal (and otherwise only zeros). Also, note that in this case the overlap matrix $\mathbf{C}_{XY}$ has rank $\leq \min(\tilde{n}, D)$, which implies that for

our optimal weights $\mathbf{W}_*$ we are only sensitive to those output noise contributions that are in the linear span of the expected output signals. In particular, the noise covariance then generalizes to

$$\tilde{\mathbf{Q}}_\xi \coloneqq (\boldsymbol{D}^{\frac{1}{2}})^+ \boldsymbol{V}^T \mathbf{Q}_\xi \boldsymbol{V} (\boldsymbol{D}^{\frac{1}{2}})^+ = \overline{\langle \Delta\tilde{\boldsymbol{X}} \Delta\tilde{\boldsymbol{X}}^T \rangle} \tag{5.22}$$

which may have lower rank than $\mathbf{Q}_\xi$.

Chapter 6

A Theory of Direct Randomized Benchmarking [147]

6.1 Abstract

Randomized benchmarking (RB) protocols are widely used to measure an average error rate for a set of quantum logic gates. However, the standard version of RB is limited because it only benchmarks a processor's native gates indirectly, by using them in composite n-qubit Clifford gates. Standard RB's reliance on n-qubit Clifford gates restricts it to the few-qubit regime, because the fidelity of a typical composite n-qubit Clifford gate decreases rapidly with increasing n. Furthermore, although standard RB is often used to infer the error rate of native gates, by rescaling standard RB's error per Clifford to an error per native gate, this is an unreliable extrapolation. Direct RB is a method that addresses these limitations of standard RB, by directly benchmarking a customizable gate set, such as a processor's native gates. Here we provide a detailed introduction to direct RB, we discuss how to design direct RB experiments, and we present two complementary theories for direct RB. The first of these theories uses the concept of error propagation or scrambling in random circuits to show that direct RB is reliable for gates that experience stochastic Pauli errors. We prove that the direct RB decay is a single exponential, and that the decay rate is equal to the average infidelity of the benchmarked gates, under broad circumstances. This theory shows that group twirling is not required for reliable RB. Our second theory proves that direct RB is reliable for gates that experience general gate-dependent Markovian errors, using similar techniques to contemporary theories for standard RB. Our two theories for direct RB have complementary regimes of applicability, and they provide complementary perspectives on why direct RB works.

Together these theories provide comprehensive guarantees on the reliability of direct RB.

6.2 Introduction

Reliable, efficient, and flexible methods for benchmarking quantum computers are becoming increasingly important as 5-50+ qubit processors become commonplace. Isolated qubits or coupled pairs can be studied in detail with tomographic methods [148–150], but tomography of general n-qubit processes requires resources that scale exponentially with n. Randomized benchmarking (RB) methods [151–165] were introduced partly to avoid the scaling problems that afflict tomography. RB avoids these scaling problems because, in the most commonly used RB protocols [154, 155, 160], both the necessary number of experiments [166] and the complexity of the data analysis [155] are independent of n. RB methods efficiently estimate a single figure of merit—the average infidelity of a gate set—by (i) running random circuits of varied depths (d) that should implement an identity operation, (ii) observing the probability that the input is successfully returned (S_d) versus d, and then (iii) fitting this data to an exponential. The decay rate of this exponential (r) is an estimate of the gates' average infidelity.

However, many RB methods have a different scaling problem, introduced by a **gate compilation** step. Most RB protocols benchmark an n-qubit gate set that forms a group [154–164], so that they can use group twirls as their theoretical foundations. But a quantum processor's native operations do not normally form a group. Instead, subsets of a processor's native gates can be used to **generate** a variety of different groups (e.g., the unitary group, the Clifford group, the Pauli group, etc). This is a problem for implementing group-based RB methods because, for many n-qubit groups, the number of basic gates required to implement a typical group element is very large, meaning that many group elements can only be implemented with low fidelity.

The standard RB protocol—which we will call "Clifford group RB"—benchmarks the n-qubit Clifford group [154, 155], estimating an error rate (r) that corresponds closely to the average infidelity of a gate from the n-qubit Clifford group [168, 169]. This protocol runs random sequences of $d+2$ composite gates from that group (see the upper circuit in Fig. 6.1), with a variable benchmark

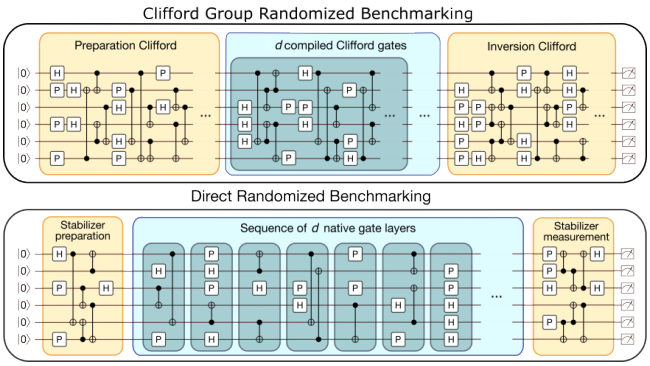


Figure 6.1: An illustration (adapted from [167]) comparing the varied-length random circuits used in the standard RB protocol of Magesan **et al.** [154, 155] ("Clifford group RB") and the streamlined direct RB protocol that was introduced in [167] and that we investigate further here. Direct RB can be implemented on almost any n-qubit gate set (a set of circuit "layers" or "cycles") that generates a unitary 2-design. The case shown here is when that gate set is a set of Clifford gates (CNOT, Hadamard, and the phase gate) that generate the n-qubit Clifford group. The circuits of direct RB can be shallower than those of Clifford group RB (e.g., consider the case of $d = 0$ in each circuit), allowing for RB on more qubits.

depth (d, with $d \geqslant 0$). The first $d+1$ gates are sampled uniformly from the Clifford group, and the last gate is selected to invert the proceeding $d+1$ gates. For a single qubit there are only 24 Clifford gates, so each Clifford gate can typically be compiled into a short sequence of native gates. But the Clifford group grows quickly with the number of qubits n: there are $2^{O(n^2)}$ n-qubit Clifford gates [170, 171], and a single typical n-qubit Clifford gate can be compiled into no fewer than $O(\frac{n^2}{\log n})$ one- and two-qubit gates [172–174]. This means that the fidelity of a random Clifford gate degrades rapidly with n for a fixed quality of one- and two-qubit gates, i.e., $r \to 1$ quickly with increasing n. This makes it infeasible to reliably estimate Clifford group RB's error rate r on many qubits—even with very high fidelity one- and two-qubit gates—because as n increases even the shortest Clifford group RB circuits ($d = 0$, corresponding to a random n-qubit Clifford gate and its inverse) typically have such small success probabilities that they cannot be distinguished from $\frac{1}{2^n}$ with a reasonable number of circuit repetitions (see Fig. 6.2). Indeed, one- and two-qubit Clifford group RB has been widely implemented [175–188] but we are aware of only two publications presenting Clifford group RB experiments on more than two qubits [189, 190], and none presenting Clifford group RB experiments on six or more qubits.

An alternative to Clifford group RB is to benchmark a gate set that is a smaller group [156–159, 161, 163], enabling group elements to be implemented with fewer native gates. The gate compilation overhead can then be reduced from Clifford group RB, but often at a cost—the data analysis and the experiments become more complex. The simple behavior of Clifford group RB, i.e., the exponential decay of its success probability data, is guaranteed by a particular property of the Clifford group: conjugation of an error channel by a uniformly random Clifford group element "twirls" that channel into an n-qubit depolarizing channel. This follows because the Clifford group is a unitary 2-design [191–193] (or, equivalently, because the superoperator representation of Clifford gates consists of the direct sum of two irreducible representations of the Clifford group). The less like a unitary 2-design a group is (i.e., the more irreducible representations of the group its superoperator representation decomposes into) the weaker the effect of twirling over it is, and so the more complex the RB decay curve can be [156–159, 161, 163] (see [163] for a theory of RB over general groups). For general groups, the RB decay curve is a sum of many exponentials, rather than the single exponential of Clifford group RB.

Another consequence of the gate compilation required in group-based RB protocols is that they measure an error **per compiled group element**. There are infinitely many ways to compile each group element into native gates, and the observed error rate depends strongly on which group compilations is used. This is desirable if we aim to estimate the fidelity with which gates in that group can be implemented. But normally group-based RB error rates are used as a proxy for native gate performance [184–186, 189]. In particular, it is common to estimate a native gate error rate from a Clifford group RB error rate, e.g., by dividing Clifford group RB's r by the average circuit size of a compiled gate [184–186] or by the average number of two-qubit gates in a compiled gate [189] (possibly after correcting for the estimated contribution of one-qubit gates). However, these rescaling methods have little theoretical justification [167, 194]: in general, the resultant error rate is not a reliable estimate of the average infidelity of the native gates.

One solution to the limitations of Clifford group RB is **direct randomized benchmarking** [167]. Direct RB is a technique that is designed to retain the conceptual and experimental simplicity

and efficiency of Clifford group RB while enabling direct benchmarking of a processor's native gates. Like all RB protocols, direct RB utilizes random circuits of variable length. Furthermore, like Clifford group RB, direct RB's data analysis simply consists of (1) fitting success probability data to a single exponential decay, and (2) rescaling the fit decay rate to obtain an estimate of the benchmarked gates' average infidelity. But, unlike Clifford group RB, the varied-length portion of direct RB circuits can consist of randomly sampled layers of a processor's native gates, rather than compiled group operations (see the lower circuit in Fig. 6.1). Direct RB therefore enables estimating the average infidelity of native gate layers. The only restriction on the gate set that can be benchmarked with direct RB is that it must **generate** a group that is a unitary 2-design (e.g., the n-qubit Clifford group).

Direct RB addresses the main limitations of Clifford group RB, but it has been lacking a theory that proves that it is reliable. Those existing theories for RB that are applicable to direct RB [167, 195, 196] provide few guarantees on direct RB's behavior. In this paper we provide a detailed introduction to direct RB and two complementary theories of direct RB. These theories show that direct RB is reliable—the direct RB success probability follows an exponential decay and the direct RB error rate is the infidelity of the benchmarked gates—and explain why direct RB works.

This paper is structured as follows. In Section 6.3 we introduce our notation and review the necessary background material. In Section 6.4 we provide a detailed definition and introduction to the direct RB protocol, and we compare it to other RB protocols for benchmarking native gate sets [153, 196, 197]. The definition for direct RB presented here generalizes that in [167]. This section includes examples that illustrate how direct RB can be used to benchmark a variety of physically relevant gate sets—including universal gate sets on few qubits. In Section 6.5 we present a theory for direct RB on gates that experience stochastic Pauli errors, which is the relevant error model for gates that have undergone Pauli frame randomization [198, 199] or randomized compilation [200, 201]. This theory provides a physically intuitive underpinning for direct RB, by using the intuitive concepts of error propagation and scrambling in random circuits. Interestingly, this theory

shows that group twirling is not necessary for reliable RB—direct RB can be reliable even when the benchmarked gate set is very far from approximating a unitary 2-design. This enables us to show that direct RB is reliable even in the $n \gg 1$ regime, where short sequences of layers of native gates cannot approximate a unitary 2-design.

One of the contributions of this paper is a theory for direct RB with general gate-dependent Markovian errors that is similar in nature to the group twirling (and Fourier analysis) theories for group-based RB methods. This theory uses the concept of a **sequence-asymptotic unitary 2-design**, which we introduce in Section 6.6. A sequence-asymptotic unitary 2-design is any gate set for which length d random sequences of gates from that set create a distribution over unitaries that converges to a unitary 2-design as $d \to \infty$. Sequence-asymptotic unitary 2-design are closely related to various existing notions of approximate or asymptotic unitary designs and the general theory of scrambling in random circuits [191, 192, 202–208]. However, to our knowledge, the particular concept and technical results that we require for our theory of direct RB do not appear in the literature.

In Section 6.7 we present our theory for direct RB under general gate-dependent Markovian errors, which is based on twirling superchannels (or Fourier transforms on groups). First, we show that the depth-d direct RB success probability can be computed exactly by taking the d^{th} power of a superchannel (R) constructed from the permutation matrix representation of the group generated by the benchmarked gate set. Then we show how the direct RB decay can be (approximately) rewritten in terms of a smaller superchannel (L) that is constructed from the superoperator representation of the generated group. These results correspond to equivalent theories for Clifford group RB presented in [168, 169, 209, 210]. We then use the spectral decomposition of these superchannels, together with the theory of sequence-asymptotic unitary 2-designs, to show that the direct RB decay is approximately a singe exponential. Finally, we use the L superchannel to show that the direct RB error rate is equal to a gauge-invariant version of the mean fidelity of the benchmarked gates, building on Wallman's derivation of an equivalent result for Clifford group RB [169]. Our two theories of direct RB are complementary, providing different perspectives as well as different regimes

of applicability. Our superchannel-based theory (Section 6.7) is more mathematically rigorous than our scrambling-based theory (Section 6.5), and it applies to a more general class of errors. However, our superchannel-based theory relies on an approximation whose size increases with n, whereas our scrambling-based theory shows that direct RB is reliable even when $n \gg 1$. Therefore, together these theories provide comprehensive evidence for the reliability of direct RB.

6.3 Definitions

In this section we introduce our notation and review the necessary background material.

6.3.1 Gates, gate sets, and circuits

An n-qubit gate G is a physical operation associated with an instruction to ideally implement a particular unitary evolution on n qubits (we use the term "gate" for consistency with the RB literature—an n-qubit gate is also often referred to as a "layer" or a "cycle"). We denote the unitary corresponding to G by $U(G) \in \mathrm{SU}(2^n)$. It will also often be convenient to use the superoperator representation of a unitary, so we define $\mathcal{G}(G)$ to be the linear map

$$\mathcal{G}(G)[\rho] := U(G)\rho U(G)^\dagger, \tag{6.1}$$

where ρ is an n-qubit density operator. We consider a gate G to be entirely defined by the unitary $U(G)$. RB methods are agnostic about how an n-qubit gate is implemented, except that an attempt must be made to faithfully implement the unitary it defines. We will use $\mathbb{G}$ to denote a set of n-qubit gates, which need not be a finite set, and, slightly abusing notation, we will also use $\mathbb{G}$ to denote the corresponding sets of unitaries (i.e., $\{U(G) \mid G \in \mathbb{G}\}$ and superoperators (i.e., $\{\mathcal{G}(G) \mid G \in \mathbb{G}\}$) with the meaning implied by the context.

A quantum circuit C over a n-qubit gate set $\mathbb{G}$ is a sequence of $d \geqslant 0$ gates from $\mathbb{G}$. We will write this as

$$C = G_d G_{d-1} \cdots G_2 G_1, \tag{6.2}$$

where each $G_i \in \mathbb{G}$, and we use a convention where the circuit is read from right to left. The circuit

C is an instruction to apply its constituent gates, G_1, G_2, ..., in sequence, and it implements the unitary

$$U(C) = U(G_d)U(G_{d-1})\cdots U(G_2)U(G_1). \tag{6.3}$$

In direct RB, and most other RB methods, multiple gates in a circuit must **not** be combined or compiled together by implementing a physical operation that enacts their composite unitary (there are "barriers" between gates).

6.3.2 Fidelity

In our theory for direct RB we will show how the direct RB error rate is related to the fidelity of the benchmarked gates. There are two commonly used version of a gate's fidelity: average gate fidelity and entanglement fidelity, defined below. Throughout this work, we will use the Markovian model for errors, whereby the imperfect implementation of a gate G is represented by a complete positive and trace preserving (CPTP) superoperator $\tilde{\mathcal{G}}(G)$ [so, for low-error gates $\tilde{\mathcal{G}}(G) \approx \mathcal{G}(G)$]. The **average gate fidelity** of $\tilde{\mathcal{G}}(G)$ to $\mathcal{G}(G)$ is defined by

$$F_{\mathrm{A}}(\tilde{\mathcal{G}}, \mathcal{G}) := \int d\psi \ \mathrm{Tr}\left\{\tilde{\mathcal{G}}[|\psi\rangle\langle\psi|]\, \mathcal{G}[|\psi\rangle\langle\psi|]\right\}, \tag{6.4}$$

where $d\psi$ is the normalized Haar measure [211], and where here we have dropped the dependence of $\mathcal{G}$ and $\tilde{\mathcal{G}}$ on G for brevity (we also do this elsewhere when convenient). The **entanglement fidelity** is defined by

$$F_{\mathrm{E}}(\tilde{\mathcal{G}}, \mathcal{G}) := \langle\varphi|(\mathbb{I}\otimes\Lambda)[|\varphi\rangle\langle\varphi|]|\varphi\rangle, \tag{6.5}$$

where φ is any maximally entangled state [211] and

$$\Lambda(G) = \tilde{\mathcal{G}}(G)\mathcal{G}(G)^{-1}. \tag{6.6}$$

The corresponding infidelities, **average gate infidelity** (ϵ_{A}) and **entanglement infidelity** (ϵ_{E}), are defined by

$$\epsilon_{\mathrm{A/E}}(\tilde{\mathcal{G}}, \mathcal{G}) = 1 - F_{\mathrm{A/E}}(\tilde{\mathcal{G}}, \mathcal{G}). \tag{6.7}$$

These infidelities are related by [211]

$$\epsilon_{\mathrm{E}} = \frac{2^n + 1}{2^n}\epsilon_{\mathrm{A}}. \tag{6.8}$$

The average gate [in]fidelity is more commonly used in the RB literature, but in this work we use the entanglement [in]fidelity. Typically we drop the subscript, letting $\epsilon \equiv \epsilon_{\mathrm{E}}$. Note that a gate's [in]fidelity is not a "gauge-invariant" property of a gate set [168], meaning that it is not technically measurable: we discuss this subtle point when presenting our theory for direct RB with general gate-dependent Markovian errors (Section 6.7).

Randomized benchmarking methods are typically designed to measure the mean infidelity of a set of gates. Direct RB is designed to measure

$$\epsilon_{\Omega} = \sum_{G\in\mathbb{G}} \Omega(G)\epsilon\left[\tilde{\mathcal{G}}(G), \mathcal{G}(G)\right], \tag{6.9}$$

where Ω is a probability distribution over the gate set $\mathbb{G}$. Note that, except where otherwise stated, $\mathbb{G}$ does not need to be a finite set, i.e., it can include gates with continuous parameters. The $\sum_{G\in\mathbb{G}}\Omega(G)$ notation is a shorthand for integrating and/or summing over $\mathbb{G}$ according to the measure $\Omega(G)$. When gates experience only stochastic Pauli errors, then ϵ_{Ω} is equal to the probability that a Pauli error occurs on a gate sampled from Ω.

6.4 Direct randomized benchmarking

In this section we first define the direct RB protocol and explain what its purpose is (Section 6.4.1). Direct RB has flexible, user-specified components, so we then provide guidance on how to choose these aspects of direct RB experiments (Sections 6.4.3-6.4.5). We then explain why the direct RB protocol is defined the way it is, and we discuss how it differs from Clifford group RB (Sections 6.4.6-6.4.8). Finally, we compare direct RB to other RB protocols for directly benchmarking native gate sets (Section 6.4.10).

6.4.1 The direct RB protocol

We now define the n-qubit direct RB protocol, and explain what it aims to measure. The definition given here is more general than that in [167]. For example, in [167], direct RB is defined for gate sets containing only Clifford gates. Our definition here permits non-Clifford gates. Note that our definition for direct RB applies to gates on a set of $n \geqslant 1$ qubits. Generalizing to $n \geqslant 1$ qudits of arbitrary dimension is conceptually simple, as is the case with Clifford group RB, but we do not pursue this here. In addition to parameters that set the number of samples (which exist in all RB protocols), direct RB on n qubits has two flexible, user-specified inputs:

(1) A set of n-qubit gates (a.k.a., layers or cycles) to benchmark ($\mathbb{G}$).

(2) A sampling distribution Ω over the gate set $\mathbb{G}$.

We denote direct RB of the gate set $\mathbb{G}$ with the sampling distribution Ω by $\mathrm{DRB}(\mathbb{G}, \Omega)$. We call $\mathbb{G}$ a "gate set" for consistency with common RB terminology, but note that $\mathbb{G}$ contains n-qubit gates—i.e., "circuit layers" or "cycles"—not one- and two-qubit gates. All n-qubit gate sets generate some n-qubit group, or a dense subset of some group [a]. This group plays an important role in the direct RB protocol, and we denote it by $\mathbb{C}$ (for brevity, for a gate set $\mathbb{G}$ that only generates a dense subset of a group $\mathbb{C}$ we also refer to $\mathbb{G}$ as generating $\mathbb{C}$). In most of our examples this group will be the n-qubit Clifford group, but this is not required (the conditions required of $\mathbb{C}$ are stated in Section 6.4.4, which include that it is a unitary 2-design).

Having introduced $\mathbb{G}$ and Ω, we are now ready to define the direct RB protocol. $\mathrm{DRB}(\mathbb{G}, \Omega)$ is a protocol for estimating ϵ_Ω [see Eq. (6.9)] and it consists of the following procedure:

(1) **Sample the circuits.** For each d in some set of user-specified **benchmark depths** all satisfying $d \geqslant 0$, and for $k = 1, 2, \ldots, K_d$ for some user-specified $K_d \geqslant 1$:

(1)1. Choose a target output n-bit string $s_{d,k}$ for this circuit. This can be set to the all-zeros

[a] Throughout the remainder of this paper we use the phrase "$\mathbb{G}$ generates the group $\mathbb{C}$" to include the case where $\mathbb{G}$ only generates a dense subset of $\mathbb{C}$.

bit string (as in the conventional implementation of Clifford group RB [154, 155]), or it can be chosen uniformly at random (our recommendation).

(1)2. Sample $F_{\rm sp}$ uniformly from $\mathbb{C}$ (the group generated by $\mathbb{G}$), and then find a circuit $C_{\rm sp}$ that creates the state

$$|\psi\rangle = U(F_{\rm sp})|0\rangle^{\otimes n}, \tag{6.10}$$

from $|0\rangle^{\otimes n}$ [$U(\cdot)$ maps a gate or circuit to the unitary it implements—see Section 6.3]. That is, find a circuit $C_{\rm sp}$ that satisfies

$$U(C_{\rm sp})|0\rangle^{\otimes n} = U(F_{\rm sp})|0\rangle^{\otimes n}, \tag{6.11}$$

meaning that $C_{\rm sp}$ is only required to faithfully implement $U(F_{\rm sp})$'s action on $|0\rangle^{\otimes n}$. The circuit $C_{\rm sp}$ can contain any gates, including gates that are not in $\mathbb{G}$, and it is ideally chosen to maximize the fidelity with which ψ is produced. The circuit $C_{\rm sp}$ is the first part of the sampled direct RB circuit, and we refer to it as the circuit's state preparation subcircuit.

(1)3. Independently sample d gates, G_1, G_2, ..., G_d, from Ω (the user-specified distribution over $\mathbb{G}$). The circuit $G_d \cdots G_2 G_1$ is the next part of the sampled direct RB circuit, and we refer to it as the circuit's "core".

(1)4. Find a circuit $C_{\rm mp}$ that, when applied after the two parts of the circuit sampled so far, maps the qubits to $|s_{d,k}\rangle$, and append it to the circuit sampled so far. That is, $C_{\rm mp}$ is a circuit that satisfies

$$U(C_{\rm mp}\, G_d \cdots G_2\, G_1\, C_{\rm sp})|0\rangle^{\otimes n} = |s_{d,k}\rangle. \tag{6.12}$$

The circuit $C_{\rm mp}$ is the final part of the sampled direct RB circuit. As with the circuit $C_{\rm sp}$, $C_{\rm mp}$ can contain any gates, and we refer to it as the circuit's measurement preparation subcircuit.

The sampling procedure of 1.1-1.4 creates the circuit

$$C_{d,k} = C_{\rm mp} G_d \cdots G_2 G_1 C_{\rm sp}, \tag{6.13}$$

that, if run on a perfect quantum computer, always outputs the bit string $s_{d,k}$. That is, $|\langle s_{d,k}|U(C_{d,k})|0\rangle^{\otimes n}|^2 = 1$.

(2) **Run the circuits.** Run each of the sampled circuits N times, for some user-specified $N \geqslant 1$. Estimate each circuit's success probability ($S_{d,k}$), as the frequency that the circuit $C_{d,k}$ returns its success bit string $s_{d,k}$ (we denote this estimate by $\hat{S}_{d,k}$). Note that the direct RB protocol does not specify the order that the circuits are run, but this is important if there could be significant drift in the system over the time period of the entire experiment [212–214]. We recommend "rastering" [215] if possible—meaning looping through all of the circuits N times, running each circuit once in each loop—as this faciliates detecting the presence of drift and a time-resolved RB analysis [215] (these analyses are not discussed further herein). If rastering is not possible then the order that circuits are run should be randomized, as this will typically reduce the impact of drift on the results.

(3) **Analyze the data.** Fit the estimated average probability of success, $\hat{S}_d = \sum_k \hat{S}_{d,k}/K_d$, versus benchmark depth (d) to the exponential decay function

$$S_d = A + Bp^d, \tag{6.14}$$

where A, B and p are fit parameters (this is the same as the fit function used in Clifford group RB). If the success bit strings were sampled uniformly then fix $A = \frac{1}{2^n}$. The estimate of the DRB$(\mathbb{G}, \Omega)$ error rate of the gates ($\hat{r}_\Omega$) is then defined to be

$$\hat{r}_\Omega = \frac{(4^n - 1)}{4^n}(1 - \hat{p}). \tag{6.15}$$

where $\hat{p}$ is the fit value of p.

Our definition for the direct RB procedure specifies fitting the data to the standard exponential decay function of Eq. (6.14), and so DRB$(\mathbb{G}, \Omega)$ is only well-behaved if the direct RB data has approximately this form. However, it is not clear **a priori** that the direct RB average success probability will decay exponentially. Moreover, even if the decay is an exponential it is not obvious

what the direct RB error rate (r_Ω) measures. Proving that $\hat{S}_d \approx A + Bp^d$ under broad conditions, providing an intuitive explanation for why $\hat{S}_d$ decays exponentially, and then explaining what r_Ω quantifies are three of the main aims of this paper. In particular, we will show that $r_\Omega \approx \epsilon_\Omega$.

6.4.2 A simple numerical demonstration of direct RB

We now demonstrate that direct RB works correctly—the decay is an exponential and $r_\Omega \approx \epsilon_\Omega$—in two simple scenarios. First, consider the case of gate-independent global depolarizing noise, and assume perfect state preparation and measurement sub-circuits. In this model, after each n-qubit gate the n qubits are mapped to the completely mixed state with some probability $1 - \lambda$. This is arguably the simplest possible error model, and we would expect the error rate measured by any well-formed RB protocol to be the infidelity of this global depolarizing channel, which is $\epsilon = (4^n - 1)(1 - \lambda)/4^n$. An explicit calculation confirms that this is the case here:

$$S_d = \frac{1}{2^n} + (1 - \frac{1}{2^n})\lambda^d, \tag{6.16}$$

which implies that $p = \lambda$ and $r_\Omega = (4^n - 1)(1 - \lambda)/4^n$.

Global depolarizing errors are physically unrealistic, so we also consider another error model that is more realistic but that is still simple to understand: gate-independent **local** depolarization. We simulated n-qubit direct RB for a hypothetical n-qubit processor (with $n = 2, 4, 6, \ldots, 14$) whereby after each n-qubit gate is applied every qubit experiences independent uniform depolarization with an error rate of 0.1%, i.e., the infidelity of each single-qubit depolarizing channel is $\epsilon = 0.001$. This error model has the convenient property that we can easily compute ϵ_Ω. The (entanglement) fidelity of a tensor product of n error channels $\mathcal{E}_1, \ldots, \mathcal{E}_n$ is the product of $\mathcal{E}_1$ to $\mathcal{E}_n$'s fidelities, i.e.,

$$F(\mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n, I \otimes \cdots \otimes I) = F(\mathcal{E}_1, I) \cdots F(\mathcal{E}_n, I), \tag{6.17}$$

so ϵ_Ω is given by $\epsilon_\Omega = 1 - (1 - 0.001)^n \approx n \times 0.001$. Note that, because the error rates are gate-independent, ϵ_Ω is again independent of the sampling distribution Ω in this example. Figure 6.2 (upper plot) shows S_d versus d (black circles) as well as the fit exponential (solid lines) and the

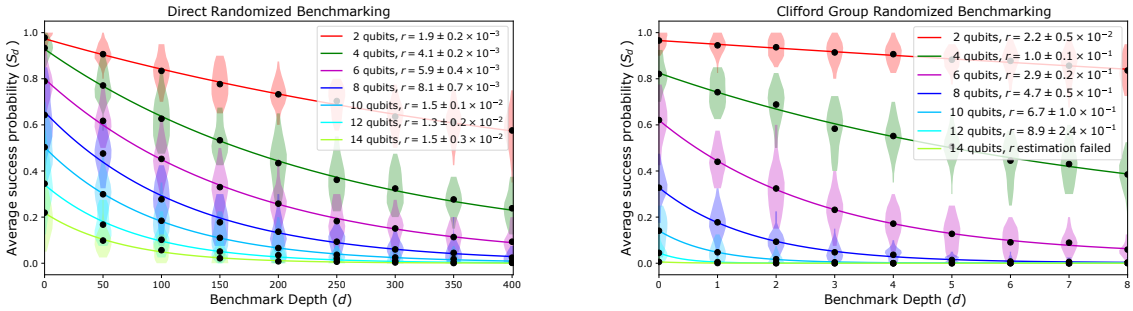


Figure 6.2: Simulated n-qubit direct RB and Clifford group RB for $n = 2, 4, 6, \ldots, 14$. Here, direct RB is benchmarking n-qubit gates that consist of a single layer of parallel one- and two-qubit gates. Clifford group RB is benchmarking the n-qubit Clifford group generated out of these layers. In this simulation, an imperfect n-qubit layer is modelled by the perfect unitary followed by independent uniform depolarization on each qubit at a rate of 0.1%. The points and violin plots are the means and the distributions of the estimated circuit success probabilities versus benchmark depth (d), respectively, and the lines are fits of the means to $S_d = A + Bp^d$. The estimated RB error rates ($\hat{r}$), which are reported in the legends, are obtained from the fit decay rates, using Eq. (6.15). We observe that the direct RB average circuit success probabilities (S_d) decay exponentially for all n, and that the direct RB error rate is given by $\hat{r} \approx 1 - (1 - 0.001)^n \approx n \times 0.001$. Direct RB is therefore accurately estimating the error rate of the n-qubit gates it is benchmarking. The Clifford group RB error rate grows very quickly with n—as it is benchmarking the n-qubit Clifford group, whose elements must be compiled into the available layers of parallel one- and two-qubit gates. The Clifford group RB $d = 0$ intercept quickly decays to approximately $\frac{1}{2^n}$ as n increases, at which point r is close to 100% and it cannot be estimated to reasonable precision with a practical amount of data. This demonstrates that direct RB is feasible on substantially more qubits than standard Clifford group RB.

estimates for r_Ω (in the legend), for each n. We observe that S_d decays exponentially, and that $\hat{r}_\Omega \approx \epsilon_\Omega$. We specify the gate set and sampling distribution used in this simulation in Example 2 of Section 6.4.4.

6.4.3 Direct RB's standard sampling parameters

The direct RB protocol has three user-specified parameters in common with all RB protocols, which we now briefly discuss. These are the user-specified benchmark depths (the values for d), the number of repetitions of each circuit (N), and the number of randomly sampled circuits at each benchmark depth (K_d). These parameters predominantly control purely statistical aspects of the experiment—specifically, the precision with which the direct RB experiment estimates the under-

lying $N, K_d \to \infty$ direct RB error rate. How these parameters control the statistical uncertainty in an estimated RB error rate, and how to choose these parameters, has been studied in some detail for Clifford group RB [155, 166, 193, 194, 216, 217]. Much of this work can be applied to direct RB. We will therefore not discuss these parameters in detail herein. Instead, we will provide some simple numerical evidence that the direct RB error rate can be estimated from reasonable amounts of data. The simulated direct RB experiments of Fig. 6.2 use a practical amount of data: we used 9 benchmarking depths, $K_d = 30$, and $N = 40$, for a total of approximately 10^4 samples for the direct RB instance at each value of n. Furthermore, the estimated direct RB error rates have low uncertainty (the uncertainties reported in the legend are 2σ and they are estimated using a standard bootstrap). Throughout the remainder of this paper we are predominantly interested in studying the behaviour of direct RB in the limit of $N, K_d \to \infty$; we denote the average circuit success probability and the direct RB error rate in this limit by S_d and r_Ω, respectively.

6.4.4 Selecting the gate set to benchmark

One of the defining characteristics of direct RB is that the user specifies the n-qubit gate set ($\mathbb{G}$) that is to be benchmarked. This is in contrast with Clifford group RB, where the benchmarked gate set is necessarily the n-qubit Clifford group—or another group that is a unitary 2-design. However there is not total freedom in choosing $\mathbb{G}$ in direct RB. The direct RB gate set has to satisfy the following conditions, some of which depend on the performance of the processor that is to be benchmarked:

- The group $\mathbb{C} \subseteq \mathrm{SU}(2^n)$ generated by $\mathbb{G}$ must be a unitary 2-design over $\mathrm{SU}(2^n)$. The requirement that $\mathbb{C}$ is a unitary 2-design can likely be relaxed (with appropriate adaptions to the direct RB experiments and/or data analysis) using techniques from RB for general groups [156–159, 161–164]. This is an interesting open area of research, as these generalizations have the potential to enable direct RB on any gate set $\mathbb{G}$. Our formal theory for direct RB (Section 6.7) relies on a slightly stronger condition on $\mathbb{G}$ than simply generating a unitary 2-design—it requires that $\mathbb{G}$ induces a **sequence-asymptotic unitary 2-design**.

The concept of a sequence-asymptotic unitary 2-design is introduced in Section 6.6, where we state necessary and sufficient conditions for a gate set to induce a sequence-asymptotic unitary 2-design. Here we state a simple condition that is sufficient (but not necessary) for a gate set $\mathbb{G}$ to induce a sequence-asymptotic unitary 2-design: the gate set $\mathbb{G}$ generates a unitary 2-design and contains the identity operation.

- The n qubits to be benchmarked must have sufficiently high fidelity operations for there to exist a circuit to prepare any random state from $\{|\psi\rangle = U(C)|0\rangle^{\otimes n} \mid C \in \mathbb{C}\}$ with non-negligible fidelity. To actually run direct RB, the user also needs an explicit algorithm for finding these circuits. This is required so that errors in the state and measurement preparation subcircuits do not mean that $S_d \approx \frac{1}{2^n}$ at $d = 0$ (how far above $\frac{1}{2^n}$ it is necessary for S_0 to be depends on how much data the user is willing to collect).

- It must be feasible to sample the direct RB circuits using a classical computer, which requires, e.g., multiplying together arbitrary elements in $\mathbb{C}$, and sampling uniformly from $\mathbb{C}$. This condition is also required to implement standard RB over $\mathbb{C}$, and it is satisfied if $\mathbb{C}$ is the Clifford group.

None of the three conditions on $\mathbb{G}$, above, require any efficient scaling with the number of qubits n. Instead, it suffices that each condition is satisfied at the values for n of interest. For example, perhaps only benchmarking one- or two-qubit gate sets is of interest (note that most RB experiments are one- or two-qubit Clifford group RB). The conditions also do not require that $\mathbb{G}$ is finite. This is illustrated by our first example, below, of a practically interesting type of gate set that can be benchmarked with direct RB.

Example gate set 1 [$\mathbb{G}_1(n)$]: An n-qubit gate set constructed from all possible combinations of parallel applications of CNOT gates on connected qubits and the one-qubit gates $X(\frac{\pi}{2})$ and $Z(\theta)$ for $\theta \in (-\pi, \pi]$ [where $X(\theta)$ and $Z(\theta)$ denote rotations by θ around σ_x and σ_z, respectively].

This gate set is universal—i.e., it generates $\mathrm{SU}(2^n)$—and so $\mathbb{G}_1(n)$ does not satisfy all our

conditions for $n \gg 1$ [generating a uniformly (Haar) random pure state is not efficient in n]. However, this gate set does satisfy our conditions for small n. For example, generating any unitary in SU(4) requires only three CNOT gates [218, 219]. The particular n-qubit universal gate set given above is just an example, and this reasoning holds for (almost) all few-qubit universal gate sets. Direct RB can therefore be used to benchmark almost any universal gate set over a few qubits [b]. This is arguably substantially simpler than the RB method used by Garion **et al.** [220] to benchmark a two-qubit gate set containing a controlled $Z(\frac{\pi}{4})$ gate (which is not a Clifford gate). Note that, recently, direct RB of one- and two-qubit universal gate sets has been experimentally demonstrated in [221].

Example gate set 2 [$\mathbb{G}_2(n)$]: An n-qubit gate set constructed from CNOT gates and any set of generators for the single-qubit Clifford group (such as the Hadamard and phase gates).

This is the type of gate set used in the simulations of Fig. 6.2 (and shown in the schematic of Fig. 6.1). In that simulation the gate set consisted of the Hadamard and phase gates as well as CNOT gates between any pair of qubits. That is, this simulation uses all-to-all connectivity. However, note that direct RB with this gate set can be applied to a processor with any connectivity, as direct RB does not have to be applied to a processor's native gate set (applying direct RB to a standardized set of n-qubit layers could be useful for comparing different processors). In that case, CNOT gates between distant qubits would be synthesized via SWAP gate chains.

Example gate set 3 [$\mathbb{G}_3(n)$]: An n-qubit gate set constructed from a maximally entangling two-qubit Clifford gate (e.g., CNOT or CPHASE) and the full single-qubit Clifford group.

Direct RB of a gate set with this form is particularly robust and simple to understand from a theoretical perspective (this is the type of gate set that was used in the direct RB experiments of [167]). As with $\mathbb{G}_2(n)$, $\mathbb{G}_3(n)$ specifies a family of gate sets. We now specify a particular convenient

[b] It is only almost any universal gate set rather than every universal gate set as, e.g., it is possible to construct universal gate sets that require arbitrarily long sequences of gates to implement some one-qubit gates—but such gate sets are of no practical relevance

gate set within this family (which is entirely specified given a processor's connectivity).

Example gate set 4 [$\mathbb{G}_4(n)$]: An n-qubit gate set consisting of all n-qubit gates composed from applying a layer L_1 and then a layer L_2 where these layers have the following forms: L_1 is a layer containing one of the 24 single-qubit Clifford gates on each qubit, and L_2 is a layer containing non-overlapping CNOT gates on connected pairs of qubits.

A random gate from $\mathbb{G}_4(n)$ locally randomizes the basis of each qubit (and applies some random arrangement of two-qubit gates)—if the marginal distribution over L_1 layers is uniform it implements **local** 2-design randomization (composed with a more complex multi-qubit randomization step). A gate set of this form is assumed in parts of Section 6.5, in order to simplify the theory presented there. Note that direct RB circuits for $\mathbb{G}_4(n)$ have much in common with the random circuits of Google's quantum supremacy experiments [32], which are used in cross-entropy benchmarking [222]. However, these direct RB circuits contain only Clifford gates (making them efficient to simulate), and they contains additional structure (the state preparation and measurement preparation subcircuits).

6.4.5 Selecting the sampling distribution

There are two main customizable aspects of a direct RB experiment: the gate set ($\mathbb{G}$) and the sampling distribution (Ω). We now discuss the role of Ω, and how to choose it. The direct RB sampling distribution Ω can be any probability distribution over $\mathbb{G}$ that has support on a subset $\mathbb{G}'$ of $\mathbb{G}$ that also satisfies all the requirements for a direct RB gate set (e.g., $\mathbb{G}'$ also generates $\mathbb{C}$). The sampling distribution Ω and the gate set $\mathbb{G}$ control what direct RB measures—direct RB estimates the mean infidelity of an n-qubit gate sampled from Ω. Therefore, the primary consideration when selecting Ω is to choose a distribution that defines an error rate ϵ_Ω of interest. There are infinite valid choices for Ω, and so we do not attempt to discuss all interesting choices for Ω here. One category of sampling distributions we have found useful in experiments is one-parameter families of distributions Ω_ξ in which ξ sets the expected two-qubit gate density of the sampled gate. There

are many possible such families of distributions—in Appendix 6.10 we briefly describe a family of probability distributions that has been used in direct RB experiments (the "edge grab" sampler, introduced in [223]).

The sampling distribution Ω defines the error rate that direct RB measures, but it also impacts the reliability of direct RB, i.e., whether S_d decays exponentially and how close r_Ω is to ϵ_Ω. So a sampling distribution should be chosen for which direct RB will be reliable. For every sampling distribution satisfying the above requirements, direct RB will be reliable for **sufficiently low error** gates. This is because, for any $(\mathbb{G}, \Omega)$ satisfying the criteria of direct RB, any Pauli error will be randomized by a depth l Ω-random circuit for sufficiently large l (see the theory in Section 6.5). But this randomization can be very slow, i.e., the required depth l can be very large, and so the error randomization can be much slower than the rate of errors—which, in Section 6.5, we explain can result in unrealiable direct RB (e.g., multi-exponential decays). For example, we could have $\Omega(G) = 1-\delta$ for one gate G and some $\delta \ll 1$. In this case, a length $l \ll \frac{1}{\delta}$ sequence of gates sampled from Ω is almost certainly just l repetitions of G—so sequences of this length do not randomize errors at all. The theory presented in Section 6.5 will help to explain how to choose a sampling distribution that guarantees reliable direct RB.

6.4.6 The reason for randomized 2-design state preparation and measurement in direct RB circuits

Direct RB is built on the simple idea of directly benchmarking a gate set $\mathbb{G}$ by running varied-depth circuits whose layers are sampled independently from a distribution over $\mathbb{G}$—a class of circuits that have been termed **Ω-distributed random circuits** [221]. However, the depth d direct RB circuits do not just consist of a depth d Ω-distributed random circuits. They surround the core direct RB circuit—the Ω-distributed random circuit—with additional structure (see Fig. 6.1). We now explain the purpose of this additional structure. The direct RB circuits begin with a randomized state preparation sub-circuit ($C_{\rm sp}$) and end with a measurement preparation sub-circuit ($C_{\rm mp}$). The purpose of $C_{\rm mp}$ is simply to return the qubits to the computational basis, maximizing error

visibility and simplifying the data analysis. The subcircuit has the same purpose as the group inversion element in Clifford group RB.

The reasons for beginning a direct RB circuit with a randomized state preparation circuit (C_{sp}) are now explained. First, together C_{sp} and C_{mp} implement an approximate (state) 2-design twirl on the error in the core circuit. This is because the state preparation subcircuit generates a sample from a (state) 2-design, which (if implemented perfectly) twirls the overall error channel of the core circuit, so that it behaves as though it is a global depolarizing channel. This will be shown in each of our complementary theories for direct RB (Section 6.5 and Section 6.7). An alternative interpretation of C_{sp} is that it seeds the random walk over the group $\mathbb{C}$ induced by a random sequence of group generators (gates from $\mathbb{G}$) with an initial uniformly random group element (that has been compiled into the state preparation for efficiency).

There is also a more practical but mundane reason for including the state preparation subcircuit C_{sp}. If C_{sp} is **not** included in a direct RB circuit, the contribution of errors in C_{mp} to a direct RB circuit's total failure probability can be strongly dependent on the depth of the core circuit (d). This would pollute r_Ω, i.e., r_Ω would not accurately quantify the error per layer in the core of the direct RB circuit (ϵ_Ω). To see this, consider direct RB without the randomized state preparation, and assume an efficient compiler for creating C_{mp}. Further, assume that $\mathbb{G}$ is small (compared to $\mathbb{C}$), so that the distribution over unitaries produced by a short sequence of gates sampled from Ω must be far from the uniform distribution over $\mathbb{C}$. When d is small, the average depth of C_{mp} will grow approximately linearly with d under many circumstances. This is because C_{mp} is likely to look very much like the depth-d core circuit that it inverts, with the order reversed and individual gates inverted. However, as d gets large, the core circuit converges to a random group element. So the mean depth (and all other properties) of the C_{mp} circuits asymptote to a fixed value that does not depend on d. Critically, this means that the depth of C_{mp} would have a nontrivial dependence on d. This variation will pollute the decay rate (and so r_Ω), and may even cause non-exponential decays—its impact is uncontrolled, as the direct RB protocol is agnostic to how C_{mp} is implemented (just as the Clifford group RB is agnostic about the Clifford compiler). This effect could instead

be mitigated by carefully designing the compiler for $C_{\rm mp}$, but this is a complicated undertaking. Instead, the problem is simply solved by the inclusion of $C_{\rm sp}$. With $C_{\rm sp}$ included, the average depth of $C_{\rm mp}$ is independent of d. The only residual d dependence outside of the core circuit is in **correlations** between the state that is prepared by $C_{\rm sp}$ and the state that must be mapped to the computational basis by $C_{\rm mp}$ (they are perfectly correlated at $d = 0$ and uncorrelated as $d \to \infty$). These correlations can have an effect on r_Ω, in principle, but in realistic settings they appear to have no observable effect on r_Ω (see Section 6.5).

6.4.7 The reason for conditional compilation: improved scalability

The state preparation and measurement sub-circuits within a direct RB circuit are very similar to the first and last gates in a standard Clifford group RB circuit (see Fig. 6.1)—and they play essentially the same roles (see above). However, they differ in a practically important way. Both Clifford group RB and direct RB begin with a circuit that implements a group element $F_{\rm sp}$ sampled uniformly from $\mathbb{C}$, but whereas Clifford group RB demands that this circuit implements $U(F_{\rm sp})$ on any input state, direct RB only requires that the circuit generates the same state as $U(F_{\rm sp})$ when applied to $|0\rangle^{\otimes n}$. We call this **unconditional** and **conditional** compilation, respectively. The same distinction holds for the final inversion step, used in direct and Clifford group RB.

We choose to use conditional compilation in direct RB circuits because it substantially increases the number of qubits on which direct RB is feasible—as illustrated by the Clifford and direct RB simulations shown in Fig. 6.2 (this difference between direct and Clifford group RB is the only reason for the difference in their S_0 values, which is the primary factor setting the number of qubits that it is feasible to benchmark). This is because conditional compilation results in circuits that are shallower and contain fewer two-qubit gates [172]. This is demonstrated in Fig. 6.3, which compares the mean number of CNOT gates in the circuits generated by a conditional and unconditional n-qubit Clifford compiler. To generate this plot we used open-source compilation algorithms [224] (that we also used for all the simulations herein). These algorithms are unlikely to generate circuits with minimal two-qubit gate counts (see [172–174] for work on optimal Clifford compilation) although

we have found that they perform reasonably well [c].

Note that the error rate measured by direct RB is (approximately) independent of the details of the compiler used to generate these states. The property of the compilation algorithm that is of importance to direct RB is the efficiency with which it generates these states—the higher the fidelity of these states the more qubits direct RB will be feasible on. This is convenient, as algorithms for generating many-qubit states/unitaries are typically "black-box" and the properties of the circuits they generate cannot be easily controlled. In contrast, with Clifford group RB the compiler entirely defines the physical meaning of the RB error rate—a Clifford group RB error rate cannot be used to quantify **native gate** performance without a detailed understanding of the compiled circuits.

[c] The algorithm we use is based on Gaussian elimination; the compiled circuits contain $O(n^2)$ two-qubit gates for both the state and unitary compilations. This scaling is not optimal: their exist algorithms that generate circuits containing $O(n^2/\log(n))$ two-qubit gates [172]. But in the $n = O(10)$ regime we have found that our compilations contain many fewer two-qubit gates than the $O(n^2/\log(n))$ algorithms of [172]. We have not compared our compilation algorithms to more recent work on Clifford compilation [174].

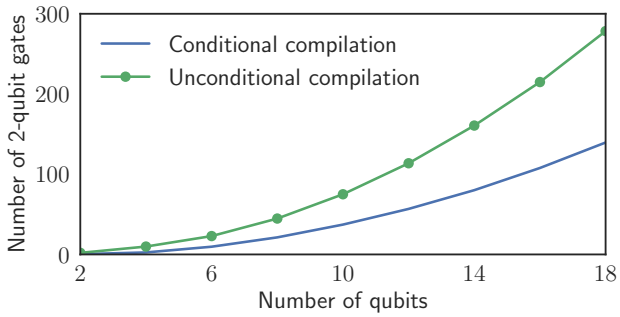


Figure 6.3: A comparison of the mean number of two-qubit gates in circuits for implementing a uniformly random n-qubit Clifford gate either unconditionally (green connected points) or conditionally on the input of $|0\rangle^{\otimes n}$ (blue line). In the latter case, this is a circuit mapping $|0\rangle^{\otimes n}$ to a uniformly random n-qubit stabilizer state. This demonstrates the significant increase in the number of qubits on which direct RB is feasible—for a fixed two-qubit gate error rate—compared to Clifford group RB. This data was generated using the same open-source compilers that we use for all our direct RB and Clifford group RB simulations [224], and here we compiled into one-qubit gates and CNOT gates between any pair of qubits (i.e., all-to-all connectivity).

6.4.8 The reason for randomizing the success outcome

Direct RB specifically allows for uniform randomization of the success bit string (this is also possible with Clifford group RB [225], but it is not standard practice). This bit string randomization is not essential, but in our view it is preferable. This is because it guarantees that the $d \to \infty$ asymptote of the average RB success probability S_d is $\frac{1}{2^n}$ (for all Markovian errors), so we can fix $A = \frac{1}{2^n}$ in $S_d = A + Bp^d$. As Harper **et al.** [225] discuss the motivation and statistical impact of this bit string randomization (in the context of Clifford group RB), we do not do so further here.

6.4.9 The error rate convention: choosing the decay rate scaling factor

For both direct and Clifford group RB, data are analyzed by fitting the average success probabilities to $S_d = A + Bp^d$ and then mapping p to an error rate r (or a fidelity $1 - r$). In Clifford group RB (and other RB protocols), p is conventionally mapped to an error rate defined by $r' = (2^n - 1)(1 - p)/2^n$ [154, 155]. This is not the definition that we use. As specified in Eq. (6.15), we define the direct RB error rate to be $r = (4^n - 1)(1 - p)/4^n$. For $n \gg 1$ the difference between r and r' is negligible, but it is substantial for $n \sim 1$. Our decision to use Eq. (6.15) to define r is not specific to direct RB. The RB error rate r' corresponds to the gate set's mean **average gate infidelity**, whereas r corresponds to the mean **entanglement infidelity**. We choose to use a definition for the RB error rate that correspond to entanglement infidelity because of entanglement infidelities convenient properties. For example, for Pauli stochastic errors the entanglement infidelity corresponds to the probability of any Pauli error occurring (see Section 6.3) and the entanglement fidelity of multiple gates used in parallel is equal to the product of their entanglement fidelities (assuming no additional errors when gates are parallelized)—see Eq. (6.17). Due to these convenient properties, this choice has also been made with other scalable benchmarking techniques (e.g., cycle benchmarking [226]).

6.4.10 Comparison to other methods for native gate randomized benchmarking

Direct RB is not the first or only proposal for RB directly on a gate set that generates a group. Below we explain the relationship between RB and each of these protocols:

- The method of Knill **et al.** [153] benchmarks a set of gates that generate the one-qubit Clifford group, and this has been widely used as an alternative to Clifford group RB (e.g., see the references within [227]). That method consists of uniform sampling over a specific set of one-qubit Clifford group generators. So it is a particular example of direct RB (as pointed out by Boone **et al.** [227]) **except** that it does not include the initial stabilizer state preparation step (which is of little importance in the single-qubit setting). So our theory for direct RB is further evidence that the protocol of [153] is just as reliable and arguably as well-motivated as Clifford group RB, although it measures a different error rate.

- The extensions of the method of Knill **et al.** up to three qubits [197] also fit within the framework of direct RB except that, again, in that method there is no stabilizer state initialization.

- Independently of the development of direct RB, França and Hashagen proposed "generator RB" [196], which is also direct RB without the stabilizer state preparation step and without user-configurable sampling (but note that some of the theory within [196] also applies to direct RB).

- Since the development of direct RB, a protocol called mirror RB [190, 221] has been introduced that adapts direct RB to improve its scalability. Mirror RB methods replace the stabilizer state preparation and measurement sub-circuits from direct RB with a mirror circuit reflection structure, meaning that they contain no large subroutines. Mirror RB techniques are more scalable than direct RB, and they are designed to measure the same error rate (ϵ_Ω). However, the error rate measured by existing mirror RB methods is a less reliable estimate of ϵ_Ω (it is typically a slight underestimate—see [190, 221] for further de-

tails), so mirror RB is not a strict improvement on direct RB. Much of the theory presented in this paper is also applicable, or adaptable, to mirror RB.

6.5 Understanding direct RB using error scrambling

Elided.

6.6 Sequence-asymptotic unitary designs

Our theory for direct RB with general Markovian errors (Section 6.7) relies on the concept of **sequence-asymptotic unitary 2-designs**. In this section we define sequence-asymptotic unitary 2-designs and we prove some results about their properties. Informally, a sequence-asymptotic unitary 2-design is a gate set $\mathbb{G}$ and a sampling distribution Ω over $\mathbb{G}$ such that the set of unitaries created by length l sequences of gates sampled from Ω converges to a unitary 2-design as $l \to \infty$. The concept of sequence-asymptotic unitary 2-designs is related to approximate unitary 2-designs [191, 192], and the broader theory of convergence to unitary t-designs and other scrambling properties of random circuits [202–208]. But, because we only use sequence-asymptotic unitary 2-designs as a tool for studying direct RB, we do not discuss the relationship with this prior work.

A plausibly sufficient condition for $(\mathbb{G}, \Omega)$ to constitute a sequence-asymptotic unitary 2-design is that $\mathbb{G}$ generates a unitary 2-design (assuming that Ω has support on all of $\mathbb{G}$). This is because, for a wide class of generating sets and probability distributions, a random sequence of group generators converges to a random group element with increasing length sequences [228–230]. However, this convergence does not happen for all probability distributions over all generating sets: the long-length distribution can oscillate between the uniform distribution on different parts of the group (an example is given towards the end of this section, and this is only possible if $\mathbb{G}$ is contained within a coset of a proper subgroup of $\mathbb{C}$). In this section, we will show the following: given a set of generators for any group that is a unitary 2-design, length-l sequences of Ω-distributed and independently sampled generators converge to a unitary 2-design as $l \to \infty$, if and only if the set of length-l sequences also generate unitary 2-designs for all l.

6.6.1 Notation

First we need to define some notation. In this section we consider a quantum system of dimension d, without the restriction to $d = 2^n$ used throughout the rest of this paper. Let $\mathfrak{L}(H_d)$ denote the space of linear operators on the d-dimensional Hilbert space H_d. Let $\mathbb{T}(H_d)$ denote the space of superoperators on H_d, which is the set of all linear operators $\mathcal{S}$ whereby $\mathcal{S} : \mathfrak{L}(H_d) \to \mathfrak{L}(H_d)$. In the following proofs, we will use Hilbert-Schmidt space, and superoperator "stacking". Let

$$\mathrm{B} = \{B_0, B_1, \ldots, B_{d^2-1}\} \tag{6.18}$$

be an orthonormal basis for $\mathfrak{L}(H_d)$, with respect to the Hilbert-Schmidt inner product

$$\langle X, Y \rangle = \mathrm{Tr}\left(X^\dagger Y\right). \tag{6.19}$$

We can always choose this basis such that $B_0 = \mathbb{I}/\sqrt{d}$ and $\mathrm{Tr}(B_j) = 0$ for $j > 0$, and this will be assumed herein. For example, with $d = 2^n$ the elements of this basis can be the n-qubit Pauli operators multiplied by $1/\sqrt{2^n}$.

Any $\rho \in \mathfrak{L}(H_d)$ can be expanded with respect to the basis B as

$$\rho = \sum_j \langle B_j,\ \rho \rangle B_j. \tag{6.20}$$

As such, ρ may be represented by the column vector

$$|\rho\rangle\rangle = \left(\langle B_0, \rho\rangle, \langle B_1, \rho\rangle, \ldots, \langle B_{d^2-1}, \rho\rangle\right)^T. \tag{6.21}$$

Similarly, a superoperator $\mathcal{X}$ may be represented as the matrix $\mathcal{X}^{\mathrm{HS}}$ with elements $\mathcal{X}^{\mathrm{HS}}_{ij} = \langle B_i, \mathcal{X} B_j \rangle$. In the following we will drop the superscript; we will not use a distinct notation for a superoperator and the matrix representation of the superoperator acting on Hilbert-Schmidt space.

In this and the next section we will be considering linear maps from superoperators to superoperators, which we will refer to as "superchannels" and denote using the script font (e.g., L). When considering superchannels it is often convenient to represent matrices acting on Hilbert-Schmidt space (i.e., superoperators) as vectors in a vector space, and superchannels as matrices

acting on that vector space. To achieve this we use the invertible linear "stacking" map

$$\mathrm{vec}\left(|B_j B_k|\right) = |B_k \otimes |B_j, \tag{6.22}$$

which stacks the columns of a matrix. We denote the inverse "unstacking" operation by

$$\mathrm{unvec}\left(|B_k \otimes |B_j\right) = |B_j B_k|. \tag{6.23}$$

From this definition, it follows that

$$\mathrm{vec}(\mathcal{A}\mathcal{B}\mathcal{C}) = \left(\mathcal{C}^T \otimes \mathcal{A}\right) \mathrm{vec}(\mathcal{B}) \tag{6.24}$$

for Hilbert-Schmidt matrices $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$. Therefore, any superchannel K with the action $K(\mathcal{B}) = \sum_i \mathcal{A}_i \mathcal{B} \mathcal{C}_i$ may be rewritten as a matrix

$$\mathrm{mat}[K] = \sum_i \left(\mathcal{C}_i^T \otimes \mathcal{A}_i\right), \tag{6.25}$$

that acts on vectors $|\mathcal{B}) \equiv \mathrm{vec}(\mathcal{B})$, where we will use the $|\cdot)$ notation to succinctly denote stacked superoperators while distinguishing them from pure states in Hilbert space (denoted $|\cdot\rangle$) and Hilbert-Schmidt vectors (denoted $|\cdot\rangle\rangle$).

6.6.2 Ordinary unitary designs

Before introducing sequence-asymptotic unitary designs, we first review the definition of an "ordinary" unitary t-design [191, 192]. Although not the original definition [192], an equivalent definition for a unitary t-design is [191]:

Definition 1. *The gate set $\mathbb{G} \subseteq \mathbb{U}$ is a unitary t-design if*

$$\frac{1}{|\mathbb{G}|} \sum_{\mathcal{G} \in \mathbb{G}} \mathcal{G}^{\otimes t} = \int_{\mathcal{U} \in \mathbb{U}} \mathcal{U}^{\otimes t} d\mu \tag{6.26}$$

where $d\mu$ is the Haar measure on $\mathbb{U}$.

The class of designs important herein are the unitary 2-designs. The definition of a unitary t-design implies that an equivalent characterization of a unitary 2-design is the following: a gate set

is a unitary 2-design if and only if, when it "twirls" any channel, it projects this channel onto the space of depolarizing channels. Specifically, define the $\mathbb{G}$-twirl superchannel $T_{\mathbb{G}} : \mathbb{T}(H_d) \to \mathbb{T}(H_d)$ by

$$T_{\mathbb{G}}(\mathcal{E}) = \frac{1}{|\mathbb{G}|} \sum_{\mathcal{G} \in \mathbb{G}} \mathcal{G}\mathcal{E}\mathcal{G}^\dagger. \tag{6.27}$$

The twirl superchannel is a linear operator on superoperators. If $\mathbb{G}$ is a unitary 2-design then

$$T_{\mathbb{G}}(\mathcal{E}) = \mathcal{D}_{\lambda(\mathcal{E})}, \quad \lambda(\mathcal{E}) = \frac{\mathrm{Tr}[\mathcal{E}] - 1}{d^2 - 1}, \tag{6.28}$$

for any trace-preserving superoperator $\mathcal{E} \in \mathbb{T}(H_d)$ [156], where $\mathcal{D}_{\lambda(\mathcal{E})}$ is a "global" depolarizing channel (i.e, uniform depolarization on the d-dimensional space).

Eq. (6.28) implies that the twirl superchannel for any unitary 2-design is a rank 2 projector, with support on the depolarizing subspace $\mathbb{S}_{\mathrm{dep}}$ spanned by the identity channel and any non-identity depolarizing channel. Explicitly,

$$\mathbb{S}_{\mathrm{dep}} = \mathrm{span}\left\{\mathcal{D}_0, \mathbb{I}\right\}, \tag{6.29}$$

where $\mathbb{I}$ denotes the identity superoperator and $\mathcal{D}_0$ is the completely depolarizing channel.

6.6.3 Sequence-asymptotic unitary designs

We now introduce the notion of a sequence-asymptotic unitary 2-design. This concept is distinct from the "asymptotic unitary 2-designs" of Gross **et al.** [191], which are asymptotically unitary 2-designs with respect to scaling of the dimension d. This definition will be in terms of a set $\mathbb{G}$ of unitaries and a probability distribution Ω over $\mathbb{G}$. Note that we do not assume that $\mathbb{G}$ is finite. Instead, we assume the weaker condition that $\mathbb{G}$ contains a finite number of "gates" that are each either fixed (i.e., a constant matrix) or continuously parameterized. For example, a one-qubit gate set of this sort $\mathbb{G}$ is the Hadamard gate and a continuous family of rotations around σ_z. The probability distribution Ω should be interpreted as an arbitrary weighted mixture of probability distributions over these fixed or parametrized gates. That is, sampling from Ω consists of first selecting one of the gates, and if this selected gate is continuously parameterized we then further sample a value for the continuous parameter(s) according to some distribution. We also assume that

all of the probability density functions over any continuous parameter(s) are continuous function. So when we use the notation $\sum_{\mathcal{G}\in\mathbb{G}}$ this should be interpreted as a sum over the finite number of (perhaps parameterized) gates and an integral over any continuous parameters in each gate.

Definition 2. *Let $\mathbb{G} \subseteq \mathbb{U}$ be a unitary gate set and $\Omega : \mathbb{G} \to [0,1]$ be a probability distribution over $\mathbb{G}$. The tuple $(\mathbb{G}, \Omega)$ is a sequence-asymptotic unitary t-design if*

$$\lim_{m\to\infty} \left(\sum_{\mathcal{G}\in\mathbb{G}} \Omega(\mathcal{G}) \mathcal{G}^{\otimes t} \right)^m = \int_{\mathcal{U}\in\mathbb{U}} \mathcal{U}^{\otimes t} d\mu. \tag{6.30}$$

If $\mathbb{G}$ is a unitary t-design then $(\mathbb{G}, \Pi)$ is also a sequence-asymptotic unitary t-design, where Π is the uniform distribution (this can be shown by induction).

Definition 3. *The $(\mathbb{G}, \Omega)$-twirl superchannel is the linear map $T_{\mathbb{G},\Omega} : \mathbb{T}(H_d) \to \mathbb{T}(H_d)$ with the action*

$$T_{\mathbb{G},\Omega}(\mathcal{E}) = \sum_{\mathcal{G}\in\mathbb{G}} \Omega(\mathcal{G}) \mathcal{G}\mathcal{E}\mathcal{G}^\dagger. \tag{6.31}$$

If $(\mathbb{G}, \Omega)$ is a sequence-asymptotic unitary 2-design, Definition 2 implies that the m^{th} power of the $(\mathbb{G}, \Omega)$ twirl superchannel $T_{\mathbb{G},\Omega}$ converges to a rank 2 projector as $m \to \infty$. As such, the absolute values of the eigenvalues of $T_{\mathbb{G},\Omega}$ outside of this subspace must be strictly bounded above by 1. The converse is also true: this condition on the eigenvalues of $T_{\mathbb{G},\Omega}$ is sufficient for any $(\mathbb{G}, \Omega)$ to be a sequence-asymptotic unitary 2-design. We state this more formally in the following proposition, for which we will require the notion of the fixed points of a linear map. The fixed points of a linear map $L : V \to V$, for some vector space V, are those $v \in V$ such that $L(v) = v$. That is, they are the eigenvectors with eigenvalue 1. For example, the space of all fixed points of the twirling map for any unitary 2-design is $\mathbb{S}_{\text{dep}}$ [as defined in Eq. (6.29)]. The following are formal requirements on properties of the map $T_{\mathbb{G},\Omega}$, that will allow us to characterize the sequence asymptotic 2-designs.

Proposition 1. *Let $\mathbb{G} \subseteq \mathbb{U}$ be a unitary gate set and $\Omega : \mathbb{G} \to [0,1]$ be a probability distribution over $\mathbb{G}$. The tuple $(\mathbb{G}, \Omega)$ is a sequence-asymptotic unitary 2-design if and only if the following two conditions hold:*

(1) The space of all fixed points of $T_{\mathbb{G},\Omega}$ is the subspace of depolarizing channels $\mathbb{S}_{\text{dep}}$.

(2) If $T_{\mathbb{G},\Omega}(\mathcal{E}) = \lambda\mathcal{E}$ *with* $\mathcal{E} \notin \mathbb{S}_{\text{dep}}$ *then* $|\lambda| < 1$.

Proof. We first prove that if $(\mathbb{G}, \Omega)$ is a sequence-asymptotic unitary 2-design then (1) and (2) are implied. If $(\mathbb{G}, \Omega)$ is a sequence-asymptotic unitary 2-design then the fixed points of $T^m_{\mathbb{G},\Omega}$ converge to $\mathbb{S}_{\text{dep}}$ as $m \to \infty$. The superchannel $T_{\mathbb{G},\Omega}$ has complex eigenvalues of modulus at most 1. One way to see this is by observing that $\text{mat}\,[T_{\mathbb{G},\Omega}] := \sum_{\mathcal{G}\in\mathbb{G}} \Omega(\mathcal{G})\mathcal{G}^* \otimes \mathcal{G}$, which has the same spectrum as $T_{\mathbb{G},\Omega}$, is a convex sum of unitaries, and so it is a CPTP map. Hence, any fixed point of $T^m_{\mathbb{G},\Omega}$ is an eigenchannel of $T_{\mathbb{G},\Omega}$ with an eigenvalue of unit modulus, and so it is also a fixed point of $T_{\mathbb{G},\Omega}$ if this eigenvalue is 1. Moreover, all of the unit modulus eigenvalues of $T_{\mathbb{G},\Omega}$ must be equal to 1, because if this was not the case then $T^m_{\mathbb{G},\Omega}$ would not have a convergent spectrum as $m \to \infty$, which would imply that $(\mathbb{G}, \Omega)$ was not a sequence-asymptotic unitary 2-design. Hence, the space of fixed points of $T_{\mathbb{G},\Omega}$ is $\mathbb{S}_{\text{dep}}$, and the eigenvalues for all eigenchannels outside of this subspace have absolute value strictly less than 1.

It remains to prove that if (1) and (2) hold then $(\mathbb{G}, \Omega)$ is a sequence-asymptotic unitary 2-design. If (1) and (2) hold this immediately implies that $T^m_{\mathbb{G},\Omega}$ converges to a rank-2 projector onto $\mathbb{S}_{\text{dep}}$ as $m \to \infty$. Projection onto this subspace is the action of the right hand side of Definition 2 for $t = 2$ [156], so (1) and (2) imply that Definition 2 holds for $t = 2$. □

We now prove a lemma that will be of use later. Here and later, we let $\mathbb{G}^k = \{g_k \dots g_1 \mid g_i \in \mathbb{G}\}$ be the gate set consisting of the set of all length k sequences of gates from $\mathbb{G}$ and Ω^k be the distribution over $\mathbb{G}^k$ obtained by k convolutions of Ω, i.e., $\Omega^k(g_k \dots g_1) = \Omega(g_k) \cdots \Omega(g_1)$.

Lemma 1. *The tuple* $(\mathbb{G}, \Omega)$ *is a sequence-asymptotic unitary 2-design if and only if, for all* $k > 0$, $(\mathbb{G}^k, \Omega^k)$ *is a sequence-asymptotic unitary 2-design.*

Proof. Suppose $(\mathbb{G}, \Omega)$ is a sequence-asymptotic unitary 2-design. The composite of twirling maps is a twirling map. Thus by taking the composite map, (1) and (2) of Proposition 1 hold. The reverse direction is trivial. □

Proposition 1 allows us to test whether or not a given $(\mathbb{G}, \Omega)$ tuple is a sequence-asymptotic

unitary 2-design, by explicitly calculating the spectrum of $T_{\mathbb{G},\Omega}$. However, for an n-qubit gate set, this brute force calculation becomes quickly intractable as n increases—using the stacked Hilbert-Schmidt representation, $T_{\mathbb{G},\Omega}$ is a $16^n \times 16^n$ dimensional matrix. Moreover, this proposition is not particularly instructive for the task of finding asymptotic unitary 2-designs. In the following proposition and theorem, we give an alternative characterization of sequence-asymptotic unitary 2-designs.

Proposition 2. *Let $\mathbb{G} \subseteq \mathbb{U}$ be a unitary gate set and $\Omega : \mathbb{G} \to [0,1]$ be a probability distribution over $\mathbb{G}$ with support on all of $\mathbb{G}$. Then $(\mathbb{G}, \Omega)$ is a sequence-asymptotic unitary 2-design if and only if the subspace of $\mathbb{T}(H_d)$ containing all superoperators that are preserved up to a (possibly superoperator-dependent) phase under conjugation by all elements of $\mathbb{G}$ is the space of depolarizing channels $\mathbb{S}_{\text{dep}}$.*

Proof. Assume that $(\mathbb{G}, \Omega)$ is a sequence-asymptotic unitary 2-design. Let $\mathbb{S}$ denote the set of all superoperators preserved under conjugation by all elements of $\mathbb{G}$ up to a phase, meaning the set of all $\mathcal{S} \in \mathbb{T}(H_d)$ satisfying $\mathcal{G}\mathcal{S}\mathcal{G}^{-1} = e^{i\theta_{\mathcal{S}}}\mathcal{S}$ for all $\mathcal{G} \in \mathbb{G}$, where $\theta_{\mathcal{S}}$ is a phase that depends on the superoperator being conjugated. If the superoperator $\mathcal{S}$ is preserved under conjugation up to a phase by all elements of $\mathbb{G}$ then it is a unit-modulus eigenvector of the $(\mathbb{G}, \Omega)$-twirl superchannel $T_{\mathbb{G},\Omega}$. Thus, by (2) in Proposition 1, it must have $\theta_{\mathcal{S}} = 0$ because $(\mathbb{G}, \Omega)$ is a sequence-asymptotic unitary 2-design by assumption. Hence, $\mathbb{S}$ is a subspace of the set of all fixed points of $T_{\mathbb{G},\Omega}$. If $\mathbb{G}_\Omega$ is a sequence-asymptotic unitary 2-design then Proposition 1 states that the space of all fixed points of $T_{\mathbb{G},\Omega}$ is $\mathbb{S}_{\text{dep}}$. Therefore, as all depolarizing maps are preserved under conjugation by any unitary gate, if $(\mathbb{G}, \Omega)$ is a sequence-asymptotic unitary 2-design then $\mathbb{S} = \mathbb{S}_{\text{dep}}$.

It remains to show that $\mathbb{S} = \mathbb{S}_{\text{dep}}$ implies that $(\mathbb{G}, \Omega)$ is a sequence-asymptotic unitary 2-design. Assume that $\mathbb{S} = \mathbb{S}_{\text{dep}}$, let $\mathcal{R}$ be any element of $\mathbb{G}$, and consider the gate set $\mathcal{R}^{-1}\mathbb{G} = \{\mathcal{R}^{-1}\mathcal{G} \mid \mathcal{G} \in \mathbb{G}\}$. Now consider any superoperator $\mathcal{V}$ satisfying $\mathcal{G}\mathcal{V}\mathcal{G}^{-1} = e^{i\theta_{\mathcal{V}}}\mathcal{V}$ for all $\mathcal{G} \in \mathcal{R}^{-1}\mathbb{G}$ and some $\theta_{\mathcal{V}}$. The identity superoperator $\mathbb{I}$ is in $\mathcal{R}^{-1}\mathbb{G}$, and hence $e^{i\theta_{\mathcal{V}}} = 1$ for any such $\mathcal{V}$. As such, the only superoperators preserved up to phase under conjugation by all elements of $\mathcal{R}^{-1}\mathbb{G}$ are

preserved with a phase of 1. Thus, because $\mathbb{S} = \mathbb{S}_{\text{dep}}$ by assumption, the space of all superoperators preserved under conjugation by all elements of $\mathcal{R}^{-1}\mathbb{G}$ contains $\mathbb{S}_{\text{dep}}$ since we see, for $\mathcal{S} \in \mathbb{S}_{\text{dep}}$, $(\mathcal{R}^{-1}\mathbb{G})\mathcal{S}(\mathcal{R}^{-1}\mathbb{G})^{-1} = e^{-i\theta_{\mathcal{S}}}\mathcal{R}\mathcal{S}\mathcal{R}^{-1} = e^{-i\theta_{\mathcal{S}}}\mathcal{S}$. The space of superoperators preserved up to a given phase under conjugation by all elements of $\mathbb{G}\mathcal{R}^{-1}$ is related via unitary conjugation to the space of superoperators preserved up to the same phase under conjugation by all elements of $\mathbb{G}$. Hence, if $\mathcal{V}$ is preserved up to phase under conjugation by all elements of $\mathbb{G}$, it is preserved with a phase of 1 and $\mathcal{V} \in \mathbb{S}_{\text{dep}}$.

Now, consider the $(\mathbb{G}, \Omega)$-twirl superchannel $T_{\mathbb{G},\Omega}$, and consider any $\mathcal{V}$ satisfying $T_{\mathbb{G},\Omega}(\mathcal{V}) = e^{i\theta}\mathcal{V}$ for some $e^{i\theta}$. For clarity, let us express this as a matrix acting on a vector, by using the "stack" operation. We then have

$$\sum_{\mathcal{G}\in\mathbb{G}} \Omega(\mathcal{G})[\mathcal{G}^* \otimes \mathcal{G}] \operatorname{vec}(\mathcal{V}) = e^{i\theta} \operatorname{vec}(\mathcal{V}) \tag{6.32}$$

where the matrix acting on $\operatorname{vec}(\mathcal{V})$ is a convex sum of unitaries, or a weighted sum and integral over unitaries with the total "weighting" integrating to 1. Because unitaries preserve the length of vectors, this can only hold if $[\mathcal{G} \otimes \mathcal{G}] \operatorname{vec}(\mathcal{V}) = e^{i\theta} \operatorname{vec}(\mathcal{V})$ for all $\mathcal{G} \in \mathbb{G}$ for which $\Omega(\mathcal{G}) > 0$, and this is all $\mathcal{G} \in \mathbb{G}$ by the conditions of the proposition. Hence, returning to the unstacked representation, we have that

$$\mathcal{G}\mathcal{V}\mathcal{G}^{-1} = e^{i\theta}\mathcal{V} \tag{6.33}$$

for all $\mathcal{G} \in \mathbb{G}$. But we have shown above that, if $\mathbb{S} = \mathbb{S}_{\text{dep}}$, this holds if and only if $\mathcal{V} \in \mathbb{S}_{\text{dep}}$ and $e^{i\theta} = 1$. As such, any eigen-operator of $T_{\mathbb{G},\Omega}$ outside of the depolarizing subspace has an eigenvalue with absolute value strictly less than 1 in modulus. Proposition 1 then implies that $(\mathbb{G}, \Omega)$ is a sequence-asymptotic unitary 2-design. □

Proposition 2 implies the following theorem:

Theorem 1. *Let $\mathbb{G} \subseteq \mathbb{U}$ be a unitary gate set that either generates a group or generates a dense subset of a group, $\mathbb{C}$, and let $\Omega : \mathbb{G} \to [0, 1]$ be a probability distribution over $\mathbb{G}$ with support on all of $\mathbb{G}$. Let $\mathbb{G}^k = \{g_1 \dots g_k \mid g_i \in \mathbb{G}\}$, which generates a group, or a dense subset of a group, $\mathbb{C}^k$. Then $(\mathbb{G}, \Omega)$ is a sequence-asymptotic unitary 2-design if and only if, for all k, $\mathbb{C}^k$ is a unitary 2-design.*

Proof. First we show that if each $\mathbb{C}^k$ is a unitary 2-design, then $\mathbb{G}$ is a sequence asymptotic 2-design. If $\mathbb{G}$ generates a unitary 2-design, then the depolarizing subspace is a subspace of the space preserved up to a phase by $\mathbb{G}$. Suppose there is some element in the preserved space that is not depolarizing. There is a k-string (a length k sequence of $G \in \mathbb{G}$) that brings this phase arbitrarily close to one. Then, in the group generated by these k-strings, this element is also preserved by $\mathbb{C}^k$ with a phase arbitrarily close to 1. However, we also have that $\mathbb{C}^k$ is a unitary 2-design and thus element must also belong to the depolarizing subspace, and thus we have a contradiction.

Next, we prove that if there is some $\mathbb{C}^k$ that is not a unitary 2-design, then $\mathbb{G}$ is not a sequence asymptotic 2-design. Let $\mathbb{S}_{\mathbb{G}^k}$ and $\mathbb{S}_{\mathbb{C}^k}$ denote the space of all superoperators preserved under conjugation up to a phase by all elements of $\mathbb{G}^k$ and $\mathbb{C}^k$, respectively. Because $\mathbb{G}^k$ generates $\mathbb{C}^k$, or generates a dense subset of $\mathbb{C}^k$, then $\mathbb{S}_{\mathbb{G}^k} = \mathbb{S}_{\mathbb{C}^k}$. If $\mathbb{C}^k$ is not a unitary 2-design then since $\mathbb{S}_{\mathbb{C}^k} = \mathbb{S}_{\mathbb{G}^k}$, by Proposition 2 the elements of $\mathbb{S}_{\mathbb{G}^k}$ preserve some superoperator up to a phase that is not a depolarizing channel and thus $\mathbb{S}_{\mathbb{G}^k} \neq \mathbb{S}_{\text{dep}}$. But if $\mathbb{S}_{\mathbb{G}^k} \neq \mathbb{S}_{\text{dep}}$, then $\mathbb{S}_{\mathbb{C}^k} \neq \mathbb{S}_{\text{dep}}$. Therefore, $(\mathbb{G}^k, \Omega^k)$ is not a sequence-asymptotic unitary 2-design, and hence by Lemma 1, $(\mathbb{G}, \Omega)$ is not a sequence-asymptotic unitary 2-design. □

Theorem 1 characterizes the class of all gate sets, and probability distributions over those gate sets, that form sequence-asymptotic unitary 2-designs. Perhaps surprisingly, a necessary and sufficient condition is that the generated group by $\mathbb{G}^k$ is a unitary 2-design for every k. As an example of a consequence of this theorem, consider the gate set $\mathbb{G} = \{\sqrt{Z}H, Z\sqrt{Z}H\}$ for which both elements have a period of 3. $\mathbb{G}$ generates a subgroup of the Clifford group that is a unitary 2-design. However, $\mathbb{G}^3 = \{I, X, Y, Z\}$ is the Pauli group, which does not form a 2-design. Hence, $\mathbb{G}$ does not induce a sequence asymptotic 2-design.

In any instance where a gate set $\mathbb{G}$ does generate a unitary 2-design but does not induce a sequence-asymptotic unitary 2-design (for any Ω), $\mathbb{G}$ can be altered so that it can induce a sequence-asymptotic unitary 2-design by adding a single element to $\mathbb{G}$. In particular, any gate set $\mathbb{G}$ that generates a unitary 2-design and contains the identity gate is a sequence-asymptotic unitary

2-design. This is a practically reasonable and easy-to-verify condition on a gate set.

6.7 A theory of Direct RB with gate-dependent errors

In this section we present two closely-related theories for direct RB under general Markovian errors that build on (and are of similar rigor to) the most accurate theories of Clifford group RB under general gate-dependent errors [168, 169, 209, 210]. These theories show that the direct RB decay is an exponential ($S_d \approx A+Bp^d$) for sufficiently small gate errors, and that the direct RB error rate (r_Ω) is equal to a gauge-invariant definition for the Ω-weighted average infidelity (ϵ_Ω) of the benchmarked gate set. This section begins with some preliminaries: in Section 6.7.1 we introduce some assumptions and notation that are used in this section but not elsewhere in the paper; and in Section 6.7.2 we review the concept of gauge, and gauge-invariant metrics. In Section 6.7.3 we introduce the first of the two closely-related theories for direct RB that we present in this section. This theory—which we refer to as the **R-matrix theory** for direct RB—combines the regular representation of the group $\mathbb{C}$ with the superoperators for the imperfect gates to construct a matrix R that can be used to exactly compute the direct RB decay (S_d).

In Sections 6.7.4-6.7.7 we then present our second theory for direct RB with general gate-dependent errors—which we refer to as the **L-matrix theory** for direct RB—that enables us to show that $r_\Omega \approx \epsilon_\Omega$. This theory combines the superoperator representation of the group $\mathbb{C}$ with the superoperators for the imperfect gates to construct a matrix L that is an imperfect version of the twirling superchannel that appears in our theory of sequence-asymptotic unitary 2-designs (Section 6.6). In Section 6.7.4 we show how S_d can be approximated by a function of the d^{th} power of this L matrix. In Section 6.7.5 we then show how S_d can be expanded into a linear combination of many exponentials, using the spectral decomposition of L. In Section 6.7.6 we use the properties of sequence-asymptotic unitary 2-designs (Section 6.6) to show that all but two terms in this linear combination are negligable, i.e., $S_d \approx A + B\gamma^d$ where γ is an eigenvalue of L that is close to 1. This implies that direct RB approximately measures $r_\gamma = (4^n - 1)(1 - \gamma)/4^n$, i.e., $r_\Omega \approx r_\gamma$. In Section 6.7.7 we conclude our theory, by building on a proof of Wallman's [169] to show that r_γ is

equal to a gauge-invariant definition for the Ω-weighted average infidelity (ϵ_Ω) of the benchmarked gate set. Finally, in Section 6.7.8 we use the exact R-matrix theory for direct RB, and numerical simulations, to validate our approximate L-matrix theory for direct RB—by demonstrating that $|(r_\Omega - r_\gamma)/r_\gamma|$ is small for a selection of randomly generated error models on a single-qubit gate set. The theories presented in this section have a close relationship to the Fourier transforms over a group, as is the case for Clifford group RB [209].

6.7.1 Assumptions and notation

Here we state assumptions and notation that are used throughout this section but that do not necessarily apply elsewhere in the paper. We will assume that the state preparation and measurement components of direct RB circuits contain only gates from the benchmarked gate set ($\mathbb{G}$). This is without loss of generality, as the sampling distribution Ω need not have support on all of $\mathbb{G}$. We also assume (1) unconditional compilation of the initial group element and final inversion group element, and (2) no randomization of the output bit string. These two assumptions mean that the state preparation circuit implements a random element of $\mathbb{C}$, and the measurement preparation subcircuit implements the unique element in $\mathbb{C}$ that inverts the preceding circuit (as in Clifford group RB). These assumptions are not necessary for much of theory in this section, but they greatly simplify notation.

Unlike elsewhere in this paper, we will use $G \in \mathbb{G}$ and $C \in \mathbb{C}$ to represent group elements [elsewhere we use, e.g., $U(G)$] rather than instructions to implement logic operations. Throughout this paper, we denote the superoperator representing the imperfect [perfect] implementation of $G \in \mathbb{G}$ by $\tilde{\mathcal{G}}(G)$ [$\mathcal{G}(G)$]. Similarly, in this section we will denote the superoperator representing the imperfect [perfect] implementation of $C \in \mathbb{C}$ (which are used in the state preparation and measurement stages of direct RB circuits) by $\tilde{\mathcal{C}}(G)$ [$\mathcal{C}(G)$]. Note that $\mathcal{C}(\cdot)$ is a representation of the group $\mathbb{C}$. We will assume a low-error gate set, so that each $\tilde{\mathcal{G}}(G)$ is a small deviation from $\mathcal{G}(G)$ (i.e., $\tilde{\mathcal{G}}(G) \approx \mathcal{G}(G)$) for all $G \in \mathbb{G}$ and similarly $\tilde{\mathcal{C}}(C) \approx \mathcal{C}(C)$ for all $C \in \mathbb{C}$. We will quantify this assumption when used.

6.7.2 Background: gauge-invariant metrics

In this section we will show that the direct RB error rate (r_Ω) is related to a gauge-invariant version of the average infidelity of an n-qubit gate sampled from Ω (ϵ_Ω). We will therefore first review the concept of gauge in gate sets, as well as gauge-variant and gauge-invariant properties of a gate set. The sets of all as-implemented and ideal operations in direct RB circuits can be represented by

$$\mathbb{G}_{\text{all}} = \{\langle\langle x|\}_{x\in\mathbb{B}_n} \cup \{\mathcal{G}(G)\}_{G\in\mathbb{G}} \cup \{\mathcal{C}(C)\}_{C\in\mathbb{C}} \cup \{|0^n\rangle\rangle\}\,, \tag{6.34}$$

$$\tilde{\mathbb{G}}_{\text{all}} = \{\langle\langle E_x|\}_{x\in\mathbb{B}_n} \cup \{\tilde{\mathcal{G}}(G)\}_{G\in\mathbb{G}} \cup \{\tilde{\mathcal{C}}(C)\}_{C\in\mathbb{C}} \cup \{|\rho\rangle\rangle\}\,, \tag{6.35}$$

respectively, where $\mathbb{B}_n$ is the set of all length n bit strings, E_x is a measurement effect representing the x measurement outcome (so $\langle\langle E_x| \approx \langle\langle x|$), and ρ represents imperfect state initialization (so $|\rho\rangle\rangle \approx |0^n\rangle\rangle$). (Furthermore, due to our assumption in this section that the target bit string is always the all-zeros bit string, we only need the $x = 0^n$ effect).

All ideal and actual outcomes of circuits in direct RB can be computed from $\mathbb{G}_{\text{all}}$ and $\tilde{\mathbb{G}}_{\text{all}}$, respectively. However, the predictions of these models are unchanged under a gauge transformation—all operation sets of the form

$$\tilde{\mathbb{G}}_{\text{all}}(\mathcal{M}) = \left\{\langle\langle E_x|\mathcal{M}^{-1}\right\}_{x\in\mathbb{B}_n} \cup \left\{\mathcal{M}\tilde{\mathcal{G}}_i\mathcal{M}^{-1}\right\} \cup \{\mathcal{M}|\rho\rangle\rangle\}\,, \tag{6.36}$$

make identical predictions for the outcome distributions of all circuits, where $\mathcal{M}$ is any invertible linear map, called a gauge transformation [148, 149, 231, 232]. If f is a property of $\tilde{\mathbb{G}}_{\text{all}}$, it is only experimentally observable if

$$f\left(\tilde{\mathbb{G}}_{\text{all}}\right) = f\left(\tilde{\mathbb{G}}_{\text{all}}(\mathcal{M})\right), \tag{6.37}$$

for all $\mathcal{M}$, i.e., f must be gauge-invariant. Many metrics of error for a gate or set of gates that are defined as functions of process matrices are not gauge-invariant [148, 149, 168, 231, 232]. This includes the [in]fidelity of a superoperator, so ϵ_Ω is not gauge-invariant. This implies that an RB error rate (which is observable) cannot be equal to the standard, gauge-variant definition of the

average infidelity of a gate set [168], i.e., to connect RB error rates to infidelities we need a gauge-invariant definition for infidelity [168, 169, 210].

6.7.3 An exact theory of direct randomized benchmarking

Here we present a theory for direct RB that is exact, generalizing a theory for Clifford group RB presented in [168]. This theory, which we call the R-matrix theory for direct RB, provides an exact formula for the average success probability (S_d), as a function of an imperfect gate set $\tilde{\mathbb{G}}$ and a sampling distribution Ω. This formula is efficient in d (but not n) to compute. We will find this theory useful for verifying the accuracy of the approximate theory (that shows that $r_\Omega \approx \epsilon_\Omega$) introduced later in this section. Our exact theory requires one additional assumptions: we require that $\mathbb{G}$ generates a finite group $\mathbb{C}$.

Our theory starts from the formula for S_d that follows directly from the definition of S_d and the direct RB circuits. Specifically,

$$S_d = \langle\langle E_0 | \tilde{\mathcal{S}}_d | \rho \rangle\rangle, \tag{6.38}$$

where

$$\tilde{\mathcal{S}}_d = \sum_{G_d \in \mathbb{G}} \cdots \sum_{G_1 \in \mathbb{G}} \sum_{C \in \mathbb{C}} [\Omega(G_d) \cdots \Omega(G_1) \Pi(C) \ \times \ \tilde{\mathcal{C}}(C^{-1} G_1^{-1} \cdots G_d^{-1}) \tilde{\mathcal{G}}(G_d) \cdots \tilde{\mathcal{G}}(G_1) \tilde{\mathcal{C}}(C)], \tag{6.39}$$

where (as above) Π is the uniform distribution. To expresses $\tilde{\mathcal{S}}_d$ (and so S_d) in an efficient-in-d form, we use the regular representation $[R(\cdot)]$ of the group $\mathbb{C}$. To specify these matrices explicitly, we index the elements of $\mathbb{C}$ by $i = 1, 2, \ldots, |\mathbb{C}|$, with $i = 1$ corresponding to the identity group element. Then $R(G)$ is the $\mathbb{C} \times \mathbb{C}$ matrix with elements given by $R(G)_{jk} = 1$ if $GC_k = C_j$ and $R(G)_{jk} = 0$ otherwise. So, $R(G)$ is a permutation matrix in which the i^{th} row encodes the product of G and the i^{th} group element. We now use the regular representation to define a matrix $[R_{\mathbb{G},\tilde{\mathbb{G}},\Omega}]$ that can be used to model sequences of gates from $\tilde{\mathbb{G}}$ that are independent samples from Ω. We

define $R_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$ by

$$R_{\mathbb{G},\tilde{\mathbb{G}},\Omega} = \sum_{G\in\mathbb{G}} \Omega(G) R(G) \otimes \tilde{\mathcal{G}}(G). \tag{6.40}$$

We now express $\tilde{\mathcal{S}}_d$ in terms of R matrices, and to do so we will need the vector

$$\boldsymbol{v} = (1, 0, 0, \dots)^T \otimes I, \tag{6.41}$$

where I is the $4^n \times 4^n$ dimensional identity superoperator (and the first entry of $\boldsymbol{v}$ corresponds to the identity group element). We can express $\tilde{\mathcal{S}}_d$ in terms of R matrices and $\boldsymbol{v}$ as follows:

$$\tilde{\mathcal{S}}_d = \boldsymbol{v}^T R_{\mathbb{C},\tilde{\mathbb{C}},\Pi} R^d_{\mathbb{G},\tilde{\mathbb{G}},\Omega} R_{\mathbb{C},\tilde{\mathbb{C}},\Pi} \boldsymbol{v}, \tag{6.42}$$

noting that $R_{\mathbb{C},\tilde{\mathbb{C}},\Pi}$ represents the application of a uniformly random element from $\tilde{\mathbb{C}}$. To understand why this equation holds, consider

$$\boldsymbol{w} = R_{\mathbb{C},\tilde{\mathbb{C}},\Pi} R^d_{\mathbb{G},\tilde{\mathbb{G}},\Omega} R_{\mathbb{C},\tilde{\mathbb{C}},\Pi} \boldsymbol{v}. \tag{6.43}$$

This is a vector of superoperators where the i^{th} element of $\boldsymbol{w}$ consists of an average over all sequences of superoperators consisting of (1) a uniformly random element of $\tilde{\mathbb{C}}$, (2) d elements of $\tilde{\mathbb{G}}$ sampled from Ω, and (3) a uniformly random element of $\tilde{\mathbb{C}}$, **conditioned** on the overall ideal (error-free) action of the sequence being equal to the group element $\mathcal{C}(C_i)$. To average over all direct RB sequences of length d we simply need to condition on the overall action of the sequence being the identity—which is $\boldsymbol{v}^T \boldsymbol{w}$.

To obtain our final result—a formula for S_d—we substitute Eq. (6.42) into Eq. (6.38) to obtain

$$S_d = \langle\langle E_0 | \boldsymbol{v}^T R_{\mathbb{C},\tilde{\mathbb{C}},\Pi} R^d_{\mathbb{G},\tilde{\mathbb{G}},\Omega} R_{\mathbb{C},\tilde{\mathbb{C}},\Pi} \boldsymbol{v} \, | \rho \rangle\rangle. \tag{6.44}$$

This formula for S_d is efficient in d to calculate (simply via multiplication of R matrices, or via the spectral decomposition of the two R matrices). The R matrices grow quickly with both the size of the generated group (i.e., with $|\mathbb{C}|$) and n. However, the R matrices are sufficiently small when $\mathbb{C}$ is the single-qubit Clifford group (they are 96×96 matrices) to use this theory to numerical evaluate S_d (without resorting to sampling) for any single-qubit gate set $\mathbb{G}$ that generates the Clifford group.

As such, this theory is useful for numerically studying direct RB in the single-qubit setting—and this will allow us to validate our theory that shows that $r_\Omega \approx \epsilon_\Omega$ (see Section 6.7.8).

Equation (6.44) can be rewritten in terms of the d^{th} power of the eigenvalues of $R_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$. However, this does not immediately imply that $S_d \approx A + Bp^d$ (each R matrix has $4^n \times |\mathbb{C}|$ eigenvalues).

6.7.4 Modelling the direct RB decay with twirling superchannels

In this and the subsequent subsections, we present our **L-matrix theory** for direct RB. In this theory, we will show how the direct RB decay, and error rate, can be approximately expressed in terms of the following matrix

$$L_{\mathbb{G},\tilde{\mathbb{G}},\Omega} = \sum_{G\in\mathbb{G}} \Omega(G)\mathcal{G}(G) \otimes \tilde{\mathcal{G}}(G). \tag{6.45}$$

Like the R matrix [Eq. (6.40)], this L matrix consists of summing over a tensor product of (1) a representation of the group $\mathbb{C}$ generated by $\mathbb{C}$ and (2) the as-implemented superoperators. In the case of the R matrix, this representation of the group was the regular representation $[R(\cdot)]$, and in the case of the L matrix it is the standard superoperator representation $[\mathcal{G}(\cdot)]$ of the group.

To derive our theory, we first show how the direct RB success probability (S_d) can be approximately expressed as a function of twirling superchannels Q whereby $\text{mat}[Q] = L$. In subsequent subsections, this will enable us to show that the direct RB decay is approximately exponential, and that $r_\Omega \approx \epsilon_\Omega$. Our theory will use **error maps** for the as-implemented group elements $\tilde{\mathbb{C}}$. Let

$$\Lambda(C) = \mathcal{C}(C)^\dagger \tilde{\mathcal{C}}(C), \tag{6.46}$$

for $C \in \mathbb{C}$, which is the "pre-gate" error map for the superoperator $\tilde{\mathcal{C}}(C)$. We also use the average error map ($\bar{\Lambda}$), defined by

$$\bar{\Lambda} = \sum_{C\in\mathbb{C}} \Pi(C)\Lambda(C). \tag{6.47}$$

The first step to deriving our L-matrix theory for direct RB is to approximate the imperfect initial group element using a gate-independent error channel. To do so, we first note that Eq. (6.39)

can be rewritten as

$$\tilde{\mathcal{S}}_d = \sum_{C\in\mathbb{C}} \sum_{G_d\in\mathbb{G}} \cdots \sum_{G_1\in\mathbb{G}} [\Pi(C)\Omega(G_d)\cdots\Omega(G_1) \;\times \\ \tilde{\mathcal{C}}(C)\tilde{\mathcal{G}}(G_d)\cdots\tilde{\mathcal{G}}(G_1)\tilde{\mathcal{C}}(\{CG_d\cdots G_1\}^{-1})], \quad (6.48)$$

i.e., we have rewritten the first group element in terms of the G_i and the inversion group element (which is now denoted by C). Now we have that

$$\tilde{\mathcal{C}}(\{CG_d\cdots G_1\}^{-1}) = \\ \mathcal{G}^\dagger(G_1)\cdots\mathcal{G}^\dagger(G_d)\mathcal{C}^\dagger(C)\Lambda(\{CG_d\cdots G_1\}^{-1}), \quad (6.49)$$

and by substituting this into Eq. (6.48), we obtain

$$\tilde{\mathcal{S}}_d = \sum_{C\in\mathbb{C}} \sum_{G_d\in\mathbb{G}} \cdots \sum_{G_1\in\mathbb{G}} [\Pi(C)\Omega(G_d)\cdots\Omega(G_1) \;\times \\ \tilde{\mathcal{C}}(C)\tilde{\mathcal{G}}(G_d)\cdots\tilde{\mathcal{G}}(G_1) \;\times \\ \mathcal{G}^\dagger(G_1)\cdots\mathcal{G}^\dagger(G_d)\mathcal{C}^\dagger(C)\Lambda((\{G_d\cdots G_1\}^{-1})]. \quad (6.50)$$

By replacing the error maps in this equation (i.e., the error maps for each initial group element) with their average, we obtain

$$\tilde{\mathcal{S}}_d = \sum_{C\in\mathbb{C}} \sum_{G_d\in\mathbb{G}} \cdots \sum_{G_1\in\mathbb{G}} [\Pi(C)\Omega(G_d)\cdots\Omega(G_1) \;\times \\ \tilde{\mathcal{C}}(C)\tilde{\mathcal{G}}(G_d)\cdots\tilde{\mathcal{G}}(G_1) \;\times \\ \mathcal{G}^\dagger(G_1)\cdots\mathcal{G}^\dagger(G_d)\mathcal{C}^\dagger(C)]\bar{\Lambda} + \Delta_d, \quad (6.51)$$

where Δ_d absorbs the approximation error. This approximation breaks the dependencies, enabling independent averaging, as we can rearrange Eq. (6.51) to

$$\tilde{\mathcal{S}}_d = \Bigg\{ \sum_{C\in\mathbb{C}} \Pi(C)\tilde{\mathcal{C}}(C)\Bigg\{ \sum_{G_d\in\mathbb{G}} \Omega(G_d)\tilde{\mathcal{G}}(G_d)\cdots \\ \Big\{ \sum_{G_1\in\mathbb{G}} \Omega(G_1)\tilde{\mathcal{G}}(G_1)\mathcal{G}^\dagger(G_1)\Big\}\cdots\mathcal{G}^\dagger(G_d)\Bigg\}\mathcal{C}^\dagger(C)\Bigg\}\bar{\Lambda} + \Delta_d. \quad (6.52)$$

This can be more concisely expressed using two superchannels:

$$\tilde{\mathcal{S}}_d = \left(Q_{\mathbb{C},\tilde{\mathbb{C}},\Pi} \circ Q^d_{\mathbb{G},\tilde{\mathbb{G}},\Omega}\right)[I]\bar{\Lambda} + \Delta_d, \tag{6.53}$$

where I is the identity superoperator, and $Q_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$ and $Q_{\mathbb{C},\tilde{\mathbb{C}},\Pi}$ are imperfect twirling superchannels, defined by

$$Q_{\mathbb{G},\tilde{\mathbb{G}},\Omega}[\mathcal{E}] = \sum_{G\in\mathbb{G}} \Omega(G)\tilde{\mathcal{G}}(G)\mathcal{E}\mathcal{G}^\dagger(G), \tag{6.54}$$

$$Q_{\mathbb{C},\tilde{\mathbb{C}},\Pi}[\mathcal{E}] = \sum_{C\in\mathbb{C}} \Pi(C)\tilde{\mathcal{C}}(C)\mathcal{E}\mathcal{C}^\dagger(C). \tag{6.55}$$

By substituting Eq. (6.53) into Eq. (6.38) we obtain

$$S_d = \langle\langle E_0| \left(Q_{\mathbb{C},\tilde{\mathbb{C}},\Pi} \circ Q^d_{\mathbb{G},\tilde{\mathbb{G}},\Omega}\right)[I]|\,\rho'\rangle\rangle + \delta_d, \tag{6.56}$$

where $|\,\rho'\rangle\rangle = \bar{\Lambda}|\,\rho\rangle\rangle$ and

$$\delta_d = \langle\langle E_0|\,\Delta_d|\,\rho\rangle\rangle. \tag{6.57}$$

Equation (6.56) represents the average direct RB success probability in terms of imperfect twirling superchannels, which builds on similar results for Clifford group RB [168, 169, 233].

Later we will apply the theory of sequence-asymptotic unitary 2-designs (Section 6.6) to understand S_d, and to do so we will find it useful to replace the Q superchannels in Eq. (6.56) with the closely related L matrices [see Eq. (7.14)]. Consider

$$\mathrm{vec}(Q_{\mathbb{C},\tilde{\mathbb{C}},\Pi} \circ Q^d_{\mathbb{G},\tilde{\mathbb{G}},\Omega}[I]) = \mathrm{mat}(Q_{\mathbb{C},\tilde{\mathbb{C}},\Pi})\mathrm{mat}(Q_{\mathbb{G},\tilde{\mathbb{G}},\Omega})^d|I), \tag{6.58}$$

where, as introduced in Section 6.6, vec[·] denotes the "stacking" map, mat[·] turns a superchannel into a matrix and $|\cdot)$ turns a superoperator into a vector. So, e.g., we find that

$$\mathrm{mat}[Q_{\mathbb{G},\tilde{\mathbb{G}},\Omega}] = \sum_{G\in\mathbb{G}} \Omega(G)\mathcal{G}(G)\otimes\tilde{\mathcal{G}}(G) = L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}, \tag{6.59}$$

where $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$ is the matrix introduced in Eq. (7.14) [here we have assumed a Hermitian basis for superoperators, so that $\mathcal{G}(G)$'s element are real numbers]. We can therefore rewrite Eq. (6.56) in terms of two L matrices:

$$S_d = \langle\langle E_0|\,\mathrm{unvec}\left[L_{\mathbb{C},\tilde{\mathbb{C}},\Pi}L^d_{\mathbb{G},\tilde{\mathbb{G}},\Omega}|I)\right]|\,\rho'\rangle\rangle + \delta_d. \tag{6.60}$$

Equation (7.12) contains an approximation error term (δ_d) and we now bound this term. Using basic properties of CPTP maps and the diamond norm it may be shown that

$$|\delta_d| \leqslant \|\Delta_d\|_\diamond \leqslant \sum_{C\in\mathbb{C}} \Pi(C) \left\|\bar{\Lambda} - \Lambda(C)\right\|_\diamond \equiv \Delta, \tag{6.61}$$

where $\|\cdot\|_\diamond$ is the diamond norm [d]. Error maps are not gauge-invariant, so the size of δ_d (and the size of our upper-bound on $|\delta_d|$) depends on the representation of $\tilde{\mathbb{G}}_{\text{all}}$. Even for low-error gates—here meaning that every $\Lambda(C)$ is close to the identity in some representation—Δ and δ_d can always be made large by choosing a "bad" gauge in which to express $\tilde{\mathbb{G}}_{\text{all}}$. But our bound holds for any CPTP representation, and so $|\delta_d| \leqslant \Delta_{\min}$ with $\Delta_{\min}$ the minimum of Δ over all CPTP representations of $\tilde{\mathbb{G}}_{\text{all}}$. In the following, we will assume we are using a representation in which δ_d is small, which always exists with sufficiently low-error gates. Finally, we note that it is possible to improve our bound on δ_d, using an analysis similar to Wallman's [169] treatment of Clifford group RB, which bounds a quantity that is closely related to δ_d from above by a function that decays exponentially in d.

6.7.5 Modelling the direct RB decay using the spectrum of twirling superchannels

We have shown that the direct RB average success probability (S_d) can be approximated by a function of the d^{th} power of $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$ [see Eq. (7.12)]. This implies that S_d can be approximated using the spectral decomposition of $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$. We now derive the functional form of this spectral decomposition, which in the subsequent subsection we will use to show that $S_d \approx A + Bp^d$. Let $\tilde{\gamma}_i$ for $i = 1, 2, \ldots, 16^n$ denote the eigenvalues of $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$ ordered by descending absolute value, i.e., $|\tilde{\gamma}_i| \geqslant |\tilde{\gamma}_{i+1}|$ for all i. Then Eq. (7.12) implies that S_d may be written as

$$S_d = \tilde{\omega}_1\tilde{\gamma}_1^d + \tilde{\omega}_2\tilde{\gamma}_2^d + \cdots + \tilde{\omega}_{16^n}\tilde{\gamma}_{16^n}^d + \delta_d, \tag{6.62}$$

[d] An explicit derivation that immediately implies this relation is given in [168] in the context of an equivalent relation in Clifford group RB (see the supplemental material in [168], and also see [155, 233] for similar work). Note that this upper-bound is d independent, unlike the similar bound for approximating Clifford group RB decay curves in [155] that grow with d and which is therefore not able to provide useful guarantees on S_d [168, 169].

where the $\tilde{\omega}_k$ are d-independent parameters. Note that the eigenvalues of an L matrix are gauge-invariant—because gauge transformations act on a L matrix by matrix conjugation, with a matrix of the form $I \otimes \mathcal{M}$—so the $\tilde{\gamma}_i$ are gauge-invariant.

We now derive a formula for $\tilde{\omega}_k$. It is convenient to rewrite Eq. (7.12) as

$$S_d = \langle\langle E_0 | \tilde{S}'_d | \rho' \rangle\rangle + \delta_d, \tag{6.63}$$

where

$$\tilde{\mathcal{S}}'_d = \text{unvec}\left[L_{\mathbb{C},\tilde{\mathbb{C}},\Pi} L^d_{\mathbb{G},\tilde{\mathbb{G}},\Omega} | I)\right]. \tag{6.64}$$

We now write $\tilde{\mathcal{S}}'_d$ in terms of the eigenvectors and eigenvalues of the two L matrices, $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$ and $L_{\mathbb{C},\tilde{\mathbb{C}},\Pi}$. Let $\tilde{\eta}_i$ for $i = 1, 2, \ldots, 16^n$ denote the eigenvalues of $L_{\mathbb{C},\tilde{\mathbb{C}},\Pi}$ ordered by descending absolute value, i.e., $|\tilde{\eta}_i| \geqslant |\tilde{\eta}_{i+1}|$ for all i (noting that $\tilde{\gamma}_i$ denote the ordered eigenvalues of $L_{\mathbb{G}\tilde{\mathbb{G}},\Omega}$). Moreover, let $\mathcal{L}_{\mathbb{C},\tilde{\mathbb{C}},i}$ and $\mathcal{R}_{\mathbb{C},\tilde{\mathbb{C}},i}$ ($\mathcal{L}_{\mathbb{G},\tilde{\mathbb{G}},i}$ and $\mathcal{R}_{\mathbb{G},\tilde{\mathbb{G}},i}$) be superoperators that, when stacked, are mutually orthonormal left and right eigenvectors of the matrix $L_{\mathbb{C},\tilde{\mathbb{C}},\Pi}$ (the matrix $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$) with eigenvalue $\tilde{\eta}_i$ (eigenvalue $\tilde{\gamma}_i$). That is

$$L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}\Big|\mathcal{R}_{\mathbb{G},\tilde{\mathbb{G}},i}\Big) = \tilde{\gamma}_i\Big|\mathcal{R}_{\mathbb{G},\tilde{\mathbb{G}},i}\Big), \tag{6.65}$$

$$(\mathcal{L}_{\mathbb{G},\tilde{\mathbb{G}},i}\Big|L_{\mathbb{G},\tilde{\mathbb{G}},\Omega} = (\mathcal{L}_{\mathbb{G},\tilde{\mathbb{G}},i}\Big|\tilde{\gamma}_i, \tag{6.66}$$

$$L_{\mathbb{C},\tilde{\mathbb{C}},\Pi}\Big|\mathcal{R}_{\mathbb{C},\tilde{\mathbb{C}},i}\Big) = \tilde{\eta}_i\Big|\mathcal{R}_{\mathbb{C},\tilde{\mathbb{C}},i}\Big), \tag{6.67}$$

$$(\mathcal{L}_{\mathbb{C},\tilde{\mathbb{C}},i}\Big|L_{\mathbb{C},\tilde{\mathbb{C}},\Pi} = (\mathcal{L}_{\mathbb{C},\tilde{\mathbb{C}},i}\Big|\tilde{\eta}_i, \tag{6.68}$$

with

$$\Big(\mathcal{L}_{\mathbb{G},\tilde{\mathbb{G}},i} \mid \mathcal{R}_{\mathbb{G},\tilde{\mathbb{G}},j}\Big) = \delta_{ij}, \quad \Big(\mathcal{L}_{\mathbb{C},\tilde{\mathbb{C}},i} \mid \mathcal{R}_{\mathbb{C},\tilde{\mathbb{C}},j}\Big) = \delta_{ij}. \tag{6.69}$$

Writing the two L matrices in terms of their eigenvalues and their left and right eigenvectors, and substituting this decomposition of the L matrices into Eq. (6.64), we obtain

$$\tilde{\mathcal{S}}'_d = \sum_{jk} \tilde{\eta}_j \tilde{\gamma}_k^d \text{unvec}\left[|\mathcal{R}_{\mathbb{C},\tilde{\mathbb{C}},j})(\mathcal{L}_{\tilde{\mathbb{C}},\mathbb{C},j}|\mathcal{R}_{\mathbb{G},\tilde{\mathbb{G}},k})(\mathcal{L}_{\mathbb{G},\tilde{\mathbb{G}},k}||I)\right]. \tag{6.70}$$

By substituting Eq. (6.70) into Eq. (6.63) and comparing the resultant equation to Eq. (6.62), it follows that

$$\tilde{\omega}_k = \left\langle\left\langle E'_0\right| \mathcal{S}'_{d,k} |\rho\right\rangle\right\rangle \tag{6.71}$$

where

$$\mathcal{S}'_{d,k} = \sum_j \tilde{\eta}_j \text{ unvec } \left[|\mathcal{R}_{\mathbb{C},\tilde{\mathbb{C}},j})(\mathcal{L}_{\mathbb{C},\tilde{\mathbb{C}},j}|\mathcal{R}_{\mathbb{G},\tilde{\mathbb{G}},k})(\mathcal{L}_{\mathbb{G},\tilde{\mathbb{G}},k}||I) \right]. \tag{6.72}$$

This formula shows that we can quantify $\tilde{\omega}_k$ by studying the overlap in the eigenchannels of the two L matrices, as well as the spectrum of $L_{\mathbb{C},\tilde{\mathbb{C}},\Pi}$.

6.7.6 The direct RB decay is approximately exponential

We have now shown that the average success probability (S_d) in direct RB can be written as a linear combination of the d^{th} powers of the 16^n eigenvalues of the L matrix ($\tilde{\gamma}_k$), and we have derived formula for the coefficients in this sum ($\tilde{\omega}_k$). We now use the theory of sequence-asymptotic unitary 2-designs (Section 6.6), as well as the theory of ordinary unitary 2-designs, to show that $S_d \approx A + B\gamma^d$ where $\gamma \equiv \tilde{\gamma}_2$ is the second largest eigenvalue of $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$. Our theory relies on two properties of $\mathbb{G}$, Ω, and $\tilde{\mathbb{G}}$: (1) $(\mathbb{G}, \Omega)$ is a sequence-asymptotic unitary 2-design, and (2) the gates are low error, i.e., $\tilde{\mathbb{G}}$ is close to $\mathbb{G}$ and $\tilde{\mathbb{C}}$ is close to $\mathbb{C}$.

This second condition, above, implies that both the L matrices that appear in Eq. (6.64) [$L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$ and $L_{\mathbb{C},\tilde{\mathbb{C}},\Pi}$] will be small perturbations on the L matrices for perfect, error-free gate sets [i.e., $L_{\mathbb{G},\mathbb{G},\Omega}$ and $L_{\mathbb{C},\mathbb{C},\Pi}$]. The group generated by $\mathbb{G}$, i.e, $\mathbb{C}$, is a unitary 2-design. Therefore

$$L_{\mathbb{C},\mathbb{C},\Pi} = \sum_{C\in\mathbb{C}} \Pi(C)\mathcal{C}(C) \otimes \mathcal{C}(C), \tag{6.73}$$

is a projection onto the space spanned by a completely depolarizing channel [$\mathcal{D}_0$] and the identity superoperator [I] (see Section 6.6). Letting η_i denote the ordered eigenvalues of $L_{\mathbb{C},\mathbb{C},\Pi}$, we have that (1) $\eta_1 = \eta_2 = 1$, with the two-dimensional eigen-space for this eigenvalue spanned by $\mathcal{D}_0$ and I, and (2) $\eta_i = 0$ for $i \geqslant 3$. Similarly, Propopsition 1 implies that, because $(\mathbb{G}, \Omega)$ is a sequence-asymptotic unitary 2-design,

$$L_{\mathbb{G},\mathbb{G},\Omega} = \sum_{G\in\mathbb{G}} \Omega(G)\mathcal{G}(G) \otimes \mathcal{G}(G), \tag{6.74}$$

has two unit eigenvalues, which also correspond to the two-dimensional depolarizing subspace, and all other eigenvalues have absolute value strictly less than one. So, letting γ_i denote the ordered eigenvalues of $L_{\mathbb{G},\mathbb{G},\Omega}$, we have that (1) γ_i $\gamma_1 = \gamma_2 = 1$, with the two-dimensional eigen-space for

this eigenvalue spanned by $\mathcal{D}_0$ and I, and (2) $|\gamma_i| \leqslant 1$ for all $i \geqslant 3$. Using Eqs. (6.71)-(6.72), this then implies that if $\tilde{\mathbb{G}} = \mathbb{G}$ and $\tilde{\mathbb{C}} = \mathbb{C}$, i.e., the gates are perfect, then $\tilde{\omega}_k = 0$ for $k \geqslant 3$.

Consider now the case of imperfect but low-error gates, i.e., $\tilde{\mathbb{G}} \approx \mathbb{G}$ and $\tilde{\mathbb{C}} \approx \mathbb{C}$. As long as the errors are small, the eigenvalues of $L_{\mathbb{G},\mathbb{G},\Omega}$ will be close to those of $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$. This is demonstrated with an example in Fig. 6.4. Furthermore, as long as the errors are small enough so that the gap between the 2^{nd} largest eigenvalue and the smaller eigenvalues of $L_{\mathbb{G},\mathbb{G},\Omega}$ does not close, the span of the eigenvectors of $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ is a small perturbation on the depolarizing subspace. Similarly, as long as the errors are small enough so that the gap between the 2^{nd} largest eigenvalue and the zero eigenvalues of $L_{\mathbb{C},\mathbb{C},\Pi}$ does not close, the span of the eigenvectors of $\tilde{\eta}_1$ and $\tilde{\eta}_2$ is a small perturbation on the depolarizing subspace. This then implies that

$$|\tilde{\omega}_k| \ll 1 \quad \forall k \geqslant 3. \tag{6.75}$$

Substituting this into Eq. (6.62), and by noting that 1 is always an eigenvalue of an L matrix for trace-preserving superoperators (we are assuming CPTP superoperators throughout), we find that

$$S_d = A + B\gamma^d + \delta_d + \delta'_d \tag{6.76}$$

where $A = \omega_1$, $B = \omega_2$, $\gamma = \tilde{\gamma}_2$ and $\delta'_d = \sum_{k\geqslant 3}^{16^n} \tilde{\gamma}_k^d \omega_k$ consists of a sum of $16^n - 2$ small terms. Therefore, for small n, we have that

$$S_d \approx A + B\gamma^d, \tag{6.77}$$

i.e., the direct RB average success probability decay is approximately an exponential (plus a constant).

Interestingly, note that the above theory does not show that $|\delta'_d|$ is small for $n \gg 1$. This is because we have only shown that δ'_d is the sum of $16^n - 1$ small terms. Furthermore, the above theory has limited applicability in the $n \gg 1$ regime for another reason: for any $\mathbb{G}$ consisting of gates constructed from parallel one- and two-qubit gates, the upper spectral gap in the $L_{\mathbb{G},\mathbb{G},\Omega}$ matrix must decrease in size as n increases. This is because the spectral gap is closely related to convergence to a unitary 2-design, and the number of layers of one- and two-qubit gates needed to

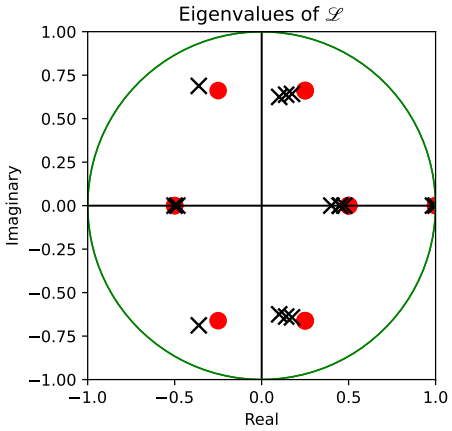


Figure 6.4: The eigenvalues of the $L_{\tilde{\mathbb{G}},\mathbb{G},\Omega}$ matrix for the single qubit gate set $\mathbb{G} = \{X(\frac{\pi}{2}), Y(\frac{\pi}{2})\}$ without errors [red circles], i.e., $\tilde{\mathbb{G}} = \mathbb{G}$, and with a small coherent over-rotation error on each gate [black crosses], i.e. $\tilde{\mathbb{G}} = \{X(\frac{\pi}{2} + \epsilon), Y(\frac{\pi}{2} + \epsilon)\}$, with $\epsilon = 0.1$. There are two unit eigenvalues for the error-free gate set, and all other eigenvalues have absolutely value strictly less than 1. Here the upper spectral gap—meaning the difference in the magnitudes of the second and third largest eigenvalues of L—is $\frac{1}{2}$ for the error-free gate set, and it is still close to $\frac{1}{2}$ for the imperfect gates.

implement a good approximation to unitary 2-design increases with n. Yet our less formal theory for direct RB—presented in Section 6.5—shows that direct RB **is** reliable in the $n \gg 1$ setting, i.e., we do have $S_d \approx A + Bp^d$ for some A, B and p, even when $n \gg 1$. It is an interesting open question as to how to obtain this result from the R or L matrix theories presented in this section.

6.7.7 The direct RB measures a version of infidelity

We have now shown that the direct RB success probability (S_d) is approximately given by $S_d \approx A + B\gamma^d$ where γ is the second largest eigenvalue of the superchannel $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$. The direct RB error rate (r_Ω) is defined to be $r_\Omega = (4^n - 1)(1 - p)/4^n$, where p is obtained by fitting S_d to $S_d = A + Bp^d$. The theory presented so far therefore implies that $r_\Omega \approx r_\gamma$ where

$$r_\gamma = (4^n - 1)(1 - \gamma)/4^n. \tag{6.78}$$

The final part of our theory is to show that r_γ is (exactly) equal to a gauge-invariant version of the mean infidelity of a gate sampled from Ω (ϵ_Ω). That is, we now show that direct RB measures the Ω-weighted average infidelity of the gate set $\tilde{\mathbb{G}}$ it is used to benchmark, in the same sense that Clifford group RB measures the mean infidelity of an implementation of the Clifford group [168, 169].

First we introduce a term for a gate set $\tilde{\mathbb{G}}$ that has sufficiently small errors for the upper spectral gap in the $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$ to not close (which is a condition that we required, above, to show that $S_d \approx A + B\gamma^d$).

Definition 4. *Let $(\mathbb{G}, \Omega)$ be a sequence-asymptotic unitary 2-design, and let $\tilde{\mathbb{G}}$ represent some implementation of $\mathbb{G}$. This implementation is $(\mathbb{G}, \Omega)$-benchmarking compatible if the upper spectral gap of $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$ does not close.*

Next, we present a rather technical proposition, that generalizes a result derived by Wallman [169] (see Theorem 2 therein) for Clifford group RB:

Proposition 3. *Let $(\mathbb{G}, \Omega)$ be a sequence-asymptotic unitary 2-design, and let $\tilde{\mathbb{G}}$ represent an implementation of $\mathbb{G}$ that is sufficiently low error to be $(\mathbb{G}, \Omega)$-benchmarking compatible. Let γ be the*

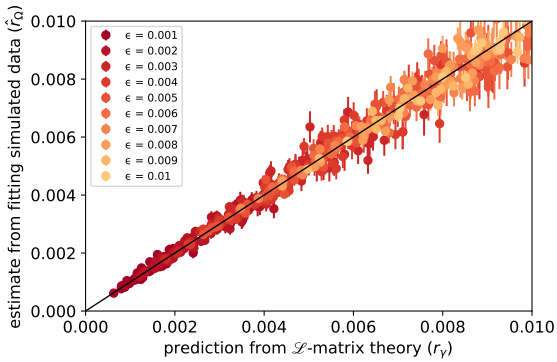


Figure 6.5: Verifying the accuracy of our L matrix theory for direct RB, using numerical simulations. We show the theory's prediction for the direct RB error rate (r_γ) versus an estimate $\hat{r}_\Omega$ of the true direct RB error rate (r_Ω) obtained by simulating direct RB. Each data point shows r_γ versus $\hat{r}_\Omega$ for a different randomly generated Markovian error models (see main text for error model details). We find that r_γ and $\hat{r}_\Omega$ are consistent, but note that small systematic deviations between r_γ and r_Ω would not be visible due to the finite sample error on $\hat{r}_\Omega$.

second largest eigenvalue of $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$ *by absolute value, and let* $\mathcal{E}_1$ *and* $\mathcal{E}_\gamma$ *be eigenoperators of* $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$ *with eigenvalues 1 and* γ*, respectively. Then* $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_\gamma$ *satisfies*

$$L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}(\mathcal{E}) = \mathcal{D}_\gamma \mathcal{E}, \tag{6.79}$$

where $\mathcal{D}_\gamma$ *is the depolarization channel with decay constant* γ*.*

Proof. See Appendix 6.9. □

We now use Proposition 3 to prove that $r_\gamma = \epsilon_\Omega(\tilde{\mathbb{G}}(\mathcal{E}), \mathbb{G})$. Here

$$\epsilon_\Omega(\tilde{\mathbb{G}}(\mathcal{E}), \mathbb{G}) = \sum_{G \in \mathbb{G}} \Omega(G)\epsilon(\mathcal{E}\tilde{\mathcal{G}}(G)\mathcal{E}^{-1}, \mathcal{G}(G)), \tag{6.80}$$

where $\mathcal{E}$ is as defined in Proposition 3. Equation (6.80) is the Ω-weighted average infidelity between the superoperators in $\mathbb{G}$ and $\tilde{\mathbb{G}}(\mathcal{E})$, where $\tilde{\mathbb{G}}(\mathcal{E})$ expresses the imperfect operations in a particular gauge [e]:

$$\tilde{\mathbb{G}}(\mathcal{E}) := \left\{ \mathcal{E}\tilde{\mathcal{G}}(G)\mathcal{E}^{-1} \mid G \in \mathbb{G} \right\}. \tag{6.81}$$

[e] Note that the $\mathcal{E}$ transformation is also applied to the SPAM whenever any probabilities are to be calculated, but for simplicity we have dropped the SPAM operations from the notation throughout this subsection.

Note that the superoperators in $\tilde{\mathbb{G}}(\mathcal{E})$ are not necessarily completely positive (this follows because it is true in the case of Clifford group RB [168, 169], and Clifford group RB is a special case of direct RB). Furthermore, note that $\tilde{\mathbb{G}}(\mathcal{E})$ is ill-defined if $\mathcal{E}$ is singular, and in any such case we take $\tilde{\mathbb{G}}(\mathcal{E})$ to instead be defined using a small perturbation on $\mathcal{E}$. In the following proposition, which is a generalization of a result for Clifford group RB shown in [168, 169], we prove that $r_\gamma = \epsilon_\Omega(\tilde{\mathbb{G}}(\mathcal{E}), \mathbb{G})$.

Proposition 4. *Let $\mathbb{G}, \Omega, \tilde{\mathbb{G}}, \gamma$ and $\mathcal{E}$ be as defined in Proposition 3. Then $r_\gamma = (d^2-1)(1-\gamma)/d^2$ satisfies*

$$r_\gamma = \epsilon_\Omega(\tilde{\mathbb{G}}(\mathcal{E}), \mathbb{G}). \tag{6.82}$$

Proof. From Proposition 3 we have that $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}(\mathcal{E}) = \mathcal{D}_\gamma \mathcal{E}$, and so $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}(\mathcal{E})\mathcal{E}^{-1} = \mathcal{D}_\gamma$. Taking the entanglement infidelity to the identity superoperator of both sides of this equality, and using the definition of $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$, we have that

$$\epsilon\left(\sum_{G\in\mathbb{G}} \Omega(G)\mathcal{G}^{-1}(G)\mathcal{E}\tilde{\mathcal{G}}(G)\mathcal{E}^{-1}, \mathbb{I}\right) = \epsilon(\mathcal{D}_\gamma, \mathbb{I}) \tag{6.83}$$

Now using the well-known relation [155]

$$\epsilon(\mathcal{D}_\lambda, \mathbb{I}) = \frac{(4^n-1)(1-\lambda)}{4^n} \tag{6.84}$$

and the definition of r_γ, it follows that $\epsilon(\mathcal{D}_\gamma, \mathbb{I}) = r_\gamma$. Using this equality, the linearity of ϵ, and the easily verified relation $\epsilon(B^{-1}A, \mathbb{I}) = \epsilon(A, B)$ for any superoperators A and B with B invertible, Eq. (6.83) implies that

$$\sum_{G\in\mathbb{G}} \Omega(\mathcal{G})\epsilon\left(\mathcal{E}\tilde{\mathcal{G}}(G)\mathcal{E}^{-1}, \mathcal{G}(G)\right) = r_\gamma \tag{6.85}$$

By noting that $\tilde{\mathbb{G}}(\mathcal{E}) = \{\mathcal{E}\tilde{\mathcal{G}}(G)\mathcal{E}^{-1} \mid G \in \mathbb{G}\}$, it follows immediately from the definition of ϵ_Ω, which is the average infidelity of a gate sampled from Ω, that the left hand side of this equation is the $\epsilon_\Omega(\tilde{\mathbb{G}}(\mathcal{E}), \mathbb{G})$. As such $r_\gamma = \epsilon_\Omega(\tilde{\mathbb{G}}(\mathcal{E}), \mathbb{G})$, as required. □

We now summarize the results of our theory for direct RB in the following lemma:

Lemma 2. *The direct RB error rate (r_Ω) is approximately equal to the Ω-weighted average infidelity of the benchmarked gate set: $r_\Omega \approx \epsilon_\Omega(\tilde{\mathbb{G}}(\mathcal{E}), \mathbb{G})$.*

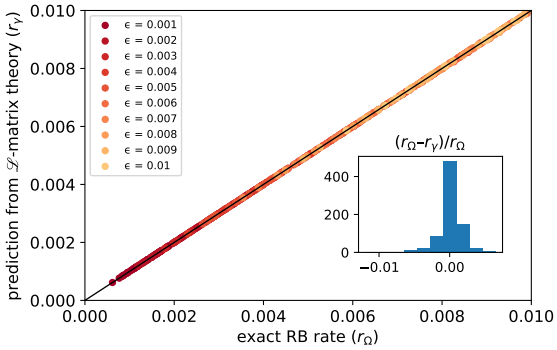


Figure 6.6: Verifying the accuracy of our L matrix theory for direct RB, using our exact R matrix theory. We show the L matrix theory's prediction for the direct RB error rate (r_γ) versus r_Ω computed from our exact R matrix theory (the R matrix theory enables computing S_d exactly, and we then fit S_d to $S_d = A + Bp^d$). Each data point shows r_γ versus r_Ω for the same randomly generated Markovian error models as in Figure 6.5. The inset shows a histogram of the relative error $[(r_\Omega - r_\gamma)/r_\Omega]$, demonstrating that our L matrix theory is extremely accurate. We observe extremely close agreement between r_Ω and r_γ, with no more than $\pm 1\%$ relative error (see inset histogram).

6.7.8 Validating the theory using numerical simulation

In this section we verify the correctness of our L matrix theory for direct RB, using numerical simulations. We simulated direct RB on a single qubit with a gate set consisting of $X(\frac{\pi}{2})$ and $Y(\frac{\pi}{2})$ gates. We choose an error model consisting of random Markovian errors with rates that contribute no more than ε to the gate's infidelity, for some small $\varepsilon > 0$, using the error generator formalism in [234]. That is, the superoperator $\tilde{\mathcal{G}}(G)$ for each gate is modelled as

$$\tilde{\mathcal{G}}(G) = \exp\left(\sum_{i=1}^{12} v_i(G)\mathcal{V}_i\right)\mathcal{G}(G) \tag{6.86}$$

where the $\mathcal{V}_i$ form a basis for the space of Markovian errors. For this example, the 12 error generators consist of three each of C, A, S and H type generators defined in [234]—where S and H are Pauli stochastic and Hamiltonian error generators respectively, while the remaining "Pauli-correlation" and "active" error generators augment the stochastic errors. For Hamiltonian errors, the leading order contribution to a gate's infidelity is at second order, whereas for stochastic errors it is at first order. Therefore, the Hamiltonian error rates (the v_i) are chosen uniformly at random from choose $[0, \sqrt{\varepsilon}]$, while the rates (the v_i) for all other errors are chosen uniformly from $[0, \varepsilon]$. We do this

for 785 different error models, varying ε from $\varepsilon = .001$ to $\varepsilon = .01$. The results of the numerical experiments are shown in Figs. 6.5 and 6.6.

For each error model, Fig. 6.5 shows the prediction for r_Ω from the L matrix theory (which is r_γ) versus an estimate $\hat{r}_\Omega$ for r_Ω obtained by simulating direct RB circuits, and fitting the data to $S_d = A + Bp^d$. Error bars are 1σ, and they are computed using a standard bootstrap. We observe good agreement between r_γ (the theory's prediction) and r_Ω. However, quantifying any small underlying discrepancies between r_Ω and r_γ is difficult due to the uncertainties in the $\hat{r}_\Omega$—which arise from simulating a finite set of direct RB circuits. To remove this uncertainty, we computed the exact S_d values, using our exact R matrix theory for direct RB, and we then fit these exact values for S_d to $S_d = A + Bp^d$ to obtain r_Ω. Figure 6.6 shows r_Ω versus r_γ, and a histogram (inset) of the relative error $(r_\Omega - r_\gamma)/r_\Omega$. We find extremely close agreement between r_Ω and r_γ, with no more than $\pm 1\%$ relative error. This therefore demonstrates the accuracy of our L matrix theory for direct RB.

6.8 Conclusion

Direct RB [167] is a method for benchmarking a set of quantum gates that does not require the inflexible circuit sampling and the complex gate-compilation of standard Clifford group RB [154]. This means that direct RB can benchmark more qubits than Clifford group RB, and direct RB can reliably extract information about the performance of the native gates of a device. In this paper we have presented a general definition for the direct RB protocol, and theories that explain how and why direct RB is broadly reliable. Our theory shows that direct RB is reliable as long as, whenever an error occurs, it almost certainly spreads on to many qubits before another error occurs. This proof that direct RB is reliable does not rely on convergence of a sequence of gates to a uniformly element from a unitary 2-design. This is crucial for proving that direct RB can be reliable on $n \gg 1$ qubits with realistic error rates, as the number of layers required for such convergence necessarily increases with n. Furthermore, our theories show that the direct RB error rate can be interpreted as the average infidelity of the benchmarked gates, building on similar results for Clifford group

RB [168, 169].

Clifford group RB has been extended to an entire suite of methods, including methods for estimating the error rates of individual gates [160, 235, 236], for quantifying the coherent component of gate errors [237–239], for estimating leakage and loss [233, 240, 241], for calibrating gates [187, 188], and for quantify crosstalk [165, 242]. We anticipate that all of these methods can be adapted to the more flexible direct RB framework, and fully developing these techniques is an interesting direction for future research. One limitation of direct RB is that it is not indefinitely scalable—as shown in Fig. 6.3—because initialization and measurement of a random stabilizer state requires $O(n^2/\log n)$ two-qubit gates. Recent work on mirror RB [190, 221, 243] has shown that it is possible to adapt direct RB to remove these expensive steps, and much of the theory and many of the analysis techniques used in this paper are also applicable to that more scalable protocol.

6.9 Proof of Proposition 3

Proof. By linearity $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}\left(\mathcal{E}_1 + \mathcal{E}_\gamma\right) = \mathcal{E}_1 + \gamma\mathcal{E}_\gamma$. If $\gamma = 1$ then the proposition holds trivially, so consider the case of $\gamma < 1$. If $\gamma < 1$ then, because $\tilde{\mathbb{G}}$ is a low-error implementation of $\mathbb{G}$, both 1 and γ are non-degenerate eigenvalues of $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$. Therefore, up to constant scalings, $\mathcal{E}_1$ and $\mathcal{E}_\gamma$ are unique. Define

$$\tilde{\mathcal{G}}_{\Omega-\text{avg}} \equiv \sum_{\mathcal{G}\in\mathbb{G}} \Omega(\mathcal{G})\tilde{\mathcal{G}}, \quad \mathcal{G}_{\Omega-\text{avg}} \equiv \sum_{\mathcal{G}\in\mathbb{G}} \Omega(\mathcal{G})\mathcal{G}, \tag{6.87}$$

where we abuse notation and consider all gates as these as superoperators. Let $\langle\langle V|$ denote some left eigenvector of the Hilbert-Schmidt space matrix $\tilde{\mathcal{G}}_{\Omega-\text{ avg}}$ with eigenvalue 1, i.e., $\langle\langle V|\tilde{\mathcal{G}}_{\Omega-\text{ avg}} = \langle\langle V|$. This eigenvector exists because the elements of $\tilde{\mathbb{G}}$ are trace-preserving maps, which implies that $\tilde{\mathcal{G}}_{\Omega-\text{avg}}$ is also a trace-preserving map, and all trace-preserving maps have 1 as an eigenvalue. Because the the elements of $\mathbb{G}$ are the superoperators for unitary gates, $\mathcal{G}^{-1}|B_0\rangle\rangle = |B_0\rangle\rangle$ for all $\mathcal{G} \in \mathbb{G}$, where $B_0 = \mathbb{I}_d/\sqrt{d}$ (here we are using the basis for Hilbert-Schmidt space introduced in

Section 6.6.1). This is the statement that unitary maps are unital. Hence we have that

$$\begin{aligned} L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}(|\,B_0\rangle\rangle\langle\langle V\,|) &= \sum_{\mathcal{G}\in\mathbb{G}} \Omega(\mathcal{G})[\mathcal{G}^{-1}|\,B_0\rangle\rangle\langle V|\tilde{\mathcal{G}}] \\ &= |\,B_0\rangle\rangle\langle\langle V|\,\tilde{\mathcal{G}}_{\Omega-\text{avg}}, \\ &= |\,B_0\rangle\rangle\langle\langle V| \end{aligned} \tag{6.88}$$

and therefore

$$\mathcal{E}_1 = |\,B_0\rangle\rangle\langle\langle V|\,. \tag{6.89}$$

Using the "stacked" representation discussed in Section 6.6.1, we may turn $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}(\mathcal{E}_\gamma) = \gamma\mathcal{E}_\gamma$ into a more standard eigenvector equation.

Let $\mathcal{G}_{\text{u}}^{-1} = \mathcal{G}^{-1} - |\,B_0\rangle\rangle\langle\langle B_0|$ for all $\mathcal{G}\in\mathbb{G}$, which is known as the unital component of $\mathcal{G}^{-1}$. Because the elements in $\mathbb{G}$ are unitary gates

$$\langle\langle B_0|\,\mathcal{G}_{\text{u}}^{-1} = \mathcal{G}_{\text{u}}^{-1}|\,B_0\rangle\rangle = 0, \tag{6.90}$$

for all $\mathcal{G}\in\mathbb{G}$. Therefore, for any superoperator $\mathcal{X}$,

$$\begin{aligned} \text{vec}(L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}(\mathcal{X})) &= \sum_{\mathcal{G}\in\mathbb{G}} \Omega(\mathcal{G})[\tilde{\mathcal{G}}^T\otimes\mathcal{G}^{-1}]\,\text{vec}(\mathcal{X}) \\ &= \sum_{\mathcal{G}\in\mathbb{G}} \Omega(\mathcal{G})[\tilde{\mathcal{G}}^T\otimes(|\,B_0\rangle\rangle\langle B_0| + \mathcal{G}_{\text{u}}^{-1})]\,\text{vec}(\mathcal{X}) \\ &= [\mathcal{L}_0 + \mathcal{L}_\perp]\,\text{vec}(\mathcal{X}) \end{aligned} \tag{6.91}$$

with $\mathcal{L}_0 = \tilde{\mathcal{G}}_{\Omega-\text{avg}}^T\otimes|\,B_0\rangle\rangle\langle B_0|$ and

$$\mathcal{L}_\perp = \sum_{\mathcal{G}\in\mathbb{G}} \Omega(\mathcal{G})(\tilde{\mathcal{G}}^T\otimes\mathcal{G}_{\text{u}}^{-1}). \tag{6.92}$$

Because $\mathcal{G}_{\text{u}}^{-1}|\,B_0\rangle\rangle\langle B_0| = |\,B_0\rangle\rangle\langle B_0|\mathcal{G}_{\text{u}}^{-1} = 0$, we have that the $\mathcal{L}_0$ and $\mathcal{L}_\perp$ matrices satisfy

$$\mathcal{L}_0\mathcal{L}_\perp = \mathcal{L}_\perp\mathcal{L}_0 = 0. \tag{6.93}$$

As such, the set of eigenvalues of L is the union of the sets of eigenvalues of $\mathcal{L}_0$ and $\mathcal{L}_\perp$. Moreover, if $\mathcal{V}$ is an eigen-operator of L and if the associated eigenvalue is not an eigenvalue of $\mathcal{L}_0$, then

$$\text{vec}(\mathcal{V}) = \sum_j\sum_{k>0} v_{jk}|\,B_j\rangle\rangle\otimes|\,B_k\rangle\rangle, \tag{6.94}$$

for some v_{jk}, and so $\langle\langle B_0|\,\mathcal{V} = 0$.

We now show that γ is not an eigenvalue of $\mathcal{L}_0$, which then implies that $\langle\langle B_0|\,\mathcal{E}_\gamma = 0$. The eigenvalues of $\mathcal{L}_0$ are the eigenvalues of $\tilde{\mathcal{G}}_{\Omega-\text{avg}}$, together with 0. $\mathcal{G}_{\Omega-\text{avg}}$ is a sequence-asymptotic 2-design, and thus has 1 as a non-degenerate eigenvalue. We will show this is also true for $\tilde{\mathcal{G}}_{\Omega-\text{avg}}$.

In [169], the author shows that, if the generating set is a 2-design, $\tilde{\mathcal{G}}_{\Omega-\text{avg}}$ will generally have a unique largest eigenvalue. We extend this argument to sequence-asymptotic 2-designs. The limiting map of the generator twirl is a 2-design twirl, and so in the limit the result from [169] holds. Now, suppose that it did not hold that the generator twirl had a non-degenerate 1-eigenspace. At no point in the limiting sequence could it acquire one, thus it has to have one to begin with.

Now, suppose by way of contradiction that γ is an eigenvalue of $\tilde{\mathcal{G}}_{\Omega-\text{avg}}$. Since γ is a perturbation of 1, in the limit that the perturbation goes to 0, $\tilde{\mathcal{G}}_{\Omega-\text{avg}}$ would acquire 1 as a degenerate eigenvalue, thus γ cannot be an eigenvalue of $\tilde{\mathcal{G}}_{\Omega-\text{avg}}$.

Thus, $\mathcal{E}_\gamma$ is not an eigenvector of $\mathcal{L}_0$ and hence

$$\langle\langle B_0|\,\mathcal{E}_\gamma = 0. \tag{6.95}$$

This equality implies that

$$\mathcal{E}_\gamma = \sum_i \sum_{j>0} \alpha_{ij} |\,B_j\rangle\rangle\langle\langle B_i|\,, \tag{6.96}$$

for some α_{ij}. A depolarization channel $\mathcal{D}_\lambda$ maps $\rho \to \rho$ with probability λ and $\rho \to \mathbb{I}/d$ with probability $1-\lambda$. As such, in Hilbert-Schmidt space with the basis considered here

$$\mathcal{D}_\lambda = |\,B_0\rangle\rangle\langle\langle B_0|\,| + \lambda \sum_{i\neq 0} |\,B_i\rangle\rangle\langle\langle B_i|\,. \tag{6.97}$$

Hence, from Eqs. (6.96) and (6.89) we have that $\mathcal{E}_1 + \gamma\mathcal{E}_\gamma = \mathcal{D}_\gamma(\mathcal{E}_1 + \mathcal{E}_\gamma)$. Finally, using this relation and considering the action of $L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}$ on $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_\gamma$, we have that

$$\begin{aligned} L_{\mathbb{G},\tilde{\mathbb{G}},\Omega}(\mathcal{E}) &= \mathcal{E}_1 + \gamma\mathcal{E}_\gamma, \\ &= \mathcal{D}_\gamma(\mathcal{E}_1 + \mathcal{E}_\gamma), \\ &= \mathcal{D}_\gamma\mathcal{E}, \end{aligned} \tag{6.98}$$

which concludes the proof. □

Note that the $\mathcal{E}$ in this proposition is not necessarily a completely positive map, and generically it is not.

6.10 Sampling distributions

In this appendix we present one possible family of sampling distributions (Ω), called the **edge grab** sampler and first introduced in [223]. This sampling distribution is designed for sampling from an n-qubit gate set $\mathbb{G}$ containing gates consisting of a single layer of parallel one- and two-qubit gates, with one-qubit gates from some set $\mathbb{G}_1$ and two-qubit gates only between a connected qubits as specified by an edge list E (that can contain directed edges). The sampling distribution is parameterized by the mean two-qubit gate of the sampled gates $\bar{\xi}$. The two-qubit gate density of an n-qubit layer is simply $\xi = \frac{2\alpha}{n}$ where α is the number of two-qubit gates in the layer. The edge grab sampling distribution $\Omega_{\bar{\xi}}$ is most easily described by an algorithm for drawing a sample from $\Omega_{\bar{\xi}}$. Drawing a sampling consists of the following procedure

(1) **Select a candidate set of edges** E_c. Initialize E_c to the empty set, and initialize E_r to the set of all edges E. Then, until E_r is the empty set:

1.1 Select an edge v uniformly at random from E_r.

1.2 Add v to E_c and remove all edges that have a qubit in common with v from E_r.

(2) **Select a subset of the candidate edges**. For each edge in E_c, include it in a final edge set E_f with a probability of $n\bar{\xi}/(2|E_c|)$ where $|E_c|$ is the total number of edges in E_c. If $n\bar{\xi}/(2|E_c|) > 1$ construct a new edge set sample E_c.

(3) **Construct the sampled gate**. The sampled gate $G \in \mathbb{G}$ is then constructed by adding a two-qubit gate on each edge in E_f and for all remaining qubits independently and uniformly sample a single-qubit gate from $\mathbb{G}_1$ to include in G.

The expected number of two-qubit gates in G is $n\bar{\xi}/2$, so this sampler generates an n-qubit layer with an expected two-qubit gate density of $\bar{\xi}$. One useful property of this sampling is that

the probability of sampling any particular n-qubit gate G is non-zero for every $G \in \mathbb{G}$ (except if $\bar{\xi} = 0$ or $\bar{\xi} = 1$). Finally, note that the edge grab algorithm is invalid if $n\bar{\xi}/(2|E_c|) > 1$ for any possible candidate edge set E_c. For an even number of fully-connected qubits, $\bar{\xi}$ can take any value between 0 and 1. But for any other connectivity the maximum achievable value of $\bar{\xi}$ is smaller. In our experiments, we set $\bar{\xi} = \frac{1}{8}$. This is an achievable value of $\bar{\xi}$ in the edge grab algorithm for all the processors that we benchmarked.

Chapter 7

The Impact of Markovian Errors on Random Circuits

7.1 Abstract

In addition to the characterization of quantum gate sets, an important task for understanding the performance of quantum computers is the characterization of quantum circuits. In general, this is a challenging problem to solve efficiently, given the complexity of arbitrary circuits. In this work, we make progress towards a solution of this problem by studying the impact of Markovian errors on random circuits, paying careful attention to Hamiltonian errors. By considering the factorization of a noisy circuit into a perfect circuit composed with a noise process, we find a perturbative expression for the survival probabilities of the circuits we study. From this expression we learn the precise way in which the gate set errors impact the deviations of the circuits from their ideal behavior.

7.2 Introduction

It is relatively unknown how errors, especially coherent errors, affect the success probabilities of quantum circuits. Ensuring that circuits have high success probabilities, in turn, is necessary for reliable quantum computation. For these reasons, developing detailed models of how errors affect the success probability of quantum circuits is one of the most important next steps in quantum characterization. One simple technique for characterizing quantum circuits is to first assess individual gates in a gate set, and then to use those errors to predict properties of circuits constructed from them, such as the success probability. While this idea is simple, even its best answers provide surprisingly little insight and place a large burden on experimental efforts. Bounds using the dia-

mond norm are often overly conservative, and rough heuristics such as multiplying the fidelities can fail to take advantage of circuit structure. As a simple example, coherent errors, if known precisely, can often be echoed away.

In this paper we further develop this line of research. To do this, we will constrain our model of errors and circuits. Despite this restriction, we will still be able to observe several interesting behaviors, and understand general features of the structure and strength of the noise. The first restriction will be to Markovian errors. In the case of a single gate with an error model that consists solely of Markovian errors, there are at least two natural ways to represent errors that are equivalently expressive. These are pre- and post- circuit factorizations [244]. These factorizations have the advantage of being interpretable, since they model the action of the imperfect operation as being the perfect operation composed with an error process. For circuits, such a decomposition is also desirable for the same reason. In this work, we will consider the post-circuit representation.

For certain kinds of circuits, finding this decomposition may be easier than for others. Specifically, random circuits have been shown to scramble errors, causing cancellation, and producing error channels that can be effectively stochastic and even depolarizing. Circuits with repeated elements, such as those used for amplitude amplification in Grover's algorithm, may additionally amplify coherent errors resulting in post-circuit error channels that may themselves be coherent with similar structure to the original gate's error channel. Other circuits yet, such as error correction circuits, will in general be more complicated.

To circumvent some of the challenges of analyzing specific circuits, and to take advantage of many of the properties of random circuits, many theoretical and experimental works use randomized compiling techniques [245] to simplify the structure of the noise channels in a circuit. Randomized compiling refers to the introduction of classical randomness into a quantum circuit in such a way that non-stochastic errors are converted into nearly stochastic ones. Randomized compiling, however, remains undesirable for two related reasons. First, it fails to take advantage of the coherence of quantum systems, in the sense that coherent errors can be echoed away if understood precisely. Second, in ignoring the coherence of the errors and converting them into stochastic errors, the error

rate of the circuit is in general increased.

In this paper, we study in detail the propagation of small, Markovian errors through a class of random quantum circuits that are similar to, but more general than, those used in direct randomized benchmarking (DRB) [246]. After introducing the required technical background, we will prove a few interesting technical results about the leading order contributions to the effective error rate of these circuits. We will conclude by analyzing the family of DRB experiments that draw from the two-gate gate set consisting of $RX(\pi/2)$ and $RY(\pi/2)$.

7.3 Errors, Gate Sets and Gauge in Quantum Circuits

In this section, we provide the necessary definitions and background for the rest of the paper. We start by defining a quantum circuit. A quantum circuit is a program for a quantum computer described as a sequence of layers. A layer is a collection of quantum gates on one or two qubits to be applied simultaneously. A quantum gate is generally described by a unitary operator that acts on a vector in the Hilbert space spanned by the states of the qubits in the quantum computer, and computing the success probabilities is done via Born's rule, the details of which will be assumed to be known by the reader.

In general, a quantum computer will have imperfections and to model these imperfections we describe quantum gates with completely positive trace preserving (CPTP) maps, which are an equivalent description to using Kraus operators for open system dynamics. While the unitary description of the ideal quantum gates is a bounded operator acting on the Hilbert space of states of the quantum computer, the CPTP maps are instead maps on bounded operators themselves. This space is known as Hilbert-Schmidt space, and vectors in this space are denoted with double angle brackets $|\cdot\rangle\rangle$.

In addition to imperfect gates, the state preparation and measurement will also be imperfect. To model imperfect state preparation and measurement (SPAM), we write $\langle\langle E_i|$ and $|\rho_j\rangle\rangle$ [246]. The advantages of moving to this superoperator space is that quantum gates act as linear operators on the states, and the probabilities associated with the different measurement outcomes are given

by inner products with the SPAM effects $\langle\langle E_i|$. We call the collection of gates and SPAM effects available to a quantum computer a gate set.

We will use $\langle\langle 0|$ and $|0\rangle\rangle$ to mean the measurement of the zero state and the preparation of the zero state respectively and $\{\mathcal{G}_k\}$ to denote a gate set. A circuit $\mathcal{C}$, consisting of gates as well as a measurement effect E_i and state preparation ρ_j, therefore has associated with it a success probability p_{ij} given by

$$p_{ij} = \langle\langle E_i|\mathcal{C}|\,\rho_j\rangle\rangle = \langle\langle E_i|\mathcal{G}_{k_n}...\mathcal{G}_{k_1}|\,\rho_j\rangle\rangle, \tag{7.1}$$

where the indices k_1 through k_n index into the gate set.

When the gates are noisy, as considered in this paper, the probability will instead be given as

$$\tilde{p}_{ij} = \langle\langle E_i|\tilde{\mathcal{C}}|\,\rho_j\rangle\rangle = \langle\langle E_i|\tilde{\mathcal{G}}_{k_n}...\tilde{\mathcal{G}}_{k_1}|\,\rho_j\rangle\rangle, \tag{7.2}$$

where $\tilde{\mathcal{G}}$ is meant to represent an imperfect implementation of $\mathcal{G}$.

In particular, in this paper, we will consider the specific noise model of Markovian errors, and so we are able to express $\tilde{\mathcal{G}}$ as

$$\tilde{\mathcal{G}} = \mathcal{G}e^L, \tag{7.3}$$

for an error generator L. Characterizing gates in this way, as a perfect gate composed with an error process is a convenient way of understanding errors in quantum gates. Therefore, in this paper, we are interested in the related question of how to understand quantum circuits as perfect circuits, followed by error processes. Unfortunately, the general difficulty of characterizing circuits is further confounded by a gauge degree of freedom. To introduce what we mean by a gauge degree of freedom, we first need to give some representation of the gates in our gate set. Because our errors are Markovian, we can choose our particular representation as the **Pauli transfer matrices** [245].

As discussed in [245], such a process e^L can be expressed as

$$e^L = \left(\begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \boldsymbol{m} & R \end{array}\right), \tag{7.4}$$

where the top row of all trace-preserving (TP) error maps is fixed to $\{1, 0, 0, 0, \cdots\}$ and the remainder of the first column, $\boldsymbol{m}$, describes any deviations from unitality.

Error generators, L, as introduced in Eq. (7.3) are discussed in detail in [244] and are the logarithms of error processes. Thus unitary error processes can be generated by Hamiltonian error generators, while stochastic error processes can be generated by stochastic error generators. The classification is in general more complicated, and for this paper it will suffice to give a short introduction of the relevant sectors of error generators. Returning to the taxonomy in [244], the H sector is most familiar from the Von Neumann equation and is given by Hermitian operators that act via commutation

$$H_P(\rho) = -i[P, \rho], \tag{7.5}$$

while the stochastic generators are given by

$$S_P(\rho) = P(\rho)P - \rho, \tag{7.6}$$

where P is a Pauli matrix in both equations. The stochastic errors appear as diagonal elements in the transfer matrix representation.

The final class of errors which we will talk about are active errors. These generate parts of non-unital errors, such as amplitude damping, and have the specific form

$$a_{PQ}(\rho) = i(P\rho Q - Q\rho P + \frac{1}{2}\{[P, Q], \rho\}), \tag{7.7}$$

for Pauli matrices P, Q. To gain intuition for these errors, note that we can express a small amplitude damping channel, $\Gamma_{1\to 0}(\rho)$ as

$$\Gamma_{1\to 0}(\rho) = e^{s_x + s_y + a_{x,y}} |\rho\rangle\rangle. \tag{7.8}$$

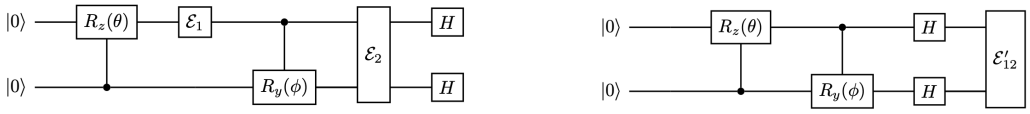


Figure 7.1: On the left, the circuit has three layers, with errors (denoted as $\mathcal{E}_1$ and $\mathcal{E}_2$) after each of the first two layers. It is possible to factor the errors to the right hand side of the circuit (the third layer), but the errors will in general be transformed. This is denoted on the right with the error $\mathcal{E}'_{12}$ after the third layer of the circuit.

It then becomes clear, upon writing out each circuit, that there is a gauge degree of freedom given by any invertible matrix M, which we call a gauge transformation. The probabilities in Eq. (7.1) transform as

$$p_{ij} = \langle\langle E_i | MM^{-1} G_{k_n} ... G_{k_2} MM^{-1} G_{k_1} MM^{-1} | \rho_j \rangle\rangle, \tag{7.9}$$

so that the gates in the gate set transform as $\mathcal{G} \to M^{-1}\mathcal{G}M$ and the SPAM effects transform as $\langle\langle E_i| \to \langle\langle E_i|M$ and $|\rho_j\rangle\rangle \to M^{-1}|\rho_j\rangle\rangle$. Because these gauge transformations do not affect any observable quantities, like the survival probabilities discussed in this paper, they are themselves unobservable and unphysical. Because of this, directions that are variant under these transformations are also not physically well-defined directions. We call directions which are first-order gauge invariant directions - directions that remain the same under gauge transformations - FOGI directions [247]. For a local quantity on the tangent space to be physical, it must be supported on the FOGI subspace.

In particular, consider a gauge transformation $\mathcal{M}$ as before. A function is first-order gauge invariant at 0 if

$$\mathcal{D}\left(f(\tilde{\mathbb{G}}_{\text{all}}\,)\right)|_{\boldsymbol{g}=0}\boldsymbol{g} = \mathcal{D}\left(f(\tilde{\mathbb{G}}_{\text{all}}\,(\mathcal{M}))\right)|_{\boldsymbol{g}=0}\boldsymbol{g}, \tag{7.10}$$

where $\mathcal{D}$ is the derivative, and $\boldsymbol{g}$ is a vector that parameterizes the errors on the gate set.

7.4 Error Factorization in Quantum Circuits

As we have discussed, a common method for understanding the impact of errors on a gate is to factor the noisy gate into an ideal gate followed by an error channel. In Sec. 7.3 we have

given an explicit example of this in Eq. 7.3 for the Markovian errors we consider in this paper. In this section we will introduce and discuss the generalization of this idea to factoring errors out of quantum circuits. An example of this is shown in Fig. 7.1, with the errors denoted as $\mathcal{E}_k$, with k denoting the layer of the error and $\mathcal{E}'_k$ with the prime denoting the error having been factored to the end of the circuit. For the errors discussed in this paper it is always possible to factor the errors to the end of the circuit, but they will in general be modified. Formally, for any noisy circuit $\tilde{\mathcal{C}}$ as before, we have

$$\tilde{\mathcal{C}} = \tilde{\mathcal{G}}_{k_n}...\tilde{\mathcal{G}}_{k_1} = \mathcal{G}_{k_n}\mathcal{E}_n...\mathcal{G}_{k_1}\mathcal{E}_1 = \mathcal{C}\mathcal{E}' \tag{7.11}$$

where we have defined $\mathcal{E}' \coloneqq \mathcal{C}^\dagger\mathcal{G}_{k_n}\mathcal{E}_n...\mathcal{G}_{k_1}\mathcal{E}_1$. A strong motivation for such a representation is given by quantum error correction [248]. While this is not the topic of this paper, we elaborate on the idea to highlight the similarities to our problem. The noisy gate model of quantum error correction considers the circuit and the error correction circuit as consisting of faulty gates. Thresholds for this model, while potentially complicated to compute, are the most accurate. The channel coding model is used largely in quantum information, and considers the computation being error-corrected to be a channel, and the error correction circuit's syndrome extraction to introduce no errors. This model, while unphysical, is equivalent to the first model in expressive strength if we can factor the errors out of the circuit, and has the advantage of being simpler to work with.

While this is a single important example, such an error factorization technique would find utility in nearly every area of quantum information science where noisy quantum circuits are studied, such as in metrology or characterization. This is due in part to the interpretability of the model. By factoring the error to the end of the circuit, it is physically intuitive to tune parameters of the noise model and understand their effect as corrupting an otherwise perfect circuit. In the remainder of this section, we discuss how to perform this factorization on a specific family of random circuits.

7.4.1 Factoring Errors out of Random Circuits

Efficiently factoring errors out of circuits can in general, however, be quite challenging. A cursory analysis would suggest that at least the circuit structure or noise structure must be simple.

We have chosen a relatively simple noise model in this paper, thus we are now interested in choosing a simple circuit model. One way to simplify the circuit structure is to consider ensembles of circuits. To this end, in the remainder of this paper we will consider collections of random circuits.

There are many choices of random ensembles of circuits in quantum information science, such as matchgate circuits, XEB circuits, RB circuits and more recently DRB circuits [246]. DRB gives a particularly powerful family of circuits for random circuits as compared to traditional Clifford randomized benchmarking, since it can in fact consider (almost) any set of gates that generate the Clifford group, which in turn is precisely the gates that are used by stabilizer error correction circuits.

The reason we might expect random circuits to make the factorization of errors easier is in part because they rapidly scramble errors - it is this rapid error scrambling that washes out coherent errors and generally affects a twirl superchannel (discussed in [246] and the previous chapter, see Eq. (3)). Intuitively, while a state twirl scrambles the information in a state, a twirling superchannel randomizes a channel. By considering a Markovian error model, we will see how the DRB error rates comes from this error scrambling of the Markovian errors.

7.4.2 DRB as Random Circuits

While the theory developed in [246] was motivated by DRB, it encompasses a larger class of random circuits. The restriction of DRB circuits to Clifford gates is so that a randomly sampled sequence can be efficiently inverted. However, the analysis itself does not depend on this property, and can be applied to the fidelity of *any* random circuit. In particular, we will consider circuits whose layers consist of gates sampled from a gate set $\{\mathcal{G}_k\}$ given by a sequence asymptotic 2-design [246]. To estimate the fidelity, however, it is necessary to introduce state preparation and measurement circuits - these circuits in turn will allow us to represent the effective noise in the circuit in a very simple way. A simple subset of these circuits are those that constitute DRB experiments, however almost any randomly sampled circuit that composes to the identity is in this class. This is because, in particular, almost any gate set is universal [249]. It therefore generates a 2-design and thus is

also a sequence asymptotic 2-design.

To use the work in [246] we will, however, consider DRB circuits and assume that these circuits have perfect stabilizer preparation, however that work additionally shows that the circuits are robust to imperfections in this SPAM, and in general, without the SPAM preparation, the form of the noise is altered to non-uniform stochastic noise. To see this, consider that in [246] the authors show that the survival probabilities are given as

$$S_d = \langle\langle 0| L_{\mathbb{C},\tilde{\mathbb{C}},\Pi}(L^d_{\mathbb{G},\tilde{\mathbb{G}},\Omega}(I))|0\rangle\rangle + \delta_d, \tag{7.12}$$

where we have fixed the SPAM effects to preparing and measuring in the zero state, which is possible for DRB, as we will discuss in the remainder of this paper. In this equation, we have used the $L-$matrix

$$L_{\mathbb{G},\tilde{\mathbb{G}},\Omega} = \sum_{G\in\mathbb{G}} \Omega(G)\mathcal{G}(G)\otimes\tilde{\mathcal{G}}(G), \tag{7.13}$$

for a gate set $\mathbb{G}$ and

$$L_{\mathbb{C},\tilde{\mathbb{C}},\Omega} = \sum_{C\in\mathbb{C}} \Omega(C)\mathcal{C}(C)\otimes\tilde{\mathcal{C}}(C), \tag{7.14}$$

for the Clifford group $\mathbb{C}$ and stabilizer circuits $\mathcal{C}$. In Eq. (7.12) we have interpreted these matrices as operators on the space of process matrices, otherwise known as "super-duper" operators.

The final superchannel twirl comes from the stabilizer preparation in DRB. If we remove this, as would be the case in truly random circuits, then as mentioned previously, we find that the channel is in general stochastic. We leave the study of this more general setting to future work. For this work, we consider the circuits in [246], which give the explicit form of the post-factored noise as a depolarizing channel D_γ. That is, the circuits can be rewritten approximately as

$$\langle\langle 0|\tilde{\mathcal{C}}|0\rangle\rangle \approx \langle\langle 0| D_\gamma S\mathcal{G}_{i_n}...\mathcal{G}_{i_1}|0\rangle\rangle = \langle\langle 0| D_\gamma\mathcal{C}|0\rangle\rangle, \tag{7.15}$$

where S is the stabilizer preparation. Thus what remains interesting for this work is determining γ.

7.5 The Impact of Markovian Errors on the Depolarization Rate

The random circuits we consider in this section are characterized by the sampling distribution Ω, the gate set $\{\mathcal{G}_k\}$ and the error model, which is described by the $3|\mathbb{G}|$ coefficients of the Hamiltonian contributions to the error rate. The idea that random quantum circuits, especially over gates that have some property close to being a two-design, average or twirl errors into a depolarizing channel is not a new idea. However, the normal intuition that this twirling depolarizes or dephases coherent errors in a simple way can easily be shown to be wrong. For a single error $\sum \alpha_i L_i$, we expect that we can twirl the Hamiltonian errors separately so that, for instance, the Hamiltonian errors h_x are sent to stochastic s_x errors of order h_x^2, and so that the Hamiltonian error rates add in quadrature.

Contrast this with the the gate set we will consider in Sec. 7.5.4, consisting of an $RX(\pi/2)$ and $RY(\pi/2)$ gate on a single qubit. In these random circuits there is a single free sampling parameter, p, corresponding to the probability of sampling the $RX(\pi/2)$ gate (with $1-p$ being the rate of sampling $RY(\pi/2)$). Intuitively, assuming that coherent errors on the gates add in quadrature, we would expect a formula for the decay rate r_γ, to be

$$r_\gamma = \sum_j p{h_x^j}^2 + (1-p){h_y^j}^2, \tag{7.16}$$

where h_x^j is the rate [244] of Hamiltonian j errors on the $RX(\pi/2)$ gate and h_y^j is the rate of Hamiltonian j errors on the $RY(\pi/2)$ gate.

This is a simple linear function in p, and is demonstrated in Fig. 7.2 to be incorrect. Fig. 7.2 demonstrates the survival probabilities of a DRB experiment on these two gates with a hamiltonian error model given as $h_x^y - h_x^z$ on the $RX(\pi/2)$ gate and $h_y^x + h_y^z$ on the $RY(\pi/2)$ gate. The figure shows the survival probabilities as a function of p using two methods. First, the points are from numerically simulating DRB experiments, and fitting the decay constant. The solid line is given by using the $L-$matrix theory from [246]. This plot thus shows good agreement between the theory in [246] and fits from numerical experiments even near the extremal points where the theory is not proven to be valid from the results in the previous chapter and [246]. The error we have chosen

that contrasts with our intuition is a relational error [247] and motivates a further understanding of how precisely Hamiltonian errors affect the DRB decay rate.

Using the theory in [246] we now prove a few technical results about the eigenvalue of the $L-$matrix in the presence of perturbative Markovian errors. In particular, in what follows we will consider an expansion of the form

$$\gamma \approx \sum_j c_j {}_j^2 \tag{7.17}$$

where j ranges over the two gates, and c is the coefficients for the rates ${}_j$ associated with the Hamiltonian error generators [244]. Because linear terms in these rates are amplitudes, we have assumed (and it can be shown) that these rates do not enter linearly in the error rate as amplitudes are not observable. After proving these results, we will look at an example using the ideas we have developed, to further understand the physics of Markovian errors on random quantum circuits.

7.5.1 Active Errors Don't Contribute to the Error Rate

In this section, we refer to the representation in Sec. 7.3 of Markovian processes with process matrices. The top row of all trace-preserving (TP) error maps is fixed to $\{1, 0, 0, 0, \cdots\}$ and the remainder of the first column, $\boldsymbol{m}$, describes any deviations from unitality, which arise from and only from active errors. Because of the zeroes in the top row, any composition of these maps will only ever have contributions from the first column of the matrix in the first columns.

There is a gauge in which the errors are completely depolarizing, and thus the survival probabilities are given by a trace, since for depolarizing noise the survival probabilities are given by the fidelity. Because the trace is basis independent, and active errors don't contribute to the diagonal elements in the Pauli transfer basis, we find that in the depolarizing gauge the contribution from the active errors is identically zero - the expression for the survival probabilities, a gauge invariant quantity, is functionally independent of the active errors. Because gauge transformations act as affine transformations on the vector of active generators, it remains that in any gauge the contribution from the active generators will be identically zero.

In other words, the contribution to the survival probabilities from the active generators is given

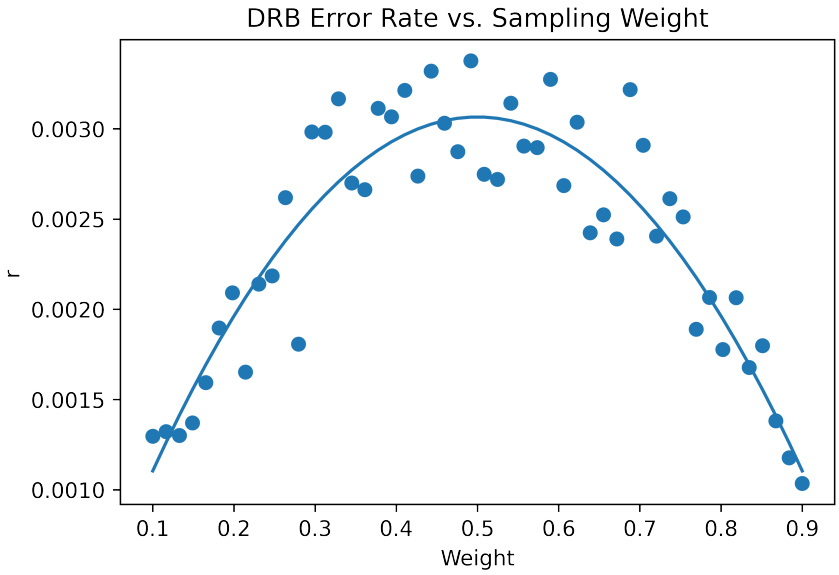


Figure 7.2: A simple gate set with a single relational error [247]. Intrinsic errors on either gate increase the decay rate in proportion with the fraction of the gate being introduced. Relational errors, on the other hand, can form an "inverse arch" - they contribute the most when ratio of gates is uniform. This is evidence that the errors are truly properties of the gate set, rather than of any individual gate. This figure also demonstrates that the $L-$matrix theory is descriptive in this case despite not being "benchmarking-compatible" [246] near $p = 0$ and $p = 1$, giving evidence that the theory in that work is more reliable than proven in the previous chapter, and is applicable to this work.

by a power series. Since the survival probabilities are functionally independent of these generators, the coefficients in this power series must be zero. A gauge transformation acts only by an affine shift on the generator values, and thus the power series still evaluates to zero in any gauge.

7.5.2 The Ground Space

We note that the dimension of the ground space of the Hessian is independent of the number of gates in the gate set. This is because the ground space corresponds to the set of directions in the vector space of error generators that the DRB error rate is insensitive to, which includes the gauge directions. This is because we expect that non-trivial randomized circuits will have some sensitivity to all non-gauge directions. By counting, we see that the physical error directions are an intrinsic error on each gate, and the relative error between each pair of gates. After the first two gates, this is set by specifying a relative error to two different gates in the gate set, and a single intrinsic error, and therefore contributes three physical directions per gate. Thus, after the first two gates, additional gates add on no additional gauge directions. In the case of two gates, there are three gauge directions since there are only three physical errors - two intrinsic, and one relational.

To understand this in more detail, we recall that gauge transformations act as affine transformations on the vector of error generators. This is essentially because the gauge transformations have essentially linear structure, but do not necessarily map 0 to 0. So if you consider (e.g.) the 6-dimensional h vector, then a gauge transformation t does not act as a linear transformation that maps $\boldsymbol{h} \rightarrow \mathbf{G}(t)\boldsymbol{h}$ where $\mathbf{G}(t)$ is a matrix. Instead, it maps $\boldsymbol{h} \rightarrow \boldsymbol{h} + \boldsymbol{g}(t)$, where $\boldsymbol{g}(t)$ is a vector. The same is true for the s, a and c vectors. The only way we can be sure that such affine shifts by $\boldsymbol{g}(t)$ will leave the DRB decay rate r invariant, given the formula $r = \boldsymbol{h}^T\mathbf{M}\boldsymbol{h}$, is if every possible g(t) lies in the kernel of $\mathbf{M}$. Or, to put the other way around, if $\mathbf{M} = 0$ on the subspace spanned by all "gauge directions" $\boldsymbol{g}(t)$.

7.5.3 Gauge Action Determines Lower Eigenvalues

In the previous argument we assumed that $r = \boldsymbol{h}^T \mathbf{G} \boldsymbol{h}$. In the case of the stochastic sector we have instead $r = \boldsymbol{s}^T \boldsymbol{g} + \boldsymbol{h}^T \mathbf{G} \boldsymbol{h}$. Indeed, if the action of the gauge transformations is to produce affine shifts of the error generators, and the RB error rate is gauge invariant, then we have that

$$\boldsymbol{s}^T \boldsymbol{h} + \boldsymbol{h}^T \mathbf{G} \boldsymbol{h} = \boldsymbol{s}^T (\boldsymbol{h} + \boldsymbol{g}) + (\boldsymbol{h} + \boldsymbol{g})^T \mathbf{G} (\boldsymbol{h} + \boldsymbol{g}), \tag{7.18}$$

where s is the first-order coefficients for the error generators, h is the v sector of the generators, and g is the gauge action.

For sectors that aren't stochastic $s = 0$ and so for all pairs g, h, generated by gauge group,

$$0 = \boldsymbol{g}^T \mathbf{G} \boldsymbol{h} + \boldsymbol{h}^T \mathbf{G} \boldsymbol{g} + \boldsymbol{h}^T \mathbf{G} \boldsymbol{h} \tag{7.19}$$

Now we find

$$-\boldsymbol{s}^T \boldsymbol{g} = 2 \boldsymbol{g} \mathbf{G} \boldsymbol{h} + \lambda \boldsymbol{h}^T \boldsymbol{h} \tag{7.20}$$

for all pairs g, h, where we have written h in terms of its eigenvalue and used the symmetry of G. h is a normalized eigenvector, so

$$\lambda = \frac{-\boldsymbol{s}^T \boldsymbol{g}}{2 \boldsymbol{g}^T \boldsymbol{h} + 1} \tag{7.21}$$

for an eigenvector $\boldsymbol{h}$ that transforms under a gauge transformation by an affine shift $\boldsymbol{g}$. Thus we see that two eigenvalues in the gauge directions are determined by the gauge action on the eigenvectors. In particular, if $\boldsymbol{s} = 0$ then the gauge directions correspond to zero eigenvalues.

7.5.4 Example: RX($\pi/2$), RY($\pi/2$)

7.5.4.1 The model

In this section we consider a simple gate set on a single qubit, consisting of $RX(\pi/2)$ and $RY(\pi/2)$. We will consider the Hamiltonian sector consisting of just gate has 3 error generators, on each gate, h_i, $i \in \{x, y, z\}$. We sample with the distribution $\Omega = (p, 1-p)$.

7.5.4.2 Discussion of Eigenvectors and Eigenvalues

We start with a simple gate set, consisting of a RX($\pi/2$) gate and a RY($\pi/2$) on a single qubit, with all Markovian error generators, corresponding to the C, A, S and H sectors, for equal weighting. While we discuss only the Hamiltonian sector in this section for the reasons previously mentioned in this paper, we numerically verify our full formula from perturbation theory in Fig. 7.3.

We analyze the spectrum of this matrix to see which direction it picks out as special in FOGI space. We focus on the Hamiltonian sector because

(1) Active errors do not contribute.

(2) Stochastic error contribute trivially at first order.

(3) Pauli-correlated errors, which introduce correlations between stochastic errors, are qualitatively similar to stochastic errors.

If we consider the Hamiltonian sector of the Hessian, we find that the three nonzero eigenvectors are supported completely by the Hamiltonian FOGI vectors. The intrinsic FOGI directions are gauge invariant to all orders, however these three directions are the directions in which the DRB error rate is diagonal - by describing the system with these error rates, it is possible to additively and independently compute the error rates. In other words, by increasing one of the error rates, it is possible to increase the DRB error rate independent of the others. Labeling these directions as F_1, F_2, F_3, at $p = 1/2$, these combinations are:

(1) The sum of the FOGI directions, $\sum F_i$.

(2) The difference of the intrinsic FOGI directions. $F_1 - F_2$

(3) $2F_3 - (F_1 + F_2)$

We can interpret the eigenvectors at equal weighting via the random walk discussion given previously. We do this below:

(1) The easiest to understand is the bottom eigenvalue. The bottom eigenvalue corresponds to the direction of least variance in the random walk space. This occurs when there is cancellation in the trajectories. So rather than coherent cancellation on a single gate, we are interesting in seeing the expected coherent cancellation on any trajectory. This must be balanced against the actual error introduced from the axis error.

(2) The top eigenvalue is the maximal variance is the maximal ravine is formed by always over-rotating, and narrowing the ravine so that the errors can flow more rapidly. Exactly in the same way that the error can cancel if the relative error is made larger, they can amplify when the error is made larger.

(3) This leaves one more direction, which is the difference of intrinsic errors. This is of course the unique remaining orthogonal FOGI direction, and hence we are done. However, it also suffices to show that this is a local minimum via the following argument. The coefficient on each must be equal by symmetry. The error rate is proportional to the variance of the random walk carried out by the gates. By increasing the angle, the errors line up and therefore the variance is decreased. With the angles at ninety degrees, there is opportunity for coherent amplification and therefore ballistic trajectory. By decreasing the angle, the random walk becomes closer to a classical random walk and therefore the variance is also decreased. Because the error rate is proportional to the variance, this shows that this is a local optima.

While this decomposition is dependent on the weighting, we might expect that these are natural from symmetry properties at $\Omega = (\frac{1}{2}, \frac{1}{2})$, and this give operational meaning to the FOGI directions. However, any choice of $\Omega = (p, 1-p)$ gives, a priori, a physically meaningful operational definition. We can parameterize this by the sampling distribution, and get the following Hessian for the Hamiltonian sector of errors.

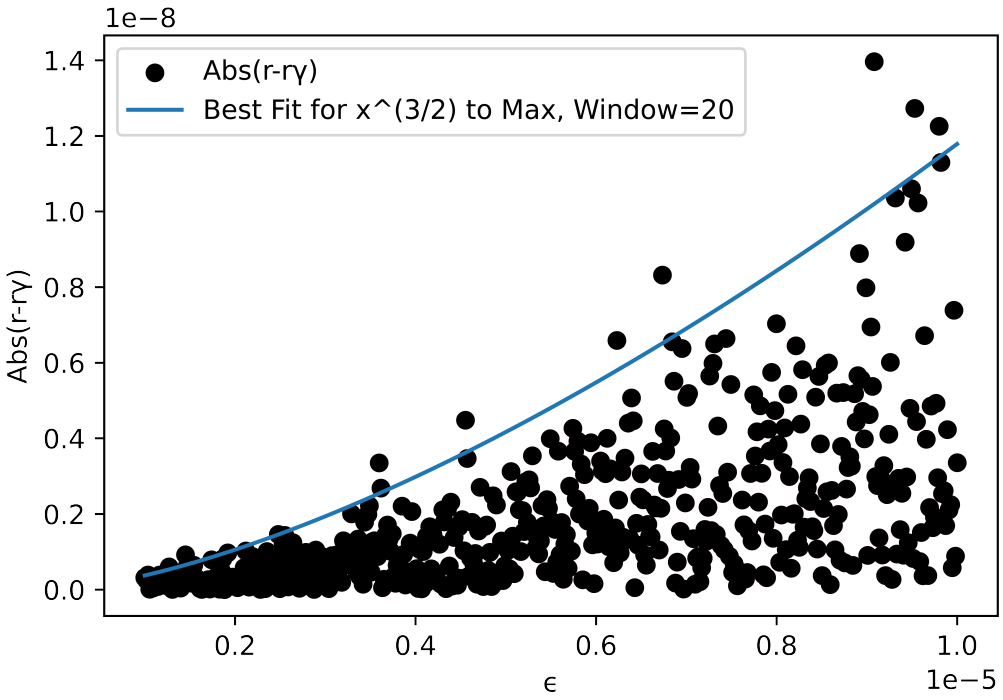


Figure 7.3: We sample 500 random Markovian error models, with leading order errors of size ϵ, for the $RX(\pi/2), RY(\pi/2)$ gate set, at $\Omega = (1/2, 1/2)$, and demonstrate that our formula is correct. The residuals are quadratic, demonstrating that we have the correct formula to first order. The fit line is proportional to $x^{3/2}$, demonstrating that the error in our estimate is $\mathcal{O}(\epsilon^{3/2})$.

$$H = \begin{pmatrix} \frac{2}{3}(p-1)p & -\frac{1}{3}\left(2p^2\right) & \frac{1}{3}(-2)(p-1)p & \frac{1}{3}(-2)(p-1)^2 & \frac{2}{3}(p-1)p & \frac{2}{3}(p-1)p \\ -\frac{1}{3}\left(2p^2\right) & \frac{2\left(p^3+p\right)}{3(p-1)} & \frac{2p^2}{3} & \frac{2}{3}(p-1)p & -\frac{1}{3}\left(2p^2\right) & -\frac{1}{3}\left(2p^2\right) \\ \frac{1}{3}(-2)(p-1)p & \frac{2p^2}{3} & \frac{2}{3}(p-1)p & \frac{2}{3}(p-1)^2 & \frac{1}{3}(-2)(p-1)p & \frac{1}{3}(-2)(p-1)p \\ \frac{1}{3}(-2)(p-1)^2 & \frac{2}{3}(p-1)p & \frac{2}{3}(p-1)^2 & \frac{2(p-1)((p-2)p+2)}{3p} & \frac{1}{3}(-2)(p-1)^2 & \frac{1}{3}(-2)(p-1)^2 \\ \frac{2}{3}(p-1)p & -\frac{1}{3}\left(2p^2\right) & \frac{1}{3}(-2)(p-1)p & \frac{1}{3}(-2)(p-1)^2 & \frac{2}{3}(p-1)p & \frac{2}{3}(p-1)p \\ \frac{2}{3}(p-1)p & -\frac{1}{3}\left(2p^2\right) & \frac{1}{3}(-2)(p-1)p & \frac{1}{3}(-2)(p-1)^2 & \frac{2}{3}(p-1)p & \frac{2}{3}(p-1)p \end{pmatrix} \tag{7.22}$$

We plot the spectral properties of this matrix in Fig. 7.4. One immediate strength of this model which is clear to see from these figures is that we can tune the probabilities to near 0 or 1. In this limit, the model becomes closer to a more tightly constrained family of circuits, whose errors consist primarily of over-rotations.

As mentioned, non-zero eigenvectors are supported directly on the (physical) FOGI directions. In the first three plots, we show the decomposition of the three eigenvectors with non-zero eignvalues into the their FOGI components. This analysis demonstrates exactly what each quadrature in the DRB experiment measures, as a function of the sampling distribution.

In addition to this, we plot the eigenvalue of each eigenvector, normalized across all values of the sampling parameter p in each plot, and give the relative contributions of each eigenvalue in the final plot. The final plot shows that at the extremal points the lower gap closes at 0. This is of course obvious - the sole contributing component becomes the divergently large coherent over rotations on the single gate being sampled.

The middle eigenvector is also particularly interesting, as it measures purely the relational error at the extremes of the sampling distribution, which naively there should be no sensitivity to. Moreover, we see that the contribution is opposite in sign at the two extremes, which we can take advantage of to construct a measure of relational errors, which we have shown previously corrupt our intuitive notion of average gate fidelity.

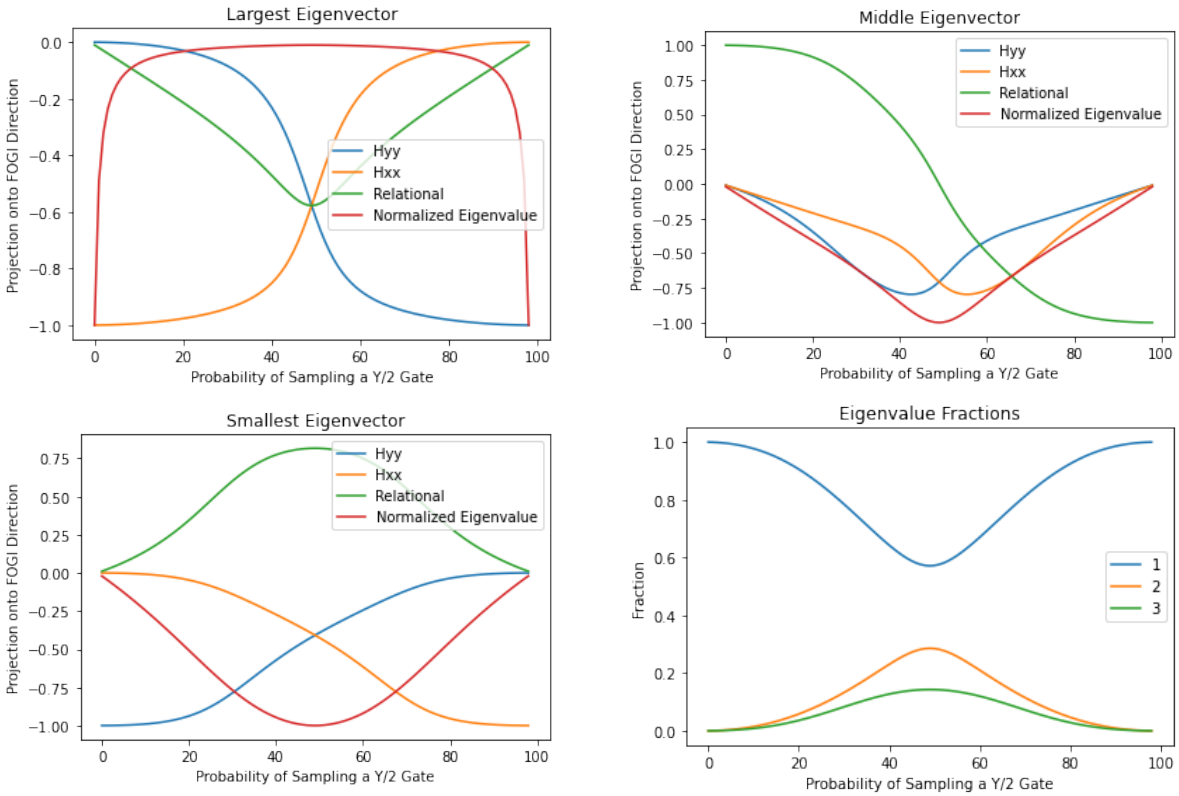


Figure 7.4: Interestingly, in the limit of $p = 0$ or $p = 1$, each eigenvector has support on a single FOGI direction. The largest eigenvalue diverges at the extremal sampling distributions, since in these limits coherent over-rotations on a single gate can coherently amplify to very large errors. The smallest eigenvalue, conversely, is seen to have support only on the other gate in the extremal limits, since these errors appear vanishingly often. Finally, the intermediate eigenvalue is completely supported by the relational error at the extremal points.

7.6 Conclusion

In this paper, we have studied random quantum circuits that asymptotically form two-designs. By assuming a perturbative Markovian error model, we were able express the impact of the noise on the circuit as a single post-circuit channel. This allowed us to produce intuitive error models for the impact of physical errors on the performance of the circuit. The characterization of circuits is an important next step for characterizing quantum computers. Developing techniques, such as those discussed in this paper, that take advantage of existing gate-level characterizations as well as the structure of the circuits being studied, are particularly useful. For instance, while conventional wisdom suggests that Hamiltonian errors should coherently accumulate linearly in the depth of a circuit, we've shown that can be false for simple relational errors [247].

Furthermore, even in the simple example considered in this paper, we have been able to argue that the entire class of active errors do not change the error rate, and have been able to further understand how the gauge freedom on a gate set can affect the contributions of physical errors to the decay rate. Thus the results and example we have worked through in this paper, for this particular family of circuits and error models, make substantial progress towards the useful characterization of quantum circuits. In particular, our results also make progress towards the characterization of errors on error correction circuits, since these consist of the gates used in conventional DRB experiments.

However, this analysis leaves open several interesting further steps. By including a more complex error model, we may be able to understand the connection between experimental system parameters, such as the anharmonicity in superconducting transmons, or laser noise in atomic systems, and circuit survival probabilities. Removing the randomness or idealized SPAM effects would give a more faithful model of error correction circuits, which would then let us understand the functional dependence of the performance of error correction on simple, experimentally controllable parameters. For general algorithmic circuits, we would also like to develop a technique for factoring the errors out of specific circuit structures or families. This is particularly useful for both designing diagnostic circuits for characterizing quantum computing systems and for predicting how gate set

errors will impact specific use cases of quantum computers.

Chapter 8

Summary and conclusions

In this thesis we have explored a wide range of topics, primarily in the subfields of quantum information that are quantum computation, quantum metrology and quantum characterization. In the first two sections, our work on quantum computation in both Penning traps and with neutral atoms demonstrates just some of the challenges associated with trying to perform quantum computations on some technologies that are purportedly scalable. Furthermore, given the fidelity and performance requirements in these works, we see that near term devices are still in their infancy.

Despite this, in chapter three we propose a quantum error correction experiment that, with today's devices, should be able to demonstrate bona fide fault tolerance - a milestone for quantum information technologies. Chapter four and five give a brief venture into quantum metrology and show that there is still much work to be done, and opportunities to be found, in using quantum systems for metrological purposes. By using qubits as sensors for tasks like axion detection and noise spectroscopy, we find that the rapid detection of signals may be possible. Furthermore, we have demonstrated the dependence of the maximal performance of a reservoir computer on the noise the reservoir experiences - suggesting a place where quantum reservoir computing might yield advantages. These works show that there is still fertile ground to explore in using quantum systems and devices for more than just quantum computation. We concluded with two projects exploring direct randomized benchmarking. By rigorously defining the properties of direct randomized benchmarking in chapter six, we were able to leverage those mathematical details in chapter seven to study the propagation of a simple class of errors, Markovian errors, through a simple family of

circuits, random circuits.

In all seven of these chapters, we find the same important conclusion - by using ideas in quantum information to understand physical systems we are able to extract more insight than we otherwise might. In the case of the last chapters, this is obvious - direct randomized benchmarking itself is a construction of quantum information. However, for the metrological examples with axion detection and reservoir computation, as well as for manipulating atomic and ionic systems with many degrees of freedom, we see that by thinking of the systems as qubits, controlled by gates, we are able to observe interesting properties of the systems that we may otherwise be blind to.

References

[1] A. M. Polloreno, A. M. Rey, and J. J. Bollinger, "Individual qubit addressing of rotating ion crystals in a penning trap", Physical Review Research **4** (2022).

[2] D. Kielpinski, C. Monroe, and D. J. Wineland, "Architecture for a large-scale ion-trap quantum computer", Nature **417** (2002).

[3] K. R. Brown, J. Kim, and C. Monroe, "Co-designing a scalable quantum computer with trapped atomic ions", npj Quantum Information **2** (2016).

[4] A. Bermudez, X. Xu, R. Nigmatullin, J. O'Gorman, V. Negnevitsky, P. Schindler, T. Monz, U. G. Poschinger, C. Hempel, J. Home, F. Schmidt-Kaler, M. Biercuk, R. Blatt, S. Benjamin, and M. Müller, "Assessing the progress of trapped-ion processors towards fault-tolerant quantum computation", Phys. Rev. X **7** (2017).

[5] C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage, "Trapped-ion quantum computing: progress and challenges", Applied Physics Reviews **6** (2019).

[6] J. Gaebler, T. Tan, Y. Lin, Y. Wan, R. Bowler, A. Keith, S. Glancy, K. Coakley, E. Knill, D. Leibfried, and D. Wineland, "High-fidelity universal gate set forBe9ion qubits", Physical Review Letters **117** (2016).

[7] C. Ballance, T. Harty, N. Linke, M. Sepiol, and D. Lucas, "High-fidelity quantum logic gates using trapped-ion hyperfine qubits", Physical Review Letters **117** (2016).

[8] R. Srinivas, S. C. Burd, H. M. Knaack, R. T. Sutherland, A. Kwiatkowski, S. Glancy, E. Knill, D. J. Wineland, D. Leibfried, A. C. Wilson, D. T. C. Allcock, and D. H. Slichter, "High-fidelity laser-free universal control of two trapped ion qubits",

[9] B. C. Sawyer and K. R. Brown, "Wavelength-insensitive, multispecies entangling gate for group-2 atomic ions", Physical Review A **103** (2021).

[10] C. R. Clark, H. N. Tinkey, B. C. Sawyer, A. M. Meier, K. A. Burkhardt, C. M. Seck, C. M. Shappert, N. D. Guise, C. E. Volin, S. D. Fallek, et al., "High-fidelity bell-state preparation with ca+ 40 optical qubits", Physical Review Letters **127** (2021).

[11] P. Wang, C.-Y. Luan, M. Qiao, M. Um, J. Zhang, Y. Wang, X. Yuan, M. Gu, J. Zhang, and K. Kim, "Single ion qubit with estimated coherence time exceeding one hour", Nature Communications **12** (2021).

[12] C. Monroe, R. Raussendorf, A. Ruthven, K. R. Brown, P. Maunz, L.-M. Duan, and J. Kim, "Large-scale modular quantum-computer architecture with atomic memory and photonic interconnects", Phys. Rev. A **89** (2014).

[13] M. J. Biercuk, H. Uys, A. P. Vandevender, N. Shiga, W. M. Itano, and J. J. Bollinger, "High-fidelity quantum control using ion crystals in a penning trap", arXiv preprint arXiv:0906.0398 (2009).

[14] M. D'Onofrio, Y. Xie, A. Rasmusson, E. Wolanski, J. Cui, and P. Richerme, "Radial two-dimensional ion crystals in a linear paul trap", Physical Review Letters **127** (2021).

[15] M. Qiao, Z. Cai, Y. Wang, B. Du, N. Jin, W. Chen, P. Wang, C. Luan, E. Gao, X. Sun, et al., "Observing frustrated quantum magnetism in two-dimensional ion crystals", arXiv preprint arXiv:2204.07283 (2022).

[16] S. Jain, J. Alonso, M. Grau, and J. P. Home, "Scalable arrays of micro-penning traps for quantum computing and simulation", Physical Review X **10** (2020).

[17] M. Mielenz, H. Kalis, M. Wittemer, F. Hakelberg, U. Warring, R. Schmied, M. Blain, P. Maunz, D. L. Moehring, D. Leibfried, et al., "Arrays of individually controlled ions suitable for two-dimensional quantum simulations", Nature communications **7** (2016).

[18] J. W. Britton, B. C. Sawyer, A. C. Keith, C.-C. J. Wang, J. K. Freericks, H. Uys, M. J. Biercuk, and J. J. Bollinger, "Engineered two-dimensional ising interactions in a trapped-ion quantum simulator with hundreds of spins", Nature **484** (2012).

[19] P. Schindler, D. Nigg, T. Monz, J. T. Barreiro, E. Martinez, S. X. Wang, S. Quint, M. F. Brandl, V. Nebendahl, C. F. Roos, et al., "A quantum information processor with trapped ions", New Journal of Physics **15** (2013).

[20] H. C. Naègerl, D. Leibfried, H. Rohde, G. Thalhammer, J. Eschner, F. Schmidt-Kaler, and R. Blatt, "Laser addressing of individual ions in a linear ion trap", Physical Review A **60** (1999).

[21] J. G. Bohnet, B. C. Sawyer, J. W. Britton, M. L. Wall, A. M. Rey, M. Foss-Feig, and J. J. Bollinger, "Quantum spin dynamics and entanglement generation with hundreds of trapped ions", Science **352** (2016).

[22] M. Gärttner, J. G. Bohnet, A. Safavi-Naini, M. L. Wall, J. J. Bollinger, and A. M. Rey, "Measuring out-of-time-order correlations and multiple quantum spectra in a trapped-ion quantum magnet", Nature Physics **13** (2017).

[23] H. Ball, C. D. Marciniak, R. N. Wolf, A. T.-H. Hung, K. Pyka, and M. J. Biercuk, "Site-resolved imaging of beryllium ion crystals in a high-optical-access penning trap with inbore optomechanics", Review of Scientific Instruments **90** (2019).

[24] S. Mavadia, J. F. Goodwin, G. Stutter, S. Bharadia, D. R. Crick, D. M. Segal, and R. C. Thompson, "Control of the conformations of ion coulomb crystals in a penning trap", Nature Communications **4** (2013).

[25] B. J. McMahon and B. C. Sawyer, "Second-scale $^{9}Be^{+}$ spin coherence in a compact penning trap", Phys. Rev. Applied **17** (2022).

[26] J. W. Evans, B. Macintosh, L. Poyneer, K. Morzinski, S. Severson, D. Dillon, D. Gavel, and L. Reza, "Demonstrating sub-nm closed loop mems flattening", Optics Express **14** (2006).

[27] C. Blain, O. Guyon, C. Bradley, and O. Lardière, "Fast iterative algorithm (fia) for controlling mems deformable mirrors: principle and laboratory demonstration", Optics express **19** (2011).

[28] M. Born and E. Wolf, "Principles of optics: electromagnetic theory of propagation, interference and diffraction of light", in (Elsevier, 2013), p. 986.

[29] J. P. Covey, L. De Marco, Ó. L. Acevedo, A. M. Rey, and J. Ye, "An approach to spin-resolved molecular gas microscopy", New Journal of Physics **20** (2018).

[30] X.-P. Huang, J. J. Bollinger, T. B. Mitchell, W. M. Itano, and D. H. E. Dubin, "Precise control of the global rotation of strongly coupled ion plasmas in a penning trap", Physics of Plasmas **5** (1998).

[31] S. S. Hong, A. T. Papageorge, P. Sivarajah, G. Crossman, N. Didier, A. M. Polloreno, E. A. Sete, S. W. Turkowski, M. P. da Silva, and B. R. Johnson, "Demonstration of a parametrically activated entangling gate protected from flux noise", Physical Review A **101** (2020).

[32] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, et al., "Quantum supremacy using a programmable superconducting processor", Nature **574** (2019).

[33] T. Harty, D. Allcock, C. J. Ballance, L. Guidoni, H. Janacek, N. Linke, D. Stacey, and D. Lucas, "High-fidelity preparation, gates, memory, and readout of a trapped-ion quantum bit", Physical review letters **113** (2014).

[34] A. Polloreno and G. Smith, "The qaoa with slow measurements", arXiv preprint arXiv:2205.06845 (2022).

[35] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm", arXiv preprint arXiv:1411.4028 (2014).

[36] S. Hadfield, Z. Wang, B. O'gorman, E. G. Rieffel, D. Venturelli, and R. Biswas, "From the quantum approximate optimization algorithm to a quantum alternating operator ansatz", Algorithms **12** (2019).

[37] J. Wurtz and P. J. Love, "Maxcut qaoa performance guarantees for p$>$ 1", arXiv preprint arXiv:2010.11209 (2020).

[38] G. González-Garcıa, R. Trivedi, and J. I. Cirac, "Error propagation in nisq devices for solving classical optimization problems", arXiv preprint arXiv:2203.15632 (2022).

[39] S. Bravyi, A. Kliesch, R. Koenig, and E. Tang, "Obstacles to variational quantum optimization from symmetry protection", Physical review letters **125** (2020).

[40] E. Farhi, D. Gamarnik, and S. Gutmann, "The quantum approximate optimization algorithm needs to see the whole graph: a typical case", arXiv preprint arXiv:2004.09002 (2020).

[41] L. Bittel and M. Kliesch, "Training variational quantum algorithms is np-hard", Physical Review Letters **127** (2021).

[42] M. B. Hastings, "Classical and quantum bounded depth approximation algorithms", arXiv preprint arXiv:1905.07047 (2019).

[43] M. B. Hastings, "A classical algorithm which also beats $\frac{1}{2} + \frac{2}{\pi}\frac{1}{\sqrt{D}}$ for high girth max-cut", arXiv preprint arXiv:2111.12641 (2021).

[44] V. Akshay, H. Philathong, M. E. Morales, and J. D. Biamonte, "Reachability deficits in quantum approximate optimization", Physical review letters **124** (2020).

[45] G. G. Guerreschi and A. Y. Matsuura, "Qaoa for max-cut requires hundreds of qubits for quantum speed-up", Scientific reports **9** (2019).

[46] M. P. Harrigan, K. J. Sung, M. Neeley, K. J. Satzinger, F. Arute, K. Arya, J. Atalaya, J. C. Bardin, R. Barends, S. Boixo, et al., "Quantum approximate optimization of non-planar graph problems on a planar superconducting processor", Nature Physics **17** (2021).

[47] J. S. Otterbach, R. Manenti, N. Alidoust, A. Bestwick, M. Block, B. Bloom, S. Caldwell, N. Didier, E. S. Fried, S. Hong, et al., "Unsupervised machine learning on a hybrid quantum computer", arXiv preprint arXiv:1712.05771 (2017).

[48] T. Graham, Y. Song, J. Scott, C. Poole, L. Phuttitarn, K. Jooya, P. Eichler, X. Jiang, A. Marra, B. Grinkemeyer, et al., "Demonstration of multi-qubit entanglement and algorithms on a programmable neutral atom quantum computer", arXiv preprint arXiv:2112.14589 (2021).

[49] G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, et al., "Quantum approximate optimization of the long-range ising model with a trapped-ion quantum simulator", Proceedings of the National Academy of Sciences **117** (2020).

[50] M. Hodson, B. Ruck, H. Ong, D. Garvin, and S. Dulman, "Portfolio rebalancing experiments using the quantum alternating operator ansatz", arXiv preprint arXiv:1911.05296 (2019).

[51] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, "Quantum approximate optimization algorithm: performance, mechanism, and implementation on near-term devices", Physical Review X **10** (2020).

[52] R. Shaydulin and Y. Alexeev, "Evaluating quantum approximate optimization algorithm: a case study", in 2019 tenth international green and sustainable computing conference (igsc) (IEEE, 2019), pp. 1–6.

[53] Y. Alexeev, S. Khairy, R. Shaydulin, L. Cincio, and P. Balaprakash, "Reinforcement learning for finding qaoa parameters", Bulletin of the American Physical Society **65** (2020).

[54] R. Shaydulin, I. Safro, and J. Larson, "Multistart methods for quantum approximate optimization", in 2019 ieee high performance extreme computing conference (hpec) (IEEE, 2019), pp. 1–8.

[55] M. Streif, S. Yarkoni, A. Skolik, F. Neukart, and M. Leib, "Beating classical heuristics for the binary paint shop problem with the quantum approximate optimization algorithm", Physical Review A **104** (2021).

[56] R. Tate, M. Farhadi, C. Herold, G. Mohler, and S. Gupta, "Bridging classical and quantum with sdp initialized warm-starts for qaoa", arXiv preprint arXiv:2010.14021 (2020).

[57] J. M. Baker, A. Litteken, C. Duckering, H. Hoffmann, H. Bernien, and F. T. Chong, "Exploiting long-distance interactions and tolerating atom loss in neutral atom quantum architectures", in 2021 acm/ieee 48th annual international symposium on computer architecture (isca) (IEEE, 2021), pp. 818–831.

[58] M. Saffman, "Quantum computing with atomic qubits and rydberg interactions: progress and challenges", Journal of Physics B: Atomic, Molecular and Optical Physics **49** (2016).

[59] M. Saffman, personal communication, Apr. 1, 2022.

[60] E. Farhi, D. Gamarnik, and S. Gutmann, **The quantum approximate optimization algorithm needs to see the whole graph: worst case examples**, 2020, eprint: arXiv:2005.08747.

[61] D. Bertsimas and J. Tsitsiklis, "Simulated annealing", Statistical science **8** (1993).

[62] S. B. Gelfand and S. K. Mitter, "Simulated annealing with noisy or imprecise energy measurements", Journal of Optimization Theory and Applications **62** (1989).

[63] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I.

Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python", Nature Methods **17** (2020).

[64] Y. Xiang, D. Sun, W. Fan, and X. Gong, "Generalized simulated annealing algorithm and its application to the thomson model", Physics Letters A **233** (1997).

[65] B. Hajek, "Cooling schedules for optimal annealing", Mathematics of operations research **13** (1988).

[66] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell, "Optimal inapproximability results for max-cut and other 2-variable csps?", SIAM Journal on Computing **37** (2007).

[67] E. Mossel, R. O'Donnell, and K. Oleszkiewicz, "Noise stability of functions with low influences: invariance and optimality", in 46th annual ieee symposium on foundations of computer science (focs'05) (IEEE, 2005), pp. 21–30.

[68] L. Trevisan, "On khot's unique games conjecture.", Bulletin (New Series) of the American Mathematical Society **49** (2012).

[69] D. Coppersmith, D. Gamarnik, M. Hajiaghayi, and G. B. Sorkin, "Random max sat, random max cut, and their phase transitions", Random Structures & Algorithms **24** (2004).

[70] D. Gamarnik and Q. Li, "On the max-cut of sparse random graphs", Random Structures & Algorithms **52** (2018).

[71] M. R. Garey and D. S. Johnson, ""strong"np-completeness results: motivation, examples, and implications", Journal of the ACM (JACM) **25** (1978).

[72] S. T. McCormick, M. R. Rao, and G. Rinaldi, "Easy and difficult objective functions for max cut", Mathematical programming **94** (2003).

[73] J. Gray, "Quimb: a python library for quantum information and many-body calculations", Journal of Open Source Software **3** (2018).

[74] Z. Wang, S. Hadfield, Z. Jiang, and E. G. Rieffel, "Quantum approximate optimization algorithm for maxcut: a fermionic view", Physical Review A **97** (2018).

[75] J. Wurtz and P. Love, "Maxcut quantum approximate optimization algorithm performance guarantees for p> 1", Physical Review A **103** (2021).

[76] D. W. Scott, **Multivariate density estimation: theory, practice, and visualization** (John Wiley & Sons, 2015).

[77] A. Shlosberg, A. M. Polloreno, and G. Smith, "Towards demonstrating fault tolerance in small circuits using bacon-shor codes", arXiv preprint arXiv:2108.02079 (2021).

[78] Z. Chen, K. J. Satzinger, J. Atalaya, A. N. Korotkov, A. Dunsworth, D. Sank, C. Quintana, M. McEwen, R. Barends, P. V. Klimov, et al., "Exponential suppression of bit or phase flip errors with repetitive error correction", arXiv preprint arXiv:2102.06132 (2021).

[79] C. K. Andersen, A. Remm, S. Lazar, S. Krinner, N. Lacroix, G. J. Norris, M. Gabureac, C. Eichler, and A. Wallraff, "Repeated quantum error detection in a surface code", Nature Physics **16** (2020).

[80] L. Egan, D. M. Debroy, C. Noel, A. Risinger, D. Zhu, D. Biswas, M. Newman, M. Li, K. R. Brown, M. Cetina, et al., "Fault-tolerant operation of a quantum error-correction code", arXiv preprint arXiv:2009.11482 (2020).

[81] N. M. Linke, M. Gutierrez, K. A. Landsman, C. Figgatt, S. Debnath, K. R. Brown, and C. Monroe, "Fault-tolerant quantum error detection", Science advances **3** (2017).

[82] J. F. Marques, B. M. Varbanov, M. S. Moreira, H. Ali, N. Muthusubramanian, C. Zachariadis, F. Battistel, M. Beekman, N. Haider, W. Vlothuizen, A. Bruno, B. M. Terhal, and L. DiCarlo, **Logical-qubit operations in an error-detecting surface code**, 2021, eprint: arXiv:2102.13071.

[83] P. W. Shor, "Fault-tolerant quantum computation", in Proceedings of 37th conference on foundations of computer science (IEEE, 1996), pp. 56–65.

[84] D. Nigg, M. Müller, E. A. Martinez, P. Schindler, M. Hennrich, T. Monz, M. A. Martin-Delgado, and R. Blatt, "Quantum computations on a topologically encoded qubit", Science **345** (2014), eprint: https://science.sciencemag.org/content/345/6194/302.full.pdf.

[85] A. Córcoles, E. Magesan, S. J. Srinivasan, A. W. Cross, M. Steffen, J. M. Gambetta, and J. M. Chow, "Demonstration of a quantum error detection code using a square lattice of four superconducting qubits", Nature Communications **6** (2015).

[86] P. Campagne-Ibarcq, A. Eickbusch, S. Touzard, E. Zalys-Geller, N. E. Frattini, V. V. Sivak, P. Reinhold, S. Puri, S. Shankar, R. J. Schoelkopf, L. Frunzio, M. Mirrahimi, and M. H. Devoret, **Quantum error correction of a qubit encoded in grid states of an oscillator**, 2020, arXiv:1907.12487 [quant-ph].

[87] C. Flühmann, T. L. Nguyen, M. Marinelli, V. Negnevitsky, K. Mehta, and J. P. Home, "Encoding a qubit in a trapped-ion mechanical oscillator", Nature **566** (2019).

[88] S. Puri, A. Grimm, P. Campagne-Ibarcq, A. Eickbusch, K. Noh, G. Roberts, L. Jiang, M. Mirrahimi, M. H. Devoret, and S. M. Girvin, "Stabilized cat in a driven nonlinear cavity: a fault-tolerant error syndrome detector", Phys. Rev. X **9** (2019).

[89] C. Ryan-Anderson, J. G. Bohnet, K. Lee, D. Gresh, A. Hankin, J. P. Gaebler, D. Francois, A. Chernoguzov, D. Lucchetti, N. C. Brown, T. M. Gatterman, S. K. Halit, K. Gilmore, J. Gerber, B. Neyenhuis, D. Hayes, and R. P. Stutz, **Realization of real-time fault-tolerant quantum error correction**, 2021, eprint: arXiv:2107.07505.

[90] D. Gottesman, "Quantum fault tolerance in small experiments", arXiv preprint arXiv:1610.03507 (2016).

[91] P. Aliferis, D. Gottesman, and J. Preskill, "Quantum accuracy threshold for concatenated distance-3 codes", Quantum Information & Computation **6** (2006).

[92] P. Aliferis and A. W. Cross, "Subsystem fault tolerance with the bacon-shor code", Phys. Rev. Lett. **98** (2007).

[93] D. Bacon, "Operator quantum error-correcting subsystems for self-correcting quantum memories", Physical Review A **73** (2006).

[94] D. Gottesman, **Stabilizer codes and quantum error correction**, 1997, eprint: arXiv:quant-ph/9705052.

[95] L. Vaidman, L. Goldenberg, and S. Wiesner, "Error prevention scheme with four particles", Physical Review A **54** (1996).

[96] M. Grassl, T. Beth, and T. Pellizzari, "Codes for the quantum erasure channel", Physical Review A **56** (1997).

[97] A. M. Polloreno, J. L. Beckey, J. Levin, A. Shlosberg, J. K. Thompson, M. Foss-Feig, D. Hayes, and G. Smith, "Opportunities and limitations in broadband sensing", Physical Review Applied **19** (2023).

[98] V. Giovannetti, "Quantum-enhanced measurements: beating the standard quantum limit", Science **306** (2004).

[99] K. M. Backes, D. A. Palken, S. A. Kenany, B. M. Brubaker, S. B. Cahn, A. Droster, G. C. Hilton, S. Ghosh, H. Jackson, S. K. Lamoreaux, and et al., "A quantum enhanced search for dark matter axions", Nature **590** (2021).

[100] W. Wasilewski, K. Jensen, H. Krauter, J. J. Renema, M. V. Balabas, and E. S. Polzik, "Quantum noise limited and entanglement-assisted magnetometry", Phys. Rev. Lett. **104** (2010).

[101] I. D. Leroux, M. H. Schleier-Smith, and V. Vuleti ć, "Orientation-dependent entanglement lifetime in a squeezed atomic clock", Phys. Rev. Lett. **104** (2010).

[102] B. K. Malia, J. Martınez-Rincón, Y. Wu, O. Hosten, and M. A. Kasevich, "Free space ramsey spectroscopy in rubidium with noise below the quantum projection limit", Phys. Rev. Lett. **125** (2020).

[103] G. P. Greve, C. Luo, B. Wu, and J. K. Thompson, "Entanglement-enhanced matter-wave interferometry in a high-finesse cavity", Nature **610** (2022).

[104] D. Leibfried, E. Knill, S. Seidelin, J. Britton, R. B. Blakestad, J. Chiaverini, D. B. Hume, W. M. Itano, J. D. Jost, C. Langer, et al., "Creation of a six-atom 'schrödinger cat' state", Nature **438** (2005).

[105] T. Monz, P. Schindler, J. T. Barreiro, M. Chwalla, D. Nigg, W. A. Coish, M. Harlander, W. Hänsel, M. Hennrich, and R. Blatt, "14-qubit entanglement: creation and coherence", Phys. Rev. Lett. **106** (2011).

[106] B. J. Lawrie, P. D. Lett, A. M. Marino, and R. C. Pooser, "Quantum sensing with squeezed light", ACS Photonics **6** (2019).

[107] R. C. Pooser, N. Savino, E. Batson, J. L. Beckey, J. Garcia, and B. J. Lawrie, "Truncated nonlinear interferometry for quantum-enhanced atomic force microscopy", Phys. Rev. Lett. **124** (2020).

[108] C. M. Caves, "Quantum-mechanical noise in an interferometer", Phys. Rev. D **23** (1981).

[109] C. Helstrom, "Minimum mean-squared error of estimates in quantum statistics", Physics Letters A **25** (1967).

[110] A. Fujiwara and H. Nagaoka, "Quantum fisher metric and estimation for pure state models", Physics Letters A **201** (1995).

[111] S. L. Braunstein and C. M. Caves, "Statistical distance and the geometry of quantum states", Physical Review Letters **72** (1994).

[112] J. M. Weiner, K. C. Cox, J. G. Bohnet, Z. Chen, and J. K. Thompson, "Superradiant raman laser magnetometer", Applied Physics Letters **101** (2012).

[113] L. J. Fiderer, J. M. E. Fraisse, and D. Braun, "Maximal quantum fisher information for mixed states", Phys. Rev. Lett. **123** (2019).

[114] J. Clarke and A. I. Braginski, **The squid handbook**, Vol. 1 (Wiley Online Library, 2004).

[115] D. Budker and M. Romalis, "Optical magnetometry", Nature physics **3** (2007).

[116] H. Dang, A. C. Maloof, and M. V. Romalis, "Ultrahigh sensitivity magnetic field and magnetization measurements with an atomic magnetometer", Applied Physics Letters **97** (2010).

[117] C. Degen, "Scanning magnetic field microscope with a diamond single-spin sensor", Applied Physics Letters **92** (2008).

[118] O. Arcizet, V. Jacques, A. Siria, P. Poncharal, P. Vincent, and S. Seidelin, "A single nitrogen-vacancy defect coupled to a nanomechanical oscillator", Nature Physics **7** (2011).

[119] S. Kolkowitz, A. C. B. Jayich, Q. P. Unterreithmeier, S. D. Bennett, P. Rabl, J. Harris, and M. D. Lukin, "Coherent sensing of a mechanical resonator with a single-spin qubit", Science **335** (2012).

[120] M. Bal, C. Deng, J.-L. Orgiazzi, F. Ong, and A. Lupascu, "Ultrasensitive magnetic field detection using a single artificial atom", Nature communications **3** (2012).

[121] W. D. Phillips, W. E. Cooke, and D. Kleppner, "Magnetic moment of the proton in h2o in bohr magnetons", Metrologia **13** (1977).

[122] L. Viola, E. Knill, and S. Lloyd, "Dynamical decoupling of open quantum systems", Physical Review Letters **82** (1999).

[123] L. Viola and E. Knill, "Robust dynamical decoupling of quantum systems with bounded controls", Physical Review Letters **90** (2003).

[124] K. Khodjasteh and L. Viola, "Dynamical quantum error correction of unitary operations with bounded controls", Physical Review A **80** (2009).

[125] L. F. Santos and L. Viola, "Enhanced convergence and robust performance of randomized dynamical decoupling", Physical review letters **97** (2006).

[126] J. Bylander, S. Gustavsson, F. Yan, F. Yoshihara, K. Harrabi, G. Fitch, D. G. Cory, Y. Nakamura, J.-S. Tsai, and W. D. Oliver, "Noise spectroscopy through dynamical decoupling with a superconducting flux qubit", Nature Physics **7** (2011).

[127] S. Meiboom and D. Gill, "Modified spin-echo method for measuring nuclear relaxation times", Review of scientific instruments **29** (1958).

[128] H. Y. Carr and E. M. Purcell, "Effects of diffusion on free precession in nuclear magnetic resonance experiments", Physical review **94** (1954).

[129] Ł. Cywi ński, R. M. Lutchyn, C. P. Nave, and S. Das Sarma, "How to enhance dephasing time in superconducting qubits", Phys. Rev. B **77** (2008).

[130] M. Tsang, H. M. Wiseman, and C. M. Caves, "Fundamental quantum limit to waveform estimation", Physical Review Letters **106** (2011).

[131] S. Pang and A. N. Jordan, "Optimal adaptive control for quantum metrology with time-dependent hamiltonians", Nature Communications **8** (2017).

[132] M. Holland and K. Burnett, "Interferometric detection of optical phase shifts at the heisenberg limit", Physical review letters **71** (1993).

[133] J. J. Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen, "Optimal frequency measurements with maximally correlated states", Physical Review A **54** (1996).

[134] L. Allen and J. H. Eberly, **Optical resonance and two-level atoms**, Vol. 28 (Courier Corporation, 1987).

[135] Z. Chen, J. G. Bohnet, J. M. Weiner, and J. K. Thompson, "General formalism for evaluating the impact of phase noise on bloch vector rotations", Phys. Rev. A **86** (2012).

[136] D. Poulin, A. Qarry, R. Somma, and F. Verstraete, "Quantum simulation of time-dependent hamiltonians and the convenient illusion of hilbert space", Physical review letters **106** (2011).

[137] D. Budker, P. W. Graham, M. Ledbetter, S. Rajendran, and A. O. Sushkov, "Proposal for a cosmic axion spin precession experiment (casper)", Physical Review X **4** (2014).

[138] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, "Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms", Journal of Magnetic Resonance **172** (2005).

[139] J. Liu and H. Yuan, "Quantum parameter estimation with optimal control", Physical Review A **96** (2017).

[140] G. S. Uhrig, "Keeping a quantum bit alive by optimized π-pulse sequences", Phys. Rev. Lett. **98** (2007).

[141] X. Feng and L. Wei, "Quantifying quantum coherence with quantum fisher information", Scientific Reports **7** (2017).

[142] G. Tóth and I. Apellaniz, "Quantum metrology from a quantum information science perspective", Journal of Physics A: Mathematical and Theoretical **47** (2014).

[143] J. Huyghebaert and H. D. Raedt, "Product formula methods for time-dependent schrodinger problems", Journal of Physics A: Mathematical and General **23** (1990).

[144] A. M. Polloreno, R. R. W. Wang, and N. A. Tezak, "A note on noisy reservoir computation", (2023), eprint: arXivpreprintarXiv:2302.10862.

[145] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, "Information processing capacity of dynamical systems", Scientific Reports **2** (2012).

[146] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, "Recent advances in physical reservoir computing: a review", Neural Networks **115** (2019).

[147] A. M. Polloreno, A. Carignan-Dugas, J. Hines, R. Blume-Kohout, K. Young, and T. Proctor, "A theory of direct randomized benchmarking", (2023), eprint: arXiv:2302.13853.

[148] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, "Self-consistent quantum process tomography", Phys. Rev. A **87** (2013).

[149] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, "Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography", Nature communications **8** (2017).

[150] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, "Gate set tomography", arXiv preprint arXiv:2009.07301 (2020).

[151] J. Emerson, R. Alicki, and K. Życzkowski, "Scalable noise estimation with random unitary operators", J. Opt. B Quantum Semiclass. Opt. **7** (2005).

[152] 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** (2007).

[153] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. Blakestad, J. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. Wineland, "Randomized benchmarking of quantum gates", Phys. Rev. A **77** (2008).

[154] E. Magesan, J. M. Gambetta, and J. Emerson, "Scalable and robust randomized benchmarking of quantum processes", Phys. Rev. Lett. **106** (2011).

[155] E. Magesan, J. M. Gambetta, and J. Emerson, "Characterizing quantum gates via randomized benchmarking", Phys. Rev. A **85** (2012).

[156] A. Carignan-Dugas, J. J. Wallman, and J. Emerson, "Characterizing universal gate sets via dihedral benchmarking", Phys. Rev. A **92** (2015).

[157] A. W. Cross, E. Magesan, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, "Scalable randomised benchmarking of non-clifford gates", NPJ Quantum Inf. **2** (2016).

[158] W. G. Brown and B. Eastin, "Randomized benchmarking with restricted gate sets", Phys. Rev. A **97** (2018).

[159] A. K. Hashagen, S. T. Flammia, D. Gross, and J. J. Wallman, "Real randomized benchmarking", Quantum **2** (2018).

[160] 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, et al., "Efficient measurement of quantum gate error by interleaved randomized benchmarking", Phys. Rev. Lett. **109** (2012).

[161] J. Helsen, X. Xue, L. M. Vandersypen, and S. Wehner, "A new class of efficient randomized benchmarking protocols", arXiv preprint arXiv:1806.02048 (2018).

[162] J. Helsen, S. Nezami, M. Reagor, and M. Walter, "Matchgate benchmarking: scalable benchmarking of a continuous family of many-qubit gates", (2020), arXiv:2011.13048 [quant-ph].

[163] J. Helsen, I. Roth, E. Onorati, A. H. Werner, and J. Eisert, "A general framework for randomized benchmarking", (2020), arXiv:2010.07974 [quant-ph].

[164] J. Claes, E. Rieffel, and Z. Wang, "Character randomized benchmarking for non-multiplicity-free groups with applications to subspace, leakage, and matchgate randomized benchmarking", (2020), arXiv:2011.00007 [quant-ph].

[165] D. C. McKay, A. W. Cross, C. J. Wood, and J. M. Gambetta, "Correlated randomized benchmarking", (2020), arXiv:2003.02354 [quant-ph].

[166] J. Helsen, J. J. Wallman, S. T. Flammia, and S. Wehner, "Multi-qubit randomized benchmarking using few samples", arXiv preprint arXiv:1701.04299 (2017).

[167] T. J. Proctor, A. Carignan-Dugas, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young, "Direct randomized benchmarking for multiqubit devices", Phys. Rev. Lett. **123** (2019).

[168] T. Proctor, K. Rudinger, K. Young, M. Sarovar, and R. Blume-Kohout, "What randomized benchmarking actually measures", Phys. Rev. Lett. **119** (2017).

[169] J. J. Wallman, "Randomized benchmarking with gate-dependent noise", Quantum **2** (2018).

[170] M. Ozols, "Clifford group", Essays at University of Waterloo, Spring (2008).

[171] R. Koenig and J. A. Smolin, "How to efficiently select an arbitrary Clifford group element", J. Math. Phys. **55** (2014).

[172] S. Aaronson and D. Gottesman, "Improved simulation of stabilizer circuits", Phys. Rev. A **70** (2004).

[173] K. N. Patel, I. L. Markov, and J. P. Hayes, "Efficient synthesis of linear reversible circuits", Quantum Inf. Comput. **8** (2008).

[174] S. Bravyi and D. Maslov, "Hadamard-free circuits expose the structure of the clifford group", arXiv preprint arXiv:2003.09412 (2020).

[175] J. Yoneda, K. Takeda, T. Otsuka, T. Nakajima, M. R. Delbecq, G. Allison, T. Honda, T. Kodera, S. Oda, Y. Hoshi, et al., "A quantum-dot spin qubit with coherence limited by charge noise and fidelity higher than 99.9%", Nat. Nanotechnol. **13** (2018).

[176] 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** (2018).

[177] 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** (2018).

[178] J. M. Nichol, L. A. Orona, S. P. Harvey, S. Fallahi, G. C. Gardner, M. J. Manfra, and A. Yacoby, "High-fidelity entangling gate for double-quantum-dot spin qubits", NPJ Quantum Inf. **3** (2017).

[179] M. Veldhorst, J. Hwang, C. Yang, A. Leenstra, B. De Ronde, J. Dehollain, J. Muhonen, F. Hudson, K. M. Itoh, A. Morello, et al., "An addressable quantum dot qubit with fault-tolerant control-fidelity", Nat. Nanotechnol. **9** (2014).

[180] 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", Phys. Rev. A **87** (2013).

[181] T. Xia, M. Lichtman, K. Maller, A. Carr, M. Piotrowicz, L. Isenhower, and M. Saffman, "Randomized benchmarking of single-qubit gates in a 2D array of neutral-atom qubits", Phys. Rev. Lett. **114** (2015).

[182] A. Córcoles, E. Magesan, S. J. Srinivasan, A. W. Cross, M. Steffen, J. M. Gambetta, and J. M. Chow, "Demonstration of a quantum error detection code using a square lattice of four superconducting qubits", Nat. Commun. **6** (2015).

[183] 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", Phys. Rev. Lett. **116** (2016).

[184] J. Muhonen, A. Laucht, S. Simmons, J. Dehollain, R. Kalra, F. Hudson, S. Freer, K. M. Itoh, D. Jamieson, J. McCallum, et al., "Quantifying the quantum gate fidelity of single-atom spin qubits in silicon by randomized benchmarking", J. Phys. Condens. Matter **27** (2015).

[185] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. White, J. Mutus, A. Fowler, B. Campbell, et al., "Superconducting quantum circuits at the surface code threshold for fault tolerance", Nature **508** (2014).

[186] J. Raftery, A. Vrajitoarea, G. Zhang, Z. Leng, S. Srinivasan, and A. Houck, "Direct digital synthesis of microwave waveforms for quantum computing", arXiv preprint arXiv:1703.00942 (2017).

[187] M. Rol, C. Bultink, T. O'Brien, S. de Jong, L. Theis, X. Fu, F. Luthi, R. Vermeulen, J. de Sterke, A. Bruno, et al., "Restless tuneup of high-fidelity qubit gates", Phys. Rev. Appl. **7** (2017).

[188] J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. Fowler, I.-C. Hoi, E. Jeffrey, et al., "Optimal quantum control using randomized benchmarking", Phys. Rev. Lett. **112** (2014).

[189] D. C. McKay, S. Sheldon, J. A. Smolin, J. M. Chow, and J. M. Gambetta, "Three qubit randomized benchmarking", Phys. Rev. Lett. **122** (2019).

[190] T. Proctor, S. Seritan, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young, "Scalable randomized benchmarking of quantum computers using mirror circuits", arXiv preprint arXiv:2112.09853 (2021).

[191] D. Gross, K. Audenaert, and J. Eisert, "Evenly distributed unitaries: on the structure of unitary designs", J. Math. Phys. **48** (2007).

[192] C. Dankert, R. Cleve, J. Emerson, and E. Livine, "Exact and approximate unitary 2-designs and their application to fidelity estimation", Phys. Rev. A **80** (2009).

[193] J. J. Wallman and S. T. Flammia, "Randomized benchmarking with confidence", New J. Phys. **16** (2014).

[194] J. M. Epstein, A. W. Cross, E. Magesan, and J. M. Gambetta, "Investigating the limits of randomized benchmarking protocols", Phys. Rev. A **89** (2014).

[195] J. Helsen, I. Roth, E. Onorati, A. H. Werner, and J. Eisert, "A general framework for randomized benchmarking", arXiv preprint arXiv:2010.07974 (2020).

[196] D. S. França and A.-L. Hashagen, "Approximate randomized benchmarking for finite groups", arXiv preprint arXiv:1803.03621 (2018).

[197] C. Ryan, M. Laforest, and R. Laflamme, "Randomized benchmarking of single-and multi-qubit control in liquid-state nmr quantum information processing", New J. Phys. **11** (2009).

[198] E. Knill, "Quantum computing with realistically noisy devices", Nature **434** (2005).

[199] M. Ware, G. Ribeill, D. Ristè, C. A. Ryan, B. Johnson, and M. P. da Silva, "Experimental demonstration of pauli-frame randomization on a superconducting qubit", (2018), arXiv:1803.01818 [quant-ph].

[200] J. J. Wallman and J. Emerson, "Noise tailoring for scalable quantum computation via randomized compiling", Phys. Rev. A **94** (2016).

[201] A. Hashim, R. K. Naik, A. Morvan, J.-L. Ville, B. Mitchell, J. M. Kreikebaum, M. Davis, E. Smith, C. Iancu, K. P. O'Brien, I. Hincks, J. J. Wallman, J. Emerson, and I. Siddiqi, "Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor", (2020), arXiv:2010.00215 [quant-ph].

[202] W. Brown and O. Fawzi, "Decoupling with random quantum circuits", Commun. Math. Phys. **340** (2015).

[203] Y. Sekino and L. Susskind, "Fast scramblers", J. High Energy Phys. **2008** (2008).

[204] F. G. Brandão, A. W. Harrow, and M. Horodecki, "Local random quantum circuits are approximate polynomial-designs", Commun. Math. Phys. **346** (2016).

[205] M. Oszmaniec, A. Sawicki, and M. Horodecki, "Epsilon-Nets, unitary designs, and random quantum circuits", IEEE Transactions on Information Theory **68** (2022).

[206] N. Hunter-Jones, "Operator growth in random quantum circuits with symmetry", arXiv preprint arXiv:1812.08219 (2018).

[207] C. W. von Keyserlingk, T. Rakovszky, F. Pollmann, and S. L. Sondhi, "Operator hydrodynamics, OTOCs, and entanglement growth in systems without conservation laws", Physical Review X **8** (2018).

[208] A. Nahum, S. Vijay, and J. Haah, "Operator spreading in random unitary circuits", Physical Review X **8** (2018).

[209] 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).

[210] A. Carignan-Dugas, K. Boone, J. J. Wallman, and J. Emerson, "From randomized benchmarking experiments to gate-set circuit fidelity: how to interpret randomized benchmarking decay parameters", New J. Phys. **20** (2018).

[211] M. A. Nielsen, "A simple formula for the average gate fidelity of a quantum dynamical operation", Phys. Lett. A **303** (2002).

[212] S. Mavadia, C. Edmunds, C. Hempel, H. Ball, F. Roy, T. Stace, and M. Biercuk, "Experimental quantum verification in the presence of temporally correlated noise", NPJ Quantum Inf. **4** (2018).

[213] M. Fogarty, M. Veldhorst, R. Harper, C. Yang, S. Bartlett, S. Flammia, and A. Dzurak, "Nonexponential fidelity decay in randomized benchmarking with low-frequency noise", Phys. Rev. A **92** (2015).

[214] B. H. Fong and S. T. Merkel, "Randomized benchmarking, correlated noise, and ising models", arXiv preprint arXiv:1703.09747 (2017).

[215] T. Proctor, M. Revelle, E. Nielsen, K. Rudinger, D. Lobser, P. Maunz, R. Blume-Kohout, and K. Young, "Detecting and tracking drift in quantum information processors", Nat. Commun. **11** (2020).

[216] C. Granade, C. Ferrie, and D. Cory, "Accelerated randomized benchmarking", New J. Phys. **17** (2015).

[217] I. Hincks, J. J. Wallman, C. Ferrie, C. Granade, and D. G. Cory, "Bayesian inference for randomized benchmarking protocols.", arXiv preprint arXiv:1802.00401 (2018).

[218] G. Vidal and C. M. Dawson, "Universal quantum circuit for two-qubit transformations with three controlled-NOT gates", Phys. Rev. A **69** (2004).

[219] V. V. Shende, S. S. Bullock, and I. L. Markov, "Recognizing small-circuit structure in two-qubit operators", Phys. Rev. A **70** (2004).

[220] S. Garion, N. Kanazawa, H. Landa, D. C. McKay, S. Sheldon, A. W. Cross, and C. J. Wood, "Experimental implementation of non-clifford interleaved randomized benchmarking with a controlled-s gate", (2020), arXiv:2007.08532 [quant-ph].

[221] J. Hines, M. Lu, R. K. Naik, A. Hashim, J.-L. Ville, B. Mitchell, J. M. Kriekebaum, D. I. Santiago, S. Seritan, E. Nielsen, et al., "Demonstrating scalable randomized benchmarking of universal gate sets", arXiv preprint arXiv:2207.07272 (2022).

[222] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, "Characterizing quantum supremacy in near-term devices", Nat. Phys. **14** (2018).

[223] T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, "Measuring the capabilities of quantum computers", arXiv preprint arXiv:2008.11294 (2020).

[224] E. Nielsen, R. Blume-Kohout, L. Saldyt, J. Gross, T. Scholten, K. Rudinger, T. Proctor, and J. K. Gamble, **PyGSTi version 0.9.7.2**, 10.5281/zenodo.2598776, Mar. 2019.

[225] R. Harper, I. Hincks, C. Ferrie, S. T. Flammia, and J. J. Wallman, "Statistical analysis of randomized benchmarking", Physical Review A **99** (2019).

[226] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, "Characterizing large-scale quantum computers via cycle benchmarking", Nat. Commun. **10** (2019).

[227] K. Boone, A. Carignan-Dugas, J. J. Wallman, and J. Emerson, "Randomized benchmarking under different gatesets", arXiv preprint arXiv:1811.01920 (2018).

[228] P. Diaconis, **Group representations in probability and statistics**, edited by S. S. Gupta, Vol. 11, Lecture Notes-Monograph Series (Institute of Mathematical Statistics, Hayways, CA, 1988).

[229] P. Diaconis and L. Saloff-Coste, "Random walks on finite groups: a survey of analytic techniques", Probability measures on groups and related structures, XI (Oberwolfach, 1994) (1995).

[230] L. Saloff-Coste, "Random walks on finite groups", in **Probability on discrete structures** (Springer, 2004), pp. 263–346.

[231] D. Greenbaum, "Introduction to quantum gate set tomography", arXiv preprint arXiv:1509.02921 (2015).

[232] R. Blume-Kohout, J. K. Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, "Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit", arXiv preprint arXiv:1310.4492 (2013).

[233] T. Chasseur and F. Wilhelm, "Complete randomized benchmarking protocol accounting for leakage errors", Phys. Rev. A **92** (2015).

[234] R. Blume-Kohout, M. P. da Silva, E. Nielsen, T. Proctor, K. Rudinger, M. Sarovar, and K. Young, "A taxonomy of small markovian errors", PRX Quantum **3** (2022).

[235] R. Harper and S. T. Flammia, "Estimating the fidelity of T gates using standard interleaved randomized benchmarking", Quantum Sci. Technol. **2** (2017).

[236] T. Chasseur, D. M. Reich, C. P. Koch, and F. K. Wilhelm, "Hybrid benchmarking of arbitrary quantum gates", Phys. Rev. A **95** (2017).

[237] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, "Estimating the coherence of noise", New J. Phys. **17** (2015).

[238] G. Feng, J. J. Wallman, B. Buonacorsi, F. H. Cho, D. K. Park, T. Xin, D. Lu, J. Baugh, and R. Laflamme, "Estimating the coherence of noise in quantum control of a solid-state qubit", Phys. Rev. Lett. **117** (2016).

[239] S. Sheldon, L. S. Bishop, E. Magesan, S. Filipp, J. M. Chow, and J. M. Gambetta, "Characterizing errors on qubit operations via iterative randomized benchmarking", Phys. Rev. A **93** (2016).

[240] C. J. Wood and J. M. Gambetta, "Quantification and characterization of leakage errors", Phys. Rev. A **97** (2018).

[241] J. J. Wallman, M. Barnhill, and J. Emerson, "Robust characterization of loss rates", Phys. Rev. Lett. **115** (2015).

[242] J. M. Gambetta, A. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, et al., "Characterization of addressability by simultaneous randomized benchmarking", Phys. Rev. Lett. **109** (2012).

[243] K. Mayer, A. Hall, T. Gatterman, S. K. Halit, K. Lee, J. Bohnet, D. Gresh, A. Hankin, K. Gilmore, and J. Gaebler, "Theory of mirror benchmarking and demonstration on a quantum computer", arXiv preprint arXiv:2108.10431 (2021).

[244] R. Blume-Kohout, M. P. da Silva, E. Nielsen, T. Proctor, K. Rudinger, M. Sarovar, and K. Young, "A taxonomy of small markovian errors", arXiv preprint arXiv:2103.01928 (2021).

[245] A. M. Polloreno and K. C. Young, "Robustly decorrelating errors with mixed quantum gates", Quantum Science and Technology **7** (2022).

[246] A. Polloreno, A. Carignan-Dugas, K. Young, R. Blume-Kohout, and T. Proctor, "A theory of direct randomized benchmarking", (2023).

[247] M. T. Mądzik, S. Asaad, A. Youssry, B. Joecker, K. M. Rudinger, E. Nielsen, K. C. Young, T. J. Proctor, A. D. Baczewski, A. Laucht, et al., "Precision tomography of a three-qubit donor quantum processor in silicon", Nature **601** (2022).

[248] A. J. Landahl, J. T. Anderson, and P. R. Rice, **Fault-tolerant quantum computing with color codes**, 2011, eprint: arXiv:1108.5738.

[249] S. Lloyd, "Almost any quantum logic gate is universal", Physical review letters **75** (1995).