
A walk through quantum noise: a study of error signatures and characterization methods

by

Arnaud CARIGNAN-DUGAS

*A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement of a
Doctor of Philosophy*

in

Applied Mathematics

Waterloo, Ontario, Canada, 2019

©Arnaud Carignan-Dugas 2019

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Prof. Jens Eisert
Supervisor:	Prof. Joseph Emerson
Internal Member:	Prof. Achim Kempf
Internal-External Member:	Prof. David Cory
Other Member:	Prof. Joel Wallman

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This thesis contains literal transcriptions of articles for which my contribution was major, by which I mean that I heavily contributed to the derivation of central results. The chapters based on those articles begin with a short contribution statement and a reference to the work in question.

Abstract

The construction of large scale quantum computing devices might be one of the most exciting and promising endeavors of the 21st century, but it also comes with many challenges. As quantum computers are supplemented with more registers, their error profile generally grows in complexity, rendering the enterprise of quantifying the reliability of quantum computations increasingly difficult through naive characterization techniques. In the last decade, a lot of efforts has been directed toward developing highly scalable benchmarking schemes. A leading family of characterization methods built upon scalable principles is known as randomized benchmarking (RB).

In this thesis, many tools are presented with the objective of improving the scalability, and versatility of RB techniques, as well as demonstrating their reliability under various error models.

The first part of this work investigates the connection between the error of individual circuit components and the error of their composition. Before reasoning about intricate circuit constructions, it is shown that there exists a well-motivated way to define decoherent quantum channels, and that every channel can be factorized into a unitary-decoherent composition. This dichotomy carries to the circuit evolution of important error parameters by assuming realistic error scenarios. Those results are used to improve the confidence interval of RB diagnoses and to reconcile experimentally estimated parameters with physically and operationally meaningful quantities.

In the second part of this thesis, various RB schemes are either developed or more rigorously analyzed. A first result consists of the introduction of “dihedral benchmarking”, a technique which, if performed in conjunction with standard RB protocols, enables the characterization of operations that form a universal gate-set. Finally, rigorous analysis tools are provided to demonstrate the reliability of a highly scalable family of generator-based RB protocols known as direct RB.

Acknowledgements

First, I would like to thank Joseph Emerson and Joel J. Wallman who both provided me with guidance during the past years. I would also like to thank Timothy J. Proctor for being an understanding and insightful research collaborator.

All the rest of my acknowledgments minus one go to my family, friends and academic peers, who supported me directly or indirectly: Gigi, Lulu, Max, Tobi, Bohdan, Beenie, Bouhmi, Caro, Cary, Dan, David, Hammam, Hina, Jeremy, Joel, John, John, Jonas, JP, Katie, Lindsay, Matthew, Piers, Sasha, Tamara, Thom, Yuval, and many more.

Finally, I would like to thank Margot (and her garden) who provided me support, and sometimes food.

Table of Contents

Examining Committee Membership	ii
Author's Declaration	iii
Statement of Contributions	iv
Abstract	v
Acknowledgements	vi
List of Figures	xi
List of Tables	xiv
List of Abbreviations	xv
List of Symbols	xvi
1 Introduction	1
2 Introductory Material	5
2.1 Inner product and norms	5
2.1.1 Common operator bases	6
2.2 Quantum states and measurements	7
2.3 Quantum operations	8
2.3.1 Definition	8
2.4 Mapping abstract quantum operations to more familiar sets	8
2.4.1 Process matrix (Liouville representation)	8
2.4.2 Chi matrix	11
2.4.3 Choi matrix	11
2.4.4 Canonical Kraus decomposition	12
2.5 Elementary types of quantum channels	12
2.5.1 Unitary errors	12
2.5.2 Convexity of the set of CPTP maps	14
2.5.3 Incoherent channels	14
2.5.4 Stochastic errors	14
2.5.5 Depolarizing errors	15
2.5.6 Dephasing/phase damping errors	15
2.5.7 Amplitude damping channels (population transfer)	16
2.6 Quantum circuits and the Markovian assumption	17

2.7	Characterizing quantum operations	19
2.8	Reference bases and gauge transformations	21
I	Error signatures in quantum circuits	23
3	Bounding the average gate fidelity of composite channels using the unitarity	25
3.1	Foreword	25
3.2	Compendium	25
3.3	Introduction	26
3.4	Noisy quantum processes	27
3.5	Composite infidelities in terms of component infidelities	29
3.6	Improved bounds on the infidelity using the unitarity	32
3.7	Application: Interleaved RB	35
3.8	Summary and outlook	38
3.9	Proofs	39
3.9.1	Matrix inequalities on the real field	39
3.10	Afterword	41
4	A polar decomposition for quantum channels	43
4.1	Foreword	43
4.2	Compendium	43
4.3	Introduction	43
4.4	Channel properties captured by the leading Kraus operator	46
4.5	The LK approximation and two evolution theorems	51
4.6	A polar decomposition for quantum channels	53
4.6.1	Defining decoherence	53
4.6.2	The dynamics induced from decoherent channels as infinitesimal generators	55
4.6.3	Further justifying our notion of decoherence	57
4.7	Behavioral signatures of coherence and decoherence	58
4.7.1	Extremal dephasers, extremal unitaries, and equable error channels	59
	Extremal dephasers	59
	Extremal unitaries	59
	Equable error channels	61
4.7.2	Reasoning about Y	62
4.7.3	The coherence level	65
4.7.4	Bounding the worst and best case fidelity of a circuit	66
4.7.5	Decoherence-limited operations	68
4.7.6	Limitations	69
4.8	Conclusion	70
4.9	Proofs	74
4.9.1	A noteworthy trace inequality	74
4.9.2	Proofs of the main results	76
	Notation and remarks	76
	Proof of the evolution theorem 7	76
	Proof of the evolution theorem 8	77

Proof of theorem 12	82
Proof of theorem 11	84
Proof of theorem 13	87
Proof of theorem 10	88
Proof of theorem 14	89
Proof of theorem 15	92
4.10 Afterword	93
4.10.1 A bound on the spectral radius of channel deviations	93
4.10.2 A bound on the diamond distance	97
4.10.3 End of the first part	100
II Quantum characterization through randomized benchmarking	101
5 From randomized benchmarking experiments to gate-set circuit fidelity: how to interpret randomized benchmarking decay parameters	103
5.1 Foreword	103
5.2 Compendium	103
5.3 Introduction	103
5.4 The dynamics of the gate-set circuit fidelity	105
5.5 Finding the appropriate set of targeted gate realizations for specific noise models	109
5.6 Conclusion	111
5.7 Supplementary material	112
5.7.1 An expression for the total change in the gate-set circuit fidelity	112
5.7.2 Varying the ideal gate-set of comparison	114
5.8 Afterword	120
5.8.1 Extending the results to all dimensions	120
5.8.2 Comparing data from different reference bases	122
6 Characterizing Universal Gate Sets via Dihedral Benchmarking	123
6.1 Foreword	123
6.2 Compendium	123
6.3 Introduction	123
6.4 Characterizing single-qubit unitary groups	125
6.5 Characterizing the $\pi/8$ gate	126
6.6 Analysis	128
6.7 Numerical simulation	130
6.8 Conclusion	132
6.9 Afterword	133
6.9.1 Group-based randomized benchmarking	133
6.9.2 Generator-based randomized benchmarking	134
7 Generator-based randomized benchmarking	137
7.1 Foreword	137
7.2 Randomized Benchmarking under Different Gate-sets	137
7.2.1 Compendium	137
7.2.2 Introduction	138

7.2.3	Background and Motivation	138
7.2.4	Results	141
	Standard RB vs NIST RB	141
	Theoretical Analysis of NIST RB under a gate-independent error model	144
7.3	Direct randomized benchmarking	148
7.3.1	The protocol	148
7.3.2	Generalizing twirls	150
	The Clifford twirl	151
7.3.3	The convergence of sequence-asymptotic 2-designs	153
7.3.4	Noisy twirls over sequence-asymptotic 2-designs	155
	Acceptable regime for direct RB	155
	Interpretation of the decay constant p	156
7.3.5	A signal analysis for direct RB under gate-dependent effects	157
8	Summary and conclusion	161
	Bibliography	165

List of Figures

- 3.1 (Color online) The average survival probability, $p_{\text{surv.}}(m) = |\mathbf{G}|^{-m} \sum_i \langle 0 | \mathcal{S}_m^{(i)} (|0\rangle\langle 0|) |0\rangle$ over all sequences $\mathcal{S}_m^{(i)}$ of length m , as a function of the sequence length for two simulated interleaved RB experiments (see Ref. [Mag+12] for more details) with two different individual gate errors \mathcal{E}_h , but a common average error \mathcal{E} with fidelity 0.9975. Orange squares represent an error model with high fidelity ($F(\mathcal{E}_h) = 0.9991$) that interacts coherently with \mathcal{E} . Blue triangles represent an error model with lower fidelity ($F(\mathcal{E}_h) = 0.9975$), but that is purely stochastic. See section 3.7 for more details. 37
- 3.2 (Color online) Bounds on the fidelity $F(\mathcal{E}_h)$ of an individual gate h as a function of the composite $F(\mathcal{E}_h\mathcal{E})$ with $F(\mathcal{E})$ fixed and varying $u(\mathcal{E})$: a) $u(\mathcal{E}) = 1.00000$, b) $u(\mathcal{E}) = 0.99300$, c) $u(\mathcal{E}) = 0.99030$, d) $u(\mathcal{E}) = 0.99003 \approx p(\mathcal{E})^2$. The numerical data points correspond to the values $F(\mathcal{E}_h)$ and $F(\mathcal{E}_h\mathcal{E})$ for randomly-generated channels $\{\mathcal{E}_h, \mathcal{E}\}$ satisfying $F(\mathcal{E}) = 0.9975$ and with the appropriate value of $u(\mathcal{E})$. As illustrated by the color, the unitarity $u(\mathcal{E}_h)$ is minimal in the center of the shaded region and maximal when the data points approach our bound. 37
- 3.3 (Color online) Bounds on various 2-qubit gate infidelities $\delta F(\mathcal{E}_h)$ based on various experimental data [Gae+12; Cór+13; Kel+14; Bar+14; Cas+16; McK+16; She+16; Tak+16b; McK+17b]. The time refers to the dates of submission. The top plot uses eq. (3.42) with a maximal coherence angle $\theta(\mathcal{E})$, which yields in bounds spanning up to two orders of magnitude. The bottom plot assumes a purely stochastic error model, by which we mean that $u(\mathcal{E}) \approx p(\mathcal{E})^2$. For every data point, some statistical error is taken into account, hence the non-zero error bars in the bottom plot. 38
- 4.1 Representation of the spiraling action of a normal matrix acting on a 2×2 subspace. The polar decomposition, in this case, separates the azimuthal and radial components of the action. Quantum dynamics on $d > 2$ can generate spiraling actions on the Bloch space for which the rotation factor can't be interpreted as a physical unitary operation. In this sense, spiraling, despite generating some rotating action, is inherent to some decoherent dynamics. . . 54

- 4.2 Singular values σ_i – plotted as purple circles – of the $10^3 \times 10^3$ LK operator A_1 of an extremal dephaser \mathcal{A} . The dashed line corresponds to the average $\mathbb{E}_i[\sigma_i] = 0.9989(1)$. The green shaded region covers a standard deviation $\text{SD}[\sigma_i] = 0.0032(1)$ below the average. In this example, the standard deviation is roughly three times greater than the average deviation $1 - \mathbb{E}_i[\sigma_i] = 0.0011(1)$; that is, the WSE decoherence constant (see definition 7) is $\gamma_{\text{decoh}} = 3.0(1)$, which is an order of magnitude smaller than $1/\sqrt{\mathbb{E}[1 - \sigma_i]} = 30.7(1)$. From eq. (4.53), \mathcal{A} is equable in the wide-sense. There are five singular values situated around 0.955, meaning that $1 - \sigma_j$ can be more than forty times larger than the average deviation (i.e. $\Gamma_{\text{decoh}} = 41(1)$). While these extreme deviations are excluded by the equability condition, their small impact on the standard deviation allows \mathcal{A} to be WSE. 60
- 4.3 Process fidelity of three different error channel compositions as a function of the circuit length. The hard lines correspond to the three process fidelities $\Phi(\mathcal{A}_{m:1}, \mathcal{I})$, and the dotted lines correspond to the bounds given by theorem 14. The color map illustrates the margin of freedom given by the RHS of eq. (4.64), and the gray shaded area in the bottom plot corresponds to the limits of the top plot. In the top figure, shade variations indicate increments of 10^{-4} , and in the bottom figure, increments of 10^{-2} . The individual channels are of the form $\mathcal{A}_i = \mathcal{V} \circ \mathcal{D}$, where \mathcal{D} is a dephasing channel with $\Phi(\mathcal{D}, \mathcal{I}) = 10^{-4}$, and \mathcal{V} is a small unitary error. The dashed line is the decaying envelope $\Phi^m(\mathcal{D}, \mathcal{I})$. The three compositions differ by the level of coherence of their elements \mathcal{A}_i , which are 10%, 1% and 0.01%. The lowest level of coherence corresponds to a decoherence-limited scenario, in which case the decoherent envelope stays within the bounds. 72
- 5.1 Absolute value of the deviation $\delta(m, \mathcal{V}) \equiv \epsilon(m, V)$ (the label of the y -axis corresponds to the original notation appearing in [Car+18]), described in eq. (5.14) (also see eq. (5.37)), as function of circuit length m with noise model generated by $\Lambda(X) = L(P(\sigma_z, 10^{-1}))L(X)$ and $\Lambda(Y) = L(P(\sigma_z, 10^{-1}))L(Y)$, $\Lambda(CZ) = L(P(\sigma_z^1 \sigma_z^2 - \sigma_z^1 - \sigma_z^2, \pi/2 + 10^{-1}))$ (see eq. (5.21)). The red triangles are obtained with the choice of basis $V = I$, while the blue circles are obtained with the choice $V = U$ where U is found through eq. (5.20). The purple horizontal dashed line corresponds to $(1 - p)^2$, while the full green line corresponds to $(1 - F(\Lambda(\mathbf{G}), L(\mathbf{G}), 1))^2$. For both ideal gate-sets $L(\mathbf{G})$ and $L(UGU^\dagger)$, the deviation becomes quickly negligible as the sequence length increases. In fact, in the case $V = U$ (blue circles), the deviation is always below $(1 - p)^2$ 115

- 5.2 gate-set circuit fidelity $F(\Lambda(\mathbf{G}), L(\mathcal{V}\mathbf{G}\mathcal{V}^\dagger), m) \equiv \mathcal{F}(\tilde{\mathbf{G}}, \mathcal{V}\mathbf{G}\mathcal{V}^\dagger, m)$ (the label of the y -axis corresponds to the original notation appearing in [Car+18]) as function of circuit length m with noise model generated by $\Lambda(X) = L(P(\sigma_z, 10^{-1}))L(X)$ and $\Lambda(Y) = L(P(\sigma_z, 10^{-1}))L(Y)$, $\Lambda(CZ) = L(P(\sigma_z^1\sigma_z^2 - \sigma_z^1 - \sigma_z^2, \pi/2 + 10^{-1}))$ (see eq. (5.21)). The different colors portray choices of basis; the yellow circles $V = I$, the blue stars $V = U$ where U is found through eq. (5.20), and the green squares $V = U^2$. Here the lines correspond to the fit for sequence lengths of $m=5$ to 10. The choice $V = U$ produces the evolution prescribed by theorem 18, which through extrapolation has an intercept of 1. 116
- 5.3 $1 - \mathcal{F}(\tilde{\mathbf{G}}, \mathcal{V}L(\mathbf{G})\mathcal{V}^\dagger, m = 1) \equiv \delta F(\Lambda(\mathbf{G}), L(\mathcal{V}\mathbf{G}\mathcal{V}^\dagger), m = 1)$ (the label of the y -axis corresponds to the original notation appearing in [Car+18]) as function of circuit length m with noise model generated by $\Lambda(X) = L(P(\sigma_z, 10^{-1}))L(X)$ and $\Lambda(Y) = L(P(\sigma_z, 10^{-1}))L(Y)$, $\Lambda(CZ) = L(P(\sigma_z^1\sigma_z^2 - \sigma_z^1 - \sigma_z^2, \pi/2 + 10^{-1}))$ (see eq. (5.21)), with $\mathcal{V} = \mathcal{I}$ (green squares) and $\mathcal{V} = \mathcal{U}$ (blue circles) where \mathcal{U} is found through eq. (5.20). The red crosses correspond to $(1 - p)/2$ obtained through RB experiments. 117
- 6.1 The orbit, under the action of the dihedral group \mathbf{D}_8 , of an input state located at a 45° degree latitude on the Bloch sphere. $R_8(z)$ are the rotations of the octagon, while X is a reflection (or a rotation in 3 dimensions, with the rotation axis parallel to the octagon's surface). The $\pi/8$ gate corresponds to the smallest rotation $R_8(1)$ 125
- 6.2 (Color online) Decay curves corresponding to eqns. 6.5 and 6.6 for a standard randomized benchmarking simulation with $A_2 \approx \frac{1}{4}$ and $A_1 \approx \frac{1}{2}$ respectively. The shallow (blue) and steep (orange) lines correspond to eqn. 6.5 and eqn. 6.6 respectively. Each data point is obtained after averaging 500 sequences of fixed length. A weighted non-linear regression (performed using the scipy Python package) gives an estimate of 0.9924(1) for the average fidelity over \mathbf{D}_8 , compared to the analytic value 0.9925. See *Numerical simulation* for details. 131
- 6.3 Decay curves corresponding to eqns. 6.5 and 6.6 for an interleaved randomized benchmarking simulation with $A_2 \approx \frac{1}{4}$ and $A_1 \approx \frac{1}{2}$ respectively. The shallow (blue) and steep (orange) lines correspond to eqn. 6.5 and eqn. 6.6 respectively. Each data point is obtained after averaging 500 sequences of fixed length. Figure *a*) corresponds to high fidelity Clifford operations and a relatively noisy $\pi/8$ gate. Figure *b*) corresponds to errors of the same magnitude on the Clifford operations and the $\pi/8$ gate. See *Numerical simulation* for details. 132
- 7.1 Probability distribution over the Clifford gates \mathbf{C} (labelled as in [Bar+14]) after m gates (i.e clock cycles) of NIST RB drawn uniformly at random from $\mathbf{N} \subset \mathbf{C}$. This leads to a *non-uniform* sampling over the Cliffords that varies as m increases. Asymptotically, for a sequence of even (or odd) length, the probability distribution tends toward a uniform distribution over \mathbf{C}_{12} (or $\sqrt{\mathbf{Z}}\mathbf{C}_{12}$). The grey line indicates an equal probability over the full 24 Clifford group \mathbf{C} . 143

List of Tables

3.1	Linear relations between the fidelity (F), the infidelity ($r \equiv \delta F$), the effective depolarizing constant (p), and $\Phi \equiv \chi_{00}$	29
4.1	Categorization of different well-known error channels. Many canonical error mechanisms fall under the “decoherent” appellation, except for unitary errors, of course. The coherence level is negligible for decoherent channels, and 1 for coherent errors. In the intermediate regime, the coherence level can vary between 0 and 1. It only makes sense to discuss about the coherence level when errors are equable (at least in the wide sense).	62
4.2	Summary of the main concepts addressed in this paper.	73

List of Abbreviations

CP	Completely-Positive
H.O.T.	Higher Order Terms
IFF	IF and only iF
LHS	Left Hand Side
POVM	Positive Operator-Valued Measure
RHS	Right Hand Side
SSE	Strict-Sense Equable
SPAM	State Preparation And Measurement
S.T.	Such That
TP	Trace-Preserving
WOLOG	WithOut Loss Of Generality
W.R.T.	With Respect To
WSE	Wide-Sense Equable

List of Symbols

$M_d(\mathbb{C})$	$d \times d$ matrices with complex entries.
$H_d(\mathbb{C})$	$d \times d$ Hermitian matrices.
$\varrho_d(\mathbb{C})$	$d \times d$ density matrices (i.e. $\{\rho \in H_d(\mathbb{C}) \mid \rho \geq 0, \text{Tr} \rho = 1\}$)
CPTP_d	Completely positive, trace preserving maps acting on $M_d(\mathbb{C})$.
$\ \cdot\ _p$	Schatten p -norm.
$\ \cdot\ _{p \rightarrow p}$	Induced p -norm.
$ \cdot $	Absolute value. For a matrix $A \in M_d(\mathbb{C})$, $ A = \sqrt{A^\dagger A}$.
\dagger	Dual operator. For matrices, it corresponds to the conjugate transpose.
$*$	Complex conjugate.
T	Transposition.
Tr	Trace operator, $\text{Tr} : M_d(\mathbb{C}) \rightarrow \mathbb{C}$. Given $A \in M_d(\mathbb{C})$, $\text{Tr} A := \sum_i [A]_{ii}$.
$ i\rangle$	Ket vector; notation for the i^{th} canonical vector in \mathbb{C}^d .
$\langle i $	Bra vector; notation for the dual of $ i\rangle$. $\langle i = i\rangle^\dagger$
E_{ij}	Shorthand for $ i\rangle\langle j $.
\otimes	Kronecker product.

“Nothing is built on stone; All is built on sand, but we must build as if the sand were stone.”

Jorge Luis Borges

Chapter 1

Introduction

As quantum computing devices gain in size and precision, the task of characterizing their quality indubitably grows in complexity as well. A first challenge is to identify a few commendable parameters that both capture the caliber of elementary operational components and that can be used to infer the reliability of computations based on those building blocks. A second challenge is to invent experimental procedures that provide robust estimates of those figures of merit in a reasonable amount of time. Those two challenges are coupled: a commendable parameter that is not extractable by any known experimental means is no more valuable than an efficiently estimated quantity that has no important role in asserting computational reliability.

To remain tractable through the means of classical computation, characterization methods are often based around the construction of quantum circuits for which the ideal outcome can be efficiently predicted classically. The elementary constituents used to build simple circuits are the same as the ones used for more involved computations. Hence, even if a given quantum circuit remains tractable by classical means, the deviation between its results and the predicted outcomes provide information regarding the operational components used in more complex circuits.

This idea of connecting circuit outcomes and circuit component characterization is concretized in randomized benchmarking (RB), a leading experimental method conceived to characterize large-scale quantum computers [EAZ05; Dan+06; Gae+12; Cór+13; Kel+14; Bar+14; Xia+15; Muh+15; Tar; Cas+16; McK+16; She+16; Tak+16b; McK+17b]. In terms of raw data, RB protocols yield a collection of “success” and “failure” events related to many implemented computations sampled from a given family of circuit constructions. Once collected, the raw data is processed into valuable information regarding sets of quantum operations.

The active research involving RB-type protocols revolves around three main axes:

- i. The design of new circuit families and sampling distributions, for which the resulting raw data could carry novel information.
- ii. The development of more advanced analytical and statistical tools to translate the raw data into valuable error parameters.
- iii. The study of inference techniques that can be used to leverage the knowledge of a few figures of merit to quantify the outcome reliability of simple or complex computations.

Those three directions are interdependent: the relevance of designing new data collection routines rests on data interpretation tools; the notion of “commendable parameter” implicitly relies on its aptitude to be leveraged and experimentally estimated; understanding how

figures of merit evolve through specific circuit constructions provides insight for designing novel data collection procedures.

In this thesis, various advancements in the three above directions are presented. Each chapter includes a research paper for which my contribution was important. Since each chapter is self-introductory, the rest of this introduction will limit itself to a brief blueprint.

The thesis is split in two parts. The first one, titled *“Error signatures in quantum circuits”*, mostly treats of the two-way relation between the caliber of individual circuit components and the quality of compositions thereof.

Chapter 3 starts with the derivation of the worst-case error accumulation scenario (in terms of operational fidelity) in quantum circuits. This worst-case scenario is achieved by (infinitely) many different processes, but is modulated by a figure of merit labeled as the *“coherence angle”*, which combines the process fidelity and the unitarity, two quantities that both appear from the data processing of RB experiments [CWE16]. For two-fold compositions, a saturated best-case scenario is also derived in terms of the coherence angle, which can be used to tightly bound the fidelity of two-fold composite operations (in even dimensions). The composition bounds find a direct application in RB, as they can be used to connect the results of three distinct experiments – standard, interleaved, and unitarity RB [EAZ05; Dan+06; MGE12; Mag+12; Wal+15]– to improve the confidence interval on the fidelity of isolated operations.

Chapter 4, which ends the first part of the thesis, generalizes the ideas presented in chapter 3. Through the introduction of a channel approximation referred to as the leading Kraus (LK) approximation, it is first shown that there exists a natural way to decompose any non-catastrophic quantum channel into a product of a uniquely defined physical unitary and a decoherent error [CAE19]. This structural dichotomy is then shown to translate into a dynamical dichotomy: the process fidelity of a given circuit can be factored in two inherently different behaviors. The first factor corresponds to a multiplicative decay dictated by the decoherent component of individual circuit elements. The second factor, which can exhibit non-linear features, depends solely on the coherent (i.e. unitary) component of circuit elements. The close connection between the structural channel factorization and the factorization in the fidelity evolution is only ensured when well-identified unrealistic error scenarios, referred to as extremal channels, are ruled out through the *“equability condition”*. The notion of equability is also used to relate experimentally estimated parameters obtained via RB to other commendable quantities such as the diamond distance and the superoperator spectral norm.

The second part of this thesis, titled *“Quantum characterization through randomized benchmarking”* treats more directly of the analysis and protocol advancements related to RB experiments. Chapter 5 establishes a strong relationship between the raw data obtained from RB experiments and a commendable physical parameter that can be used to characterize sets of operations known as gate-sets. Previously to the work developed in chapter 5, which mostly corresponds to the work found in [Car+18], it was shown that standard RB protocols yield data than can be fit to an exponential decay, even in the advent of gate-dependent Markovian errors [MGE11; MGE12; Wal17]. The decay parameter was shown to correspond to a fidelity-like quantity of generally unclear nature [Pro+17a; Wal17]. Chapter 5 precisely addresses this issue. In particular, it is first shown that, in the single qubit case, the decay parameter obtained through experiment can be reconciled with a physically meaningful fidelity. Then, based on the results derived in chapter 4, the same type of reconciliation is

extended to arbitrary system size.

Chapter 6 describes a variant of the original randomized benchmarking scheme, labeled as dihedral benchmarking, which allows the reliable characterization of new gate-sets [CWE15]. The proposed protocol escapes the original 2-design restriction originally considered in [EAZ05; Dan+06; Dan+09; MGE11; MGE12]. In dihedral benchmarking, the main advantage of loosening the restrictions on the randomizing gate-set is that it allows the precise characterization of the celebrated $\pi/8$ gate, an operation often proposed to supplement Clifford circuits in order to gain universality. The characterization of the $\pi/8$ gate is of special interest because leading fault-tolerant schemes propose its physical implementation via magic state distillation and gate injection [BK05], which involves drastically different mechanisms than the implementation of Clifford gates.

In chapter 7, a mathematical machinery is developed with the goal of generalizing the analysis of group-based RB to generator-based RB protocols. The first section, which is derived from the work in [Boo+19], covers a simple gate-independent analysis of the protocol proposed in [Kni+08]. The protocol in question, referred to as “NIST RB” makes use of a randomizing gate-set that neither constitutes a 2-design, nor a group, but does generate the Clifford group under composition. In this sense, the experimentally implemented scheme provides a historical antecedent to generator-based RB. In the second section of chapter 7, a generalization of the NIST RB protocol, referred to as direct RB [Pro+18], is presented and analyzed. In a nutshell, direct RB invokes the native circuit elements used to generate 2-designs as the randomizing gate-set. This simplification improves even further the scalability of standard RB techniques. Moreover, the resulting figures of merit are expected to connect more closely with the characterization of primitive circuit components. The next chapter (chapter 2) covers most notions and notation required to make sense of the subsequent chapters.

Chapter 2

Introductory Material

2.1 Inner product and norms

Throughout this thesis, matrix norms and matrix inner products will be frequently used. Given two matrices $A, B \in M_d(\mathbb{C})$, the Hilbert-Schmidt inner product between A and B is defined as

$$\langle A, B \rangle := \text{Tr } A^\dagger B. \quad (2.1)$$

From there, the Schatten 2-norm (which is also known as the Frobenius norm, and as the Hilbert-Schmidt norm) is naturally defined as

$$\|A\|_2^2 := \langle A, A \rangle = \text{Tr } A^\dagger A. \quad (2.2)$$

In this thesis, $A^\dagger A$ is also noted as $|A|^2$. The absolute value symbol is used on elements of $M_d(\mathbb{C})$ as a shorthand to denote

$$|A| := \sqrt{A^\dagger A}. \quad (2.3)$$

$|A|$ is positive semi-definite and its eigenvalues are the singular values of A . From this convention and from the singular value decomposition theorem, it follows that any matrix $A \in M_d(\mathbb{C})$ can be expressed as

$$A = U|A|, \quad (2.4)$$

where $U \in U(d)$ is a unitary matrix. Equation (2.4) is commonly referred to as a polar decomposition of A . If A is full rank, then the unitary U is unique.

The Schatten 2-norm of A is the familiar Euclidean norm of the singular values of A . Equipped with such inner product and norm, $M_d(\mathbb{C})$ becomes a Hilbert space (i.e. $M_d(\mathbb{C})$ is complete w.r.t. $\|\cdot\|_2$). Given an operator $A \in M_d(\mathbb{C})$, its normalized version is denoted as

$$\underline{A} := \frac{A}{\|A\|_2}. \quad (2.5)$$

Given an orthogonal operational basis $\{B_i\}$ for $M_d(\mathbb{C})$, any matrix $A \in M_d(\mathbb{C})$ can be expressed as

$$A = \sum_i \alpha_i B_i, \quad (2.6)$$

where $\alpha_i = \langle B_i, A \rangle$. The Schatten 2-norm can also be seen as the Euclidean norm (on \mathbb{C}^{d^2}) of the coefficients α_i . Equation (2.6) naturally suggests a way to vectorize matrices, that is to map elements of $M_d(\mathbb{C})$ to \mathbb{C}^{d^2} :

$$|A\rangle_{\{B_i\}} = \sum_{i=0}^{d^2-1} \alpha_i |i\rangle, \quad (2.7)$$

where $|i\rangle$ are canonical unit vectors in \mathbb{C}^{d^2} . Let $\langle i| = |i\rangle^\dagger$ and $\langle A| = |A\rangle^\dagger$. For any choice of orthonormal basis $\{B_i\}$, and for any $A, B \in M_d(\mathbb{C})$, the Hilbert-Schmidt inner product $M_d(\mathbb{C}) \times M_d(\mathbb{C}) \rightarrow \mathbb{C}$ corresponds to the usual inner product $\mathbb{C}^{d^2} \times \mathbb{C}^{d^2} \rightarrow \mathbb{C}$:

$$\langle A, B \rangle = \langle A|B \rangle. \quad (2.8)$$

A common matrix vectorization results from choosing the orthonormal operational basis $\{E_{ij}\}$, where $E_{ij} = |i\rangle\langle j|$, $|i\rangle$ are canonical unit vectors in \mathbb{C}^d . The projection on E_{ij} is assigned to the vector $|j\rangle \otimes |i\rangle \in \mathbb{C}^{d^2}$, where \otimes denotes the Kronecker product:

$$\text{col}(A) := \sum_{i,j=1}^d \langle E_{ij}, A \rangle |j\rangle \otimes |i\rangle. \quad (2.9)$$

$\text{col} : M_d(\mathbb{C}) \rightarrow \mathbb{C}^{d^2}$ is referred to as column vectorization or column stacking, as it can be seen as stacking the columns of $A \in M_d(\mathbb{C})$ on top of each other. Given three matrices $A, B, C \in M_d(\mathbb{C})$, a little index-tracking exercise suffices to show that

$$\text{col}(ABC) = C^\dagger \otimes A \text{col}(B). \quad (2.10)$$

The Schatten p -norm is defined (and denoted) as

$$\|A\|_p := [\text{Tr} |A|^p]^{1/p}. \quad (2.11)$$

It corresponds to the ℓ_p -norm of the singular values of A . $\|A\|_\infty$ is used to denote the maximal singular value of A :

$$\|A\|_\infty := \max_i \sigma_i(A) = \sigma_{\max}(A). \quad (2.12)$$

Schatten p -norms frequently appear in the so-called Hölder's inequality. Let $1/p + 1/q = 1$ with $p, q \geq 1$ ($1/\infty$ denotes 0, so that $p = \infty, q = 1$ is a valid choice). Then,

$$|\langle A, B \rangle| \leq \|A\|_p \|B\|_q \quad (\text{Hölder's ineq.})$$

The case $p = q = 2$ falls back to the so-called Cauchy-Schwarz inequality.

2.1.1 Common operator bases

Through eq. (2.6), operators in $M_d(\mathbb{C})$ can be expressed as linear combinations of d^2 orthogonal operators. A common choice of basis consists in the so-called Heisenberg-Weyl (unitary)

operators $W_{ij} = X_i Z_j \in M_d(\mathbb{C})$, defined implicitly through

$$X_k |j\rangle := |j + k \pmod d\rangle \quad (2.13a)$$

$$Z_k |j\rangle := \omega^{jk} |j\rangle \quad (2.13b)$$

$$\omega := e^{i\frac{2\pi}{d}}. \quad (2.13c)$$

The four Pauli matrices $\{\mathbb{I}_2, X, Y, Z\}$ can be obtained from the Heisenberg-Weyl operators for $d = 2$, with $X = X_1$, $Z = Z_1$ and $Y = iX_1 Z_1$. For $d = 2^n$, a common choice of unitary operational basis is the set of n -fold Paulis $\{\mathbb{I}_2, X, Y, Z\}^{\otimes n}$, where \otimes refers to the Kronecker product.

2.2 Quantum states and measurements

In this thesis, only quantum systems with *finite* dimension are considered. Quantum states of a d -dimensional system are represented by density matrices $\varrho_d(\mathbb{C}) \subset M_d(\mathbb{C})$, which are positive semi-definite matrices with unit trace. A pure state corresponds to a density matrix with rank 1, and the maximally mixed state corresponds to \mathbb{I}_d/d . Pure states are sometime denoted by unit vectors in \mathbb{C}^d . The unit vector $|\psi\rangle \in \mathbb{C}^d$ corresponds to the eigenvector with eigenvalue 1 of the associated density matrix $\rho = |\psi\rangle\langle\psi| \in \varrho_d(\mathbb{C})$. When a state is defined on two systems S_1 and S_2 of dimensions d_1 and d_2 respectively, it is denoted as $\rho_{S_1 S_2} \in \varrho_{d_1 d_2}(\mathbb{C})$. From the perspective of the system S_1 , the quantum state becomes

$$\rho_{S_1} := \text{Tr}_{S_2} \rho_{S_1 S_2} := \sum_{i,j=1}^{d_1} \sum_{k=1}^{d_2} \langle E_{ij}^{(S_1)} \otimes E_{kk}^{(S_2)}, \rho_{S_1 S_2} \rangle E_{ij}^{(S_1)}, \quad (2.14)$$

where $E_{ij}^{(S_1)} \in M_{d_1}(\mathbb{C})$ and $E_{kk}^{(S_2)} \in M_{d_2}(\mathbb{C})$. This operation is referred to as the partial trace over S_2 .

A quantum measurement is defined by a *set* of positive operators $\mathbb{M} = \{\mu_0, \mu_1, \dots\}$ (sometimes referred to as a POVM, for “positive-valued operators measure”) for which the elements $\mu_i \in M_d(\mathbb{C})$ sum to the identity \mathbb{I}_d . The subscript i attached to the POVM element μ_i corresponds to an observed event. For instance, a photon count could correspond to $i = 0$, and a dark count could correspond to $i = 1$ (I picked the subscript’s alphabet to be the non-negative integers for concreteness, but any alphabet can be used). Let E be the “observed event” random variable with possible realizations $i \in \{0, 1, \dots\}$. Such variable is defined by the measurement set \mathbb{M} . The probability of observing the event i given a state $\rho \in \varrho_d(\mathbb{C})$ is provided by Born’s rule:

$$\Pr(E = i | \rho) = \langle \mu_i, \rho \rangle. \quad (2.15)$$

A projective measurement refers to a set of positive operators \mathbb{M} for which the elements μ_i are projectors.

2.3 Quantum operations

2.3.1 Definition

A quantum operation (or quantum channel) is a completely-positive (CP), trace-preserving (TP) linear map acting on $M_d(\mathbb{C})$ ¹. That is, a linear map \mathcal{A} acting on a system S_1 of dimension d_1 is a quantum channel if, for any quantum state defined on systems S_1 and S_2 (here the ancillary system S_2 can be of arbitrary size d_2),

$$\begin{aligned} \text{Tr } \mathcal{A}(\rho_{S_1}) &= \text{Tr } \rho_{S_1}, && \text{(Trace-preserving)} \\ \mathcal{A} \otimes \mathcal{I}_{S_2}(\rho_{S_1 S_2}) &\geq 0, && \text{(Completely-positive)} \end{aligned}$$

where \mathcal{I}_{S_2} acts as the identity on the system S_2 . The set of quantum channels acting on a d -dimensional system is denoted CPTP_d .

In this thesis, quantum operations will be often denoted by the calligraphic font. The dual \mathcal{A}^\dagger of a channel \mathcal{A} is defined via Born's rule as the linear operation that obeys

$$\langle \mu_i, \mathcal{A}(\rho) \rangle = \langle \mathcal{A}^\dagger(\mu_i), \rho \rangle, \quad (2.16)$$

for any state ρ and POVM element μ_i .

An operation $\mathcal{A} : M_d(\mathbb{C}) \rightarrow M_d(\mathbb{C})$ is said to be unital if it has \mathbb{I}_d as a fixed point, that is $\mathcal{A}(\mathbb{I}_d) = \mathbb{I}_d$. Quantum channels are not generally unital maps, but their duals are.

Notice that the positivity constraint implies that CPTP maps are hermiticity preserving. Indeed, consider an Hermitian operator $H \in H_d(\mathbb{C})$. Every Hermitian operator has real eigenvalues, and can be re-expressed as a sum $H = P + Q$ where $P \geq 0$ and $-Q \geq 0$. From there,

$$\begin{aligned} \mathcal{A}(H) &= \mathcal{A}(P + Q) \\ &= \mathcal{A}(P) + \mathcal{A}(Q) && \text{(Linearity)} \\ &= P' + Q', && (2.17) \end{aligned}$$

where $P' \geq 0$ and $-Q' \geq 0$. Since $P' + Q'$ is Hermitian, \mathcal{A} must preserve hermiticity.

2.4 Mapping abstract quantum operations to more familiar sets

There are various ways to map abstract quantum operations to more familiar sets (e.g. matrices, or sets of matrices). Each mapping has its own advantages and disadvantages. A little bit like Laplace and Fourier transforms help in solving problems by recasting the way in which one can express a function, being able to exchange between different expressions of the same quantum operation makes the life of a quantum physicist much easier.

2.4.1 Process matrix (Liouville representation)

One of the most common way to express quantum operations is through their process matrix, also referred to as Liouville (semigroup) representation. A channel $\mathcal{A} \in \text{CPTP}_d$ and

¹In this thesis, the output state is assumed to be in $M_d(\mathbb{C})$ as well.

its Liouville representation $\mathcal{A} \in M_{d^2}(\mathbb{C})$ will be often denoted by the same letter, since the context in which they are employed suffices to distinguish them. If two different matrix representations of $\mathcal{A} \in \text{CPTP}_d$ are used, appropriate indices will be appended.

Consider an operational orthogonal basis $\{B_i\}$ for $M_d(\mathbb{C})$. The process matrix $\mathcal{A} \in M_{d^2}(\mathbb{C})$ of the abstract channel $\mathcal{A} \in \text{CPTP}_d$ is defined explicitly as

$$[\mathcal{A}]_{ij} := \langle \underline{B}_i, \mathcal{A}(\underline{B}_j) \rangle, \quad (2.18)$$

or implicitly through

$$\mathcal{A}|\rho\rangle_{\{B_i\}} = |\mathcal{A}(\rho)\rangle_{\{B_i\}} \quad \forall \rho \in \mathcal{Q}_d(\mathbb{C}) \quad (2.19)$$

This definition ensures that the composition of operations $\mathcal{A} \circ \mathcal{B} \in \text{CPTP}_d$ translates into the usual matrix multiplication $\mathcal{A}\mathcal{B} \in M_{d^2}(\mathbb{C})$. Given the definition of the dual, eq. (2.16), it is easy to show that the dual of the matrix representation corresponds to its conjugate transpose.

The next lemma provides some constraint on the spectrum of Liouville representations.

Lemma 1: Complex eigenvalues of process matrices come in conjugate pairs

Let $\mathcal{A} \in M_{d^2}(\mathbb{C})$ be a Liouville representation of a quantum channel, with eigenvalues $\{\lambda_i\}$. If \mathcal{A} has an eigenvalue λ_i with a non-zero imaginary component, and a multiplicity m , then \mathcal{A} also has an eigenvalue λ_i^* with multiplicativity m .

Proof. The property stated above applies to all real matrices. Indeed, if $\mathcal{A} \in M_d(\mathbb{R})$ and $\mathcal{A}v = \lambda v$, then $(\mathcal{A}v)^* = (\lambda v)^*$ which reduces to $\mathcal{A}v^* = \lambda^* v^*$. To finish the proof, it suffices to show the existence of a real representation for any given quantum channel \mathcal{A} , since a change of operator basis $\{B_i\} \rightarrow \{B'_i\}$ is equivalent to a similarity transform, which doesn't affect the spectrum. First, as a consequence of CPness, quantum operations are hermicity preserving. Hermitian operators span $M_d(\mathbb{C})$. By choosing any Hermitian operator basis $\{B_i\}$, the corresponding Liouville representation must be real, since any complex entry would map an Hermitian operator to an operator with an anti-Hermitian component. \square

When the basis $\{B_i\}$ is the set of n -fold Pauli matrices, $\mathcal{A} \in M_{d^2}(\mathbb{C})$ is referred to as the Pauli-Liouville representation (which is real due the hermicity of Pauli operators). More generally, since quantum operations are trace-preserving, it is often convenient to choose $B_0 \propto \mathbb{I}_d$. With such choice the first row take the simple form $[\mathcal{A}]_{0i} = \delta_{0i}$, which entirely captures trace-preservation. The choice $B_0 \propto \mathbb{I}_d$ provide \mathcal{A} with a 2×2 lower triangular block form. To address each of the four blocks, let me introduce two projectors. Let $A \in M_d(\mathbb{C})$,

$$\Pi_{\text{id}}(A) := \langle \mathbb{I}_d, A \rangle \mathbb{I}_d, \quad (2.20a)$$

$$\Pi_{\text{trls}}(A) := A - \Pi_{\text{id}}(A). \quad (2.20b)$$

Π_{id} projects the matrix A on its identity component, and Π_{trls} projects A on the traceless hyperplane (or Bloch space) in $M_d(\mathbb{C})$ (i.e. the hyperplane of operations with null trace). For

the Liouville representation $\mathcal{A} \in M_{d^2}(\mathbb{C})$ of a quantum channel,

$$\Pi_{\text{id}} \mathcal{A} \Pi_{\text{id}} =: [\mathcal{A}]_{\text{id}, \text{id}} = \Pi_{\text{id}}, \quad (2.21a)$$

$$\Pi_{\text{id}} \mathcal{A} \Pi_{\text{trls}} =: [\mathcal{A}]_{\text{id}, \text{trls}} = 0, \quad (2.21b)$$

$$\Pi_{\text{trls}} \mathcal{A} \Pi_{\text{id}} =: [\mathcal{A}]_{\text{trls}, \text{id}}, \quad (2.21c)$$

$$\Pi_{\text{trls}} \mathcal{A} \Pi_{\text{trls}} =: [\mathcal{A}]_{\text{trls}, \text{trls}}. \quad (2.21d)$$

$[\mathcal{A}]_{\text{trls}, \text{id}}$ is known as the non-unital vector, since it corresponds to the mapping of \mathbb{I}_d to the traceless hyperplane. A channel is unital iff $[\mathcal{A}]_{\text{trls}, \text{id}} = 0$. From there it is easy to see how trace-preservation translates into the unital condition after taking the dual of \mathcal{A} . $[\mathcal{A}]_{\text{trls}, \text{trls}}$ is a $(d^2 - 1) \times (d^2 - 1)$ matrix sometimes referred to as the unital block of \mathcal{A} . Given a vectorization $|\rho\rangle$ of a density matrix ρ , $\Pi_{\text{trls}} |\rho\rangle = |\Pi_{\text{trls}}(\rho)\rangle$ is commonly referred to as the Bloch vector of ρ [Blo46]; for $d = 2$ the set of Bloch vectors corresponding to quantum states forms a 3-dimensional unit ball. From eqs. (2.21a) to (2.21d), it follows that the action of \mathcal{A} on $|\rho\rangle$ can be decomposed as the unital block acting on the Bloch vector added to the non-unital vector:

$$\mathcal{A}|\rho\rangle = [\mathcal{A}]_{\text{trls}, \text{trls}} |\Pi_{\text{trls}}(\rho)\rangle + [\mathcal{A}]_{\text{trls}, \text{id}}. \quad (2.22)$$

The non-unital vector induces a translational action, while the unital block induces linear transformations on the Bloch space.

Another common choice of operational basis is $\{E_{ij}\}$ (notice here that this choice won't generally yield a lower triangular block structure). Given the double index ij (instead of a single index going from 0 to $d^2 - 1$), it is convenient to translate the action of \mathcal{A} on E_{ij} as the action of the matrix \mathcal{A} on the canonical unit vector $|j\rangle \otimes |i\rangle$:

$$\mathcal{A} = \sum_{i,j,k,\ell=0}^{d-1} \langle E_{k\ell}, \mathcal{A}(E_{ij}) \rangle |\ell\rangle \otimes |k\rangle \langle j| \otimes \langle i|. \quad (2.23)$$

By design, this matrix is meant to act on column-stacked density matrices $|\rho\rangle = \text{col}(\rho)$.

The Liouville representation has the advantage of providing a geometrical visualization of the action quantum operations on quantum states. For instance, in $d = 2$ the traceless hyperplane has dimension 3 and quantum states are in one-to-one correspondence with the unit ball (Bloch sphere); by construction, the action on the Bloch sphere induced by an operation \mathcal{A} is immediately seen from its 4×4 process matrix \mathcal{A} .

While the Liouville representation is certainly useful in understanding quantum operations, it also conceals certain aspects of their structure. For instance, the Liouville representation doesn't capture CPness concisely. Due to the intricate geometry of quantum states (and that process matrices are defined through the operational action on states), the set of matrices in $M_{d^2}(\mathbb{C})$ that are tied to physical operations is also subject to convoluted geometrical restrictions.

2.4.2 Chi matrix

Every linear map $\mathcal{A} : \mathcal{Q}_d(\mathbb{C}) \rightarrow \mathcal{Q}_d(\mathbb{C})$ can be expressed as follows:

$$\mathcal{A}(\rho) = \sum_{i,j=0}^{d^2-1} \chi_{ij} B_i \rho B_j^\dagger, \quad (2.24)$$

where $\rho \in \mathcal{Q}_d(\mathbb{C})$, $\{B_i\}$ is an orthonormal basis for $M_d(\mathbb{C})$ and $\chi_{ij} \in \mathbb{C}$. This is easily seen by taking the column vectorization of operators in $M_d(\mathbb{C})$, in which case eq. (2.24) becomes

$$\mathcal{A} \text{col}(\rho) = \sum_{i,j=0}^{d^2-1} \chi_{ij} B_j^* \otimes B_i \text{col}(\rho). \quad (2.25)$$

Here $*$ denote the complex conjugate operation. Notice that since $\{B_i\}$ forms an orthonormal basis for $M_d(\mathbb{C})$, $\{B_j^* \otimes B_i\}$ forms an orthonormal basis for $M_{d^2}(\mathbb{C})$; any matrix in M_{d^2} can be expressed as $\sum_{i,j=0}^{d^2-1} \chi_{ij} B_j^* \otimes B_i$.

The χ_{ij} coefficients define a $d^2 \times d^2$ matrix, referred to as the Chi matrix of \mathcal{A} (which, of course, depends on the basis $\{B_i\}$). The Chi matrix is positive semi-definite iff \mathcal{A} is CP, and has trace d if \mathcal{A} is TP or unital. Notice here that CP-ness is easily captured.

2.4.3 Choi matrix

Given a quantum channel $\mathcal{A} : M_d(\mathbb{C}) \rightarrow M_d(\mathbb{C})$, the Choi matrix of \mathcal{A} is defined as [Cho75]

$$\text{Choi}(\mathcal{A}) := \sum_{i,j=0}^{d-1} E_{ij} \otimes \mathcal{A}(E_{ij}). \quad (\text{Choi matrix})$$

Just like the Chi matrix, the Choi matrix is positive semi-definite iff \mathcal{A} is CP, and has trace d if \mathcal{A} is TP or unital. Since $\text{Choi}(\mathcal{A}) \geq 0$, it has a (possibly non-unique) spectral decomposition of the form

$$\text{Choi}(\mathcal{A}) := \sum_{i=0}^{d^2-1} \text{col}(A_i) \text{col}^\dagger(A_i), \quad (2.26)$$

$$= \sum_{i=0}^{d^2-1} \|A_i\|_2^2 \text{col}(\underline{A}_i) \text{col}^\dagger(\underline{A}_i), \quad (2.27)$$

WOLOG, the eigenvectors $\text{col}(\underline{A}_i)$ are chosen to be orthonormal, and the eigenvalues are ordered with respect to the Schatten 2-norm (Frobenius norm):

$$\|A_0\|_2^2 \geq \|A_1\|_2^2 \geq \dots \geq \|A_{d^2-1}\|_2^2 \geq 0. \quad (2.28)$$

The focus of this thesis remains on non-catastrophic channels (definition 4), for which $\|A_0\|_2^2 > \|A_1\|_2^2$. This ensures that the highest eigenvalue of the Choi matrix is non-degenerate, and is attached to a well-defined eigenvector $\text{col}(\underline{A}_0)$.

2.4.4 Canonical Kraus decomposition

Conjugating the Choi matrix by $\langle i | \otimes \mathbb{I}_d$ and $|j\rangle \otimes \mathbb{I}_d$ yields the image of E_{ij} under \mathcal{A} :

$$\mathcal{A}(E_{ij}) = \langle i | \otimes \mathbb{I}_d \text{Choi}(\mathcal{A}) |j\rangle \otimes \mathbb{I}_d . \quad (2.29)$$

By applying the conjugation on eq. (4.2) and after some algebraic gymnastic, one obtains a Kraus decomposition [Kra+83]:

$$\mathcal{A}(E_{ij}) = \sum_{k=0}^{d^2-1} A_k E_{ij} A_k^\dagger , \quad (2.30)$$

with

$$\langle A_i, A_j \rangle = \|A_i\|_2^2 \delta_{ij} . \quad (2.31)$$

The TP condition is equivalent to ensuring

$$\sum_{i=0}^{d^2-1} A_i^\dagger A_i = \mathbb{I}_d , \quad (2.32)$$

which implies that $\sum_i (\|A_i\|_2^2 / d) = 1$. CP-ness is guaranteed from the form of eq. (2.30). The matrices $A_i \in M_d(\mathbb{C})$ are referred to as ordered canonical Kraus operators. Given a channel addressed in calligraphic font, the non-calligraphic font is used to denote a corresponding choice of ordered canonical Kraus operators. The operators $\{A_i\}$ are ordered as in eq. (2.28). The adjective ‘‘canonical’’ refers to the orthogonality constraint eq. (2.31), which can be imposed WOLOG (although there might be multiple canonical forms). In this thesis, A_0 (which is associated with the highest Choi matrix eigenvalue $\|A_0\|_2^2$) will deserve special attention, and is attributed the title of ‘‘leading Kraus (LK) operator’’. In general, A_0 might be non-unique when the spectrum of the Choi matrix is degenerate. However, as mentioned earlier, the focus of this thesis remains on non-catastrophic channels (definition 4), for which A_0 is unique.

2.5 Elementary types of quantum channels

Within the vast family of quantum operations, certain mechanisms (or families of mechanisms) have obtained special appellations.

First of all, an error simply denotes a channel for which the target is the identity.

2.5.1 Unitary errors

An error \mathcal{W} is said to be unitary or purely coherent if it corresponds to a unitary channel

$$\mathcal{W}(\rho) = W\rho W^\dagger , \quad (2.33)$$

where $W \in U(d)$ is close to \mathbb{I}_d . Such error can realistically be modeled as the result of a perturbation from a targeted ‘‘Hamiltonian drive’’.

It is well-known that $U(d)$ can be obtained by exponentiating the algebra of Hermitian matrices $H_d(\mathbb{C})$. That is, the surjective map $\exp(i \cdot H)$ takes any $H \in H_d(\mathbb{C})$ to $U(d)$ ². For any $U \in U(d)$ there exists at least one $H \in H_d(\mathbb{C})$ s.t. $U = \exp(iH)$. Physically, it is useful to see discrete unitaries $U \in U(d)$ as a the result of a continuous process performed over a time T , described by a time-dependent Hamiltonian $H(t) \in H_d(\mathbb{C})$ (also referred to as a pulse function):

$$U = \lim_{dt \rightarrow 0} \exp(iH(T-dt)dt) \cdot \exp(iH(T-2dt)dt) \cdots \exp\{iH(dt)dt\} \cdot \exp\{iH(0)dt\}. \quad (2.35)$$

A general unitary error occurs when the implemented Hamiltonian is not exactly $H(t)$, but some other function $H'(t)$ (which is a perturbation away from $H(t)$). By driving $H'(t)$ rather than $H(t)$, one would end up with the unitary

$$V = \lim_{dt \rightarrow 0} \exp(iH'(T-dt)dt) \cdot \exp(iH'(T-2dt)dt) \cdots \exp\{iH'(dt)dt\} \cdot \exp\{iH'(0)dt\}. \quad (2.36)$$

Let $W = VU^\dagger$; \mathcal{W} is an example of unitary error channel. Implementing \mathcal{V} can be seen as implementing the perfect targeted operation \mathcal{U} followed by a unitary error \mathcal{W} .

Simple examples of unitary errors includes the so-called over-rotations and under-rotations. Consider the simple case where the time-dependent Hamiltonian is of the form:

$$H(t) = f(t)H, \quad (2.37)$$

where $H \in H_d(\mathbb{C})$ is fixed and $f(t) \in \mathbb{R}$. In such a case, $U = \exp(iH \int_0^T f(t)dt)$. If instead of performing $H(t) = f(t)H$, one were to perform $H'(t) = f'(t)H$, the resulting unitary would be $V = \exp(iH \int_0^T f'(t)dt)$. In the case where $\int_0^T f(t)dt = \int_0^T f'(t)dt$, then $U = V$. However, if $\int_0^T f(t)dt < \int_0^T f'(t)dt$, V corresponds to an over-rotation. Indeed, when the Hamiltonian H is driven for too long (or too strongly), the resulting eigenvalues of V correspond to over-rotated phases. A similar argument goes if the Hamiltonian is not driven for long enough, or too weakly, in which case $\int_0^T f(t)dt > \int_0^T f'(t)dt$ and V is said to be an under-rotation.

²The domain (and codomain) of a function $f : \mathbb{C} \rightarrow \mathbb{C}$ is naturally extended to normal matrices $N_d(\mathbb{C}) = \{M \in M_d(\mathbb{C}) | MM^\dagger = M^\dagger M\}$ by applying $f : \mathbb{C} \rightarrow \mathbb{C}$ to their eigenvalues. Let $N \in N_d(\mathbb{C})$ have a spectral decomposition of the form $N = \sum_i \lambda_i v_i v_i^\dagger$, then

$$f(N) := \sum_i f(\lambda_i) v_i v_i^\dagger. \quad (2.34)$$

2.5.2 Convexity of the set of CPTP maps

Quantum channels (CPTP maps) form a convex set. That is, given a probability measure μ and a set of channels $\{\mathcal{A}_x\}$, the convex combination \mathcal{B} of the \mathcal{A}_x s is also a quantum channel:

$$\mathcal{B}(\rho) = \int \mathcal{A}_x(\rho) d\mu(x). \quad (2.38)$$

The convexity of CPTP maps is quickly verified: the TP property is preserved by linearity (take the trace on each side of eq. (2.38)); the CPness of \mathcal{B} can be verified by looking at its Choi matrix (and by using the fact that a sum of positive semidefinite matrices is itself positive semidefinite).

2.5.3 Incoherent channels

Since quantum channels form a convex set (see section 2.5.2), any convex combination of unitary channels forms a valid quantum channel. This motivates the following construction.

Definition 1: incoherent channels

A channel \mathcal{A} is said to be incoherent if it's of the form

$$\mathcal{A}(\rho) = \int U_x \rho U_x^\dagger d\mu(x), \quad (2.39)$$

where μ is a probability measure and $U_x \in U(d)$.

An incoherent process can be interpreted as the result of Hamiltonian fluctuations (not necessarily the driving Hamiltonian, but also the idling Hamiltonian). By seeing the pulse $H(t)$ as a stochastic process, it follows that the implemented unitary is a random variable.

Incoherence is to be distinguished from decoherence, a wider notion that is covered in chapter 4. Incoherent channels, which are unital, form a strict subset of decoherent channels.

2.5.4 Stochastic errors

Stochastic error channels constitute a subset of incoherent channels for which there is an incoherent canonical Kraus form with a LK operator proportional to the identity.

Definition 2: stochastic errors

An error channel \mathcal{A} is said to be stochastic if it can be expressed as

$$\mathcal{A} := \sum_{i=1}^{d^2} \text{Pr}(i) \mathcal{U}_i, \quad (2.40)$$

where $\{\mathcal{U}_i\}$ (with $\mathcal{U}_1 = \mathbb{I}_d$) forms an orthogonal unitary basis for $M_d(\mathbb{C})$, and where $\text{Pr} : \{1, 2, \dots, d^2\} \rightarrow \mathbb{R}$ is a probability distribution.

A stochastic channel for which the unitary basis consists of the n -fold Paulis $\{\mathbb{I}_2, X, Y, Z\}^{\otimes n}$ is simply referred to as a *Pauli channel*.

2.5.5 Depolarizing errors

Depolarizing channels \mathcal{D}_p are a special family of stochastic error channels parameterized by $p \in [0, 1]$:

$$\mathcal{D}_p(\rho) = p\rho + (1-p)(\text{Tr } \rho)\mathbb{I}_d/d; \quad (2.41)$$

essentially, with probability p , the state ρ remains unscathed, and with probability $1-p$ it is replaced with the completely mixed state \mathbb{I}_d/d . \mathcal{D}_0 is referred to as the *completely* depolarizing channel (for obvious reasons) and corresponds to Π_{id} defined in eq. (2.20a). The appearance of $\text{Tr } \rho$ in eq. (2.41) is meant to allow the domain extension to $M_d(\mathbb{C})$. For instance, in $d = 2^n$ non-identity n -fold Pauli matrices P_i are mapped to pP_i , while \mathbb{I}_d is mapped to itself.

Consider a finite group with an irreducible representation of $d \times d$ unitaries, $\mathbf{G} = \{U_i\}$, where $U_i \in U(d)$ (e.g. the Heisenberg-Weyl operators, or the n -fold Paulis). Then, from Schur's lemma [Art], it directly follows that

$$\frac{1}{|\mathbf{G}|} \sum_i U_i \rho U_i^\dagger = (\text{Tr } \rho)\mathbb{I}_d/d, \quad (2.42)$$

which can be used to reexpress depolarizing channels as convex sum of unitary conjugations

$$\mathcal{D}_p(\rho) = p\rho + (1-p) \frac{1}{|\mathbf{G}|} \sum_{U_i} U_i \rho U_i^\dagger \quad (2.43)$$

$$= \left(p + \frac{(1-p)}{|\mathbf{G}|} \right) \rho + (1-p) \frac{1}{|\mathbf{G}|} \sum_{U_i \neq \mathbb{I}_d} U_i \rho U_i^\dagger. \quad (2.44)$$

This last form explicitly shows that depolarizing channels are special instances of stochastic error channels.

2.5.6 Dephasing/phase damping errors

A channel \mathcal{A} is said to induce a strictly dephasing (or phase damping) action between two orthogonal pure states $|\psi_i\rangle\langle\psi_i|$ and $|\psi_j\rangle\langle\psi_j|$ if

$$\mathcal{A}(|\psi_i\rangle\langle\psi_i|) = |\psi_i\rangle\langle\psi_i|, \quad (2.45a)$$

$$\mathcal{A}(|\psi_j\rangle\langle\psi_j|) = |\psi_j\rangle\langle\psi_j|, \quad (2.45b)$$

$$\mathcal{A}(|\psi_i\rangle\langle\psi_j|) = \gamma|\psi_i\rangle\langle\psi_j|, \quad (2.45c)$$

$$\mathcal{A}(|\psi_j\rangle\langle\psi_i|) = \gamma|\psi_j\rangle\langle\psi_i|, \quad (2.45d)$$

where $\gamma \in [0, 1)$ is a contraction factor. Notice that the invariant subspace of such a dephasing action is $\text{Span}(|\psi_i\rangle\langle\psi_i|, |\psi_j\rangle\langle\psi_j|)$, which includes pure states. The dephasing action progressively transforms a quantum superposition of the form $\alpha|\psi_i\rangle + \beta|\psi_j\rangle$ into a (classical) probabilistic mixture $|\alpha|^2|\psi_i\rangle\langle\psi_i| + |\beta|^2|\psi_j\rangle\langle\psi_j|$.

A common family of stochastic error channels consists of convex combinations of mutually orthogonal unitaries that all commute with each other.

Definition 3: stochastic dephasing errors

A stochastic error channel \mathcal{A} is said to be dephasing (or phase damping) if it can be expressed as

$$\mathcal{A} := \sum_{i=1}^d \Pr(i) \mathcal{U}_i, \quad (2.46)$$

where $\Pr : \{1, 2, \dots, d\} \rightarrow \mathbb{R}$ is a probability distribution and where the unitaries $U_i \in U(d)$ ($U_1 = \mathbb{I}_d$) associated with a non-zero probability are orthogonal and all commute (only d orthogonal unitaries can mutually commute).

A simple example of stochastic dephasing channel is

$$\mathcal{A}(\rho) = \sum_i^d \Pr(i) Z_i \rho Z_i^\dagger \quad (2.47)$$

where $\{Z_i\}$ are the Heisenberg-Weyl operators defined in eq. (2.13b).

The terminology “dephasing” is judicious. Since the unitaries $\{U_i\}$ all commute, they can all be simultaneously diagonalized. In the above example, the Z_i s are already diagonal in the $\{|i\rangle\}$ canonical basis. In particular, this means that the states $E_{ii} = |i\rangle\langle i|$ are left invariant under \mathcal{A} . However, the off-diagonal elements $E_{ij} = |i\rangle\langle j|$ are mapped to

$$\mathcal{A}(E_{ij}) = \gamma_{ij} E_{ij}, \quad (2.48)$$

where $\gamma_{ij} := \sum_k \Pr(k) \omega^{(i-j)k}$. By construction, $|\gamma_{ij}| \leq 1$, meaning that the phase between $|i\rangle$ and $|j\rangle$ is generally dampened.

A stochastic dephasing channel with uniform probability distribution $\Pr(i) = 1/d$ is referred to as *completely* dephasing. By Schur’s lemma (the above definition implicitly enforces the set $\{U_i\}$ to form a group), it follows that a completely dephasing error channel cancels the off-diagonal elements of any density matrix ρ expressed in the basis in which the U_i s are simultaneously diagonal. A completely dephasing channel can be seen as a projective measurement with d events for which the classical information gained from the measurement is lost.

2.5.7 Amplitude damping channels (population transfer)

Incoherent channels, as well as phase damping errors are all unital channels (i.e. for which \mathbb{I}_d is a fixed point). However, in more general error scenarios, probabilistic state transitions could prevent \mathbb{I}_d to be a fixed point. Consider the canonical example of amplitude damping acting on a two-level system, parameterized by the following Kraus operators

$$A_0 = \begin{pmatrix} \sqrt{p_{00}} & 0 \\ 0 & \sqrt{p_{11}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{p_{00}} & 0 \\ 0 & \sqrt{p_{11}} \end{pmatrix}, \quad (2.49a)$$

$$A_1 = \begin{pmatrix} 0 & \sqrt{p_{01}} \\ \sqrt{p_{10}} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{p_{10}} & 0 \\ 0 & \sqrt{p_{01}} \end{pmatrix}, \quad (2.49b)$$

where $p_{ij} \geq 0$ can be seen in the classical case as transition probabilities from $|j\rangle$ to $|i\rangle$, with $\sum_i p_{ij} = 1$. Indeed, it is easy verified that $\mathcal{A}(E_{00}) = p_{00}E_{00} + p_{10}E_{11}$ and $\mathcal{A}(E_{11}) = p_{11}E_{11} + p_{01}E_{00}$. In particular, the identity \mathbb{I}_2 is generally not a fixed point. Of course, the channels with unitarily conjugated Kraus operators UA_iU^\dagger are still considered as amplitude damping, but with respect to other orthogonal states $|\psi_i\rangle\langle\psi_i| = UE_{ii}U^\dagger$.

A famous example of amplitude damping channel is the so-called state preparation. Indeed, for instance, let the ‘‘transition probabilities’’ of a two-level amplitude damping channel obey $p_{0i} = 1 \forall i \in \{0, 1\}$. Then, it follows that the resulting amplitude channel acting on any density matrix $\rho \in \mathcal{Q}_2(\mathbb{C})$ yields

$$\mathcal{A}(\rho) = E_{00}. \quad (2.50)$$

Here E_{00} is the only fixed point in $\mathcal{Q}_2(\mathbb{C})$.

There are many different ways to extend the notion of amplitude damping to higher dimensions. A natural one, which arises from looking at the RHS of eqs. (2.49a) and (2.49b), is to consider processes for which the Kraus operators are of the form

$$A_\pi = P_\pi \cdot \sum_i \sqrt{p_{\pi(i)i}(\pi)} E_{ii}, \quad (2.51)$$

where $\pi : \{0, \dots, d-1\} \rightarrow \{0, \dots, d-1\}$ is a permutation, and P_π its associated $d \times d$ matrix (i.e. $[P_\pi]_{ji} := \delta_{j\pi(i)}$). To be trace-preserving, the $p_{\pi(i)i}(\pi) \geq 0$ must simply obey $\sum_\pi p_{\pi(i)i}(\pi) = 1$. Again, simultaneously conjugating the Kraus operators with a unitary, $UA_\pi U^\dagger$, would yield the same process in essence, but with respect to different orthogonal states.

2.6 Quantum circuits and the Markovian assumption

A particular object of study in this thesis is the construction of noisy quantum circuits. Just as classical computing relies on elementary logical gates (e.g. NAND) to implement Boolean functions, quantum computing involves circuit constructions based on limited elementary gates to generate general elements of $U(d)$.

A gate-set $\mathbf{G} = \{g\}$, denoted by the bold font, is any collection of elements $g \in U(d)$, referred to as gates. Notice that subgroups of $U(d)$ are examples of gate-sets. The canonical mapping from $U(d)$ to CPTP_d is implicitly defined by

$$L(g)[\rho] := g\rho g^\dagger, \quad (2.52)$$

where $g \in U(d)$, $\rho \in \mathcal{Q}_d(\mathbb{C})$. $L(g) \in \text{CPTP}_d$ is referred to as a target realization of the gate g , or simply as a target. In a similar fashion, the quantum channel corresponding to the physical implementation of a gate g is noted as $\Lambda(g)$, and can be factored as

$$\Lambda(g) = \mathcal{E}(g|L) \circ L(g), \quad (2.53)$$

where \mathcal{E} is labeled as the (left) error map. Notice that the error \mathcal{E} explicitly depends on the target as a different target yields a different error; sometimes the dependence will be left implicit.

The physical realization mapping Λ can include additional variables to account for the context under which the gates are being applied. For instance, the mapping could depend on the initial state of the system, on time, and on previous operations that have been applied.

Consider a target quantum circuit $\mathcal{C} \in \text{CPTP}_d$ consisting of the composition of m ideal (unitary) gates g_i :

$$L(g_{m:1}) = L(g_m) \circ L(g_{m-1}) \circ \cdots \circ L(g_2) \circ L(g_1). \quad (2.54)$$

The notation $g_{a:b}$ for $a \geq b$ is a shorthand for $g_a \cdot g_{a-1} \cdots g_b$.

The average³ implemented circuit is typically modeled as a process of the form

$$\Lambda_{\text{circuit}}(g_m, \cdots, g_1) = \Lambda_m(g_m) \circ \Lambda_{m-1}(g_{m-1}) \circ \cdots \circ \Lambda_2(g_2) \circ \Lambda_1(g_1), \quad (2.55)$$

where Λ_i are all different maps. Generally, those processes could depend on all previous operations, on the time t_i at which they occurs, as well as on the initial state ρ on which the circuit acts:

$$\Lambda_i(g_i) = \Lambda(g_i | g_{i-1}, \cdots, g_1, \rho, t_i). \quad (2.56)$$

Notice that such a model is quite general: Λ_i could simply consist of an amplitude damping channel that maps any state to ρ , followed by the ideal circuit $L(g_{i:1})$, and by an arbitrary (time-dependent) error channel. However, it is often realistic to assume that the coupling with the environment doesn't allow for long-term memory effects. In particular, if memory effects (i.e. coherent interactions with the environment) occur on a shorter timescale than the timescale between operations, it is usually reasonable to assume a Markovian model, for which errors only depend on the previous operation (and on time):

$$\Lambda_i(g_i) = \Lambda(g_i | t_i). \quad (2.57)$$

This last assumption still allows for non-stationary (time-dependent) effects such as drift. In this thesis, the focus will remain on stationary processes (or approximately stationary processes), for which it is reasonable to express Λ_i as a function of g_i alone. Essentially, elementary gates are implemented as $\Lambda_i(g_i) = \Lambda(g_i) = \mathcal{E}(g_i | L) \circ L(g_i)$, and the noisy circuit is expressed as

$$\Lambda_{\text{circuit}}(g_m, \cdots, g_1) = \Lambda(g)_{m:1} := \Lambda(g_m) \circ \Lambda(g_{m-1}) \circ \cdots \circ \Lambda(g_1). \quad (2.58)$$

Keep in mind that $\Lambda(g_{m:1}) \neq \Lambda(g)_{m:1}$. For instance, consider a simple error model in which \mathbb{I}_2 is ideally implemented, $\Lambda(\mathbb{I}_2) = L(\mathbb{I}_2)$, but the Pauli Z is attached to a depolarizing error, $\Lambda(Z) = \mathcal{P}_p \circ L(Z)$. Let $g_1, g_2 = Z$, in which case

$$\begin{aligned} \Lambda(g_{2:1}) &= \Lambda(Z \cdot Z) = \Lambda(\mathbb{I}_2) = L(\mathbb{I}_2), \\ \Lambda(g)_{2:1} &= \Lambda(Z) \circ \Lambda(Z) = \mathcal{P}_p \circ L(Z) \circ \mathcal{P}_p \circ L(Z) = \mathcal{P}_{p^2} \circ L(\mathbb{I}_2). \end{aligned} \quad (2.59)$$

Due to the gate argument “ g ”, the usage of the notation “ $\Lambda(g)$ ” and “ $L(g)$ ” is mostly useful when the notion of gate-set plays a central role. This only occurs in part II which focuses primarily characterization schemes based on specific circuit constructions. The introduction

³Here, the average is performed over infinitely many realizations of the same noisy circuit.

of the gate as an argument can be used to easily differentiate compiled implemented gates $\Lambda(g_{m:1})$ from physical compositions $\Lambda(g)_{m:1}$. Moreover, the appearance of $L(g)$, as discussed in section 2.8, will be used to label reference bases more explicitly, a subject that only becomes relevant in part II.

In part I, it is more useful to denote channels by the calligraphic font, such as $\mathcal{A} \in \text{CPTP}_d$. Indeed, many results revolve around the Kraus decomposition of \mathcal{A} , for which the Kraus operators $\{A_i\}$ can be easily labeled by the corresponding non-calligraphic font.

2.7 Characterizing quantum operations

Given a quantum channel \mathcal{A} and a target unitary channel \mathcal{U} , we can compare the overlap of their outputs given specific inputs $M \in M_d(\mathbb{C})$ through the M -fidelity:

$$f_M(\mathcal{A}, \mathcal{U}) := \frac{\langle \mathcal{A}(M), \mathcal{U}(M) \rangle}{\|M\|_2^2}. \quad (2.60)$$

When the target is the identity, the second argument can be dropped:

$$f_M(\mathcal{A}) := \frac{\langle \mathcal{A}(M), M \rangle}{\|M\|_2^2}. \quad (2.61)$$

Such abbreviation carries to all subsequent fidelity-like quantities.

The so-called average gate fidelity is obtained by averaging the M -fidelities uniformly⁴ over all physical pure states $|\psi\rangle\langle\psi|$:

$$F(\mathcal{A}, \mathcal{U}) := \mathbb{E}_{\text{Haar}} f_{|\psi\rangle\langle\psi|}(\mathcal{A}, \mathcal{U}). \quad (2.62)$$

Instead of averaging over quantum states, we could also average uniformly over all operators $M \in M_d(\mathbb{C})$. More precisely, given any orthogonal operator basis $\{B_i\}$ for $M_d(\mathbb{C})$, we can uniformly average over the M -fidelities f_{B_i} , which yields the average process fidelity⁵

$$\Phi(\mathcal{A}, \mathcal{U}) := \mathbb{E}_{\{B_i\}} f_{B_i}(\mathcal{A}, \mathcal{U}). \quad (2.63)$$

Compared to Φ , F puts a slightly higher weight over the identity component \mathbb{I}_d . The TP condition enforces this special component to take a fixed value, $f_{\mathbb{I}_d} = 1$. Hence the two quantities are closely related via [Nie02]:

$$F(\mathcal{A}, \mathcal{U}) = \frac{d\Phi(\mathcal{A}, \mathcal{U}) + 1}{d + 1}. \quad (2.64)$$

$F(\mathcal{A}, \mathcal{U})$ is the overlap between the output state $\mathcal{A}(\rho)$ of an implemented channel \mathcal{A} and its ideal output $\mathcal{U}(\rho)$, averaged over all physical pure input states $|\psi\rangle\langle\psi|$. While $F(\mathcal{A}, \mathcal{U})$ conveys a more graspable interpretation, it will often remain easier to work with $\Phi(\mathcal{A}, \mathcal{U})$

⁴“Uniformly” is to be read here as “with respect to the Haar measure”.

⁵For the readers familiar with the Chi matrix, $\Phi(\mathcal{A}, \mathcal{U})$ is a way to express the so-called χ_{00} element. Of course, the χ -matrix has to be defined with respect to an orthonormal operator basis $\{B_i\}$ with $B_0 = U$. Some might also be more familiar with the notion of entanglement fidelity, which is again Φ .

since it ties with the Kraus operators through

$$\Phi(\mathcal{A}, \mathcal{U}) = \sum_{i=1}^{d^2} \left| \left\langle \frac{A_i}{\sqrt{d}}, \underline{U} \right\rangle \right|^2 = \sum_{i=1}^{d^2} (\|A_i\|_2^2/d) |\langle \underline{A}_i, \underline{U} \rangle|^2. \quad (2.65)$$

Since $\{\underline{A}_i\}$ forms an orthonormal basis and $\|\underline{U}/\sqrt{d}\|_2 = 1$, it follows that

$$\sum_{i=1}^{d^2} |\langle \underline{A}_i, \underline{U} \rangle|^2 = 1. \quad (2.66)$$

If $\|A_i\|_2^2/d$ can be thought as the “weights” of the Kraus operators, $|\langle \underline{A}_i, \underline{U} \rangle|^2$ can be thought as normalized overlaps with the target \underline{U} .

Another common fidelity quantity is obtained by averaging the M -fidelity uniformly over the traceless hyperplane (Bloch space). More precisely, let $\{B_i\}$ be an orthogonal operator basis with $B_0 \propto \mathbb{I}_d$. Then,

$$f_{\text{trls}}(\mathcal{A}, \mathcal{U}) := (d^2 - 1)^{-1} \sum_{i \neq 0} f_{B_i}(\mathcal{A}, \mathcal{U}). \quad (2.67)$$

It relates to the process fidelity Φ and the average fidelity F via

$$f_{\text{trls}}(\mathcal{A}, \mathcal{U}) = \frac{dF(\mathcal{A}, \mathcal{U}) - 1}{d - 1} = \frac{d^2\Phi(\mathcal{A}, \mathcal{U}) - 1}{d^2 - 1}. \quad (2.68)$$

Given a depolarizing channel \mathcal{P}_p , $f_{\text{trls}}(\mathcal{P}_p, \mathcal{I}) = p$. In other words, $f_{\text{trls}}(\mathcal{A}, \mathcal{I})$ can be interpreted as the effective depolarizing constant of an error channel \mathcal{A} . For this reason, f_{trls} of \mathcal{A} to the identity is also referred to as

$$p(\mathcal{A}) \equiv f_{\text{trls}}(\mathcal{A}). \quad (2.69)$$

Recall that when the target argument is left implicit, it corresponds to the identity.

To quantify the coherence of a quantum channel, one could wonder how much the Bloch vectors (the traceless component of quantum states [Blo46]) are contracted. For instance, consider the unitarity, which is the squared length ratio of the Bloch vectors before and after the action of the channel \mathcal{A} , averaged over all physical Bloch vector inputs corresponding to pure states $|\psi\rangle\langle\psi| - \mathbb{I}_d/d$ [Wal+15]:

$$u(\mathcal{A}) := \mathbb{E}_{\text{Haar}} \frac{\|\mathcal{A}(|\psi\rangle\langle\psi| - \mathbb{I}_d/d)\|_2^2}{\| |\psi\rangle\langle\psi| - \mathbb{I}_d/d \|_2^2}. \quad (2.70)$$

Let's extend the domain of the process fidelity Φ to include a new function of \mathcal{A} :

$$Y(\mathcal{A}) := \sqrt{\Phi(\mathcal{A}, \mathcal{A})} = \sqrt{\sum_{i,j=1}^{d^2} \left| \left\langle \frac{A_i}{\sqrt{d}}, \frac{A_j}{\sqrt{d}} \right\rangle \right|^2} = \sqrt{\sum_{i=1}^{d^2} \left(\frac{\|A_i\|_2^2}{d} \right)^2}. \quad (2.71)$$

Straightforward calculations closely relate the unitarity to Y via

$$u(\mathcal{A}) = \frac{d^2 Y^2(\mathcal{A}) - 1}{d^2 - 1}. \quad (2.72)$$

Notice that the notation alludes to the connection between Greek and Latin alphabets; it relates “Phi” to “F” and “Upsilon” to “u”.

Fidelity-like quantities ($f_M, f_{\text{trls}}, F, \Phi$) are 1 in the targeted case. Sometimes, it is useful to refer to infidelity-like quantities, which are 0 in the targeted case. More formally, consider an implemented gate $\Lambda(g) \in \text{CPTP}_d$, an targeted realization $L(g) \in \text{CPTP}_d$, and a functional $f(\Lambda(g), L(g)) \in \mathbb{R}$. Then the variation δf is defined as

$$\delta f(\Lambda(g), L(g)) := f(L(g), L(g)) - f(\Lambda(g), L(g)). \quad (2.73)$$

For instance $\delta F(\Lambda(g), L(g)) = 1 - F(\Lambda(g), L(g))$ is typically referred to as the average infidelity, sometimes denoted as $r(\Lambda(g), L(g))$.

The notion of fidelity naturally extends to circuits and families of circuits. Let \mathbf{G} be a gate-set and consider all circuits resulting from an m -fold composition of the physical realizations of its elements:

$$\Lambda(\mathbf{G}, m) := \{\Lambda(g)_{m:1} | g_i \in \mathbf{G}\}, \quad (2.74)$$

with the abbreviation $\Lambda(\mathbf{G}, 1) = \Lambda(\mathbf{G})$, which refers to the “gate-set physical implementation”. Similarly, $L(\mathbf{G}, m)$ denotes the set of all targeted circuit realizations resulting from an m -fold composition of the type $L(g_{m:1})$ where $g_i \in \mathbf{G}$. $L(\mathbf{G}) := L(\mathbf{G}, 1)$ refers to the gate-set target realization. In [Car+18] (see chapter 5), the gate-set circuit fidelity is defined as the average fidelity over all possible circuits of lengths m for which the individual components are gate-set realizations $\Lambda(g_i)$:

$$F(\Lambda(\mathbf{G}), L(\mathbf{G}), m) := \frac{1}{|\mathbf{G}|^m} \sum_{g_i \in \mathbf{G}} F(\Lambda(g)_{m:1}, L(g_{m:1})) \quad (2.75)$$

The gate-set circuit fidelity could be generalized by introducing weights on the gate-set elements. Let $\Omega(g_m, \dots, g_1)$ by a probability distribution over circuits realizations $\Lambda(g)_{m:1}$, then the weighted gate-set circuit fidelity is defined as

$$F(\Lambda(\mathbf{G}), L(\mathbf{G}), m, \Omega) := \sum_{g_i \in \mathbf{G}} \Omega(g_m, \dots, g_1) F(\Lambda(g)_{m:1}, L(g_{m:1})). \quad (2.76)$$

Equations (2.75) and (2.76) are based on the average fidelity F , but can be trivially extended to fidelity-like, infidelity-like or unitarity-like quantities.

2.8 Reference bases and gauge transformations

The statistical observations of quantum mechanical phenomena are fully described by Born’s rule. Let $\rho \in \mathcal{Q}_d(\mathbb{C})$ be a quantum state, $\{\mu_i\}$ be a set of POVM elements and $L(g)$ be a CPTP

map defined by

$$L(g)[\rho] := g\rho g^\dagger, \quad (2.77)$$

where $g \in U(d)$. Then, Born's rule dictates the probability of observing the event $E = i$ given an initial state ρ and quantum process $\Lambda(g)$:

$$\Pr(E = i | \Lambda(g)[\rho]) = \langle \mu_i | \Lambda(g)[\rho] \rangle. \quad (2.78)$$

Notice that the same statistics would arise if one were to redefine

$$|\rho\rangle \rightarrow \mathcal{B}|\rho\rangle, \quad (2.79a)$$

$$\langle \mu | \rightarrow \langle \mu | \mathcal{B}^{-1}, \quad (2.79b)$$

$$\Lambda(g) \rightarrow \mathcal{B}\Lambda(g)\mathcal{B}^{-1}, \quad (2.79c)$$

where $\mathcal{B} : M_d(\mathbb{C}) \rightarrow M_d(\mathbb{C})$ is any invertible linear map. The transformation described by eq. (2.79), referred to as a gauge transformation, bears no significant meaning in the abstract picture, but does have an effect when considering matrix representations of quantum processes.

A gauge transformation will said to be physical if \mathcal{B} corresponds to a physical unitary. In this case, it maps density matrices to density matrices and POVMs to POVMs. When acting on a target gate realization $L(g)$, a physical gauge transformation can be seen as an inner automorphism of $U(d)$:

$$\mathcal{U}L(g)\mathcal{U}^\dagger = L(\mathcal{U}g\mathcal{U}^\dagger), \quad (2.80)$$

which can be interpreted as a change of reference basis to express $U(d)$ elements. The re-labeling of the Pauli matrices $X \rightarrow Y \rightarrow Z \rightarrow X$ is a simple example of physical gauge transformation.

Part I

Error signatures in quantum circuits

Chapter 3

Bounding the average gate fidelity of composite channels using the unitarity

3.1 Foreword

The present chapter mainly consists in a literal transcription of [CWE16], for which my contribution was major. I altered the notation of the original article to ensure a consistent notation throughout the thesis.

The following work focuses on relating the average fidelity of circuit constructions with the average fidelity of elementary circuit components. The relation is generally quite loose. The physical intuition behind this is that unitary errors can either coherently build up or cancel each other. The intuition is partially confirmed by obtaining a generally tighter relation when introducing the unitarity as a parameter. However, this work doesn't pinpoint unitary processes as the unique mechanism responsible for the leeway in the average fidelity of composite channels (this comes in the next chapter). While this work is mostly cast as an improvement of inference techniques for circuit constructions, it should also be seen as the seed of a more ambitious enterprise: factoring the effect of coherent errors in quantum systems.

3.2 Compendium

There is currently a significant need for robust and efficient methods for characterizing quantum devices. While there has been significant progress in this direction, there remains a crucial need to precisely determine the strength and type of errors on individual gate operations, in order to assess and improve control as well as reliably bound the total error in a quantum circuit given some partial information about the errors on the components. In this work, we first provide an optimal bound on the total fidelity of a circuit in terms of component fidelities, which can be efficiently experimentally estimated via randomized benchmarking. We then derive a tighter bound that applies under additional information about the coherence of the error, namely, the unitarity, which can also be estimated via a related experimental protocol. This improved bound smoothly interpolates between the worst-case quadratic and best-case linear scaling for composite error channels. As an application we show how our analysis substantially improves the achievable precision on estimates of the infidelities of individual gates under interleaved randomized benchmarking, enabling greater precision for current experimental methods to assess and tune-up control over quantum gate operations.

3.3 Introduction

The output of a quantum computer will only be reliable if the total error in the whole computation is sufficiently small. This can be rigorously guaranteed if the error on the individual components (i.e., preparations, measurements and gate operations) is sufficiently small compared to the length of the computation. A very common experimental practice [Gae+12; Cór+13; Kel+14; Bar+14; Xia+15; Muh+15; Tar; Cas+16; McK+16; She+16; Tak+16b; McK+17b] for estimating errors on gate operations is randomized benchmarking (RB) of Clifford operations [EAZ05; Dan+06]. The experimentally measured infidelities under RB experiments have very recently been shown to give a very precise estimate of the average gate fidelity (hereafter simply the fidelity) of an error channel to the identity,

$$F(\mathcal{E}) := \int d\psi \langle \psi | \mathcal{E}(|\psi\rangle\langle\psi|) | \psi \rangle \quad (3.1)$$

under very robust and experimentally realistic conditions [MGE11; Mag+12; Wal+15; WBE15; WE16; Pro+17a; Wal17; Car+18; Har+19; DHW19], when expressed in a physical operational gauge [Pro+17a; Wal17; MPF18; Car+18]¹, resolving the concern (that RB did not reliably measure a physically meaningful fidelity) raised in [Pro+17b; QK18].

An important practical application of RB is interleaved RB (IRB) [Mag+12], a now-standard approach for estimating infidelities on individual gates [Gae+12; Cór+13; Kel+14; Bar+14; Vel+14; Muh+15; Vel+15; Bar+15; Cas+16; McK+16; She+16; Tak+16b; Tak+16a; McK+17b; McK+17a; Nic+17; Cha+18; Cal+18; Wan+18b; Wan+18a; Yon+18; Zha+18], including gates that collectively generate universality [CWE15; Cro+16; HF17; Pro+18]. However this approach is subject to a systematic error that can significantly limit the precision of the estimate and often goes unreported - a problem which we address below. As noted above, the average gate fidelity gives only very limited information about the error and error channels with the same fidelity on the component gate operations can lead to dramatically different total error for a circuit composed from these gate operations. For example, the infidelity $\delta F(\mathcal{E}) = 1 - F(\mathcal{E})$ grows linearly in the number of gates under purely stochastic errors (that is, errors that can be modeled by classical probabilities over different Pauli operators) and grows quadratically under purely unitary errors (that is, coherent errors due to small calibrations that are common in quantum control) in the limit of small infidelities [She+16]. However, realistic experimental errors are neither purely stochastic nor purely unitary, but rather some combination of the two. To adequately characterize quantum circuits, which are the result of multiple noisy operations, it is crucial to understand (and bound) how errors can accumulate given an intermediate level of coherence. In this paper, we study the impact of coherence on the fidelity of circuit constructions. An important application from our work is to provide a dramatic improvement to the achievable precision of IRB, enabling significantly more reliable experimental methods for assessing and tuning-up the individual gate operations required for quantum computing and other applications.

This paper is organized as follows. We first obtain strictly optimal upper- and lower-bounds on the total infidelity of the circuit for all parameter regimes when only the infidelities of the components are known. These bounds are saturated by unitary channels and so grow quadratically with the number of gates. Moreover, because our bounds are saturated,

¹In Dugas et al, the physicality of the gauge is proven for $d = 2$, and conjectured otherwise.

they cannot be improved without further knowledge about the errors. Because the worst-case growth of the infidelity is achieved by purely unitary channels, intuitively, quantifying how far an error channel is from purely unitary error should enable an improved bound. One such quantity is the unitarity. Thus our second contribution in this work is a proof that the unitarity

$$u(\mathcal{E}) := \frac{d}{d-1} \int d\psi \operatorname{Tr} \mathcal{E}(\psi - \frac{1}{d}\mathbb{I})^2, \quad (3.2)$$

of the components, which can be estimated using a variant of RB [Wal+15; Car+18; DHW19] (URB), can be used to obtain a tighter bound on the total infidelity. This information enables a smooth interpolation between the quadratic growth of purely unitary errors and the linear scaling of purely stochastic errors. Including the unitarity to characterize circuits can be used to quantitatively reason about an often omitted statement: elementary operations with low infidelity and highly coherent errors can rapidly compose to a worse circuit than a sequence of elementary operations with moderate infidelity but highly stochastic errors. Our bounds implicitly quantify how fast this can happen given the infidelity and unitarity of individual components. Our third contribution, noted above, goes the other way: from a composite error $\mathcal{E}_h \circ \mathcal{E}$, we bound the fidelity of one of its component \mathcal{E}_h . We demonstrate an immediate practical application of this result by providing a dramatically improved bound on the accuracy of the estimates of gate infidelities under interleaved RB [Mag+12]. This is done by substituting the estimate of the effective depolarizing constant of the individual interleaved gate $\hat{p} = p_{\text{IRB}}/p_{\text{RB}}$ by $\hat{p} = p_{\text{IRB}}p_{\text{RB}}/p_{\text{URB}}$, which requires a unitarity RB (URB) experiment. In the experiments reported in [Xue+18; Yan+18], our estimator is used to rigorously bound the infidelity of individual quantum gates via eq. (3.42).

3.4 Noisy quantum processes

Markovian quantum processes can be described by completely-positive and trace-preserving (CPTP) linear maps $\mathcal{A} : \mathcal{Q}_d(\mathbb{C}) \rightarrow \mathcal{Q}_d(\mathbb{C})$ where $\mathcal{Q}_d(\mathbb{C})$ is the set of density operators acting on \mathbb{C}^d , that is, the set of positive-semidefinite operators with unit trace. We denote quantum channels using single calligraphic capital Roman letters and the composition of channels by multiplication for brevity, so that $\mathcal{A}\mathcal{B}(\rho) = \mathcal{A} \circ \mathcal{B}(\rho)$. We also denote the composition of m channels $\mathcal{A}_m, \dots, \mathcal{A}_1$ by $\mathcal{A}_{m:1} = \mathcal{A}_m \dots \mathcal{A}_1$.

Abstract quantum channels can be represented in many ways. In this paper, we will use the Kraus operator, χ -matrix and the Liouville (or transfer matrix) representations. The Kraus operator and χ -matrix representations of a quantum channel \mathcal{A} are

$$\mathcal{A}(\rho) = \sum_j A_j \rho A_j^\dagger = d \sum_{k,l \in \mathbb{Z}_{d^2}} \chi_{kl}^A B_k \rho B_l^\dagger \quad (3.3)$$

respectively, where the A_j are the Kraus operators, $\mathbb{Z}_{d^2} = \{0, \dots, d^2 - 1\}$ and $\{B_i\}$ is a trace-orthonormal basis of $\mathbb{C}^{d \times d}$ satisfying $\langle B_j, B_k \rangle := \operatorname{Tr} B_j^\dagger B_k = \delta_{j,k}$, and $B_0 = \mathbb{I}_d$. Note that we include the dimensional factor in the definition of the χ -matrix to be consistent with the standard construction in terms of unnormalized Pauli operators.

The Kraus operators can be expanded as $A_j = \sum_{k \in \mathbb{Z}_{d^2}} \langle B_k, A_j \rangle B_k$ relative to $\{B_i\}$. Making use of the phase freedom in the Kraus operators (that is, $A_j \rightarrow e^{i\theta_j} A_j$ gives the same quantum

channel), we can set $\langle B_0, A_j \rangle \geq 0$ for all j . We can then expand the Kraus operators as

$$A_j = \sqrt{a_j d} \left(\cos(\alpha_j) B_0 + \sin(\alpha_j) \vec{v}_j \cdot \vec{B} \right) \quad (3.4)$$

where $a_j d = \langle A_j, A_j \rangle$, $\vec{B} = (B_1, \dots, B_{d^2-1})$, $\vec{v}_j \in \mathbb{C}^{d^2-1}$ with $\|\vec{v}_j\|_2 = 1$, and α_j can be chosen to be in $[0, \frac{\pi}{2}]$ by incorporating any phase into \vec{v}_j . Substituting this expansion into the Kraus operator decomposition and equating coefficients with the χ -matrix representation gives

$$\chi_{kl}^A = \frac{1}{d} \sum_j \langle B_k, A_j \rangle \langle A_j, B_l \rangle, \quad (3.5)$$

and, in particular,

$$\Phi(\mathcal{A}) \equiv \chi_{00}^A = \frac{1}{d^2} \sum_j |\text{Tr } A_j|^2 = \sum_j a_j \cos^2(\alpha_j), \quad (3.6)$$

where $\Phi(\mathcal{A})$ is the process fidelity of \mathcal{A} to the identity. Applying the trace-preserving constraint with $\langle B_j, B_k \rangle = \delta_{j,k}$ gives

$$1 = \frac{1}{d} \text{Tr} \sum_j A_j^\dagger A_j = \sum_j a_j, \quad (3.7)$$

which then implies

$$1 - \Phi(\mathcal{A}) = \sum_j a_j \sin^2(\alpha_j). \quad (3.8)$$

Alternatively, density matrices ρ and measurements μ (elements of positive-operator-valued measures) can be expanded with respect to $\{B_i\}$ as $\rho = \sum_j \langle B_j, \rho \rangle B_j$ and $\mu = \sum_j \langle B_j, \mu \rangle B_j$. The Liouville representations of ρ and μ are the column vector $|\rho\rangle$ and row vector $\langle \mu| = |\mu\rangle^\dagger$ of the corresponding expansion coefficients. The Born rule is then $\langle \mu, \rho \rangle = \langle \mu | \rho \rangle$. The Liouville representation of a channel \mathcal{A} is the unique matrix $\mathcal{A} \in M_{d^2}(\mathbb{C})$ (noted with the same symbol as $\mathcal{A} \in \text{CPTP}_d$, for the context will differentiate them) such that $\mathcal{A}|\rho\rangle = |\mathcal{A}(\rho)\rangle$, which can be written as $\mathcal{A} = \sum_j |\mathcal{A}(B_j)\rangle \langle B_j|$. With $B_0 = \mathbb{I}_d / \sqrt{d}$, the Liouville representation of any CPTP map can be expressed in block form as

$$\mathcal{A} = \begin{pmatrix} 1 & 0 \\ \mathcal{A}_n & \mathcal{A}_u \end{pmatrix} \quad (3.9)$$

where $\mathcal{A}_n \in \mathbb{C}^{d^2-1}$ is the non-unital vector and $\mathcal{A}_u \in \mathbb{C}^{(d^2-1) \times (d^2-1)}$ is the unital block. The unitarity and effective depolarizing constant can be written as

$$\begin{aligned} u(\mathcal{A}) &= \frac{\text{Tr } \mathcal{A}_u^\dagger \mathcal{A}_u}{d^2 - 1} = \frac{\|\mathcal{A}_u\|_2^2}{d^2 - 1} \\ p(\mathcal{A}) &= \frac{\text{Tr } \mathcal{A}_u}{d^2 - 1} \end{aligned} \quad (3.10)$$

with respect to the Liouville representation [Wal+15; Kim+14].

	F	$r \equiv \delta F$	p	$\Phi \equiv \chi_{00}$
F	F	$1 - \delta F$	$\frac{(d-1)p+1}{d}$	$\frac{d\Phi+1}{d+1}$
$r \equiv \delta F$	$1 - F$	δF	$\frac{d-1}{d}(1-p)$	$\frac{d}{d+1}(1-\Phi)$
p	$\frac{dF-1}{d-1}$	$1 - \frac{d}{d-1}\delta F$	p	$\frac{d^2\Phi-1}{d^2-1}$
$\Phi \equiv \chi_{00}$	$\frac{(d+1)F-1}{d}$	$1 - \frac{d+1}{d}\delta F$	$\frac{(d^2-1)p+1}{d^2}$	Φ

TABLE 3.1: Linear relations between the fidelity (F), the infidelity ($r \equiv \delta F$), the effective depolarizing constant (p), and $\Phi \equiv \chi_{00}$.

The effective depolarizing constant $p(\mathcal{A})$ and $\Phi(\mathcal{A}) \equiv \chi_{00}^{\mathcal{A}}$ are linear functions of the average fidelity that can be more convenient to work with. The relations between the various linear functions of the fidelity used in this paper are tabulated in table 3.1.

3.5 Composite infidelities in terms of component infidelities

We now prove that unitary error processes lead to the fastest growth in the total infidelity of a circuit. In particular, we obtain strict bounds on the infidelity of a composite error process in terms of the infidelities of the components and show that the bounds are saturated by unitary processes for all even-dimensional systems.

We first obtain a bound on the infidelity of the composition of two channels that strictly improves on the corresponding bound of Ref. [Kim+14]. We also show that the improved bound is saturated for all values of the relevant variables. Therefore theorem 1 gives the optimal bounds on the infidelity of the composite in terms of only the infidelities of the components, and so obtaining a more precise estimate of the composite infidelity requires further information about the errors. We then obtain an upper bound on the infidelity of the composition of m channels that inherits the tightness of the bound for the composition of two channels.

We present the following bounds in terms of the χ matrix, though the results can be rewritten in terms of other linear functions of the infidelity using table 3.1. For example, consider the composition of m noisy operations \mathcal{A}_i with equal infidelity, that is, $\delta F(\mathcal{A}_i) = \delta F$. Then by corollary 1 and table 3.1, the total infidelity of the composite process is at most

$$\delta F(\mathcal{A}_{m:1}) \leq m^2 \delta F + O(m^4 \delta^2 F), \quad (3.11)$$

which exhibits the expected quadratic scaling with m . Moreover, this upper bound is saturated and so cannot be improved without additional information about the errors.

Theorem 1

For any two quantum channels \mathcal{X} and \mathcal{Y} ,

$$\begin{aligned} & \left| \Phi(\mathcal{X}\mathcal{Y}) - \Phi(\mathcal{X})\Phi(\mathcal{Y}) - (1 - \Phi(\mathcal{X}))(1 - \Phi(\mathcal{Y})) \right| \\ & \leq 2\sqrt{\Phi(\mathcal{X})\Phi(\mathcal{Y})(1 - \Phi(\mathcal{X}))(1 - \Phi(\mathcal{Y}))}. \end{aligned} \quad (3.12)$$

Furthermore, for all even dimensions and all values of $\Phi(\mathcal{X})$, $\Phi(\mathcal{Y})$, there exists a pair of channels \mathcal{X} and \mathcal{Y} saturating both signs of the above inequality.

Proof. Let $\mathcal{X}(\rho) = \sum_j X_j \rho X_j^\dagger$ and $\mathcal{Y}(\rho) = \sum_j Y_j \rho Y_j^\dagger$ be Kraus operator decompositions of \mathcal{X} and \mathcal{Y} respectively. From eq. (3.4), we can expand the Kraus operators as

$$\begin{aligned} X_j &= \sqrt{x_j d} \left(\cos(\xi_j) B_1 + \sin(\xi_j) \vec{u}_j \cdot \vec{\mathcal{B}} \right) \\ Y_j &= \sqrt{y_j d} \left(\cos(\theta_j) B_1 + \sin(\theta_j) \vec{v}_j \cdot \vec{\mathcal{B}} \right) \end{aligned} \quad (3.13)$$

where $\vec{u}_j, \vec{v}_j \in \mathbb{C}^{d^2-1}$ are unit vectors and $\xi_j, \theta_j \in [0, \frac{\pi}{2}]$. Then a Kraus operator decomposition of $\mathcal{X}\mathcal{Y}$ is

$$\mathcal{X}\mathcal{Y}(\rho) = \sum_{j,k} X_j Y_k \rho Y_k^\dagger X_j^\dagger \quad (3.14)$$

and so, by eq. (3.6),

$$\Phi(\mathcal{X}\mathcal{Y}) = \sum_{j,k} x_j y_k |\cos \xi_j \cos \theta_k + \beta_{j,k} \sin \xi_j \sin \theta_k|^2, \quad (3.15)$$

where $\beta_{j,k} = \vec{u}_j \cdot \vec{v}_k$ and we have chosen the basis \mathcal{B} to be Hermitian so that $\text{Tr } B_j^\dagger B_k = \text{Tr } B_j B_k = \delta_{j,k}$. By the triangle and reverse-triangle inequalities,

$$|\alpha| - |\gamma| \leq |\alpha + \beta\gamma| \leq |\alpha| + |\gamma| \quad (3.16)$$

for any $\alpha, \beta, \gamma \in \mathbb{C}$ such that $|\beta| \leq 1$, which then implies

$$\left| |\alpha + \beta\gamma|^2 - |\alpha|^2 - |\gamma|^2 \right| \leq 2|\alpha\gamma|. \quad (3.17)$$

From eq. (3.6) and (3.8),

$$\begin{aligned} \sum_{j,k} x_j y_k |\cos(\xi_j) \cos(\theta_k)|^2 &= \Phi(\mathcal{X})\Phi(\mathcal{Y}) \\ \sum_{j,k} x_j y_k |\sin(\xi_j) \sin(\theta_k)|^2 &= (1 - \Phi(\mathcal{X}))(1 - \Phi(\mathcal{Y})), \end{aligned} \quad (3.18)$$

so by eq. (3.17),

$$\begin{aligned} & \left| \Phi(\mathcal{X}\mathcal{Y}) - \Phi(\mathcal{X})\Phi(\mathcal{Y}) - (1 - \Phi(\mathcal{X}))(1 - \Phi(\mathcal{Y})) \right| \\ & \leq \sum_{j,k} 2x_j y_k \cos(\xi_j) \cos(\theta_k) \sin(\xi_j) \sin(\theta_k), \end{aligned} \quad (3.19)$$

using $|\beta_{j,k}| \leq 1$ and the non-negativity of the trigonometric functions over $[0, \frac{\pi}{2}]$. Note that the above inequalities are saturated if and only if $\beta_{j,k} = \pm 1$.

By the Cauchy-Schwarz inequality with the fact that all the quantities are non-negative,

$$\begin{aligned} \sum_j x_j \sin(\xi_j) \cos(\xi_j) & \leq \sqrt{\sum_j x_j \sin^2(\xi_j)} \sqrt{\sum_j x_j \cos^2(\xi_j)} \\ & \leq \sqrt{(1 - \Phi(\mathcal{X}))\Phi(\mathcal{X})}, \end{aligned}$$

where the second line follows from eq. (3.6) and eq. (3.8). Applying this upper bound for \mathcal{X} and the corresponding one for \mathcal{Y} to eq. (3.19) gives the inequality in the theorem.

To see that both signs of the inequality are saturated for all values of $\Phi(\mathcal{X}), \Phi(\mathcal{Y})$ in even dimensions, let $\mathcal{X} = \mathcal{U}(\phi) \otimes \mathcal{I}_{d/2}$ and $\mathcal{Y} = \mathcal{U}(\theta) \otimes \mathcal{I}_{d/2}$ where

$$U(\phi) = e^{i\phi}|0\rangle\langle 0| + e^{-i\phi}|1\rangle\langle 1|. \quad (3.20)$$

By eq. (3.6), $\Phi(\mathcal{U}(\phi) \otimes \mathcal{I}_{d/2}) = \Phi(U(\phi)) = \cos^2 \phi$. As $\mathcal{X}\mathcal{Y} = \mathcal{U}(\phi + \theta) \otimes \mathcal{I}_{d/2}$, some trivial trigonometric manipulations give

$$\begin{aligned} & \Phi(\mathcal{X}\mathcal{Y}) - \Phi(\mathcal{X})\Phi(\mathcal{Y}) - (1 - \Phi(\mathcal{X}))(1 - \Phi(\mathcal{Y})) \\ & = -2 \cos \phi \sin \phi \cos \theta \sin \theta \\ & = -2 \sqrt{\Phi(\mathcal{X})\Phi(\mathcal{Y})(1 - \Phi(\mathcal{X}))(1 - \Phi(\mathcal{Y}))} \text{sign}(\sin 2\phi \sin 2\theta), \end{aligned} \quad (3.21)$$

which saturates the lower bound if the sign function is positive and the upper bound if it is negative. \square

Corollary 1

For any m quantum channels \mathcal{X}_i such that

$$\sum_{i=1}^m \arccos \sqrt{\Phi(\mathcal{X}_i)} \leq \frac{\pi}{2}, \quad (3.22)$$

the process fidelity Φ of the composite channel satisfies

$$\Phi(\mathcal{X}_{m:1}) \geq \cos^2 \left(\sum_{i=1}^m \arccos \sqrt{\Phi(\mathcal{X}_i)} \right). \quad (3.23)$$

Furthermore, this bound is saturated for all even dimensions and all values of the $\Phi(\mathcal{X}_i)$ satisfying eq. (3.22).

Proof. We can rewrite the lower bound in eq. (3.12) as

$$\sqrt{\Phi(\mathcal{X}\mathcal{Y})} \geq \sqrt{\Phi(\mathcal{X})}\sqrt{\Phi(\mathcal{Y})} - \sqrt{1-\Phi(\mathcal{X})}\sqrt{1-\Phi(\mathcal{Y})}. \quad (3.24)$$

Writing $\sqrt{\Phi} = \cos(\arccos \sqrt{\Phi})$ and $\sqrt{1-\Phi} = \sin(\arccos \sqrt{\Phi})$ and using standard trigonometric identities, the above becomes

$$\arccos \sqrt{\Phi(\mathcal{X}\mathcal{Y})} \leq \arccos \sqrt{\Phi(\mathcal{X})} + \arccos \sqrt{\Phi(\mathcal{Y})}, \quad (3.25)$$

taking note to change the direction of the inequality when taking the arccos, which follows from eq. (3.22). By induction, we have

$$\arccos \left(\sqrt{\Phi(\mathcal{X}_{m:1})} \right) \leq \sum_i \arccos \left(\sqrt{\Phi(\mathcal{X}_i)} \right) \quad (3.26)$$

for any set of m channels \mathcal{X}_i . Taking the cosine and squaring gives the bound in eq. (3.23). The saturation follows directly from the saturation of eq. (3.12). \square

A way to intuitively think about eq. (3.23) goes as follows: “the worst possible fidelity of a composition is obtained through a coherent (unitary) buildup”. Indeed, the trigonometric form of the inequality reflects this coherent nature.

3.6 Improved bounds on the infidelity using the unitarity

The bounds in theorem 1 and corollary 1 are tight for general channels if only the infidelity (or, equivalently, Φ) is known. In particular, from eq. (3.11), the infidelity increases at most quadratically in m (to lowest order in r). However, the examples that saturate the bounds are all unitary channels. If, on the other hand, the error model is a depolarizing channel

$$\mathcal{P}_p(\rho) = p\rho + \frac{(1-p)}{d}\mathbb{I}_d, \quad (3.27)$$

or a Pauli channel (that is, a channel with a diagonal χ -matrix with respect to the Pauli basis), then the infidelity increases at most linearly in m to lowest order, that is

$$\delta F(\mathcal{X}_{m:1}) \leq m\delta F + O(m^2\delta^2 F). \quad (3.28)$$

The intermediate regime between Pauli errors and unitary errors can be quantified via the unitarity [Wal+15]. In particular, we define the (positive) coherence angle to be

$$\theta(\mathcal{E}) = \arccos \left(p(\mathcal{E}) / \sqrt{u(\mathcal{E})} \right). \quad (3.29)$$

As $u(\mathcal{E}) \leq 1$ with equality if and only if \mathcal{E} is unitary, $\theta(\mathcal{E}) \in [0, \arccos p(\mathcal{E})]$ and

$$\theta(\mathcal{E}) = \begin{cases} 0 & \text{iff } \mathcal{E} \text{ is depolarizing,} \\ O(\delta F) & \text{if } \mathcal{E} \text{ is Pauli,} \\ \arccos p(\mathcal{E}) = O(\sqrt{\delta F}) & \text{iff } \mathcal{E} \text{ is unitary.} \end{cases} \quad (3.30)$$

That is, $\theta(\mathcal{E})$ quantifies the intermediate regime between Pauli and unitary errors for an isolated error process.

We now show that combining the coherence angle and the infidelity enables improved bounds on the growth of the infidelity. For example, for any m unital channels, or for any m single qubit operations \mathcal{X}_i , with equal infidelity $\delta F(\mathcal{X}_i) = \delta F$ and coherence angles $\theta(\mathcal{X}_i) = \theta$, the total infidelity is at most

$$\delta F(\mathcal{X}_{m:1}) \leq m \left(\delta F - \frac{(d-1)\theta^2}{2d} \right) + m^2 \frac{(d-1)\theta^2}{2d} \quad (3.31)$$

plus higher-order terms in δF and θ^2 by eq. (3.35). For Pauli errors, $\theta^2 = O(\delta^2 F)$, so we recover eq. (3.28). Conversely, for unitary errors $(d-1)\theta^2 = 2d\delta F + O(\delta^2 F)$, so we recover eq. (3.11) in such regime. Moreover, the above bound is saturated (to the appropriate order) in even dimensions by channels of the form

$$\mathcal{X}_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \gamma \cos \theta(\mathcal{X}_i) & -\gamma \sin \theta(\mathcal{X}_i) & 0 \\ 0 & \gamma \sin \theta(\mathcal{X}_i) & \gamma \cos \theta(\mathcal{X}_i) & 0 \\ 0 & 0 & 0 & \lambda \end{pmatrix} \otimes \mathbb{I}_{d^2/2}. \quad (3.32)$$

These include the unital action of single qubit amplitude damping and dephasing channels combined with a unitary evolution around the dampening/dephasing axis. The unitary factor is parameterized by the coherence angle: $Z_\theta = \exp i2\theta Z$ (hence the ‘‘coherence’’ qualifier). In this sense, the coherence angle portays the allowed amount of rotation in Bloch space, as opposed to contractions (quantified by γ , λ in our saturation example) due to decoherent effects.

Theorems 2 and 3 result from more general matrix inequalities that we prove in section 3.9.1. We apply the inequalities to the unital block of the Liouville representation from eq. (3.9), and substitute the expressions for the effective depolarizing constant and the unitarity from eq. (3.10). For theorem 3, we also use results from [PG+06], which state that the maximal singular value of the unital block is upper-bounded by $\sqrt{\frac{d}{2}}$ for general channels and 1 for unital channels.

Theorem 2

For any two quantum channels \mathcal{X} and \mathcal{Y} ,

$$\cos[\theta(\mathcal{X}) + \theta(\mathcal{Y})] \leq \frac{p(\mathcal{X}\mathcal{Y})}{\sqrt{u(\mathcal{X})u(\mathcal{Y})}} \leq \cos[\theta(\mathcal{X}) - \theta(\mathcal{Y})]. \quad (3.33)$$

In other words, the leeway in the effective depolarizing constant of a composition $\mathcal{X}\mathcal{Y}$ is limited by constructive and destructive coherent effects. For longer compositions, we have:

Theorem 3

For any m channels \mathcal{X}_i with $p(\mathcal{X}_i) = p$, $u(\mathcal{X}_i) = u$, the effective depolarizing constant of the composite channel satisfies

$$|p(\mathcal{X}_{m:1}) - p^m| \leq \sqrt{\frac{d}{2}} \binom{m}{2} u \sin^2(\theta). \quad (3.34)$$

Furthermore, if the \mathcal{X}_i are unital channels, the bound can be improved to

$$|p(\mathcal{X}_{m:1}) - p^m| \leq \binom{m}{2} u \sin^2(\theta). \quad (3.35)$$

Notice that the binomial factor – which indicates a quadratic behavior in m – demonstrates that the effective depolarizing constant of a large composition, $p(\mathcal{X}_{m:1})$, can quickly differ from p^m . This difference grows quicker with the coherence angle, which can be tied to coherent effects through eq. (3.32).

The bounds in theorem 2 can be made even tighter if one of the channels is guaranteed to be Pauli.

Theorem 4

Consider a Pauli channel \mathcal{X} and any quantum channel \mathcal{Y} . Then, the composite infidelity is essentially linear in the individual infidelities $\delta F(\mathcal{X})$ and $\delta F(\mathcal{Y})$:

$$\delta F(\mathcal{X}\mathcal{Y}) = \delta F(\mathcal{X}) + \delta F(\mathcal{Y}) + O(\delta F(\mathcal{X})\delta F(\mathcal{Y})). \quad (3.36)$$

This bound is to be contrasted with the naive usage of theorem 2:

$$\begin{aligned} \delta F(\mathcal{X}\mathcal{Y}) &= \delta F(\mathcal{X}) + \delta F(\mathcal{Y}) + O(\theta(\mathcal{X})\theta(\mathcal{Y})) && \text{(theorem 2)} \\ &= \delta F(\mathcal{X}) + \delta F(\mathcal{Y}) + O(\delta F(\mathcal{X})\sqrt{\delta F(\mathcal{Y})}). && \text{(eq. (3.30))} \end{aligned}$$

The improvement can be easily shown as follow. The infidelity is invariant under unitary conjugation $\delta F(\mathcal{X}\mathcal{Y}) = \delta F(\mathcal{U}\mathcal{X}\mathcal{Y}\mathcal{U}^\dagger)$ or convex combination of thereof. In particular, it is invariant under a Pauli twirl. Since \mathcal{X} is a Pauli channel, it commutes with Pauli unitaries, and the twirl gets effectively performed on \mathcal{Y} , which becomes a Pauli channel $\mathcal{Y}_{\text{Pauli}}$ with low coherence angle $\theta(\mathcal{Y}_{\text{Pauli}}) = O(\delta F(\mathcal{Y}))$ (see eq. (3.30)). From there we can apply theorem 2.

Theorems 2 and 3 implicitly suggest that using the coherence angle (rather than the infidelity) as the objective function² for optimizing operations would strongly tighten eventual assertions about the fidelity of circuit constructions.

²A more stable choice would be $\sin^2(\theta)$.

3.7 Application: Interleaved RB

The fidelity extracted from standard RB experiments typically characterizes the average error over a gate-set \mathbf{G} , defined as

$$\mathcal{E} := |\mathbf{G}|^{-1} \sum_{g \in \mathbf{G}} \mathcal{E}_g. \quad (3.37)$$

However, one might only care about the fidelity $F(\mathcal{E}_h)$ attached to a specific gate of interest $h \in \mathbf{G}$, such as one of the generators required for universal quantum computing. The interleaved RB protocol (IRB) [Mag+12] yields a fidelity estimate of $\mathcal{E}_h \mathcal{E}$ ³, the composition between the single gate error and the gate-set error, which provides bounds on the desired value $F(\mathcal{E}_h)$. An issue with this approach is that these bounds generally have a wide range, since possible coherent effects cannot be ignored. This issue is illustrated by the results of two simulations of interleaved RB experiments, plotted in fig. 3.1. In both scenarios, the fidelity of the gate error and of the composed gate were fixed at $F(\mathcal{E}) = 0.9975$ and $F(\mathcal{E}_h \mathcal{E}) = 0.9960$ respectively, hence leading to the same single gate fidelity estimate. In the first case, the interleaved gate h is unitary with high fidelity ($F(\mathcal{E}_h) = 0.9991$), whereas in the second case the error is depolarizing, with a lower fidelity ($F(\mathcal{E}_h) = 0.9975$). This example illustrates how interleaved RB, without a measure of unitarity, can only provide a loose estimate of the infidelity of an individual gate.

More generally, rearranging the bound in theorem 1 to isolate $\Phi(\mathcal{Y})$ gives

$$\begin{aligned} & \left| \Phi(\mathcal{Y}) - \Phi(\mathcal{X}\mathcal{Y})\Phi(\mathcal{X}) - (1 - \Phi(\mathcal{X}\mathcal{Y}))(1 - \Phi(\mathcal{X})) \right| \\ & \leq 2\sqrt{\Phi(\mathcal{X}\mathcal{Y})\Phi(\mathcal{X})(1 - \Phi(\mathcal{X}\mathcal{Y}))(1 - \Phi(\mathcal{X}))}. \end{aligned} \quad (3.38)$$

Moreover, this bound cannot be improved without further information. Now suppose that $\delta F(\mathcal{E}_h \mathcal{E}) \approx 2\delta F(\mathcal{E})$, so that the uncertainty of $\delta F(\mathcal{E}_h)$, obtained via eq. (3.38) and table 3.1 is

$$\Delta \delta F(\mathcal{E}_h) \approx 4\sqrt{2}\delta F(\mathcal{E}). \quad (3.39)$$

While this bound does give an estimate of the infidelity, this estimate is comparable to the following naive estimate that requires no additional experiment. As the fidelity, and hence the infidelity, is a linear function of \mathcal{E} we have

$$\delta F(\mathcal{E}) = |\mathbf{G}|^{-1} \sum_{g \in \mathbf{G}} \delta F(\mathcal{E}_g) \quad (3.40)$$

which, since $\delta F(\mathcal{E})$ is non-negative for any channel \mathcal{E} , implies

$$\delta F(\mathcal{E}_h) \leq |\mathbf{G}|\delta F(\mathcal{E}) \quad (3.41)$$

for any $h \in \mathbf{G}$. (Note also that this bound can be heuristically improved by identifying sets of gates that are expected to have comparable error.) When \mathbf{G} is chosen to be the 12-element subgroup of the Clifford group that forms a unitary 2-design, the naive bound is, at the very

³For the sake of simplicity, we assume that the protocols all provide fidelity estimates defined with respect to the same (or very close) ideal representation of gates.

worst, a factor of $3/\sqrt{2}$ worse than the bound from interleaved benchmarking and requires no additional statistical analysis or data collection.

However, if the error channels were guaranteed to be depolarizing, $F(\mathcal{E}_h)$ could be exactly estimated from an interleaved RB experiment. In general, we can use our knowledge of the unitarity of \mathcal{E} – which can be obtained from a URB experiment⁴ – to quantify how close the error model is to depolarizing noise. From theorem 2, we then have the following bounds, which can be orders of magnitude tighter as illustrated in fig. 3.2.

Corollary 2

For any two quantum channels \mathcal{E}_h and \mathcal{E} ,

$$\left| p(\mathcal{E}_h) - \frac{p(\mathcal{E}_h\mathcal{E})p(\mathcal{E})}{u(\mathcal{E})} \right| \leq \sqrt{1 - \frac{p(\mathcal{E})^2}{u(\mathcal{E})}} \sqrt{1 - \frac{p(\mathcal{E}_h\mathcal{E})^2}{u(\mathcal{E})}}. \quad (3.42)$$

Notice that this new estimate of $p(\mathcal{E}_h)$ is an amalgam of three experiments: standard RB, IRB and unitarity RB. A recommended experimental practice would be, for instance [Yan+18]:

- i. Perform standard RB over the Clifford group. Estimate the resulting decay parameter $p(\mathcal{E})$, where \mathcal{E} is tied to the average error over the Clifford group.
- ii. Perform unitarity RB over the Clifford group. Estimate the resulting decay parameter, which corresponds to the unitarity $u(\mathcal{E})$.
- iii. Perform IRB with the Clifford group as randomizing set and h as ideal gate of interest. Estimate the resulting decay constant $p(\mathcal{E}_h\mathcal{E})$, where \mathcal{E}_h is the error map attached to h .
- iv. Use eq. (3.42) to bound $p(\mathcal{E}_h)$, and use table 3.1 to convert it to the fidelity (or infidelity).

Recall that in the depolarizing case $u(\mathcal{E}) = p(\mathcal{E})^2$, for which eq. (3.42) reduces to the familiar equality $p(\mathcal{E}_h) = p(\mathcal{E}_h\mathcal{E})/p(\mathcal{E})$ ⁵. In fact, the equality remains true up to order $\delta F(\mathcal{E})^2$ in the more general case of stochastic Pauli errors, as demonstrated in theorem 4. Treating the infidelity as a linear quantity under composition is a very common assumption stemming from a classical probabilistic view of error accumulation. To take another example of a linear manipulation, the infidelity per pulse (or infidelity per primitive gate) is often obtained by implicitly dividing the infidelity of a set of composite gates by the average number of pulses used to generate them. These are only good estimates if the error is mostly stochastic. This might be a valid presumption since many error mechanisms are naturally stochastic, but is certainly not a trivial one, since coherent effects also commonly arise from faulty control. The present paper offers a means to avoid the often unrealistic stochasticity assumption by explicitly providing a confidence interval based on experimental estimates of the unitarity. To illustrate the idea, in fig. 3.3 we applied our bounds on various experimental results [Gae+12; Cór+13; Kel+14; Bar+14; Cas+16; McK+16; She+16; Tak+16b; McK+17b] and varied the value of the unitarity.

⁴The current analysis of URB is done under a gate-independent noise approximation.

⁵In the interleaved RB lingo, this relation is often expressed as $p(\mathcal{E}_h) = p_{\text{IRB}}/p_{\text{RB}}$, where \mathcal{E}_h is the error attached to the interleaved gate.

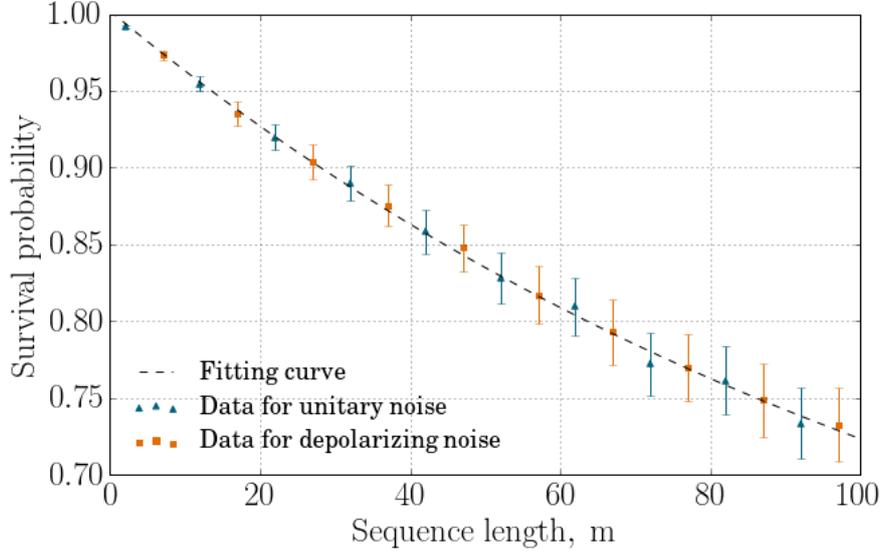


FIGURE 3.1: (Color online) The average survival probability, $p_{\text{surv.}}(m) = |\mathbf{G}|^{-m} \sum_i \langle 0 | \mathcal{S}_m^{(i)} (|0\rangle\langle 0|) |0\rangle$ over all sequences $\mathcal{S}_m^{(i)}$ of length m , as a function of the sequence length for two simulated interleaved RB experiments (see Ref. [Mag+12] for more details) with two different individual gate errors \mathcal{E}_h , but a common average error \mathcal{E} with fidelity 0.9975. Orange squares represent an error model with high fidelity ($F(\mathcal{E}_h) = 0.9991$) that interacts coherently with \mathcal{E} . Blue triangles represent an error model with lower fidelity ($F(\mathcal{E}_h) = 0.9975$), but that is purely stochastic. See section 3.7 for more details.

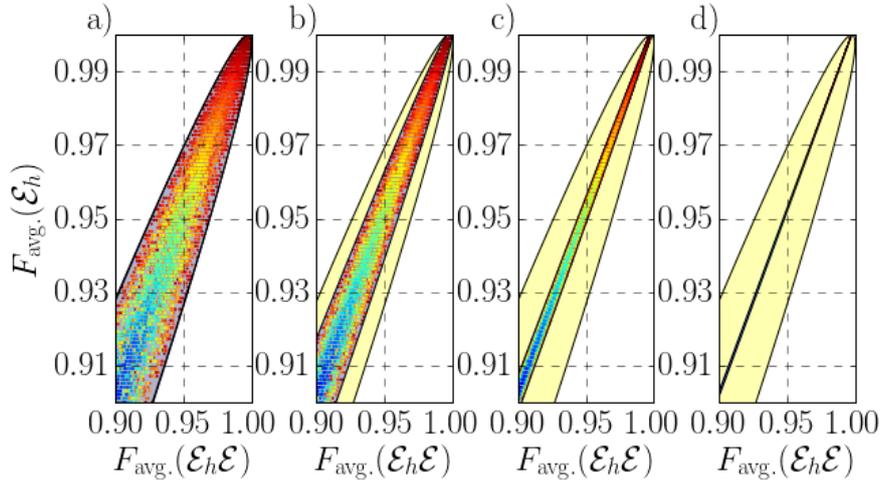


FIGURE 3.2: (Color online) Bounds on the fidelity $F(\mathcal{E}_h)$ of an individual gate h as a function of the composite $F(\mathcal{E}_h \mathcal{E})$ with $F(\mathcal{E})$ fixed and varying $u(\mathcal{E})$: a) $u(\mathcal{E}) = 1.00000$, b) $u(\mathcal{E}) = 0.99300$, c) $u(\mathcal{E}) = 0.99030$, d) $u(\mathcal{E}) = 0.99003 \approx p(\mathcal{E})^2$. The numerical data points correspond to the values $F(\mathcal{E}_h)$ and $F(\mathcal{E}_h \mathcal{E})$ for randomly-generated channels $\{\mathcal{E}_h, \mathcal{E}\}$ satisfying $F(\mathcal{E}) = 0.9975$ and with the appropriate value of $u(\mathcal{E})$. As illustrated by the color, the unitarity $u(\mathcal{E}_h)$ is minimal in the center of the shaded region and maximal when the data points approach our bound.

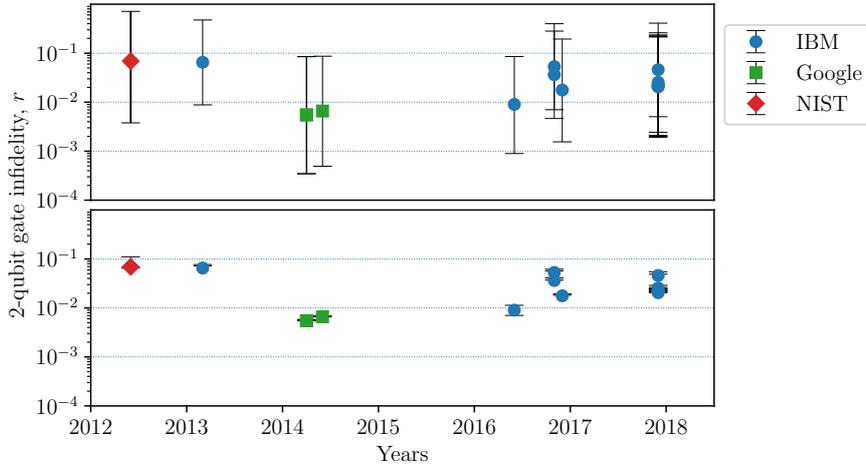


FIGURE 3.3: (Color online) Bounds on various 2-qubit gate infidelities $\delta F(\mathcal{E}_h)$ based on various experimental data [Gae+12; Cór+13; Kel+14; Bar+14; Cas+16; McK+16; She+16; Tak+16b; McK+17b]. The time refers to the dates of submission. The top plot uses eq. (3.42) with a maximal coherence angle $\theta(\mathcal{E})$, which yields in bounds spanning up to two orders of magnitude. The bottom plot assumes a purely stochastic error model, by which we mean that $u(\mathcal{E}) \approx p(\mathcal{E})^2$. For every data point, some statistical error is taken into account, hence the non-zero error bars in the bottom plot.

3.8 Summary and outlook

In this paper, we have studied the impact of coherent errors on the fidelity of quantum circuits. We first demonstrate why coherent errors are a serious concern: a coherent composition of unitary quantum channels results in the fastest decay of the fidelity. In this case, the infidelity grows quadratically in the number of gates, in contrast with the linear growth for stochastic Pauli channels. The disparity between these two regimes means that the characterization of the gate fidelities alone only allows to formulate weak statements about the fidelity of more elaborate circuit constructions.

Hence, in order to characterize circuits more precisely, we introduced a coherence angle—which corresponds to a rotation angle on the Bloch space, as opposed to a contraction (see eq. (3.29))—which enables a tighter bound on the total error in a quantum circuit in terms of robustly estimable quantities that smoothly interpolates between the linear and quadratic regimes.

Our new bound can be used upside-down: from the fidelity of a small circuit construction, we can bound the fidelity of one of its elements. As an immediate application, we demonstrated that this bound substantially improves the estimates of individual gate fidelities from interleaved randomized benchmarking, which, in the absence of the improved bound, are comparable to the naive bound obtained by noting that the infidelity from standard RB is the average of the infidelities of the individual gates. The practicality of corollary 2 relies on the implicit assumption that the unitarity obtained from RB as well as the average gate fidelities are resulting from closely related gauges [Pro+17a; Wal17; Car+18]. An open problem would be to relax this assumption by connecting more rigorously the interpretations

of different RB experiments.

Acknowledgments — This research was supported by the U.S. Army Research Office through grant W911NF-14-1-0103, CIFAR, the Government of Ontario, and the Government of Canada through CFREF, NSERC and Industry Canada.

3.9 Proofs

3.9.1 Matrix inequalities on the real field

We define the coherence angle of a matrix $M \in \mathbb{R}^{d \times d}$ to be

$$\theta(M) := \arccos \left(\frac{\text{Tr } M}{\sqrt{d} \|M\|_2} \right) \in [0, \pi]. \quad (3.43)$$

Theorem 5

For any nonzero $M_1, M_2 \in \mathbb{R}^{d \times d}$,

$$\cos[\theta(M_1) + \theta(M_2)] \leq \frac{\text{Tr } M_1 M_2}{\|M_1\|_2 \|M_2\|_2} \leq \cos[\theta(M_1) - \theta(M_2)]. \quad (3.44)$$

Moreover, both bounds are saturated for all values of $\|M_1\|_2$, $\|M_2\|_2$, $\theta(M_1)$, and $\theta(M_2)$ in even dimensions.

Proof. By the Cauchy-Schwarz inequality,

$$|\text{Tr } AB|^2 = \left(\sum_{ij} A_{ij} B_{ji} \right)^2 \leq \left(\sum_{ij} A_{ij}^2 \right) \left(\sum_{ij} B_{ij}^2 \right) = (\text{Tr } A^\dagger A) (\text{Tr } B^\dagger B) = \|A\|_2^2 \|B\|_2^2. \quad (3.45)$$

Setting $D_i := \frac{\text{Tr } M_i}{d} \mathbb{I}_d$ for $i = 1, 2$,

$$\begin{aligned} \|M_i - D_i\|_2 &= \sqrt{\text{Tr}(M_i^\dagger M_i - M_i^\dagger D_i - D_i^\dagger M_i + D_i^\dagger D_i)} \\ &= \sqrt{\|M_i\|_2^2 - d^{-1} (\text{Tr } M_i)^2} \\ &= \|M_i\|_2 \sqrt{1 - \cos^2 \theta(M)} \\ &= \|M_i\|_2 \sin \theta(M) \end{aligned} \quad (3.46)$$

using $\text{Tr } M^\dagger = \text{Tr } M$, which holds for $M \in \mathbb{R}^{d \times d}$. Setting $A = M_1 - D_1$ and $B = M_2 - D_2$ in eq. (3.45) and using eq. (3.46) gives

$$|\text{Tr}(M_1 - D_1)(M_2 - D_2)| \leq \|M_1\|_2 \|M_2\|_2 \sin \theta(M_1) \sin \theta(M_2). \quad (3.47)$$

Using eq. (3.43) on the left-hand side gives

$$\begin{aligned} |\operatorname{Tr}(M_1 - D_1)(M_2 - D_2)| &= |\operatorname{Tr} M_1 M_2 - d^{-1} \operatorname{Tr} M_1 \operatorname{Tr} M_2| \\ &= |\operatorname{Tr} M_1 M_2 - \|M_1\|_2 \|M_2\|_2 \cos[\theta(M_1)] \cos[\theta(M_2)]|. \end{aligned} \quad (3.48)$$

Combining eqs. (3.47) and (3.48) with the identity $\cos(a \pm b) = \cos(a) \cos(b) \mp \sin(a) \sin(b)$ gives both desired inequalities. For even d , the bounds of eq. (3.44) are saturated by

$$\frac{\|M_i\|_2}{\sqrt{d}} \begin{pmatrix} \cos \theta(M_i) & -\sin \theta(M_i) \\ \sin \theta(M_i) & \cos \theta(M_i) \end{pmatrix} \otimes \mathbb{I}_{\frac{d}{2}}. \quad (3.49)$$

□

We can generalize the lower bound of eq. (3.44) to matrix products $M_{m:1} := M_1 M_2 \cdots M_m$.

Theorem 6

Let $M_1, \dots, M_m \in \mathbb{R}^{d \times d}$ be such that for all j , $\theta(M_j) = \theta$, $\frac{\operatorname{Tr}(M_j)}{d} = p \leq 1$, $\frac{\|M_j\|_2^2}{d} = u \leq 1$, and $\|M_{j:1}\|_2 \leq \sigma_{\max}$. Then,

$$\left| \frac{\operatorname{Tr} M_{m:1}}{d} - p^m \right| \leq \sigma_{\max} \left(\frac{1 - mp^{m-1} - (m-1)p^m}{(1-p)^2} \right) u \sin^2(\theta) \leq \sigma_{\max} \binom{m}{2} u \sin^2(\theta). \quad (3.50)$$

Proof. Let $D := p\mathbb{I}_d$, and $M_j = D + \Delta_j$. Using a telescoping expansion twice gives

$$\begin{aligned} M_{m:1} - D^m &= \sum_{i=1}^m M_{i-1:1} (M_i - D) D^{m-i} \\ &= \sum_{i=1}^m [D^{i-1} + \sum_{j=1}^{i-1} M_{j-1:1} \Delta_j D^{i-1-j}] \Delta_i D^{m-i}. \end{aligned} \quad (3.51)$$

Taking the trace of each side and using $\operatorname{Tr} \Delta_j = 0$ gives

$$\operatorname{Tr} M_{m:1} - dp^m = \sum_{i=1}^m \sum_{j=1}^{i-1} p^{m-j-1} \operatorname{Tr} M_{j-1:1} \Delta_j \Delta_i. \quad (3.52)$$

Therefore

$$\begin{aligned}
|\mathrm{Tr} M_{m:1} - dp^m| &= \left| \sum_{i=1}^m \sum_{j=1}^{i-1} p^{m-j-1} \mathrm{Tr} M_{j-1:1} \Delta_j \Delta_i \right| \\
&\leq \sum_{i=1}^m \sum_{j=1}^{i-1} p^{m-j-1} |\mathrm{Tr} M_{j-1:1} \Delta_j \Delta_i| && (\triangle \text{ inequality}) \\
&\leq \sum_{i=1}^m \sum_{j=1}^{i-1} p^{m-j-1} \|M_{j-1:1} \Delta_j\|_2 \|\Delta_i\|_2 && (\text{Cauchy-Schwarz inequality}) \\
&\leq \sum_{i=1}^m \sum_{j=1}^{i-1} p^{m-j-1} \sigma_{\max} \|\Delta_j\|_2 \|\Delta_i\|_2 && ([\text{Bha97, Prop. 9.3.6}]) \\
&= \sigma_{\max} du \sin^2(\theta) \sum_{i=1}^m \sum_{j=1}^{i-1} p^{m-j-1}, && (3.53)
\end{aligned}$$

where we used $\|\Delta_j\|_2 = \sqrt{du} \sin(\theta)$ on the last line. Let $S := \sum_{i=1}^m \sum_{j=1}^{i-1} p^{m-j-1} = \sum_{i=1}^{m-1} ip^{i-1}$. Using a telescoping expansion leads to

$$\begin{aligned}
S - pS &= -(m-1)p^{m-1} + \sum_{i=0}^{m-2} p^i \\
&= \frac{1-p^{m-1}}{1-p} - (m-1)p^{m-1} \\
\Rightarrow S &= \frac{1-mp^{m-1} - (m-1)p^m}{(1-p)^2}. && (3.54)
\end{aligned}$$

S is maximized when $p = 1$, in which case it equals $\binom{m}{2}$. □

3.10 Afterword

The results derived in the above work, in particular theorems 2 and 3, raised additional questions.

1. The bounds provided in theorem 2 are easily shown to be saturated for even dimensions. As far as I tried to design different saturating (or approximately saturating) cases, the constructions seemed to always involve some physical unitary channels, which are almost entirely accountable for the leeway in the value of $p(\mathcal{X}\mathcal{Y})$. This is not so surprising given the trigonometric nature of the upper and lower bounds in eq. (3.33). However, it is not clear in the light of this chapter alone that physical unitary motion is indisputably the only mechanism responsible for any substantial looseness in the value of $p(\mathcal{X}\mathcal{Y})$.
2. While the simple 2-fold composition bound provided in theorem 2, and the m -fold unital composition bound provided in eq. (3.35) are saturated for even dimensions, the more general m -fold composition bound provided by eq. (3.34) (theorem 3) is not

shown to be saturated. Does there exist a saturating case or, case failing, what is the tight bound on the fidelity of m -composite channels?

In the next chapter, the two questions above will find complete answers as the payoff of a heavier mathematical machinery. I should emphasize that these are not mere curiosities: for instance, tying unitaries to an unambiguous signature could allow a robust and scalable characterization of coherent errors.

Chapter 4

A polar decomposition for quantum channels

(with applications to bounding error propagation in quantum circuits)

4.1 Foreword

The present chapter mainly consists in a literal transcription of [CAE19], for which my contribution was major. It completes and vastly extends the work introduced in [CWE16] (see previous chapter).

4.2 Compendium

Inevitably, assessing the overall performance of a quantum computer must rely on characterizing some of its elementary constituents and, from this information, formulate a broader statement concerning more complex constructions thereof. However, given the vastitude of possible quantum errors as well as their coherent nature, accurately inferring the quality of composite operations is generally difficult. To navigate through this jumble, we introduce a non-physical simplification of quantum maps that we refer to as the leading Kraus (LK) approximation. The uncluttered parameterization of LK approximated maps naturally suggests the introduction of a unitary-decoherent polar factorization for quantum channels in any dimension. We then leverage this structural dichotomy to bound the evolution – as circuits grow in depth – of two of the most experimentally relevant figures of merit, namely the average process fidelity and the unitarity. We demonstrate that the leeway in the behavior of the process fidelity is essentially taken into account by physical unitary operations.

4.3 Introduction

Just like evaluating a piano doesn't involve playing all possible pieces of music, characterizing a computer (classical or quantum) doesn't involve running all infinitely many circuits.

The natural procedure to characterize both these devices is to gather information on a restricted number of components, and based on that information make conclusions on the quality of more involved constructions (melodies, chords, circuits, magic state injections, etc). When considering the tuning of a piano, the extrapolation is not much of a problem; imperfections are typically tied to specific keys, and they don't tend to propagate over the keyboard as the music goes on, and unless there is some resonant effect, the errors don't coherently interfere. Hence, the quality of individual keys generally guarantees playability. In this sense, the characterization of a piano is similar to that of a classical computer: the well-behaved stochasticity of the noise eases the passage between an assertion of components quality to a broader assertion on the performance of more complex operations. This statement can be phrased the other way around: a limited range of behaviors simplifies the search for imperfections.

In contrast, when characterizing a quantum computer, the jump from a characterization of elementary operations to a quantified assertion on the overall device performance is more knotty; errors can coherently interfere and propagate through the entire device via multi-qubit operations. This thorny situation can be quantified, for instance, by bounding the behavior of the average process fidelity (hereafter the fidelity and its counterpart, the infidelity), an experimentally important figure of merit which captures the overlap between an implemented operation and its target. More precisely, one may ask: "What are the best and worst fidelities of a circuit given a knowledge of the fidelity of its components?" When dealing with a classical scenario, we would expect the difference between the best and worst cases to remain insignificant (remember the piano analogy). In a quantum scenario, however, it is known that the largest discrepancy, which is achieved by unitary errors, grows quickly (quadratically) in the circuit depth (see, for instance, Carignan-Dugas et al [CWE16]). Not so surprisingly, the best case corresponds to a unitary cancellation, and the worst case corresponds to a coherent buildup. This lead to another question: "What if we are guaranteed that the individual errors are not unitary?" In particular, what if we measure the degree to which the error operations are unitary, known as the unitarity[Wal+15], an experimental figure of merit which captures the coherence in the noise? Previous work has given partial answers to this question: Carignan-Dugas et al [CWE16] derive bounds that fall back to the "piano analogy" when the unitarity is minimal; additionally, they provide examples of quantum channels that saturate their bound in the intermediate regime where errors are neither purely unitary nor purely stochastic, but still unital and acting on a single qubit coupled with a system of arbitrary (but finite) size.¹ In this paper, we generalize that bound to all dimensions and show its near saturation (i.e. to second order in the infidelity or better) and also account for non-unital processes. That is, we provide a closely saturated bound for all finite-dimensional quantum channels. While this is already an interesting result, the tools that we develop to generalize the bound help us answering a far more fundamental question. In previous work, the saturation was shown through a handful of examples. Now, we provide a complete descriptive answer to:

What is the set of mechanisms responsible for the discrepancy between the best and the worst fidelity of a circuit?

¹They attribute all the error dynamics on the qubit; the intuitive geometric picture offered by parameterization of processes acting on the Bloch sphere allows showing the saturation of the bound for unital channels. The bound in the non-unital case included a dimensional factor which prevented its saturation.

This would not be much of a fundamental question if the answer didn't also unravel an important dichotomy in classifying quantum errors. Given the intricate geometry of quantum states [BZ06], the answer could have included some obscure blend of non-intuitive mechanisms, leaving us with yet another resignation in the attempt to intuitively reason about quantum dynamics. Although, for once, this is not the case: the discrepancy between the best and worst fidelity is, to high precision, entirely taken into account by unitary dynamics². Even more surprisingly, the unitary dynamics itself is precisely the product of the “unitary factors” of individual circuit components. As we demonstrate through theorem 9, every non-catastrophic channel (see definition 4) can be decomposed as a physical unitary followed or preceded by a decoherent channel. For realistic errors, the unitary is unique and is referred to as the coherent factor. This factorization is analogous to the well-known matrix polar decomposition and, as we will show, directly stems from it. The uniqueness of the coherent factor might puzzle the skeptical reader. For example, how should we unambiguously define such factor in the case of an error which consists of a mixture of near-identity unitaries (i.e. $\mathcal{A}(\rho) = \sum_i p_i U_i \rho U_i^\dagger$, where $U_i \approx \mathbb{I}_d$)? Should it be the unitary operation with the highest weight? Should it relate with some kind of ensemble average over the associated Hamiltonians? To systematically answer this type of question, we introduce the leading Kraus (LK) approximation (see definition 5), a sub-parameterization of quantum channels which, among other things, exposes a natural definition for the coherent and decoherent factors of a channel.

What allows us to really profit from the channel polar decomposition is the surprising property that the LK approximation, despite its seemingly bare structure, closely captures the evolution of the fidelity and unitarity in circuits. That is, we can mathematically replace all the channels in a circuit by their respective LK approximation and still expect to accurately bound its fidelity and unitarity (see theorems 7 and 8). Working with the uncluttered structure offered by the LK approximation helped us identify and rule out pathological error scenarios, which we refer to as “extremal” (see section 4.7.1 for more details). For all realistic noisy channels, we derive the following observations (they hold to high precision):

- i. The infidelity (the counterpart to the fidelity) of a channel can be split into two terms (see theorem 14 and the discussion that immediately follows):
 - (a) a coherent infidelity, which corresponds to the infidelity of the coherent factor to the target channel;
 - (b) a decoherent infidelity, which corresponds to the infidelity of the decoherent factor to the identity.
- ii. The decoherent infidelity of a channel is in one-to-one correspondence with its unitarity. Moreover, the decoherent infidelity corresponds to the minimum infidelity of the channel after the application of a unitary (the coherent infidelity is correctable through a composition with a unitary). (See theorem 13.)
- iii. The unitarity of a composite channel is a decay function expressed in terms of individual channels' unitarity. (See theorem 11.)

²Given realistic errors, which are properly defined in section 4.7.1, and are more formally referred to as “equable”. The equability assumption corresponds to ruling out two types of errors. 1) Extreme dephasing effects between a small set of states and the rest of the systems. 2) Extreme Hamiltonian alterations.

- iv. The fidelity of the composition of decoherent channels is a decay function expressed in terms of individual channels' fidelity. (See theorem 12.)
- v. The fidelity of a general composition is upper bounded by a decay dictated by the decoherent factors (hence by the unitarity of individual components). (See theorem 15.)
- vi. The discrepancy between the upper and the lower bound of the fidelity is captured by the fidelity of the composition of the coherent factors (to the target circuit). (See theorem 14.)

These realizations are directly applicable to the analysis and development of process characterization methods. The fidelity of various error processes can be robustly and efficiently estimated through a scalable experimental protocol known as randomized benchmarking (RB) [EAZ05; Dan+09; MGE11; MGE12] and a family of generalizations thereof [Kni+08; Mag+12; Gam+12; Gae+12; GFC14; Bar+14; WF14; CWE15; Wal+15; WBE15; She+16; Cro+16; Com+17; Has+18; BE18; SH18; Hel+18; Pro+18]. To remain efficient as quantum devices grow larger, RB experiments only extract partial information about specific sets of components. A known challenge is to leverage this limited view to formulate a more rounded understanding of the device. By looking at the fidelity of well-designed compositions, it should be possible to extract other figures of merit attached to quantum processes. The idea is that since process matrices dictate the evolution of the fidelity, conversely, the evolution of the fidelity can tell us information about process matrices. However, given the generally large amount of parameters involved in process matrices, it is not always immediately clear how the signal obtained from extracting the fidelity of various circuit compositions connects with quantities of interest. The above six enumerated observations allow to make more sense out of such signals.

We structure the paper as follows. In section 4.4, we introduce important characterization figures of merit – the average process fidelity and the unitarity – and relate them with the Kraus operator formalism. In section 4.5, we define the LK approximation and present its aptitude in capturing important characteristics of evolving quantum circuits. In section 4.6, based on the emergent mathematical structure of LK approximated channels, we show the existence of a channel polar unitary-decoherent decomposition. In section 4.7, we make use of the approximation to demonstrate key behavioral aspects of quantum circuits based on partial knowledge of their components.

For the sake of conciseness, most demonstrations are pushed to the appendix. Moreover, in the main text, certain results have been abridged by gathering higher order terms under the acronym "H.O.T.". The complete expressions – which are not any more insightful than their abbreviated analog – are provided in the appendix.

4.4 Channel properties captured by the leading Kraus operator

A quantum channel is a completely-positive (CP), trace-preserving (TP) map acting on $M_d(\mathbb{C})$. Given a quantum channel $\mathcal{A} : M_d(\mathbb{C}) \rightarrow M_d(\mathbb{C})$, the Choi matrix of \mathcal{A} is defined as [Cho75]

$$\text{Choi}(\mathcal{A}) := \sum_{ij} E_{ij} \otimes \mathcal{A}(E_{ij}), \quad (\text{Choi matrix})$$

where

$$E_{ij} := e_i e_j^\dagger, \quad (4.1)$$

and e_i are canonical orthonormal vectors. The Choi matrix is positive semi-definite iff \mathcal{A} is CP, and has trace d if \mathcal{A} is TP or unital³. Since $\text{Choi}(\mathcal{A}) \geq 0$, it has a spectral decomposition of the form

$$\text{Choi}(\mathcal{A}) := \sum_{i=1}^{d^2} \text{col}(A_i) \text{col}^\dagger(A_i), \quad (4.2)$$

$$= \sum_{i=1}^{d^2} \|A_i\|_2^2 \text{col}(\underline{A}_i) \text{col}^\dagger(\underline{A}_i), \quad (4.3)$$

where $\text{col}(A) \in \mathbb{C}^{d^2}$ denotes the column vectorization of a matrix $A \in M_d(\mathbb{C})$ ⁴, $\|\cdot\|_p$ denotes the Schatten p -norm, and $\underline{A} = A/\|A\|_2$ denotes normalized matrices with respect to the Schatten 2-norm. The eigenvectors $\text{col}(\underline{A}_i)$ are orthonormal, and without loss of generality the eigenvalues are ordered with respect to the Frobenius norm (Schatten 2-norm):

$$\|A_1\|_2^2 \geq \|A_2\|_2^2 \geq \dots \geq \|A_{d^2}\|_2^2 \geq 0. \quad (4.4)$$

Given a spectral decomposition like eq. (4.2), we can express the channel's action on states $\rho \in M_d(\mathbb{C})$ as [Kra+83]:

$$\mathcal{A}(\rho) = \sum_{i=1}^{d^2} A_i \rho A_i^\dagger, \quad (\text{Kraus decomposition})$$

with

$$\langle A_i, A_j \rangle = \|A_i\|_2^2 \delta_{ij}, \quad (4.5)$$

where the usual Hilbert-Schmidt inner product is used. Notice that the TP condition implies that $\sum_i (\|A_i\|_2^2/d) = 1$. The matrices $A_i \in M_d(\mathbb{C})$ are referred to as (ordered) canonical Kraus operators. In this work, A_1 (which is associated with the highest Choi matrix eigenvalue $\|A_1\|_2^2$) will deserve special attention, and is attributed the title of ‘‘leading Kraus (LK) operator’’. In general, A_1 might be non-unique when the spectrum of the Choi matrix is degenerate. However, in this work we focus on non-catastrophic channels (definition 4), for which A_1 is unique.

Given an operation \mathcal{A} and a target unitary channel $\mathcal{U}(\rho) = U\rho U^\dagger$ ⁵, we can compare the overlap of their outputs given specific inputs $M \in M_d(\mathbb{C})$ through the M -fidelity:

$$f_M(\mathcal{A}, \mathcal{U}) := \frac{\langle \mathcal{A}(M), \mathcal{U}(M) \rangle}{\|M\|_2^2}. \quad (4.6)$$

³A channel \mathcal{A} is unital iff $\mathcal{A}(\mathbb{I}_d) = \mathbb{I}_d$.

⁴ $\text{col}(A) := \sum_{ij} A_{ij} e_j \otimes e_i$

⁵For unitaries, we used the calligraphic font to denote the channel and the non-calligraphic one to denote its associated $d \times d$ unitary matrix.

The well-known average gate fidelity is obtained by averaging the M -fidelities uniformly (i.e. with respect to the Haar measure) over all physical pure states $|\psi\rangle\langle\psi|$:

$$F(\mathcal{A}, \mathcal{U}) := \mathbb{E}_{\text{Haar}} f_{|\psi\rangle\langle\psi|}(\mathcal{A}, \mathcal{U}). \quad (4.7)$$

The average infidelity r is simply a shorthand for $1 - F$. Instead of averaging over quantum states, we could also average uniformly over all operators $M \in M_d(\mathbb{C})$. More precisely, given any orthogonal operator basis $\{B_i\}$ for $M_d(\mathbb{C})$, we can uniformly average over the M -fidelities f_{B_i} , which yields the average process fidelity⁶

$$\Phi(\mathcal{A}, \mathcal{U}) := \mathbb{E}_{\{B_i\}} f_{B_i}(\mathcal{A}, \mathcal{U}). \quad (4.8)$$

Compared to Φ , F puts a slightly higher weight over the identity component \mathbb{I}_d . The TP condition enforces this special component to take a fixed value, $f_{\mathbb{I}_d} = 1$. Hence the two quantities are closely related via [Nie02]:

$$F(\mathcal{A}, \mathcal{U}) = \frac{d\Phi(\mathcal{A}, \mathcal{U}) + 1}{d + 1}. \quad (4.9)$$

$F(\mathcal{A}, \mathcal{U})$ is the overlap between the output state $\mathcal{A}(\rho)$ of an implemented channel \mathcal{A} and its ideal output $\mathcal{U}(\rho)$, averaged over all physical pure input states $|\psi\rangle\langle\psi|$. While $F(\mathcal{A}, \mathcal{U})$ conveys a more graspable interpretation, it will remain easier here to work with $\Phi(\mathcal{A}, \mathcal{U})$ since it ties with the Kraus operators through

$$\Phi(\mathcal{A}, \mathcal{U}) = \sum_{i=1}^{d^2} \left| \left\langle \frac{A_i}{\sqrt{d}}, \frac{U}{\sqrt{d}} \right\rangle \right|^2 = \sum_{i=1}^{d^2} (\|A_i\|_2^2/d) \left| \left\langle \underline{A}_i, U/\sqrt{d} \right\rangle \right|^2. \quad (4.10)$$

Since $\{\underline{A}_i\}$ forms an orthonormal basis and $\|U/\sqrt{d}\|_2 = 1$, it follows that

$$\sum_{i=1}^{d^2} \left| \left\langle \underline{A}_i, U/\sqrt{d} \right\rangle \right|^2 = 1. \quad (4.11)$$

If $\|A_i\|_2^2/d$ can be thought as the “weights” of the Kraus operators, $\left| \left\langle \underline{A}_i, U/\sqrt{d} \right\rangle \right|^2$ can be thought as normalized overlaps with the target U .

To quantify the coherence of a quantum channel, one could wonder how much the Bloch vectors (the traceless component of quantum states [Blo46]) are contracted. For instance, consider the unitarity, which is the squared length ratio of the Bloch vectors before and after the action of the channel \mathcal{A} , averaged over all physical Bloch vector inputs corresponding to pure states $|\psi\rangle\langle\psi| - \mathbb{I}_d/d$ [Wal+15]:

$$u(\mathcal{A}) := \mathbb{E}_{\text{Haar}} \frac{\|\mathcal{A}(|\psi\rangle\langle\psi| - \mathbb{I}_d/d)\|_2^2}{\| |\psi\rangle\langle\psi| - \mathbb{I}_d/d \|_2^2}. \quad (4.12)$$

⁶For the readers familiar with the χ -matrix, $\Phi(\mathcal{A}, \mathcal{U})$ is a way to express the well-known χ_{00} element. Of course, the χ -matrix has to be defined with respect to an orthonormal operator basis $\{B_i\}$ with $B_0 = U$. Some might also be more familiar with the notion of entanglement fidelity, which is again Φ .

Let's extend the domain of Φ to include a new function of \mathcal{A} :

$$Y(\mathcal{A}) := \sqrt{\Phi(\mathcal{A}^\dagger \mathcal{A}, \mathcal{I})} = \sqrt{\sum_{i,j=1}^{d^2} \left| \left\langle \frac{A_j^\dagger A_i}{\sqrt{d}}, \frac{I}{\sqrt{d}} \right\rangle \right|^2} = \sqrt{\sum_{i=1}^{d^2} \left(\frac{\|A_i\|_2^2}{d} \right)^2}. \quad (4.13)$$

Straightforward calculations closely relate the unitarity to Y via

$$u(\mathcal{A}) = \frac{d^2 Y^2(\mathcal{A}) - 1}{d^2 - 1}. \quad (4.14)$$

(Notice that the notation alludes to the connection between greek and latin alphabets; it relates "phi" to "F" and "upsilon" to "u".)

We are ready to express a first result:

Lemma 2

Consider a CPTP map \mathcal{A} with ordered canonical Kraus decomposition

$$\mathcal{A}(\rho) = \sum_{i=1}^{d^2} A_i \rho A_i^\dagger.$$

Then,

$$0 \leq Y^2(\mathcal{A}) - \left(\frac{\|A_1\|_2^2}{d} \right)^2 \leq (1 - Y^2(\mathcal{A}))^2. \quad (4.15)$$

Proof. $Y^2(\mathcal{A})$ can be expanded as a sum over d^2 terms:

$$Y^2(\mathcal{A}) = \sum_i \left(\frac{\|A_i\|_2^2}{d} \right)^2. \quad (4.16)$$

Using Hölder's inequality on the RHS yields

$$Y^2(\mathcal{A}) \leq \max_i \frac{\|A_i\|_2^2}{d} \sum_j \frac{\|A_j\|_2^2}{d} = \|A_1\|_2^2 / d. \quad (4.17)$$

Using this lower bound on $\|A_1\|_2$, we get

$$\begin{aligned}
Y^2(\mathcal{A}) &= \left(\frac{\|A_1\|_2^2}{d}\right)^2 + \sum_{i \neq 1} \left(\frac{\|A_i\|_2^2}{d}\right)^2 & (4.18) \\
&\leq \left(\frac{\|A_1\|_2^2}{d}\right)^2 + \left(\sum_{i \neq 1} \frac{\|A_i\|_2^2}{d}\right)^2 \\
&= \left(\frac{\|A_1\|_2^2}{d}\right)^2 + \left(1 - \frac{\|A_1\|_2^2}{d}\right)^2 & (\sum_i \frac{\|A_i\|_2^2}{d} = 1) \\
&\leq \left(\frac{\|A_1\|_2^2}{d}\right)^2 + (1 - Y^2(\mathcal{A}))^2 & (\text{eq. (4.17)})
\end{aligned}$$

From eq. (4.16), it follows that $\left(\frac{\|A_1\|_2^2}{d}\right)^2 \leq Y^2(\mathcal{A})$, which completes the proof. \square

It follows from eq. (4.15) that $Y^2(\mathcal{A}) > 1/2$ is a sufficient condition to guarantee the uniqueness of A_1 ⁷. This partially motivates the following definition:

Definition 4: non-catastrophic channels

A channel \mathcal{A} is said to be non-catastrophic if it overlaps enough with its targeted unitary channel \mathcal{U} :

$$\Phi(\mathcal{A}, \mathcal{U}) > 1/2, \quad (4.19)$$

and if it doesn't greatly contract the Bloch vectors:

$$Y^2(\mathcal{A}) > 1/2. \quad (4.20)$$

The condition described by eq. (4.19) allows us to express our second result:

Lemma 3

Consider a non-catastrophic channel \mathcal{A} with unitary target \mathcal{U} and ordered canonical Kraus decomposition

$$\mathcal{A}(\rho) = \sum_{i=1}^{d^2} A_i \rho A_i^\dagger.$$

Then,

$$0 \leq \Phi(\mathcal{A}, \mathcal{U}) - \left| \left\langle \frac{A_1}{\sqrt{d}}, \frac{\mathcal{U}}{\sqrt{d}} \right\rangle \right|^2 \leq (1 - Y^2(\mathcal{A}))(1 - \Phi(\mathcal{A}, \mathcal{U})). \quad (4.21)$$

⁷Indeed, it implies that $\|A_1\|_2^2/d > 1/\sqrt{2} > 1/2$.

Proof. Using Hölder's inequality on the RHS of eq. (4.10), we have

$$\Phi(\mathcal{A}, \mathcal{U}) \leq \max_i |\langle \underline{A}_i, \underline{U} \rangle|^2 \sum_{j=1}^{d^2} (\|A_j\|_2^2/d) = \max_i |\langle \underline{A}_i, \underline{U} \rangle|^2. \quad (4.22)$$

For non-catastrophic channels, it must be that $\max_i |\langle \underline{A}_i, \underline{U} \rangle|^2 = |\langle \underline{A}_1, \underline{U} \rangle|^2$. To see this more clearly, let $|\langle \underline{A}_1, \underline{U} \rangle|^2 = 1/2 - \epsilon_1$ and $\|A_1\|_2^2/d = 1/2 + \epsilon_2$, where $\epsilon_2 > 0$ from the non-catastrophic condition. Then, consider the following inequalities:

$$\Phi(\mathcal{A}, \mathcal{U}) = |\langle \underline{A}_1, \underline{U} \rangle|^2 (\|A_1\|_2^2/d) + \sum_{i \neq 1} |\langle \underline{A}_i, \underline{U} \rangle|^2 (\|A_i\|_2^2/d) \quad (\text{Equation (4.10)})$$

$$\begin{aligned} &\leq |\langle \underline{A}_1, \underline{U} \rangle|^2 (\|A_1\|_2^2/d) + \left(\sum_{i \neq 1} (\|A_i\|_2^2/d) \right) \left(\sum_{j \neq 1} |\langle \underline{A}_j, \underline{U} \rangle|^2 \right) \\ &= |\langle \underline{A}_1, \underline{U} \rangle|^2 (\|A_1\|_2^2/d) + (1 - |\langle \underline{A}_1, \underline{U} \rangle|^2) (1 - \|A_1\|_2^2/d) \end{aligned} \quad (4.23)$$

$$= 1/2 - 2\epsilon_1\epsilon_2. \quad (4.24)$$

From the non-catastrophic condition, $\epsilon_1 < 0$, which implies that $|\langle \underline{A}_1, \underline{U} \rangle|^2 > 1/2$.

Hence, eq. (4.22) can be reexpressed into $1 - |\langle \underline{A}_1, \underline{U} \rangle|^2 \leq 1 - \Phi(\mathcal{A}, \mathcal{U})$, which yields the following:

$$\begin{aligned} \Phi(\mathcal{A}, \mathcal{U}) &\leq |\langle \underline{A}_1, \underline{U} \rangle|^2 (\|A_1\|_2^2/d) + (1 - |\langle \underline{A}_1, \underline{U} \rangle|^2) (1 - \|A_1\|_2^2/d) \quad (\text{Equation (4.23)}) \\ &\leq (\|A_1\|_2^2/d) |\langle \underline{A}_1, \underline{U} \rangle|^2 + (1 - Y^2(\mathcal{A})) (1 - \Phi(\mathcal{A}, \mathcal{U})). \end{aligned} \quad (\text{eq. (4.17)})$$

From eq. (4.10) we also have $(\|A_1\|_2^2/d) |\langle \underline{A}_1, \underline{U} \rangle|^2 \leq \Phi(\mathcal{A}, \mathcal{U})$, which completes the proof. \square

The LK operator alone provides a very accurate approximation of $1 - \Phi$ and $1 - Y$. This only begins a list of realizations regarding the role of LK operators in quantum dynamics. As we will see, they also contain most of the information necessary to describe the *evolution* of Φ and Y .

4.5 The LK approximation and two evolution theorems

The last section naturally suggests the following channel approximation as a means to partially characterize non-catastrophic quantum dynamics:

Definition 5: the Leading Kraus (LK) approximation

Consider a channel $\mathcal{A} : M_d(\mathbb{C}) \rightarrow M_d(\mathbb{C})$ with leading Kraus operator A_1 . We define its leading Kraus (LK) approximation as:

$$\mathcal{A}^*(\rho) = A_1 \rho A_1^\dagger. \quad (4.25)$$

Notice that \mathcal{A}^* is always CP ($\text{Choi}(\mathcal{A}^*) \geq 0$), but is TP iff \mathcal{A} is unitary. Hence, \mathcal{A}^* generally fails to be physical. However, as we will see, it closely describes the dynamics of certain

physical quantities, so one may qualify this map as “quasi-dynamical”. The general specification of a map acting on a d -dimensional quantum system requires roughly d^4 parameters, and due to the intricate geometry of quantum states, the parameterization of its range of action is quite convoluted. In contrast, the LK approximation is remarkably transparent: it is fully parameterized by $d \times d$ matrices with spectral radius smaller than 1 (contractions) and Frobenius norm greater than $d/\sqrt{2}$ ⁸. If the noise is non-catastrophic, every quantum map has a corresponding LK approximation, and every $d \times d$ linear contraction corresponds to at least one quantum operator.

Given m channels \mathcal{A}_i , we denote the composition $\mathcal{A}_m \circ \mathcal{A}_{m-1} \circ \dots \circ \mathcal{A}_2 \circ \mathcal{A}_1$ as $\mathcal{A}_{m:1}$. Replacing every element of the composition by its LK approximation, $\mathcal{A}_m^* \circ \mathcal{A}_{m-1}^* \circ \dots \circ \mathcal{A}_2^* \circ \mathcal{A}_1^*$, is noted as $\mathcal{A}_{m:1}^*$. In general, the composition operation doesn't commute with the LK approximation, that is $\mathcal{A}_{m:1}^* \neq (\mathcal{A}_{m:1})^*$. To put it in other words, the LK operator of a circuit is generally not the multiplication of the LK operators of its elements. However, while $\mathcal{A}_{m:1}^*$ provides an incomplete description of $\mathcal{A}_{m:1}$, they still might share some comparable characteristics. That is, there might exist some function $f : \text{CP maps} \rightarrow \mathbb{R}$ for which $f(\mathcal{A}_{m:1}^*) \approx f(\mathcal{A}_{m:1})$. As we show, not only there exist such functions, but some of them correspond to important experimental figures of merit. From the previous section, we know that $\Phi(\mathcal{A}, \mathcal{U}) \approx \Phi(\mathcal{A}^*, \mathcal{U})$ and $Y(\mathcal{A}) \approx Y(\mathcal{A}^*)$. What may be more surprising are the following two theorems:

Theorem 7: the unitarity of a circuit after approximating its elements

Consider m non-catastrophic channels \mathcal{A}_i with respective unitary targets \mathcal{U}_i and suppose that the composition $\mathcal{A}_{m:1}$ is also non-catastrophic. Then,

$$0 \leq Y^2(\mathcal{A}_{m:1}) - Y^2(\mathcal{A}_{m:1}^*) \leq (1 - Y(\mathcal{A}_{m:1}^*))^2 \leq (1 - Y^2(\mathcal{A}_{m:1}))^2. \quad (4.26)$$

Theorem 8: the fidelity of a circuit after approximating its elements

Consider m non-catastrophic channels \mathcal{A}_i with respective unitary targets \mathcal{U}_i and suppose that the composition $\mathcal{A}_{m:1}$ is also non-catastrophic. Then,

$$\begin{aligned} 0 \leq \Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1}) - \Phi(\mathcal{A}_{m:1}^*, \mathcal{U}_{m:1}) &< \\ (1 - \Phi(\mathcal{A}_{m:1}^*, \mathcal{U}_{m:1})) \sum_{i=1}^m (1 - Y(\mathcal{A}_i^*)) + \frac{1}{2} \left(\sum_{i=1}^m (1 - Y^*(\mathcal{A}_i)) \right)^2 &\leq \\ (1 - \Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1})) \sum_{i=1}^m (1 - Y^2(\mathcal{A}_i)) + \frac{1}{2} \left(\sum_{i=1}^m (1 - Y^2(\mathcal{A}_i)) \right)^2 + \text{H.O.T.} &\quad (4.27) \end{aligned}$$

\mathcal{A}^* differs from the veritable channel \mathcal{A} in many ways as shown by comparing various M -fidelities $f_M(\mathcal{A}_{m:1}, \mathcal{U}_{m:1})$ with $f_M(\mathcal{A}_{m:1}^*, \mathcal{U}_{m:1})$ (see two animated examples at <https://youtu.be/ITrBTIJHJJM> and <https://youtu.be/A6i-k6eHsGM>). Of course, some kind of discrepancy is expected since the LK approximation contains only d^2 parameters instead

⁸This last constraint only prevents catastrophic noise scenarios.

of $\sim d^4$. Essentially, the LK operators closely dictate the evolution of the average of M -fidelities $\Phi = \mathbb{E}f_M$ (see eq. (4.8)), while the other Kraus operators add or subtract to specific M -fidelities f_M in such a way that the sum of those variations almost exactly cancels.

The evolution theorems presented in this section will greatly help classify different types of errors⁹. Indeed, they allow tying behavioral signatures in the evolution of Y and Φ to more digestible error profiles. In particular, the two theorems further motivate, as shown in section 4.7, the definition of a natural dichotomy in quantum channels (itself introduced in section 4.6).

4.6 A polar decomposition for quantum channels

4.6.1 Defining decoherence

Due to the intricate geometry of d -dimensional quantum states [BZ06], quantum processes can be delicate to dissect. One of the main reasons the single qubit Bloch sphere is frequently invoked stems from the simple picture it offers:

- i. There is a clear bijection between quantum states and the Bloch ball [Blo46].
- ii. The action on the Bloch vectors can be decomposed into a positive semi-definite contraction $|M| \leq \mathbb{I}_3$, followed by orthogonal matrix $R \in O(3)$, which corresponds to a physical unitary $U \in SU(2)$, added to a translational vector \vec{t} (the non-unital vector) [FA99; RSW02; BW04]:

$$\vec{v} \rightarrow \underbrace{R|M|}_M \vec{v} + \vec{t}, \quad (4.28)$$

where $|M|$ denotes $(M^\dagger M)^{\frac{1}{2}}$. $M = R|M|$ is referred to the unital matrix.

Not every contraction $|M|$ is physical; for instance, transforming the Bloch sphere into a disk violates CP-ness (the folkloric “no pancake” theorem [Blu+10]). A thorough analysis of CPTP maps acting on $M_2(\mathbb{C})$ is provided in [RSW02]. For higher dimensions, the Bloch sphere imagery falls apart in many ways:

- i. The generalized Bloch space is not a $(d^2 - 1)$ -ball (with respect to the 2-norm on \mathbb{R}^{d^2-1}) [BZ06].
- ii. If we express the action on the Bloch vector as in eq. (4.28) where $R \in O(d^2 - 1)$ and $|M| \geq 0$, we realize that
 - (a) R generally doesn’t correspond to a physical unitary operation in $SU(d)$ (the unitary map defined by $\vec{v} \rightarrow R\vec{v}$ is not necessarily CP).
 - (b) $|M|$ is not necessarily a contraction. Its spectrum is optimally upper-bounded by $\sqrt{\frac{d}{2}}$ for even dimensions and $(\frac{1}{d-1} + \frac{1}{d+1})^{-\frac{1}{2}}$ for odd dimensions [PG+06].

The polar decomposition of the unital matrix M generally splits it into two nonphysical constituents. Essentially, the unitary factor of M ($R \in O(d^2 - 1)$ s.t. $R^{-1}M \geq 0$) can’t generally

⁹An error channel simply refers to a channel with identity target \mathcal{I} .

be interpreted as a physically meaningful unitary operation. To see this, consider the following canonical Kraus decomposition:

$$\begin{aligned}
 A_1 &= \begin{pmatrix} \cos(\alpha) & 0 & 0 \\ 0 & \cos(\alpha/2)e^{i\alpha^3/2} & 0 \\ 0 & 0 & \cos(\alpha/2)e^{-i\alpha^3/2} \end{pmatrix}; \\
 A_2 &= \begin{pmatrix} \sin(\alpha) & 0 & 0 \\ 0 & -\sin(\alpha/2)e^{i(\alpha+\alpha^3/2)} & 0 \\ 0 & 0 & -\sin(\alpha/2)e^{-i(\alpha+\alpha^3/2)} \end{pmatrix}. \quad (4.29)
 \end{aligned}$$

The spectrum of the associated unital part M is a subset of the spectrum of $A_1^* \otimes A_1 + A_2^* \otimes A_2 \in M_{d^2}(\mathbb{C})$ ¹⁰. By expanding up to order α^4 , it is straightforward to show that the phase factors of M are all ≈ 1 except for a single conjugate pair $\phi_{\pm} \approx \exp(\pm i3\alpha^3/2)$. This single pair can't be factored into any unitary process since any non-trivial $V^* \otimes V$ contains at least two conjugate pairs. Hence, trying to cancel the rotating component of the spiraling action (see fig. 4.1) induced on \vec{v}_{\pm} by ϕ_{\pm} would merely relocate the spiraling motion on an other pair of eigenvectors \vec{v}'_{\pm} (or on multiple other pairs). To put it simply, spiraling is inherent to some decoherent processes. To explicitly show this, we constructed an example in which the rotation factors in the spirals couldn't be accounted for by any physical unitary (without creating more spirals).

Separating a quantum channel \mathcal{A} into a composition of a physical unitary \mathcal{V} and a decoherent operation \mathcal{D} (i.e. $\mathcal{A} = \mathcal{V} \circ \mathcal{D}$ or $\mathcal{A} = \mathcal{D} \circ \mathcal{V}$) demands a more careful surgery. If one were to allocate too many rotating components to the unitary factor, \mathcal{V} may fail to remain physical; on the other hand, allocating too little unitary action to \mathcal{V} may leave the allegedly decoherent factor \mathcal{D} with some physically reversible motion. In fact, depending on the definition of decoherence, it is not even clear if such surgery is even possible. Here, we propose a definition of decoherence which can be used to easily decompose any non-catastrophic quantum channel into a composition of a unitary channel with a decoherent one.

¹⁰ $A_1^* \otimes A_1 + A_2^* \otimes A_2$ is the matrix acting on the column-vectorized density matrices, and has an extra eigenvalue of 1 due the TP condition. Here the star $*$ denotes the complex conjugation, which is not to be confused with the star $*$ used for the LK approximation.

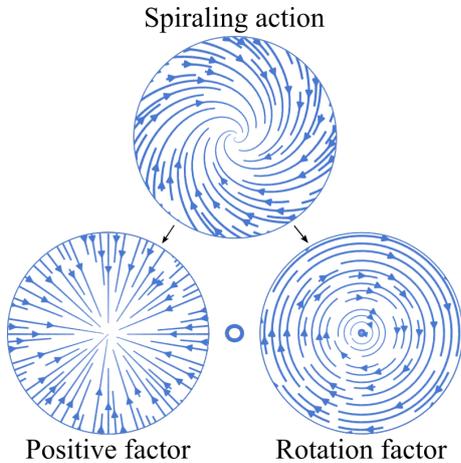


FIGURE 4.1: Representation of the spiraling action of a normal matrix acting on a 2×2 subspace. The polar decomposition, in this case, separates the azimuthal and radial components of the action. Quantum dynamics on $d > 2$ can generate spiraling actions on the Bloch space for which the rotation factor can't be interpreted as a physical unitary operation. In this sense, spiraling, despite generating some rotating action, is inherent to some decoherent dynamics.

Consider a channel \mathcal{A} . Its LK operator $A_1 \in M_d(\mathbb{C})$ can be factored into a $d \times d$ unitary component U multiplied with a positive semi-definite contraction $0 < |A_1| \leq \mathbb{I}_d$, i.e. $A_1 = V|A_1|$. This polar decomposition provides a geometric understanding of the range of action of LK approximated channels on the space of quantum states. The absence of phase factors in the spectrum of $|A_1|$ motivates the following definition:

Definition 6: decoherent channel

A non-catastrophic channel \mathcal{A} is said to be decoherent if its LK operator is positive semi-definite:

$$A_1 \geq 0. \quad (4.30)$$

From this definition immediately follows a unitary-decoherent decomposition for quantum channels:

Theorem 9: a polar decomposition for quantum channels

Any non-catastrophic quantum channel \mathcal{A} can be expressed as a composition of a unitary channel \mathcal{V} with an decoherent channel $\mathcal{D} = \mathcal{V}^\dagger \circ \mathcal{A}$ (or $\mathcal{D}' = \mathcal{A} \circ \mathcal{V}^\dagger$):

$$\mathcal{A} = \mathcal{V} \circ \mathcal{D}, \quad (4.31a)$$

$$\mathcal{A} = \mathcal{D}' \circ \mathcal{V}. \quad (4.31b)$$

In terms of LK approximation, we have:

$$\mathcal{A}^*(\rho) = A_1 \rho A_1^\dagger = V|A_1| \rho |A_1|^\dagger V^\dagger, \quad (4.32a)$$

$$\mathcal{D}^*(\rho) = |A_1| \rho |A_1|^\dagger, \quad (4.32b)$$

$$\mathcal{D}'^*(\rho) = V|A_1| V^\dagger \rho V|A_1|^\dagger V^\dagger. \quad (4.32c)$$

Proof. Under the composition $\mathcal{V}^\dagger \circ \mathcal{A}$, the canonical Kraus operators $\{A_i\}$ of \mathcal{A} are mapped to $\{V^\dagger A_i\}$, since it preserves their orthonormality. Given the polar decomposition $A_1 = V|A_1|$, it follows that the LK operator of $\mathcal{V}^\dagger \circ \mathcal{A}$ is positive semi-definite. \square

4.6.2 The dynamics induced from decoherent channels as infinitesimal generators

While the proof of theorem 9 nearly trivially follows from definition 6, it remains to show that decoherent channels as we defined them deserve such an appellation. An interesting angle to initially justify our definition of decoherence is to observe its contribution in the Gorini—Kossakowski—Sudarshan—Lindblad (GKSL) equation [Lin76; GKS76]. Consider a time evolution dictated by instantaneous CPTP channels¹¹ with (possibly time-dependent) canonical Kraus operators $\{A_k(t, dt)\}$:

$$\rho(t + dt) = \sum_k A_k(t, dt) \rho(t) A_k^\dagger(t, dt). \quad (4.33)$$

¹¹This corresponds to the well-known Markovian regime.

Since dt is infinitesimal, the instantaneous LK operator $A_1(t, dt)$ must be close to \mathbb{I} , and can be expressed as

$$\begin{aligned} A_1(t, dt) &= \exp(-iH(t)dt - P(t)dt) \\ &= \mathbb{I} - iH(t)dt - P(t)dt + O(dt^2), \end{aligned} \quad (4.34)$$

where $H(t)$ is Hermitian and $P(t)$ is positive semi-definite. The TP condition can be expressed as

$$\sum_k A_k^\dagger(t, dt) A_k(t, dt) = \mathbb{I}, \quad (4.35)$$

which combined with eq. (4.34) yields

$$P(t)dt = \frac{1}{2} \sum_{k \neq 1} A_k^\dagger(t, dt) A_k(t, dt) + O(dt^2). \quad (4.36)$$

This enforces the remaining instantaneous Kraus operators $A_{k \neq 1}(t, dt)$ to scale as \sqrt{dt} , and leaves us with

$$\frac{d}{dt} \rho(t) = -i [H(t), \rho(t)] + \sum_{k \neq 1} L_k(t) \rho(t) L_k^\dagger(t) - \frac{1}{2} \left\{ \sum_{k \neq 1} L_k^\dagger(t) L_k(t), \rho(t) \right\}, \quad (4.37)$$

where

$$L_k(t) := \lim_{dt \rightarrow 0} \frac{A_k(t, dt)}{\sqrt{dt}}, \quad (4.38)$$

and $[A, B] := AB - BA$, $\{A, B\} := AB + BA$ are respectively the well-known commutator and anticommutator. The fact that $\{A_k(t, dt)\}$ are canonical (hence orthogonal) at every moment in time implies that $\langle A_1(t, dt), A_{k \neq 1}(t, dt) \rangle = 0$, which by using eq. (4.34) results in

$$\text{Tr} A_{k \neq 1}(t, dt) = -idt \text{Tr} H(t) A_{k \neq 1}(t, dt) + dt \text{Tr} P(t) A_{k \neq 1}(t, dt) + O(dt^2 \sqrt{dt}). \quad (4.39)$$

This together with eq. (4.38) implies that

$$\text{Tr} L_k(t) = 0. \quad (4.40)$$

Notice that the Lindblad operators featuring in a master equation generally do not have a zero trace, but since the master eq. (4.37) is derived from instantaneous *canonical* Kraus operators, they do. That is, for every GKSL master equation, there exists an alternate one, giving rise to the same dynamics, for which the Lindblad operators have a zero trace. This is an important feature for what follows. Let's re-express eq. (4.37) as a differential equation acting on the column-vectorized states, $\text{col}(\rho)$.

Using the property $\text{col}(ABC) = C^T \otimes A \text{col}(B)$, we have

$$\frac{d}{dt}\text{col}(\rho(t)) = \left[\underbrace{-i \left(\mathbb{I} \otimes H(t) - H^T(t) \otimes \mathbb{I} \right)}_i - \underbrace{\frac{1}{2} \sum_{k \neq 1} \left(\mathbb{I} \otimes L_k^\dagger(t) L_k(t) + (L_k^\dagger(t) L_k(t))^T \otimes \mathbb{I} \right)}_{ii} + \underbrace{\sum_{k \neq 1} L_k^*(t) \otimes L_k(t)}_{iii} \right] \text{col}(\rho(t)). \quad (4.41)$$

A quick calculation suffices to show that the three indicated terms are mutually orthogonal. This means that their respective actions have no overlap. The first term should be familiar as it corresponds to the generator of unitary evolution. The remaining two terms are often referred to as the relaxation or decoherent part of the Lindbladian [EBW87; Hav03]. This integrates well with our notion of decoherence since the instantaneous channels are decoherent if and only if the Hamiltonian is null at every moment in time:

$$\exp(-iH(t)dt - P(t)dt) \geq 0 \Leftrightarrow H(t) = 0. \quad (4.42)$$

To formulate it otherwise, the Lindbladian consists solely of a decoherent part orthogonal to any commutator if and only if the instantaneous channels are decoherent. An additional interesting remark is that the LK approximation applied to the instantaneous channels essentially eliminates the term *iii*, leaving only the commutator (term *i*) and the anticommutator (term *ii*). In particular, the master equation with LK approximated instantaneous decoherent channels consists of an anticommutator only:

$$\frac{d}{dt}\rho(t) = -\{P(t), \rho(t)\}. \quad (4.43)$$

When considered as infinitesimal perturbations from the identity, the channels that we refer to as “decoherent” correspond to the generators of the familiar class of decoherent master equations. While our notion of decoherence connects with previous physics literature in the infinitesimal case, it remains to show that our definition is also appropriate without taking such limit.

4.6.3 Further justifying our notion of decoherence

Typically, quantum error channels are said to act decoherently if they exhibit a non-reversible deterioration. In turn, coherent error channels correspond to a mishandling of information - which can in principle be reverted - rather than a loss of information. An additional expected property of decoherent operations is that they shouldn't allow for coherent buildups such as in the case accumulating over-rotations. Given m non-catastrophic unitary channels $\mathcal{V}_i \approx \mathcal{I}$ with

$$V_i = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \quad (4.44)$$

the infidelity grows faster than linearly (let the composition $\mathcal{V}_{m:1}$ be non-catastrophic so that $m\theta \leq \pi/4$) [CWE16]:

$$1 - \sqrt{\Phi(\mathcal{V}_{m:1}, \mathcal{I})} = 1 - \cos(m\theta) \geq m(1 - \cos(\theta)) = \sum_i \left(1 - \sqrt{\Phi(\mathcal{V}_i, \mathcal{I})}\right). \quad (4.45)$$

As an intuitive pair of properties of our decoherent channels, we show that

- i. The average process fidelity of decoherent error channels cannot be substantially recovered by any unitary (quasi-monotonicity).
- ii. The evolution of the infidelity of a circuit composed of decoherent operations is (approximately) at most additive in the individual infidelities. There is no substantial coherent buildup.

Theorem 10: two features of decoherence

Consider m non-catastrophic decoherent channels \mathcal{D}_i and any non-catastrophic unitary channel \mathcal{V} . Then,

$$\begin{aligned} \Phi(\mathcal{V} \circ \mathcal{D}_{m:1}, \mathcal{I}) &\leq \min_i \Phi(\mathcal{D}_i, \mathcal{I}) \\ &+ \frac{1}{2} \left(\sum_{i=1}^m (1 - Y(\mathcal{D}_i^*)) \right)^2 + (1 - \Phi(\mathcal{V} \circ \mathcal{D}_{m:1}^*, \mathcal{I})) \sum_{i=1}^m (1 - Y(\mathcal{D}_i^*)) \\ &\hspace{15em} \text{(Quasi-monotonicity)} \end{aligned}$$

$$\begin{aligned} 1 - \Phi(\mathcal{V} \circ \mathcal{D}_{m:1}, \mathcal{I}) &\leq (1 - \Phi(\mathcal{V}, \mathcal{I})) + \sum_{i=1}^m (1 - \Phi(\mathcal{D}_i, \mathcal{I})) \\ &+ (1 - \Phi(\mathcal{V}, \mathcal{I}))^2 + \sum_{i=1}^m (1 - \Phi(\mathcal{D}_i^*, \mathcal{I}))^2 \\ &+ \sum_{i=1}^m (1 - \Phi(\mathcal{D}_i, \mathcal{I}))(1 - Y^2(\mathcal{D}_i)) \\ &\hspace{15em} \text{(Quasi-subadditivity property)} \end{aligned}$$

4.7 Behavioral signatures of coherence and decoherence

The introduction in the previous section of the dichotomy between coherence and decoherence, together with the demonstration of a polar decomposition for quantum channels wasn't void of ulterior motives. In this section, we leverage the intrinsic differences between coherent and decoherent channels to explore the behavior of the average process fidelity and the unitarity as circuits grow in depth. Before we begin such investigation, however, let's first make a side step to define various classes of operations which will harmonize with our notion of decoherence.

4.7.1 Extremal dephasers, extremal unitaries, and equable error channels

The non-catastrophic condition still leaves room for pathological noise scenarios. We highlight two extreme (unrealistic) types of channel; the first is of decoherent nature, and the second is purely unitary.

Extremal dephasers

For a channel \mathcal{A} to be non-catastrophic, the singular values of its LK operator $\sigma_i(A_1)$ must nearly average to 1, but nothing else constrains their distribution. Consider a 10-qubit error \mathcal{A} that essentially acts as identity on all operators in $M_d(\mathbb{C})$, but cancels any phase between $|0\rangle$ and $|i\rangle$ for $i \neq 0$ (that is, $|0\rangle\langle i|, |i\rangle\langle 0| \rightarrow 0$ for $i \neq 0$). It is easily shown that the LK operator is $A_1 = \sum_{i \neq 0} |i\rangle\langle i|$; this is an instance of what we call an “extremal dephaser”. An extremal dephaser is defined as a channel for which there exists a singular value $\sigma_j \in \{\sigma_i(A_1)\}$ (in our example, it is $\sigma_0 = 0$) that deviates from 1 by much more than the average perturbation:

$$1 - \sigma_j \gg 1 - \mathbb{E}_i[\sigma_i]. \quad (4.47)$$

To obey eq. (4.47), channels must involve excessively strong¹² dephasing mechanisms between a small number of states and the rest of the system¹³. Let’s come back to our example: a quick calculation shows that \mathcal{A} has an infidelity of around $O(2^{-10}) = O(10^{-3})$: extremal dephasers can have a high average fidelity; they are not ruled out by the non-catastrophic assumption. However, based on realistic grounds, one might discard such scenarios by assuming that the perturbations of the singular values $|1 - \sigma_j|$ remain comparable to the average perturbation $\mathbb{E}[1 - \sigma_i(A_1)]$. Indeed, most physically motivated noise mechanisms – such as unitary, amplitude damping and stochastic channels¹⁴ – perturb the singular values of A_1 in a rather homogeneous way (see table 4.1).

Extremal unitaries

The same argument that was made about the singular values of $A_1 = V|A_1|$, which are the eigenvalues of its positive semidefinite factor $|A_1|$, can be made for the eigenvalues of the unitary factor V . To mimic our previous example, consider a 10-qubit unitary error \mathcal{V} that essentially acts as identity on operators in $M_d(\mathbb{C})$, but maps $|0\rangle\langle i| \rightarrow -|0\rangle\langle i|$, $|i\rangle\langle 0| \rightarrow -|i\rangle\langle 0|$ for $i \neq 0$. It is easily shown that the LK operator is $V = -|0\rangle\langle 0| + \sum_{i \neq 0} |i\rangle\langle i|$; this is an instance of what we call an “extremal unitary”. An extremal unitary is defined as a unitary error \mathcal{V} for which there exists an eigenvalue $\lambda_j \in \{\lambda_i(V)\}$ (in our example, it is $\lambda_0 = -1$) that deviates from 1 by much more than the average perturbation. An easy way to make this precise is to fix the phase of V such that $\text{Tr } V \in \mathbb{R}_+$, and project the eigenvalues on the real

¹²Relative to other decoherent mechanisms.

¹³This is entirely different than: “excessively strong dephasing mechanisms between a *small subsystem* and the rest of the system”, which we already discarded through the non-catastrophic assumption.

¹⁴A stochastic channel has (up to constant factors) unitary operations as canonical Kraus operators and has a LK operator proportional to the identity. Examples of orthogonal unitary bases include the Heisenberg-Weyl operators, and the n -fold tensor product of Paulis. Standard dephasing channels are a special case of stochastic channels were the unitaries are simultaneously diagonalizable (i.e. they all commute).

axis (this is easy to picture on an Argand diagram):

$$1 - \text{Re}\{\lambda_j\} \gg 1 - \mathbb{E}_i[\text{Re}\{\lambda_i\}] = 1 - \text{Tr } V/d. \quad (4.48)$$

To obey eq. (4.48), the unitary error must result from a strong alteration made to the targeted Hamiltonian. Indeed, as a simple Taylor expansion can confirm, small perturbations from the intended Hamiltonian cannot yield an extremal unitary error. Just as for extremal dephasers, extremal unitaries can have a high average fidelity, yet can be reasonably discarded. The perturbations $1 - \text{Re}\{\lambda_i\}$ are expected to be comparable to the average perturbation $1 - \mathbb{E}_i[\text{Re}\{\lambda_i\}]$ (here, $\text{Tr } V \in \mathbb{R}_+$).

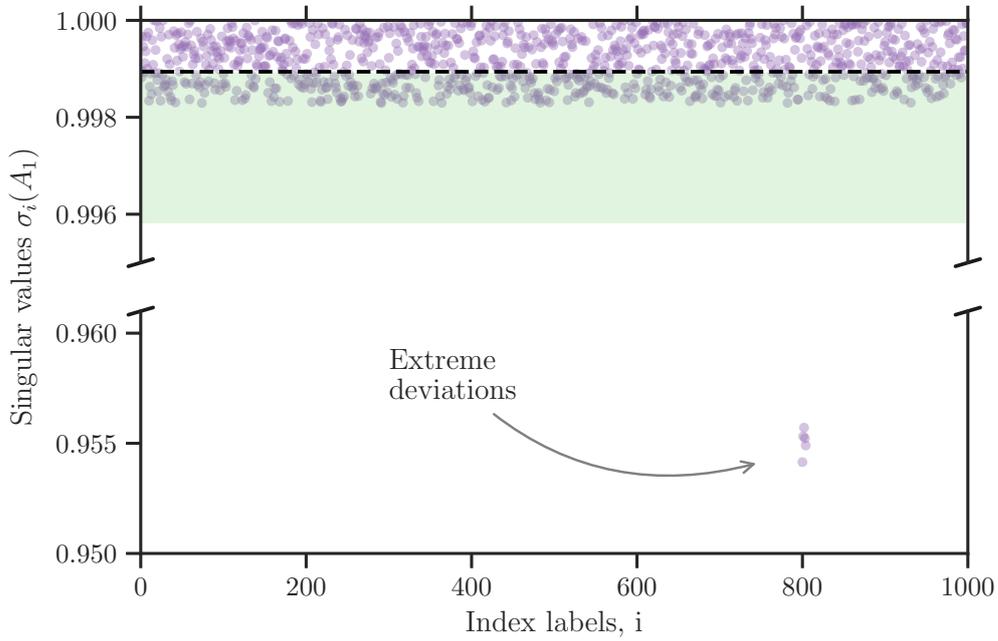


FIGURE 4.2: Singular values σ_i – plotted as purple circles – of the $10^3 \times 10^3$ LK operator A_1 of an extremal dephaser \mathcal{A} . The dashed line corresponds to the average $\mathbb{E}_i[\sigma_i] = 0.9989(1)$. The green shaded region covers a standard deviation $\text{SD}[\sigma_i] = 0.0032(1)$ below the average. In this example, the standard deviation is roughly three times greater than the average deviation $1 - \mathbb{E}_i[\sigma_i] = 0.0011(1)$; that is, the WSE decoherence constant (see definition 7) is $\gamma_{\text{decoh}} = 3.0(1)$, which is an order of magnitude smaller than $1/\sqrt{\mathbb{E}[1 - \sigma_i]} = 30.7(1)$. From eq. (4.53), \mathcal{A} is equable in the wide-sense. There are five singular values situated around 0.955, meaning that $1 - \sigma_j$ can be more than forty times larger than the average deviation (i.e. $\Gamma_{\text{decoh}} = 41(1)$). While these extreme deviations are excluded by the equability condition, their small impact on the standard deviation allows \mathcal{A} to be WSE.

Equable error channels

In this paper, we qualify as “equable” the non-catastrophic error channels $\mathcal{A} = \mathcal{V} \circ \mathcal{D}$ for which the factors \mathcal{D} and \mathcal{V} are not extremal. Notice that the equability assumption ensures a unique polar decomposition since the LK operator is guaranteed to be full rank.

While ruling out extremal error channels seems reasonable, we also define a weaker condition based on the variance of the perturbations.

Definition 7: Equable error channels

Consider a non-catastrophic error channel $\mathcal{A} = \mathcal{V} \circ \mathcal{D}$ with LK operator $A_1 = V|A_1|$. Let $\{\sigma_i\}$ be the singular values of A_1 and $\{\lambda_i\}$ be the eigenvalues of V for which the phase is fixed such that $\text{Tr } V \in \mathbb{R}_+$. We define the strict-sense equability (SSE) decoherence and coherence constants $\Gamma_{\text{decoh}}, \Gamma_{\text{coh}}$ as:

$$1 - \min_j \sigma_j = \Gamma_{\text{decoh}} \mathbb{E}[1 - \sigma_i], \quad (4.49a)$$

$$1 - \min_j \text{Re}\{\lambda_j\} = \Gamma_{\text{coh}} \mathbb{E}[1 - \text{Re}\{\lambda_i\}]. \quad (4.49b)$$

A non-catastrophic error channel is said to be equable (in the strict sense) if

$$\Gamma_{\text{decoh}} \ll 1/\sqrt{\mathbb{E}[1 - \sigma_i]}, \quad (4.50)$$

$$\Gamma_{\text{coh}} \ll 1/\sqrt{\mathbb{E}[1 - \text{Re}\{\lambda_i\}]}. \quad (4.51)$$

Analogously, we define the wide-sense equability (WSE) decoherence and coherence constants $\gamma_{\text{decoh}}, \gamma_{\text{coh}}$ as:

$$\text{SD}[\sigma_i] = \gamma_{\text{decoh}} \mathbb{E}[1 - \sigma_i], \quad (4.52a)$$

$$\text{SD}[\text{Re}\{\lambda_i\}] = \gamma_{\text{coh}} \mathbb{E}[1 - \text{Re}\{\lambda_i\}], \quad (4.52b)$$

where SD denotes the standard deviation. A non-catastrophic error channel is said to be equable in the wide sense if

$$\gamma_{\text{decoh}} \ll 1/\sqrt{\mathbb{E}[1 - \sigma_i]}, \quad (4.53)$$

$$\gamma_{\text{coh}} \ll 1/\sqrt{\mathbb{E}[1 - \text{Re}\{\lambda_i\}]}. \quad (4.54)$$

First notice that ruling out extremal errors is directly imposed by the equability condition (in the strict sense). Obviously, equability implies wide-sense equability, since by construction

$$\gamma_{\text{coh}} \leq \Gamma_{\text{coh}}, \quad (4.55a)$$

$$\gamma_{\text{decoh}} \leq \Gamma_{\text{decoh}}. \quad (4.55b)$$

Of course, the converse doesn't hold (see fig. 4.2 for an example), although such pathological cases must involve extremal channels. The motivation behind the weaker definition is not

physical, but mathematical. The results exhibited in theorems 11 to 15 solely rely on the WSE constants rather than on the realistically slightly larger SSE constants.

Error channel	Type of error	LK operator	Coherence level, r_{coh}/r
Depolarizing	Decoherent, SSE	$A_1 \propto \mathbb{I}$	$O(r)$
Standard dephasing	Decoherent, SSE	$A_1 \propto \mathbb{I}$	$O(r)$
Stochastic	Decoherent, SSE	$A_1 \propto \mathbb{I}$	$O(r)$
Amplitude damping	Decoherent, realistically SSE	$A_1 \geq 0$	$O(r)$
Unitary	Coherent, realistically SSE	$A_1 = V$	1
General SSE	Contains a coherent and decoherent factor	$A_1 = V A_1 $	$\frac{d^2 - \text{Tr } V ^2}{d^2 - \text{Tr } A_1 ^2} + O(r)$

TABLE 4.1: Categorization of different well-known error channels. Many canonical error mechanisms fall under the “decoherent” appellation, except for unitary errors, of course. The coherence level is negligible for decoherent channels, and 1 for coherent errors. In the intermediate regime, the coherence level can vary between 0 and 1. It only makes sense to discuss about the coherence level when errors are equable (at least in the wide sense).

4.7.2 Reasoning about Υ

Now that we have defined (wide-sense) equable errors, we are ready to express a first decay law:

Theorem 11: unitarity decay law

Consider m non-catastrophic channels \mathcal{A}_i . Then $Y(\mathcal{A}_{m:1})$ has the following properties:

$$Y(\mathcal{A}_{m:1}) \leq \min_i Y(\mathcal{A}_i) + (1 - Y^2(\mathcal{A}_{m:1}))^2 / \sqrt{2} \quad (\text{Quasi-monotonicity})$$

$$1 - Y(\mathcal{A}_{m:1}) \leq \sum_{i=1}^m (1 - Y(\mathcal{A}_i)) + (1 - Y^2(\mathcal{A}_i))^2 \quad (\text{Quasi-subadditivity property})$$

The quasi-monotonicity is almost saturated by extremal channels. If we introduce the WSE decoherence constants $\gamma_{\text{decoh}}(\mathcal{A}_i) \leq \gamma_{\text{decoh}}$, we obtain:

$$\begin{aligned} \left| Y(\mathcal{A}_{m:1}) - \prod_i Y(\mathcal{A}_i) \right| &\leq (1 - Y(\mathcal{A}_{m:1}^*))^2 + \sum_{j=1}^m (1 - Y(\mathcal{A}_j^*))^2 \\ &\quad + \gamma_{\text{decoh}}^2 \sum_{i=1}^m \left(1 - \sqrt{Y(\mathcal{A}_i^*)} \right)^2 \\ &\quad + 2\gamma_{\text{decoh}}^2 \left(\sum_{i=1}^m \left(1 - \sqrt{Y(\mathcal{A}_i^*)} \right) \right)^2 + \text{H.O.T.} \end{aligned} \quad (4.56)$$

If the channels are equable, $Y(\mathcal{A}_{m:1})$ is essentially a multiplicative decay.

Of course, those results can be immediately translated in terms of unitarity by using eq. (4.14). Without using the LK approximation, showing the monotonicity of the unitarity can be difficult, since quantum channels aren't contractive maps; going to the LK picture fixes this issue since Kraus operators are contractions. Quasi-multiplicativity is another way of stating that the unitarity of a composition essentially behaves as a multiplicative decay involving the unitarity of individual components:

$$u(\mathcal{A}_{m:1}) \approx \frac{d^2 \prod_{i=1}^m Y^2(\mathcal{A}_i) - 1}{d^2 - 1}. \quad (4.57)$$

Equation (4.57) should be seen as a staple of wide-sense equability; deviations from this behavior indicates the presence of extremal dephasers.

The quasi-multiplicativity of Y is not the only decay law that occurs in the equable scenario. Recall that to motivate our definition of decoherence, we initially showed the quasi-monotonicity and quasi-subadditivity property of the process fidelity of decoherent compositions (theorem 10). By introducing the equability condition we get a stronger assertion:

Theorem 12: fidelity decay law (for decoherent compositions)

Consider m non-catastrophic, decoherent channels \mathcal{D}_i (with target \mathcal{I}) with WSE decoherence constants $\gamma_{\text{decoh}}(\mathcal{D}_i) \leq \gamma_{\text{decoh}}$. Then, $\Phi(\mathcal{D}_{m:1}, \mathcal{I})$ is bounded as follows:

$$\begin{aligned} \left| \Phi(\mathcal{D}_{m:1}, \mathcal{I}) - \prod_{i=1}^m \Phi(\mathcal{D}_i, \mathcal{I}) \right| &\leq \left[\frac{1}{2} \left(\sum_{i=1}^m (1 - Y(\mathcal{D}_i^*)) \right)^2 \right. \\ &+ (1 - \Phi(\mathcal{D}_{m:1}^*, \mathcal{I})) \sum_{i=1}^m (1 - Y(\mathcal{D}_i^*)) + \sum_{i=1}^m (1 - Y(\mathcal{D}_i^*)) (1 - \Phi(\mathcal{D}_i, \mathcal{I})) \\ &\left. + \gamma_{\text{decoh}}^2 \prod_{i=1}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \right)^2 \right] + \text{H.O.T.} \end{aligned} \quad (4.58)$$

If the channels are WSE, $\Phi(\mathcal{D}_{m:1}, \mathcal{I})$ is essentially a multiplicative decay.

Using the simple relation between F and Φ (eq. (4.9)) we come to this observation: the average gate fidelity of a composition of non-catastrophic decoherent equable channels behaves almost exactly as a multiplicative decay in the average process fidelity of individual components, that is

$$F(\mathcal{D}_{m:1}, \mathcal{I}) \approx \frac{d \prod_{i=1}^m \Phi(\mathcal{D}_i, \mathcal{I}) + 1}{d + 1}. \quad (4.59)$$

The decay becomes exact with the depolarizing channel $\mathcal{P}_p(\rho) = p\rho + (1-p)(\text{Tr}\rho)\mathbb{I}_d/d$, which is a celebrated example of a decoherent operation.

The two decay laws expressed in theorems 11 and 12 are in fact describing the same observation. Let \mathcal{A} have an equable error and a polar decomposition $\mathcal{V} \circ \mathcal{D}$. As shown in the following theorem, $Y(\mathcal{A})$ can be interpreted as the maximal process fidelity of \mathcal{A} to the target \mathcal{U} under unitary corrections, or equivalently as the process fidelity of the decoherent factor \mathcal{D} to the identity:

$$Y(\mathcal{A}) \approx \Phi(\mathcal{D}, \mathcal{I}) \approx \max_{W \in SU(d)} \Phi(W \circ \mathcal{A}, \mathcal{U}). \quad (4.60)$$

Theorem 13: Y as the process fidelity of the decoherent factor

Consider a non-catastrophic channel $\mathcal{A} = \mathcal{V} \circ \mathcal{D}$ with unitary target \mathcal{U} . Then, the maximal unitary correction of \mathcal{A} (in terms of Φ) is approximately bounded by the interval $[\Upsilon^2(\mathcal{A}), \Upsilon(\mathcal{A})]$:

$$\max_{W \in SU(d)} \Phi(W \circ \mathcal{A}, \mathcal{U}) \leq \Upsilon(\mathcal{A}) + \frac{3}{2}(1 - \Upsilon^2(\mathcal{A}))^2, \quad (4.61a)$$

$$\max_{W \in SU(d)} \Phi(W \circ \mathcal{A}, \mathcal{U}) \geq \Upsilon^2(\mathcal{A}) - (1 - \Upsilon^2(\mathcal{A}))^2. \quad (4.61b)$$

Moreover, if we introduce the WSE decoherence constant γ_{decoh} , we obtain:

$$\max_{W \in SU(d)} \Phi(W \circ \mathcal{A}, \mathcal{U}) \geq \Upsilon(\mathcal{A}) - (1 + \gamma_{\text{decoh}}^2)(1 - \Upsilon^2(\mathcal{A}))^2. \quad (4.62)$$

A quasi-maximal choice of unitary correction consists in $\mathcal{W} = \mathcal{U} \circ \mathcal{V}^\dagger$.

In terms of other figures of merit, wide-sense equability ensures a quasi-one-to-one correspondence between the maximal average gate fidelity (through a unitary correction) and the unitarity through:

$$\max_{W \in SU(d)} F(W \circ \mathcal{A}, \mathcal{U}) \approx F(\mathcal{D}, \mathcal{I}) \approx \frac{\sqrt{(d^2 - 1)u(\mathcal{A}) + 1} + 1}{d + 1}. \quad (4.63)$$

4.7.3 The coherence level

Let's extend theorem 12 by appending a coherent operation to the decoherent composition:

Theorem 14: the average process fidelity of equable compositions

Consider m non-catastrophic, decoherent error channels \mathcal{D}_i (with target \mathcal{I}) with WSE decoherence constants $\gamma_{\text{decoh}}(\mathcal{D}_i) \leq \gamma_{\text{decoh}}$. Moreover, consider a non-catastrophic unitary error channel \mathcal{V} with WSE coherence constant γ_{coh} . Then, $\Phi(\mathcal{V} \circ \mathcal{D}_{m:1}, \mathcal{I})$ is bounded as follows:

$$\begin{aligned} & \left| \Phi(\mathcal{V} \circ \mathcal{D}_{m:1}, \mathcal{I}) - \Phi(\mathcal{V}, \mathcal{I}) \prod_{i=1}^m \Phi(\mathcal{D}_i, \mathcal{I}) \right| \leq \left[\frac{1}{2} \left(\sum_{i=1}^m (1 - \Upsilon(\mathcal{D}_i^*)) \right)^2 \right. \\ & + (1 - \Phi(\mathcal{V} \circ \mathcal{D}_{m:1}^*, \mathcal{I})) \sum_{i=1}^m (1 - \Upsilon(\mathcal{D}_i^*)) + \sum_{i=1}^m (1 - \Upsilon(\mathcal{D}_i^*)) (1 - \Phi(\mathcal{D}_i, \mathcal{I})) \\ & + 2\gamma_{\text{decoh}}\gamma_{\text{coh}} \left(1 - \sqrt{\Phi(\mathcal{V}, \mathcal{I})} \right) \sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \\ & \left. + \gamma_{\text{decoh}}^2 \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \right)^2 \right] + \text{H.O.T.} \end{aligned} \quad (4.64)$$

If the errors are WSE, then $\Phi(\mathcal{V} \circ \mathcal{D}_{m:1}, \mathcal{I})$ is essentially multiplicative.

Let's unfold this result one step at a time. First, consider eq. (4.64) for $m = 1$. Let \mathcal{A} be a channel with target \mathcal{U} and polar decomposition $\mathcal{V} \circ \mathcal{D}$. $\mathcal{W} := \mathcal{U}^{-1} \circ \mathcal{V}$ is a unitary error. Hence, it follows from theorems 13 and 14 that

$$\Phi(\mathcal{A}, \mathcal{U}) = \Phi(\mathcal{W} \circ \mathcal{D}, \mathcal{I}) \stackrel{\text{thm.8}}{\approx} \Phi(\mathcal{W}, \mathcal{I})\Phi(\mathcal{D}, \mathcal{I}) = \Phi(\mathcal{V}, \mathcal{U})\Phi(\mathcal{D}, \mathcal{I}) \stackrel{\text{thm.7}}{\approx} \Phi(\mathcal{V}, \mathcal{U})Y(\mathcal{A}). \quad (4.65)$$

There are two factors that compound to the average process fidelity: $\Phi(\mathcal{V}, \mathcal{U})$ relates to a coherent contribution to the total infidelity, while $\Phi(\mathcal{D}, \mathcal{I}) \approx Y(\mathcal{A})$ depicts a decoherent one. For those who are more familiar with the infidelity $r(\mathcal{A}, \mathcal{U})$, eq. (4.65) can be reformulated as¹⁵ (up to $O(r^2)$):

$$r(\mathcal{A}, \mathcal{U}) \approx \underbrace{r(\mathcal{V}, \mathcal{U})}_{\text{Coherent infidelity}} + \underbrace{r(\mathcal{D}, \mathcal{I})}_{\text{Decoherent infidelity}} = r_{\text{coh}} + r_{\text{decoh}}. \quad (4.66)$$

The channel average infidelity of a channel can be split into a sum of a coherent and decoherent terms (given equable errors). r_{decoh} is not substantially correctable through any composition, and can be obtained from the unitarity alone:

$$r_{\text{decoh}} = \frac{d - \sqrt{(d^2 - 1)u(\mathcal{A}) + 1}}{d + 1} + O(r^2) = \frac{d}{d + 1} (1 - Y(\mathcal{A})) + O(r^2). \quad (4.67)$$

r_{coh} can be corrected through a composition with a unitary (see theorem 13). Equation (4.66) motivates the definition of *coherence level* as the fraction of the infidelity that is associated to coherence. It can be obtained by combining the infidelity and the unitarity through:

$$\frac{r_{\text{coh}}}{r} = 1 - \frac{d - \sqrt{(d^2 - 1)u(\mathcal{A}) + 1}}{(d + 1)r(\mathcal{A}, \mathcal{U})} + O(r) = \frac{1 - Y(\mathcal{A})}{1 - \Phi(\mathcal{A}, \mathcal{U})} + O(r) \quad (4.68)$$

Similarly, the decoherence level is defined as r_{decoh}/r . Equation (4.66) strengthens the insight behind the notion of coherence level introduced (under different appellations) in [Yang2019; Fen+16]. In those previous works, the RHS of eq. (4.66) is generally depicted as a lower bound on the infidelity, which can be reduced to r_{decoh} through a unitary correction. The (approximate) equality – which is much more valuable since it provides an upper bound on r – is shown for single qubit case in [Fen+16] using the polar decomposition of the action on Bloch sphere. Here, we have shown the (approximate) equality (in the equable scenario) for all dimensions using the polar decomposition of LK operators.

4.7.4 Bounding the worst and best case fidelity of a circuit

Now, let's revisit theorem 14 for general circuit depth m . This will allow us to identify the worst and best case fidelity of a circuit. Consider m channels \mathcal{A}_i with target \mathcal{U}_i and polar decomposition $\mathcal{D}_i \circ \mathcal{V}_i$. The circuit $\mathcal{A}_{m:1}$ can be re-expressed as

$$\mathcal{A}_{m:1} = \mathcal{V}_{m:1} \circ (\mathcal{V}_{m:1})^\dagger \circ \mathcal{D}_m \circ \mathcal{V}_{m:1} \circ \cdots \circ (\mathcal{V}_{2:1}) \circ \mathcal{V}_1^\dagger \circ \mathcal{D}_1 \circ \mathcal{V}_1 = \mathcal{V}_{m:1} \circ \mathcal{D}'_{m:1}, \quad (4.69)$$

¹⁵The transition from eq. (4.65) to eq. (4.66) simply involves using the approximation $(1 - \delta_1)(1 - \delta_2) \approx 1 - \delta_1 - \delta_2$ for small δ_i .

where $\mathcal{D}'_k := (\mathcal{V}_{k:1})^\dagger \circ \mathcal{D}_k \circ \mathcal{V}_{k:1}$ are decoherent channels with the same fidelity as \mathcal{D}_k . This means that:

$$\Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1}) \stackrel{\text{thm.8}}{\approx} \Phi(\mathcal{V}_{m:1}, \mathcal{U}_{m:1}) \prod_{i=1}^m \Phi(\mathcal{D}_i, \mathcal{I}) \stackrel{\text{thm.7}}{\approx} \Phi(\mathcal{V}_{m:1}, \mathcal{U}_{m:1}) \prod_{i=1}^m Y(\mathcal{A}_i). \quad (4.70)$$

In this last expression, we clearly see that the evolution of Φ is factored into a decoherent decay multiplied by a function $\Phi(\mathcal{V}_{m:1}, \mathcal{U}_{m:1})$ which captures the fidelity of a purely coherent process. This is already an interesting realization: since the decoherent decay is fixed, all the freedom in the evolution of the fidelity is contained in the coherent factors. An assessment concerning the circuit's average process fidelity must rely on a characterization of coherent effects. Since we know that such effects are correctable through composition, we first get:

Theorem 15: maximal average process fidelity of channel compositions

Consider m non-catastrophic channels \mathcal{A}_i with respective unitary targets \mathcal{U}_i and polar decompositions $\mathcal{A}_i = \mathcal{V}_i \circ \mathcal{D}_i$. Let the WSE decoherence constants be $\gamma_{\text{decoh}}(\mathcal{D}_i) \leq \gamma_{\text{decoh}}$. Then, the maximal unitary correction of the composition $\mathcal{A}_{m:1}$ is bounded as follows:

$$\begin{aligned} \max_{W \in SU(d)} \Phi(W \circ \mathcal{A}_{m:1}, \mathcal{U}_{m:1}) - \prod_{i=1}^m Y(\mathcal{A}_i) &\leq \left[\frac{1}{2} \left(\sum_{i=1}^m (1 - Y(\mathcal{A}_i^*)) \right)^2 + \sum_{i=1}^m (1 - Y(\mathcal{A}_i^*))^2 \right. \\ &\quad \left. + \left(\sum_{i=1}^m (1 - Y(\mathcal{A}_i^*)) \right) \left(1 - \prod_{i=1}^m Y(\mathcal{A}_i) \right) + 2\gamma_{\text{decoh}}^2 \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \right)^2 \right] + \text{H.O.T.} \end{aligned} \quad (4.71a)$$

$$\begin{aligned} \max_{W \in SU(d)} \Phi(W \circ \mathcal{A}_{m:1}, \mathcal{U}_{m:1}) - \prod_{i=1}^m Y(\mathcal{A}_i) &\geq \left[-\gamma_{\text{decoh}}^2 \sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right)^2 - \sum_{i=1}^m (1 - Y(\mathcal{A}_i^*))^2 \right. \\ &\quad \left. - \gamma_{\text{decoh}}^2 \prod_{i=1}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \right)^2 \right] + \text{H.O.T.} \end{aligned} \quad (4.71b)$$

For equable errors, the maximal unitary correction of the composition $\mathcal{A}_{m:1}$ is essentially $\prod_{i=1}^m Y(\mathcal{A}_i)$. A quasi-optimal choice of unitary correction is $\mathcal{W} = \mathcal{U}_{m:1} \circ (\mathcal{V}_{m:1})^\dagger$.

In short, the average gate fidelity of a composite circuit is upper bounded by a decaying envelope which is closely prescribed by the decoherent factors of its individual components:

$$\max_{W \in SU(d)} F(W \circ \mathcal{A}_{m:1}, \mathcal{U}_{m:1}) \approx \frac{d \prod_{i=1}^m \Phi(\mathcal{D}_i, \mathcal{I}) + 1}{d + 1} \approx \frac{d \prod_{i=1}^m Y(\mathcal{A}_i) + 1}{d + 1}. \quad (4.72)$$

This unforgiving behavior harmonizes well with the more typical comprehension of decoherence as a limiting process.

To find the worst possible $\Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1})$, it suffices to use a lower bound for the coherent factor $\Phi(\mathcal{V}_{m:1}, \mathcal{U}_{m:1})$. This is partially done in [CWE16], where the inequality

$$\Phi(\mathcal{V}_{m:1}, \mathcal{U}_{m:1}) \geq \cos^2 \left(\sum_{i=1}^m \arccos \left(\sqrt{\Phi(\mathcal{V}_i, \mathcal{U}_i)} \right) \right) \quad (4.73)$$

is shown to be saturated in even dimensions. For odd dimensions, we find the following saturated bound:

$$\Phi(\mathcal{V}_{m:1}, \mathcal{U}_{m:1}) \geq \left(\frac{(d-1) \cos \left(\sum_{i=1}^m \arccos \left(\frac{d\sqrt{\Phi(\mathcal{V}_i, \mathcal{U}_i)} - 1}{d-1} \right) \right) + 1}{d} \right)^2. \quad (4.74)$$

Proof. The generalization to odd dimensions almost immediately follows by looking at the saturation case in even dimensions, which consists of commuting unitary errors of the form

$$\begin{pmatrix} \cos(\theta_i) & -\sin(\theta_i) \\ \sin(\theta_i) & \cos(\theta_i) \end{pmatrix} \otimes \mathbb{I}_{d/2}. \quad (4.75)$$

In the odd dimension case, it suffices to always pick the global phase to fix the first eigenvalue of $V_{m:1}(U_{m:1})^{-1}$ to 1. The minimization over $|\text{Tr } V_{m:1}(U_{m:1})^{-1}|$ then falls back to the even dimensional case, since the saturation case has a real trace. \square

By using $\Phi(\mathcal{V}_i, \mathcal{U}_i) \approx \Phi(\mathcal{A}_i, \mathcal{U}_i)/Y(\mathcal{A}_i)$ we can formulate a quasi-saturated assessment about the average process fidelity of the circuit $\mathcal{A}_{m:1}$ given a partial information about its components \mathcal{A}_i (in the equable scenario).

For even dimensions:

$$\cos^2 \left(\sum_{i=1}^m \arccos \left(\sqrt{\frac{\Phi(\mathcal{A}_i, \mathcal{U}_i)}{Y(\mathcal{A}_i)}} \right) \right) \prod_{i=1}^m Y(\mathcal{A}_i) \lesssim \Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1}) \lesssim \prod_{i=1}^m Y(\mathcal{A}_i); \quad (4.76a)$$

for odd dimensions:

$$\left(\frac{(d-1) \cos \left(\sum_{i=1}^m \arccos \left(\frac{d\sqrt{\frac{\Phi(\mathcal{A}_i, \mathcal{U}_i)}{Y(\mathcal{A}_i)} - 1}}{d-1} \right) \right) + 1}{d} \right)^2 \prod_{i=1}^m Y(\mathcal{A}_i) \lesssim \Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1}) \lesssim \prod_{i=1}^m Y(\mathcal{A}_i). \quad (4.76b)$$

The terms in the cosine function are very close to what was defined as ‘‘coherence angles’’ in [CWE16]. Their sum can be interpreted as a coherent buildup. In some sense, the coherence angle is just another way to go about the notion of coherence level: it ties r_{coh} to an optimal rotation angle.

4.7.5 Decoherence-limited operations

When individual circuit elements \mathcal{A}_i have purely decoherent equable errors, the bounds given by eqs. (4.76a) and (4.76b) reduce to the approximate equality $\Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1}) \approx \prod_i Y(\mathcal{A}_i)$. In fact, as long as the errors attached to the circuit elements \mathcal{A}_i have a negligible level of coherence, $\Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1})$ is still expected to closely behave like a multiplicative decay. More rigorously, by looking more attentively at eqs. (4.76a) and (4.76b), one should quickly realize that requiring

$$\Phi(\mathcal{A}_i, \mathcal{U}_i) = Y(\mathcal{A}_i) + O(r^2(\mathcal{A}_i, \mathcal{U}_i)) \quad (4.77)$$

is sufficient to ensure

$$\cos^2 \left(\sum_{i=1}^m \arccos \left(\sqrt{\frac{\Phi(\mathcal{A}_i, \mathcal{U}_i)}{Y(\mathcal{A}_i)}} \right) \right) = 1 + O(r^2(\mathcal{A}_{m:1}, \mathcal{U}_{m:1})), \quad (4.78a)$$

and

$$\left(\frac{(d-1) \cos \left(\sum_{i=1}^m \arccos \left(\frac{d \sqrt{\frac{\Phi(\mathcal{A}_i, \mathcal{U}_i)}{Y(\mathcal{A}_i)}} - 1}}{d-1} \right) \right) + 1}{d} \right)^2 = 1 + O(r^2(\mathcal{A}_{m:1}, \mathcal{U}_{m:1})). \quad (4.78b)$$

A channel obeying the condition described by eq. (4.77) is said to be *decoherence-limited*. The terminology is self-explanatory: a channel is decoherence-limited if the infidelity to its target is mostly limited by its decoherent infidelity r_{decoh} , which cannot be (substantially) reduced further through unitary corrections (see theorem 13). Decoherence-limited channels count decoherent channels, but also include channels for which the infidelity of the coherent factor plays a negligible role in the total infidelity, that is $r_{\text{coh}} = O(r^2)$ or, equivalently, $r_{\text{coh}}/r = O(r)$.

Decoherent channels do not form a closed set under composition; the product of two positive semidefinite matrices is not necessarily positive semidefinite. The geometric picture is that if two positive semidefinite contractions have different axes of contraction, they may induce (after composition) a small effective rotation. However, the small rotation factor resulting from such composition is ensured to be very close to the identity, otherwise theorem 12 wouldn't hold. More precisely, given two decoherent channels \mathcal{D}_1 and \mathcal{D}_2 , the composite channel $\mathcal{D}_{2:1} = \mathcal{V} \circ \mathcal{D}'$ is such that $r(\mathcal{V}, \mathcal{I}) = O(r^2(\mathcal{D}', \mathcal{I}))$. In other words, the resulting channel is decoherence-limited. It is easy to see from eqs. (4.76a) and (4.76b) that equable decoherence-limited channels form a closed set under composition; if the coherence level of every channel \mathcal{A}_i in a circuit is of order $r(\mathcal{A}_i, \mathcal{U}_i)$, then the coherence level of the total circuit is of order $r(\mathcal{A}_{m:1}, \mathcal{U}_{m:1})$.

4.7.6 Limitations

In this section, we take a closer look at the bounds appearing in theorems 7, 8 and 11 to 15 and discuss their limitations. To parse through the expressions with more ease, consider m channels $\mathcal{A}_i = \mathcal{V}_i \circ \mathcal{D}_i$ with identical decoherent infidelity $r(\mathcal{D}_i, \mathcal{I}) = r_{\text{decoh}}$. From this simplification, and by using $Y(\mathcal{A}_i) \approx Y(\mathcal{A}_i^*) \approx \Phi(\mathcal{D}_i^*, \mathcal{I}) \approx \Phi(\mathcal{D}_i, \mathcal{I})$, which holds up to order r_{decoh}^2 , the margin of freedom in the bounds presented in this work reduces to the form¹⁶

$$C_0 m r_{\text{decoh}}^2 + C_1 m^2 r_{\text{decoh}}^2 + C_2 m r_{\text{decoh}} r(\mathcal{V}_{m:1}, \mathcal{I}) + \text{H.O.T.}, \quad (4.79)$$

¹⁶Theorem 10 also contains a term of the form $r^2(\mathcal{V}_{m:1})$, but this term disappears in the equable regime.

where C_i s are non-negative constants at most of order 1 in the equable scenario. From theorem 15, the total infidelity scales at most as:

$$r(\mathcal{A}_{m:1}, \mathcal{I}) \lesssim 1 - \frac{d \left(1 - \frac{d+1}{d} r_{\text{decoh}}\right)^m + 1}{d+1} = mr_{\text{decoh}} - \frac{d+1}{2d} m^2 r_{\text{decoh}}^2 + \text{H.O.T.}, \quad (4.80)$$

meaning that eq. (4.79) is always at most of order $r^2(\mathcal{A}_{m:1}, \mathcal{I})$. Hence, the bounds presented in this work apply very well in the high-fidelity regime.

As the fidelity decreases, the leeway portrayed by eq. (4.79) starts being noticeable. The appearance of quadratic terms of the form $m^2 r_{\text{decoh}}^2$ is not surprising since most bounding techniques are based on the LK approximation, which ignores some $m^2 r_{\text{decoh}}^2$ contributions. To see this, consider m identical channels \mathcal{A}_i with canonical Kraus decomposition $\{\sqrt{1-\delta}\mathbb{I}, \sqrt{\delta}P\}$ where δ is small and P is a unitary such that $P^2 = \mathbb{I}$ and $\text{Tr} P = 0$. Simple calculations yield

$$\Phi(\mathcal{A}_{m:1}^*, \mathcal{I}) = (1-\delta)^m = 1 - m\delta + \frac{1}{2}m^2\delta^2 + \text{H.O.T.} \quad (4.81)$$

$$\Phi(\mathcal{A}_{m:1}, \mathcal{I}) = \sum_{n=0}^{\lfloor m/2 \rfloor} \binom{m}{2n} (1-\delta)^{m-2n} \delta^{2n} = 1 - m\delta + \frac{1}{2}m^2\delta^2 + \binom{m}{2}\delta^2 + \text{H.O.T.} \quad (4.82)$$

The term $\frac{1}{2}(\sum_i(1 - Y(\mathcal{A}_i^*)))^2$ featured in the bound of theorem 8 is essentially achieved by the above example. However, not all the $m^2 r_{\text{decoh}}^2$ terms appearing in the previous theorems are expected to be achieved by a composition of quantum channels.

Figure 4.3 provides a good sense of the scaling of the bounds provided in theorem 14. In the figure, the decoherent infidelity of individual operations is of order 10^{-4} . The top figure shows the bounds for circuit lengths around 10^2 , in which case $m^2 r_{\text{decoh}}^2$ is of order 10^{-4} . The bottom figure shows the bounds for circuit lengths around 10^3 , in which case $m^2 r_{\text{decoh}}^2$ is of order 10^{-2} . Once the circuit length m is comparable to r_{decoh}^{-1} (in the example given by fig. 4.3, it would be as m gets close to 10^4), the fidelity is no longer "small", and $O(m^2 r_{\text{decoh}}^2)$ becomes of order 1, which renders the bounds trivial. In other words, to gain anything valuable from the bounds in this work, the regime of consideration should be roughly $m^2 r_{\text{decoh}}^2 \lesssim 10^{-1}$ and $mr_{\text{decoh}}r(\mathcal{V}_{m:1}, \mathcal{I}) \lesssim 10^{-1}$. Notice that in such regime, as depicted by fig. 4.3, the only non-linear behavior in the composite fidelity must stem from unitary errors alone.

4.8 Conclusion

In this work, we investigated a quasi-dynamical sub-parameterization of quantum channels that we referred to as the LK approximation. A remarkable realization is that this reduced picture still can be used to closely follow the evolution of two important figures of merit, namely the average process fidelity and the unitarity (see theorems 7 and 8).

Working with a simplified portrait sets aside superfluous subtleties and typically grants new mathematical properties to the object of consideration. In our case, LK approximated mappings can be parameterized as contractions in $M_d(\mathbb{C})$; this set of matrices offers a much more intelligible categorization of error scenarios than the more abstruse full process matrix parameterization. Any matrix $A \in M_d(\mathbb{C})$ has a polar decomposition $V|A|$ where $|A| \geq 0$ and V is unitary. V corresponds to a purely coherent physical operation $\mathcal{V}(\rho) = V\rho V^\dagger$,

whereas the positive contraction $|A|$ is the LK operator belonging to what we classify as a decoherent channel (see definition 6). In a nutshell, the polar decomposition in $M_d(\mathbb{C})$ translates into a coherent-decoherent factorization for quantum channels (see theorem 9). We leveraged this dichotomy between types of noise to derive fundamental principles of behavior concerning our two considered figures of merit. Among other properties, we demonstrated, up to high precision, the general monotonicity of the unitarity as well as the monotonicity of the average process fidelity of circuits with decoherent components (see theorems 10 and 11).

To pursue our analysis further, we introduced the wide-sense equable parameters $\gamma_{\text{decoh}}, \gamma_{\text{coh}}$, which are defined through the LK parameterization (see definition 7). Equable error channels, for which $\gamma_{\text{decoh}}, \gamma_{\text{coh}}$ are not too high, include all realistic noise models (and potentially more). Under the equability condition, we make multiple interesting connections between individual channels and compositions thereof:

- i. The infidelity of any channel can be decomposed into a sum of two terms: a decoherent infidelity and a coherent one (respectively tied to the decoherent/coherent components of the channel). (See theorem 14 and the discussion that immediately follows.)
- ii. The unitarity, as well as the fidelity of circuits with decoherent elements, obey decay laws. Both these decays are closely dictated by the unitarity of individual components alone. (See theorems 11 to 13.)
- iii. The decoherent decay (that is, the decay prescribed by the decoherent factors of the circuit components) forms an upper bound to the total average process fidelity. Any substantial deviation from this upper bound is due to coherent effects alone (which gives us a lower bound). (See theorems 14 and 15.)

This work was primarily cast as a stepping-stone to formulate assessments about the performance of circuits based on partial knowledge of their constituents. While we do provide some assertion formulas, we want to emphasize that the more fundamental introduction of the LK approximation should also benefit the development of further characterization schemes. Indeed, the simple parameterization offered by the LK approximation facilitates the identification of specific noise signatures.

Acknowledgments — The authors would like to thank Joel J. Wallman for his helpful discussions. This research was supported by the U.S. Army Research Office through grant W911NF-14-1-0103, TQT, CIFAR, the Government of Ontario, and the Government of Canada through CFREF, NSERC and Industry Canada.

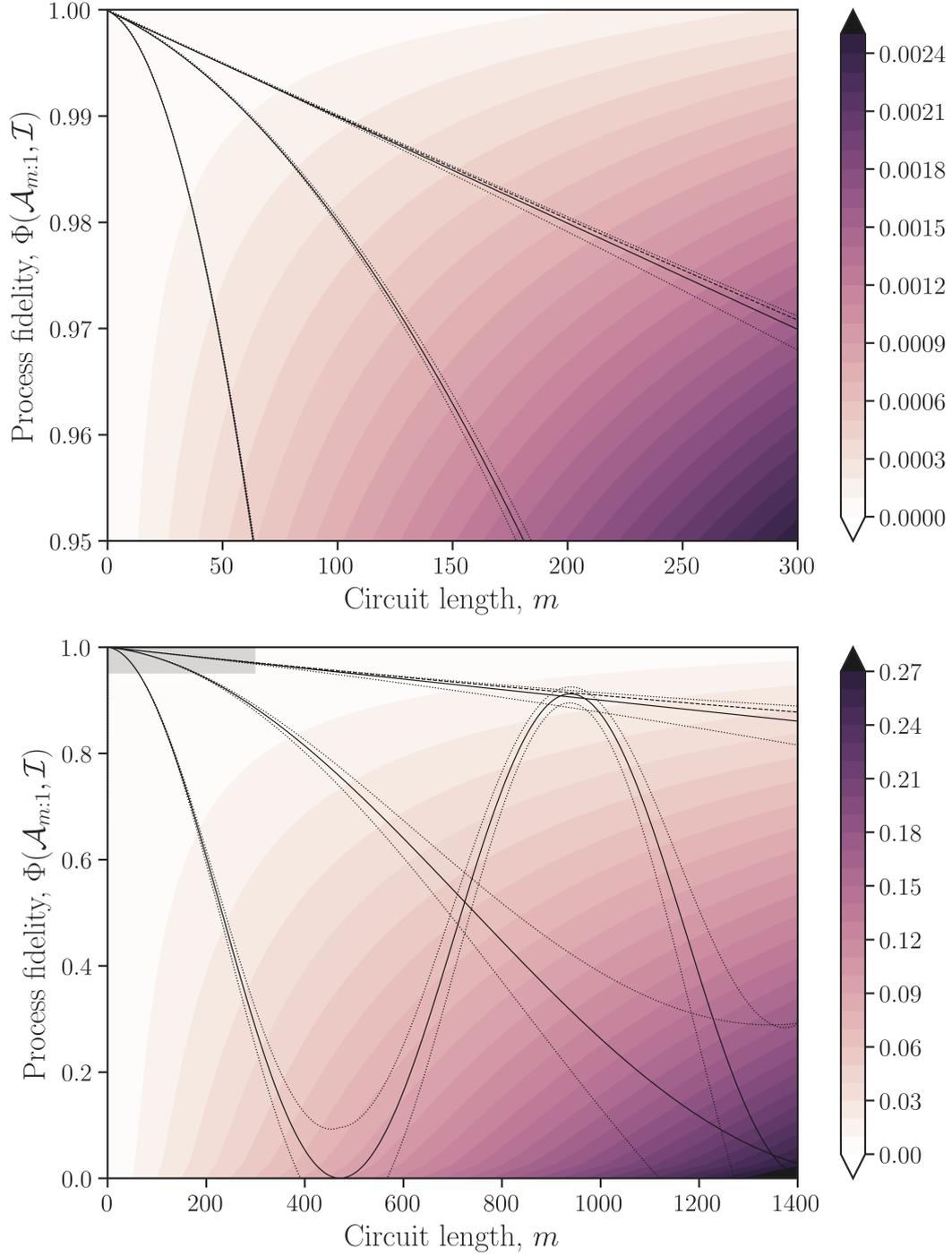


FIGURE 4.3: Process fidelity of three different error channel compositions as a function of the circuit length. The hard lines correspond to the three process fidelities $\Phi(\mathcal{A}_{m:1}, \mathcal{I})$, and the dotted lines correspond to the bounds given by theorem 14. The color map illustrates the margin of freedom given by the RHS of eq. (4.64), and the gray shaded area in the bottom plot corresponds to the limits of the top plot. In the top figure, shade variations indicate increments of 10^{-4} , and in the bottom figure, increments of 10^{-2} . The individual channels are of the form $\mathcal{A}_i = \mathcal{V} \circ \mathcal{D}$, where \mathcal{D} is a dephasing channel with $\Phi(\mathcal{D}, \mathcal{I}) = 10^{-4}$, and \mathcal{V} is a small unitary error. The dashed line is the decaying envelope $\Phi^m(\mathcal{D}, \mathcal{I})$. The three compositions differ by the level of coherence of their elements \mathcal{A}_i , which are 10%, 1% and 0.01%. The lowest level of coherence corresponds to a decoherence-limited scenario, in which case the decoherent envelope stays within the bounds.

Concept	Definition	Notes
Non-catastrophic channel	$\Phi(\mathcal{A}, \mathcal{U}), Y^2(\mathcal{A}) > 1/2$	- Guarantees a unique LK operator. - Achieved given an acceptable level of control.
LK operator, A_1	Highest weight canonical Kraus operator, A_1	- Contains remarkable information about Φ, \mathcal{Y} .
LK approximation, \mathcal{A}^*	$\mathcal{A}^*(\rho) = A_1 \rho A_1^\dagger$	- Replacing channels by their LK approximation in a circuit barely affects its fidelity and unitarity.
Decoherent channel	$A_1 \geq 0$	- Every non-catastrophic channel has a coherent-decoherent decomposition $\mathcal{A} = \mathcal{U}_A \circ \mathcal{D}_A = \mathcal{D}'_A \circ \mathcal{U}_A$. - This definition of decoherence generalizes the notion of decoherence in the Lindblad picture.
Extremal dephaser (channel)	$\exists \sigma_j \in \{\sigma_i(A_1)\}$ s.t. $1 - \sigma_j \gg 1 - \mathbb{E}[\sigma_i]$	- Strongly dephases a small set of states from the rest of the system. Since the set of states is small, extremal dephasers can still have high fidelity.
Extremal unitary (channel)	Let $\text{Tr } U \in \mathbb{R}_+$. $\exists \lambda_j \in \{\lambda_i(U)\}$ s.t. $1 - \text{Re}\{\lambda_j\} \gg 1 - \mathbb{E}[\text{Re}\{\lambda_i\}]$	- Strongly dephases a small set of states from the rest of the system. Since the set of states is small, extremal dephasers can still have high fidelity.
WSE decoherence constant, γ_{decoh}	$\text{SD}[\sigma_i] = \gamma_{\text{decoh}} \mathbb{E}[1 - \sigma_i]$	- For WSE channels, $\gamma_{\text{decoh}} \ll 1/\sqrt{\mathbb{E}[1 - \sigma_i]}$.
WSE coherence constant γ_{coh} of unitary error \mathcal{U}	Let $\text{Tr } U \in \mathbb{R}_+$. $\text{SD}[\text{Re}\{\lambda_i\}] = \gamma_{\text{coh}} \mathbb{E}[1 - \text{Re}\{\lambda_i\}]$	- For WSE channels, $\gamma_{\text{coh}} \ll 1/\sqrt{\mathbb{E}[1 - \text{Re}\{\lambda_i\}]}$.
Equable channel	Non-catastrophic, no extremal errors (dephasers and unitaries).	- Excludes pathological behaviors induced by extremal errors. - Should apply to all realistic scenarios. - Equable implies WSE.
Wide-sense equable (WSE) channel	$\gamma_{\text{decoh}} \ll 1/\sqrt{\mathbb{E}[1 - \sigma_i]}$, $\gamma_{\text{coh}} \ll 1/\sqrt{\mathbb{E}[1 - \text{Re}\{\lambda_i\}]}$	- Ensures the quasi-correspondence: $\Phi(\mathcal{D}_A, \mathcal{I}) \approx \mathcal{Y}(\mathcal{A}) \approx \max_{\mathcal{W} \in \text{SU}(d)} \Phi(\mathcal{W} \circ \mathcal{A}, \mathcal{U})$ - Ensures the simple decay of the unitarity: $\mathcal{Y}(\mathcal{A}_{m:1}) \approx \prod_i \mathcal{Y}(\mathcal{A}_i)$
Average gate fidelity, $F(\mathcal{A}, \mathcal{U})$	$\mathbb{E}_{\text{Haar}} f_{ \psi\rangle\langle\psi }(\mathcal{A}, \mathcal{U})$	- Is the overlap between noisy and ideal outputs averaged over all physical inputs.
Unitarity, $u(\mathcal{A})$	$\mathbb{E}_{\text{Haar}} \frac{\ \mathcal{A}(\psi\rangle\langle\psi - \mathbb{I}/d)\ _2^2}{\ \psi\rangle\langle\psi - \mathbb{I}/d \ _2^2}$	- Is the average contraction factor of the squared norm of the physical Bloch vectors.
$\Phi(\mathcal{A}, \mathcal{U})$	$\frac{(d+1)F(\mathcal{A}, \mathcal{U}) - 1}{d}$	- For non-catastrophic channels, $\Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1}) \approx \Phi(\mathcal{A}_{m:1}^*, \mathcal{U}_{m:1})$. - For channels $\mathcal{A}_i = \mathcal{V}_i \circ \mathcal{D}_i$ with WSE errors, $\Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1}) \approx \Phi(\mathcal{V}_{m:1}, \mathcal{U}_{m:1}) \prod_i \Phi(\mathcal{D}_i, \mathcal{I})$.
$Y^2(\mathcal{A})$	$\frac{(d^2-1)u(\mathcal{A})+1}{d^2}$	- For non-catastrophic channels, $\mathcal{Y}(\mathcal{A}_{m:1}) \approx \mathcal{Y}(\mathcal{A}_{m:1}^*)$. - In the WSE scenario, $\mathcal{Y}(\mathcal{A}_{m:1}) \approx \prod_i \mathcal{Y}(\mathcal{A}_i)$.
Infidelity, $r(\mathcal{A}, \mathcal{U})$	$1 - F(\mathcal{A}, \mathcal{U})$	- For a channel $\mathcal{A} = \mathcal{V} \circ \mathcal{D}$, (with WSE error) $r = r_{\text{coh}} + r_{\text{decoh}} + O(r^2)$, where $r_{\text{coh}} = r(\mathcal{V}, \mathcal{U})$ and $r_{\text{decoh}} = r(\mathcal{D}, \mathcal{I})$.
Coherence level	r_{coh}/r	- Quantifies the proportion to which the error is coherent.
Decoherence-limited channel	$r_{\text{coh}}/r = O(r)$	- WSE decoherence-limited channels form a closed set under composition.

TABLE 4.2: Summary of the main concepts addressed in this paper.

4.9 Proofs

4.9.1 A noteworthy trace inequality

This subsection is dedicated to demonstrating a useful trace inequality.

Lemma 4: Noteworthy trace inequality

Let $A, B \in M_d(\mathbb{C})$ be Hermitian matrices with eigenvalues of at most ρ_A, ρ_B respectively. Then,

$$\frac{\text{Tr } AB}{d} \geq \rho_B \frac{\text{Tr } A}{d} + \rho_A \frac{\text{Tr } B}{d} - \rho_A \rho_B. \quad (4.83)$$

Proof. We first show this inequality for positive semi-definite matrices with eigenvalues of at most 1, under the condition that

$$d < \lfloor \text{Tr } A \rfloor + \lfloor \text{Tr } B \rfloor + 2. \quad (4.84)$$

In such case, the inner product is minimized by the sum of eigenvalues paired in opposite order [WG93] (it's a matrix equivalent to the Hardy-Littlewood rearrangement inequality):

$$\frac{\text{Tr } AB}{d} \geq \frac{1}{d} \sum_i \lambda_i^\uparrow(A) \lambda_i^\downarrow(B). \quad (4.85)$$

This is in turn minimized when both $\{\lambda_i(A)\}$ and $\{\lambda_i(B)\}$ are maximized in terms of strong majorization. Since the eigenvalues are between zero and 1, both majorizations have a simple form:

$$\lambda_i(A) = \begin{cases} 1 & \text{for } i \leq \lfloor \text{Tr } A \rfloor \\ \text{Tr } A - \lfloor \text{Tr } A \rfloor & \text{for } i = \lfloor \text{Tr } A \rfloor + 1 \\ 0 & \text{otherwise.} \end{cases} \quad (4.86)$$

$$\lambda_i(B) = \begin{cases} 1 & \text{for } i \leq \lfloor \text{Tr } B \rfloor \\ \text{Tr } B - \lfloor \text{Tr } B \rfloor & \text{for } i = \lfloor \text{Tr } B \rfloor + 1 \\ 0 & \text{otherwise.} \end{cases} \quad (4.87)$$

With such spectrum and the condition $d < \lfloor \text{Tr } A \rfloor + \lfloor \text{Tr } B \rfloor + 2$, we are ensured that

$$\frac{1}{d} \sum_i \lambda_i^\uparrow(A) \lambda_i^\downarrow(B) = \frac{\text{Tr } A}{d} + \frac{\text{Tr } B}{d} - 1, \quad (4.88)$$

which, together with eq. (4.85), yields eq. (4.83) in this simpler case.

Now, consider the general case of Hermitian matrices A, B with eigenvalues of at most ρ_A, ρ_B respectively. Let $A = (A + n_A \mathbb{I}) - n_A \mathbb{I}$, $B = (B + n_B \mathbb{I}) - n_B \mathbb{I}$, for $n_A, n_B \in \mathbb{R}_+$, and

consider the following expansion:

$$\begin{aligned} \frac{\text{Tr } AB}{d} &= \frac{\text{Tr}(A + n_A \mathbb{I})(B + n_B \mathbb{I})}{d} - n_A \frac{\text{Tr}(B + n_B \mathbb{I})}{d} - n_B \frac{\text{Tr}(A + n_A \mathbb{I})}{d} + n_A n_B \\ &= (\rho_A + n_A)(\rho_B + n_B) \frac{\text{Tr} \left(\frac{A + n_A \mathbb{I}}{\rho_A + n_A} \right) \left(\frac{B + n_B \mathbb{I}}{\rho_B + n_B} \right)}{d} - n_A \frac{\text{Tr } B}{d} - n_B \frac{\text{Tr } A}{d} - n_A n_B \end{aligned} \quad (4.89)$$

Now, let's pick n_A, n_B large enough so that

- i. $A + n_A \mathbb{I}, B + n_B \mathbb{I} \geq 0$,
- ii. $d < \lfloor \text{Tr} \left(\frac{A + n_A \mathbb{I}}{\rho_A + n_A} \right) \rfloor + \lfloor \text{Tr} \left(\frac{B + n_B \mathbb{I}}{\rho_B + n_B} \right) \rfloor + 2$.

For i we can simply pick $n_A \geq \min \lambda(A)$, $n_B \geq \min \lambda(B)$. To see why ii is also possible, realize that

$$\lim_{n_A \rightarrow \infty} \left\lfloor \text{Tr} \left(\frac{A + n_A \mathbb{I}}{\rho_A + n_A} \right) \right\rfloor = d, \quad (4.90)$$

meaning that there exists a finite n_A such that ii is fulfilled. Moreover, realize that the maximum eigenvalue of both $\frac{A + n_A \mathbb{I}}{\rho_A + n_A}$ and $\frac{B + n_B \mathbb{I}}{\rho_B + n_B}$ is upper-bounded by 1 by construction. Combining all this, we get

$$\frac{\text{Tr} \left(\frac{A + n_A \mathbb{I}}{\rho_A + n_A} \right) \left(\frac{B + n_B \mathbb{I}}{\rho_B + n_B} \right)}{d} \geq \frac{\text{Tr} \left(\frac{A + n_A \mathbb{I}}{\rho_A + n_A} \right)}{d} + \frac{\text{Tr} \left(\frac{B + n_B \mathbb{I}}{\rho_B + n_B} \right)}{d} - 1, \quad (4.91)$$

since this corresponds to our initial simpler case. Substituting eq. (4.91) into eq. (4.89) and simplifying, we get eq. (4.83) which completes the proof. \square

This inequality pairs well with the well-known Von-Neuman's trace inequality, as when $\text{Tr } AB \geq 0$, lemma 4 provides a much better lower bound. To see this, consider the following inequality which is trivially derived from Von's Neumann's trace inequality:

Lemma 5: Flavored Von Neumann's trace inequality

Let $A, B \in M_d(\mathbb{C})$ be matrices with spectral radius of at most ρ_A, ρ_B respectively. Then,

$$\left| \frac{\text{Tr } AB}{d} \right| \leq \min \left(\rho_B \frac{\text{Tr} |A|}{d}, \rho_A \frac{\text{Tr} |B|}{d} \right). \quad (4.92)$$

Recalling that $\|A\|_2^2 = \text{Tr } A^\dagger A$ and using those two last inequalities, we get the following norm inequality:

Lemma 6: Norm inequality

Consider two matrices A, B with spectral radius of at most 1. Then,

$$\frac{\|A\|_2^2}{d} + \frac{\|B\|_2^2}{d} - 1 \leq \frac{\|AB\|_2^2}{d} \leq \min \left(\frac{\|A\|_2^2}{d}, \frac{\|B\|_2^2}{d} \right). \quad (4.93)$$

4.9.2 Proofs of the main results

Notation and remarks

Before we start proving theorems 7 and 8, let's introduce some handy notation. The i^{th} canonical Kraus operator of a channel \mathcal{A}_j is denoted A_i^j . Let $a \geq b$; we denote

$$A_{\vec{i}}^{a:b} = A_{i_{a-b+1}}^a A_{i_{a-b}}^{a-1} \dots A_{i_2}^{b+1} A_{i_1}^b, \quad (4.94)$$

where $\vec{i} \in \mathbb{N}^{a-b+1}$ simply contains indices $i_k \in \{1, \dots, d^2\}$. Finally we denote $\vec{1} = (1, \dots, 1)$ for which the dimension is left implicit.

Remark that the set $\{A_{\vec{i}}^{m:1}\}$ consist of a valid Kraus decomposition for the composite channel $\mathcal{A}_{m:1}$, and can be used to calculate $\Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1})$ and $Y(\mathcal{A}_{m:1})$ through eqs. (4.10) and (4.13) respectively. However, these Kraus operators are generally not orthogonal to one another (this is not the canonical decomposition), which prevents the same proof technique as in lemmas 2 and 3.

Proof of the evolution theorem 7

Proof. Using Hölder's inequality, we get

$$\begin{aligned} Y^2(\mathcal{A}_{m:1}) &= \sum_{\vec{i}} \left(\frac{\|A_{\vec{i}}^{m:1}\|_2^2}{d} \right)^2 && (4.95) \\ &\leq \max_{\vec{i}} \frac{\|A_{\vec{i}}^{m:1}\|_2^2}{d} \sum_{\vec{j}} \frac{\|A_{\vec{j}}^{m:1}\|_2^2}{d} && \text{(Hölder ineq.)} \\ &= \max_{\vec{i}} \frac{\|A_{\vec{i}}^{m:1}\|_2^2}{d}. && \text{(TP condition)} \end{aligned}$$

One might have a (justified) hunch that $\operatorname{argmax}_{\vec{i}} \frac{\|A_{\vec{i}}^{m:1}\|_2^2}{d} = \vec{1}$ in non-catastrophic noise scenarios. To show this, consider \vec{i} with $i_k \neq 1$ for some $k \in \{1, \dots, m\}$. Using the properties of contractions, we have

$$\begin{aligned} \frac{\|A_{\vec{i}}^{m:1}\|_2^2}{d} &\leq \frac{\|A_{i_k}^k\|_2^2}{d} && \text{(Contractions)} \\ &\leq 1 - \frac{\|A_1^k\|_2^2}{d} && \text{(TP condition)} \\ &< 1/2. && \text{(Non-catastrophic)} \end{aligned}$$

Hence, if we suppose $\operatorname{argmax}_{\vec{i}} \frac{\|A_{\vec{i}}^{m:1}\|_2^2}{d} \neq \vec{1}$, we have

$$Y^2(\mathcal{A}_{m:1}) < 1/2, \quad (4.96)$$

which cannot be respected if the channel $\mathcal{A}_{m:1}$ is non-catastrophic. Hence, by contradiction we have

$$Y^2(\mathcal{A}_{m:1}) \leq \frac{\|A_{\vec{1}}^{m:1}\|_2^2}{d} = Y(\mathcal{A}_{m:1}^*). \quad (4.97)$$

From there we get

$$\begin{aligned} Y^2(\mathcal{A}_{m:1}) &= \left(\frac{\|A_{\vec{1}}^{m:1}\|_2^2}{d} \right)^2 + \sum_{\vec{i} \neq \vec{1}} \left(\frac{\|A_{\vec{i}}^{m:1}\|_2^2}{d} \right)^2 \\ &\leq Y^2(\mathcal{A}_{m:1}^*) + \left(\sum_{\vec{i} \neq \vec{1}} \frac{\|A_{\vec{i}}^{m:1}\|_2^2}{d} \right)^2 \\ &= Y^2(\mathcal{A}_{m:1}^*) + \left(1 - \frac{\|A_{\vec{1}}^{m:1}\|_2^2}{d} \right)^2 && \text{(TP condition)} \\ &= Y^2(\mathcal{A}_{m:1}^*) + (1 - Y(\mathcal{A}_{m:1}^*))^2 && (4.98) \\ &\leq Y^2(\mathcal{A}_{m:1}^*) + (1 - Y^2(\mathcal{A}_{m:1}))^2 . && \text{(Equation (4.97))} \end{aligned}$$

□

Proof of the evolution theorem 8

Proof. We will show that the inequality eq. (4.27) holds for $m = 2^n$, $\forall n \in \mathbb{N}$. This suffices since if $N < 2^n$, then we can append $\mathcal{I}_{2^n - N:1}$ to the composition $\mathcal{A}_{N:1}$ so that $\mathcal{A}_{N:1} \circ \mathcal{I}_{2^n - N:1}$ is a composition of length 2^n . Appending $\mathcal{I}_{2^n - N:1}$ has no effect on eq. (4.27).

From the definition of Φ , we have that $\Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1}) - \Phi(\mathcal{A}_{m:1}^*, \mathcal{U}_{m:1}) \geq 0$, so it only remains to derive an upper bound on $\Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1}) - \Phi(\mathcal{A}_{m:1}^*, \mathcal{U}_{m:1})$. Our approach will be to split the sum as follows:

$$\begin{aligned} \frac{1}{d^2} \sum_{\vec{i} \neq \vec{1}} \left| \langle A_{\vec{i}}^{m:1}, U^{m:1} \rangle \right|^2 &= \frac{1}{d^2} \sum_{\vec{i} \neq \vec{1}} \left| \langle A_{\vec{i}}^{m:\frac{m}{2}+1} A_{\vec{1}}^{\frac{m}{2}:1}, U^{m:1} \rangle \right|^2 + \frac{1}{d^2} \sum_{\vec{j} \neq \vec{1}} \left| \langle A_{\vec{1}}^{m:\frac{m}{2}+1} A_{\vec{j}}^{\frac{m}{2}:1}, U^{m:1} \rangle \right|^2 \\ &\quad + \frac{1}{d^2} \sum_{\substack{\vec{i} \neq \vec{1} \\ \vec{j} \neq \vec{1}}} \left| \langle A_{\vec{i}}^{m:\frac{m}{2}+1} A_{\vec{j}}^{\frac{m}{2}:1}, U^{m:1} \rangle \right|^2 . \end{aligned} \quad (4.99)$$

The double sum (last term) can be bounded via Cauchy-Schwarz inequality followed by the usage of lemma 6:

$$\begin{aligned}
\frac{1}{d^2} \sum_{\substack{i \neq \bar{1} \\ j \neq \bar{1}}} \left| \left\langle A_{\bar{i}}^{m: \frac{m}{2}+1} A_{\bar{j}}^{\frac{m}{2}:1}, U^{m:1} \right\rangle \right|^2 &\leq \sum_{\substack{i \neq \bar{1} \\ j \neq \bar{1}}} \frac{\|A_{\bar{i}}^{m: \frac{m}{2}+1}\|_2^2}{d} \frac{\|A_{\bar{j}}^{\frac{m}{2}:1}\|_2^2}{d} && \text{(Cauchy-Schwarz ineq.)} \\
&\leq \left(1 - \frac{\|A_{\bar{1}}^{m: \frac{m}{2}+1}\|_2^2}{d}\right) \left(1 - \frac{\|A_{\bar{1}}^{\frac{m}{2}:1}\|_2^2}{d}\right) && \text{(TP condition)} \\
&\leq \left(\sum_{i=\frac{m}{2}+1}^m \left(1 - \frac{\|A_{\bar{1}}^i\|_2^2}{d}\right)\right) \left(\sum_{j=1}^{m/2} \left(1 - \frac{\|A_{\bar{1}}^j\|_2^2}{d}\right)\right) \\
&&& \text{(Lemma 6)} \\
&\leq \left(\sum_{i=\frac{m}{2}+1}^m (1 - Y(\mathcal{A}_{\bar{i}}^*))\right) \left(\sum_{j=1}^{m/2} (1 - Y(\mathcal{A}_{\bar{j}}^*))\right). && (4.100)
\end{aligned}$$

With regards to the first two terms on the RHS of eq. (4.99), let's split them both into three terms once again:

$$\begin{aligned}
\frac{1}{d^2} \sum_{i \neq \bar{1}} \left| \left\langle A_{\bar{i}}^{m: \frac{m}{2}+1} A_{\bar{i}}^{\frac{m}{2}:1}, U^{m:1} \right\rangle \right|^2 &= \frac{1}{d^2} \sum_{i \neq \bar{1}} \left| \left\langle A_{\bar{i}}^{m: \frac{3m}{4}+1} A_{\bar{i}}^{\frac{3m}{4}:1}, U^{m:1} \right\rangle \right|^2 \\
&\quad + \frac{1}{d^2} \sum_{i \neq \bar{1}} \left| \left\langle A_{\bar{i}}^{m: \frac{3m}{4}+1} A_{\bar{j}}^{\frac{3m}{4}: \frac{m}{2}+1} A_{\bar{i}}^{\frac{m}{2}:1}, U^{m:1} \right\rangle \right|^2 \\
&\quad + \frac{1}{d^2} \sum_{\substack{i \neq \bar{1} \\ j \neq \bar{1}}} \left| \left\langle A_{\bar{i}}^{m: \frac{3m}{4}+1} A_{\bar{j}}^{\frac{3m}{4}: \frac{m}{2}+1} A_{\bar{i}}^{\frac{m}{2}:1}, U^{m:1} \right\rangle \right|^2. && (4.101a)
\end{aligned}$$

$$\begin{aligned}
\frac{1}{d^2} \sum_{i \neq \bar{1}} \left| \left\langle A_{\bar{i}}^{m: \frac{m}{2}+1} A_{\bar{i}}^{\frac{m}{2}:1}, U^{m:1} \right\rangle \right|^2 &= \frac{1}{d^2} \sum_{i \neq \bar{1}} \left| \left\langle A_{\bar{i}}^{m: \frac{m}{2}+1} A_{\bar{i}}^{\frac{m}{2}: \frac{m}{4}+1} A_{\bar{i}}^{\frac{m}{4}:1}, U^{m:1} \right\rangle \right|^2 \\
&\quad + \frac{1}{d^2} \sum_{j \neq \bar{1}} \left| \left\langle A_{\bar{i}}^{m: \frac{m}{4}+1} A_{\bar{j}}^{\frac{m}{4}:1}, U^{m:1} \right\rangle \right|^2 \\
&\quad + \frac{1}{d^2} \sum_{\substack{i \neq \bar{1} \\ j \neq \bar{1}}} \left| \left\langle A_{\bar{i}}^{m: \frac{m}{2}+1} A_{\bar{i}}^{\frac{m}{2}: \frac{m}{4}+1} A_{\bar{j}}^{\frac{m}{4}:1}, U^{m:1} \right\rangle \right|^2. && (4.101b)
\end{aligned}$$

The double sums on the RHS of eqs. (4.101a) and (4.101b) can be upper bounded using the same technique as earlier, which yields

$$\begin{aligned} & \frac{1}{d^2} \sum_{\substack{i \neq 1 \\ j \neq 1}} \left| \left\langle A_i^{m: \frac{3m}{4}+1} A_j^{\frac{3m}{4}: \frac{m}{2}+1} A_1^{\frac{m}{2}: 1}, U^{m:1} \right\rangle \right|^2 + \frac{1}{d^2} \sum_{\substack{i \neq 1 \\ j \neq 1}} \left| \left\langle A_1^{m: \frac{m}{2}+1} A_i^{\frac{m}{2}: \frac{m}{4}+1} A_j^{\frac{m}{4}: 1}, U^{m:1} \right\rangle \right|^2 \\ & \leq \left(\sum_{i=\frac{3m}{4}+1}^m (1 - Y(\mathcal{A}_i^*)) \right) \left(\sum_{j=\frac{m}{2}+1}^{\frac{3m}{4}} (1 - Y(\mathcal{A}_j^*)) \right) + \left(\sum_{i=\frac{m}{4}+1}^{\frac{m}{2}} (1 - Y(\mathcal{A}_i^*)) \right) \left(\sum_{j=1}^{\frac{m}{4}} (1 - Y(\mathcal{A}_j^*)) \right). \end{aligned} \quad (4.102)$$

By iterating the same subdivision technique, we end up with

$$\begin{aligned} \frac{1}{d^2} \sum_{i \neq 1} \left| \left\langle A_i^{m:1}, U^{m:1} \right\rangle \right|^2 & \leq \sum_{i=1}^n \sum_{j=1}^{2^{i-1}} \left(\sum_{k=\frac{2^n}{2^i}(2^i-2j+1)+1}^{\frac{2^n}{2^i}(2^i-2j+2)} (1 - Y(\mathcal{A}_k^*)) \right) \left(\sum_{k=\frac{2^n}{2^i}(2^i-2j)+1}^{\frac{2^n}{2^i}(2^i-2j+1)} (1 - Y(\mathcal{A}_k^*)) \right) \\ & \quad + \frac{1}{d^2} \sum_{\substack{j=1 \\ i \neq 1}}^{j=m} \left| \left\langle A_1^{m:j+1} A_i^j A_1^{j-1:1}, U^{m:1} \right\rangle \right|^2 \end{aligned} \quad (4.103)$$

Bounding the first term on the RHS can be done by alternating between the AM-GM inequality and square completions. First let's perform the AM-GM inequality on the terms of the summation restricted to $i = n$:

$$\sum_{j=1}^{2^{n-1}} (1 - Y(\mathcal{A}_{2^n-2j+2}^*)) (1 - Y(\mathcal{A}_{2^n-2j+1}^*)) \leq \frac{1}{4} \sum_{j=1}^{2^{n-1}} \left(\sum_{k=2^n-2j+1}^{2^n-2j+2} (1 - Y(\mathcal{A}_k^*)) \right)^2. \quad (4.104)$$

Then, let's add in the terms with index $i = n - 1$ and complete the squares (taking $n = 3$ as an example is recommended):

$$\begin{aligned}
& \sum_{i=n-1}^n \sum_{j=1}^{2^{i-1}} \left(\sum_{k=\frac{2^n}{2^i}(2^i-2j+1)+1}^{\frac{2^n}{2^i}(2^i-2j+2)} (1 - Y(\mathcal{A}_k^*)) \right) \left(\sum_{k=\frac{2^n}{2^i}(2^i-2j)+1}^{\frac{2^n}{2^i}(2^i-2j+1)} (1 - Y(\mathcal{A}_k^*)) \right) \\
& \leq \frac{1}{4} \sum_{j=1}^{2^{n-1}} \left(\sum_{k=2^{n-2j+1}}^{2^n-2j+2} (1 - Y(\mathcal{A}_k^*)) \right)^2 \\
& + \sum_{j=1}^{2^{n-2}} \left(\sum_{k=2(2^{n-1}-2j+1)+1}^{2(2^{n-1}-2j+2)} (1 - Y(\mathcal{A}_k^*)) \right) \left(\sum_{k=2(2^{n-1}-2j)+1}^{2(2^{n-1}-2j+1)} (1 - Y(\mathcal{A}_k^*)) \right), \quad (\text{eq. (4.104)}) \\
& \leq \frac{1}{4} \sum_{j=1}^{2^{n-2}} \left(\sum_{k=2(2^{n-1}-2j)+1}^{2(2^{n-1}-2j+2)} (1 - Y(\mathcal{A}_k^*)) \right)^2 \\
& + \frac{1}{2} \sum_{j=1}^{2^{n-2}} \left(\sum_{k=2(2^{n-1}-2j+1)+1}^{2(2^{n-1}-2j+2)} (1 - Y(\mathcal{A}_k^*)) \right) \left(\sum_{k=2(2^{n-1}-2j)+1}^{2(2^{n-1}-2j+1)} (1 - Y(\mathcal{A}_k^*)) \right) \\
& \hspace{20em} (\text{Square completions}) \\
& \leq \frac{3}{8} \sum_{j=1}^{2^{n-2}} \left(\sum_{k=2(2^{n-1}-2j)+1}^{2(2^{n-1}-2j+2)} (1 - Y(\mathcal{A}_k^*)) \right)^2. \quad (\text{AM-GM ineq.})
\end{aligned}$$

Similarly, we can then add in the terms with index $i = n - 2$, complete the squares and use the AM-GM inequality on the leftover summation:

$$\begin{aligned}
& \sum_{i=n-2}^n \sum_{j=1}^{2^{i-1}} \left(\sum_{k=\frac{2^n}{2^i}(2^i-2j+1)+1}^{\frac{2^n}{2^i}(2^i-2j+2)} (1 - Y(\mathcal{A}_k^*)) \right) \left(\sum_{k=\frac{2^n}{2^i}(2^i-2j)+1}^{\frac{2^n}{2^i}(2^i-2j+1)} (1 - Y(\mathcal{A}_k^*)) \right) \\
& \leq \frac{7}{16} \sum_{j=1}^{2^{n-3}} \left(\sum_{k=4(2^{n-2}-2j)+1}^{4(2^{n-2}-2j+2)} (1 - Y(\mathcal{A}_k^*)) \right)^2. \quad (4.105)
\end{aligned}$$

Repeating this procedure until $i = 1$, we get

$$\begin{aligned}
& \sum_{i=1}^n \sum_{j=1}^{2^{i-1}} \left(\sum_{k=\frac{2^n}{2^i}(2^i-2j+1)+1}^{\frac{2^n}{2^i}(2^i-2j+2)} (1 - Y(\mathcal{A}_k^*)) \right) \left(\sum_{k=\frac{2^n}{2^i}(2^i-2j)+1}^{\frac{2^n}{2^i}(2^i-2j+1)} (1 - Y(\mathcal{A}_k^*)) \right) \\
& \leq \left(\frac{1}{2} - \frac{1}{2^{n+1}} \right) \left(\sum_{i=1}^m (1 - Y(\mathcal{A}_i^*)) \right)^2. \quad (4.106)
\end{aligned}$$

The last term on the RHS of eq. (4.103) is upper-bounded using an alternate technique. First, we get

$$\begin{aligned}
\sum_{\substack{j=1 \\ i \neq 1}}^{j=m} \frac{\left| \left\langle \frac{A_{\bar{1}}^{m:j+1} A_i^j A_{\bar{1}}^{j-1:1}}{d^2}, \underline{U}^{m:1} \right\rangle \right|^2}{d^2} &= \sum_{\substack{j=1 \\ i \neq 1}}^{j=m} \left| \left\langle A_{\bar{1}}^{m:j+1} \underline{A}_i^j A_{\bar{1}}^{j-1:1}, \underline{U}^{m:1} \right\rangle \right|^2 \cdot \frac{\|A_i^j\|_2^2}{d} \\
&\leq \max_{\substack{j \\ i \neq 1}} \left| \left\langle A_{\bar{1}}^{m:j+1} \underline{A}_i^j A_{\bar{1}}^{j-1:1}, \underline{U}^{m:1} \right\rangle \right|^2 \left(\sum_{\substack{k=1 \\ \ell \neq 1}}^{k=m} \frac{\|A_\ell^k\|_2^2}{d} \right) \\
&\hspace{15em} \text{(Hölder's ineq.)} \\
&= \max_{\substack{j \\ i \neq 1}} \left| \left\langle \underline{A}_i^j, (A_{\bar{1}}^{m:j+1})^\dagger \underline{U}^{m:1} (A_{\bar{1}}^{j-1:1})^\dagger \right\rangle \right|^2 \sum_{k=1}^m \left(1 - \frac{\|A_{\bar{1}}^k\|_2^2}{d} \right) \\
&\hspace{15em} \text{(TP condition)} \\
&= \max_{\substack{j \\ i \neq 1}} \left| \left\langle \underline{A}_i^j, (A_{\bar{1}}^{m:j+1})^\dagger \underline{U}^{m:1} (A_{\bar{1}}^{j-1:1})^\dagger \right\rangle \right|^2 \sum_{k=1}^m (1 - Y(\mathcal{A}_k^*)) . \\
&\hspace{15em} \text{(4.107)}
\end{aligned}$$

For fixed j , $\{\underline{A}_i^j\}$ forms an orthonormal basis. Since $\|(A_{\bar{1}}^{m:j+1})^\dagger \underline{U}^{m:1} (A_{\bar{1}}^{j-1:1})^\dagger\|_2^2 \leq 1$ (contractions), we have that, for any j :

$$\begin{aligned}
\max_{i \neq 1} \left| \left\langle \underline{A}_i^j, (A_{\bar{1}}^{m:j+1})^\dagger \underline{U}^{m:1} (A_{\bar{1}}^{j-1:1})^\dagger \right\rangle \right|^2 &\leq 1 - \left| \left\langle A_{\bar{1}}^{m:j+1} \underline{A}_{\bar{1}}^j A_{\bar{1}}^{j-1:1}, \underline{U}^{m:1} \right\rangle \right|^2 \\
&= 1 - \frac{\left| \left\langle A_{\bar{1}}^{m:j+1} \underline{A}_{\bar{1}}^j A_{\bar{1}}^{j-1:1}, \underline{U}^{m:1} \right\rangle \right|^2}{d \|A_{\bar{1}}^j\|_2^2} \\
&= 1 - \frac{\Phi(\mathcal{A}_{m:1}^*, \mathcal{U}_{m:1})}{Y(\mathcal{A}_{\bar{1}}^*)} \\
&\leq 1 - \Phi(\mathcal{A}_{m:1}^*, \mathcal{U}_{m:1}) . \\
&\hspace{15em} \text{(4.108)}
\end{aligned}$$

Combining eqs. (4.103) and (4.106) to (4.108) yields

$$\begin{aligned}
\Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1}) - \Phi(\mathcal{A}_{m:1}^*, \mathcal{U}_{m:1}) &\leq \left(\frac{1}{2} - \frac{1}{2^{n+1}} \right) \left(\sum_{i=1}^m (1 - Y(\mathcal{A}_i^*)) \right)^2 \\
&\quad + (1 - \Phi(\mathcal{A}_{m:1}^*, \mathcal{U}_{m:1})) \sum_{i=1}^m (1 - Y(\mathcal{A}_i^*)) . \\
&\hspace{15em} \text{(4.109)}
\end{aligned}$$

To obtain a lower bound that doesn't involve the LK approximation, we substitute $\Phi(\mathcal{A}_{m:1}^*, \mathcal{U}_{m:1})$ by its lower bound, and $Y(\mathcal{A}_i^*) \geq Y^2(\mathcal{A}_i)$ (see eq. (4.97)):

$$\begin{aligned}
& \Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1}) - \Phi(\mathcal{A}_{m:1}^*, \mathcal{U}_{m:1}) \\
& \leq \frac{1}{2} \left(\sum_{i=1}^m (1 - Y^2(\mathcal{A}_i)) \right)^2 + (1 - \Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1})) \sum_{i=1}^m (1 - Y^2(\mathcal{A}_i)) \\
& + \frac{1}{2} \left(\sum_{i=1}^m (1 - Y^2(\mathcal{A}_i)) \right)^3 + (1 - \Phi(\mathcal{A}_{m:1}^*, \mathcal{U}_{m:1})) \left(\sum_{i=1}^m (1 - Y^2(\mathcal{A}_i)) \right)^2. \tag{4.110}
\end{aligned}$$

□

Proof of theorem 12

The simplest route to prove theorems 11 to 14 is probably to start with the demonstration of theorem 12.

Proof. Given m decoherent channels \mathcal{D}_i with respective LK operators D_1^i , we first want to bound the behavior of

$$\sqrt{\Phi(\mathcal{D}_{m:1}^*, \mathcal{I})} = \frac{\text{Tr}(|D_1^m| \cdots |D_1^1|)}{d} \tag{4.111}$$

as a function of the $\sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})}$ s. Let's express the LK operators as $|D_1^i| = \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \mathbb{I}_d + \Delta_i$, and apply a telescopic expansion:

$$\begin{aligned}
\frac{\text{Tr}(|D_1^m| \cdots |D_1^1|)}{d} &= \prod_{i=1}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} + \sum_{j=1}^m \frac{\text{Tr}(|D_1^m| \cdots |D_1^{j+1}| \Delta_j)}{d} \prod_{i=1}^{j-1} \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \\
& \hspace{15em} \text{(Telescopic sum)} \\
&= \prod_{i=1}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} + \left(\prod_{i=1}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \sum_{j=1}^m \frac{\text{Tr} \Delta_j}{d} / \sqrt{\Phi(\mathcal{D}_j^*, \mathcal{I})} \\
&+ \sum_{j=1}^m \sum_{k=j+1}^m \frac{\text{Tr}(|D_1^m| \cdots |D_1^{k+1}| \Delta_k \Delta_j)}{d} \prod_{\substack{i=1 \\ i \neq j}}^{k-1} \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})}. \\
& \hspace{15em} \text{(Telescopic sum, again)}
\end{aligned}$$

By construction, $\text{Tr } \Delta_i = 0$, which leaves us with

$$\begin{aligned}
\left| \sqrt{\Phi(\mathcal{D}_{m:1}^*, \mathcal{I})} - \prod_{i=1}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right| &\leq \left| \sum_{j=1}^m \sum_{k=j+1}^m \frac{\text{Tr} \left(|D_1^m| \cdots |D_1^{k+1}| \Delta_k \Delta_j \right)}{d} \prod_{\substack{i=1 \\ i \neq j}}^{k-1} \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right| \\
&\leq \sum_{j=1}^m \sum_{k=j+1}^m \left| \frac{\text{Tr} \left(|D_1^m| \cdots |D_1^{k+1}| \Delta_k \Delta_j \right)}{d} \right| \prod_{\substack{i=1 \\ i \neq j}}^{k-1} \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \\
&\hspace{15em} \text{(Triangle ineq.)} \\
&\leq \sum_{j=1}^m \sum_{k=j+1}^m \left| \frac{\text{Tr} \left(|D_1^m| \cdots |D_1^{k+1}| \Delta_k \Delta_j \right)}{d} \right| \hspace{10em} (\Phi(\mathcal{D}_i^*, \mathcal{I}) \leq 1) \\
&\leq \sum_{j=1}^m \sum_{k=j+1}^m \frac{\| |D_1^m| \cdots |D_1^{k+1}| \Delta_k \|_2 \| \Delta_j \|_2}{\sqrt{d} \sqrt{d}} \\
&\hspace{15em} \text{(Cauchy-Schwarz ineq.)} \\
&\leq \sum_{j=1}^m \sum_{k=j+1}^m \frac{\| \Delta_k \|_2 \| \Delta_j \|_2}{\sqrt{d} \sqrt{d}} \hspace{10em} \text{(Contractions)}
\end{aligned}$$

This is where definition 7 (equability) comes in handy, since it essentially states that $\frac{\| \Delta_i \|_2}{\sqrt{d}} = \gamma_{\text{decoh}}(\mathcal{D}_i) \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right)$. From there, we have

$$\left| \sqrt{\Phi(\mathcal{D}_{m:1}^*, \mathcal{I})} - \prod_{i=1}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right| \leq \gamma_{\text{decoh}}^2 \sum_{i=1}^m \sum_{j=i+1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_j^*, \mathcal{I})} \right) \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \tag{4.112}$$

$$\begin{aligned}
&\leq \gamma_{\text{decoh}}^2 \sum_{i=1}^m \sum_{j=i+1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_j^*, \mathcal{I})} \right) \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \\
&+ \frac{\gamma_{\text{decoh}}^2}{2} \sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right)^2 \quad \text{(Adding a positive term)} \\
&= \frac{\gamma_{\text{decoh}}^2}{2} \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \right)^2 \tag{4.113}
\end{aligned}$$

A few straightforward algebraic manipulations on eq. (4.113) yield

$$\begin{aligned}
\left| \Phi(\mathcal{D}_{m:1}^*, \mathcal{I}) - \prod_{i=1}^m \Phi(\mathcal{D}_i^*, \mathcal{I}) \right| &\leq \gamma_{\text{decoh}}^2 \prod_{i=1}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \right)^2 \\
&+ \frac{\gamma_{\text{decoh}}^4}{4} \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \right)^4. \tag{4.114}
\end{aligned}$$

Using a simple telescopic expansion and lemma 2 and theorem 8, we have

$$\begin{aligned} \prod_{i=1}^m \Phi(\mathcal{D}_i, \mathcal{I}) - \prod_{i=1}^m \Phi(\mathcal{D}_i^*, \mathcal{I}) &= \sum_{j=1}^m \left(\prod_{i=j+1}^m \Phi(\mathcal{D}_i, \mathcal{I}) \right) \left(\Phi(\mathcal{D}_j, \mathcal{I}) - \Phi(\mathcal{D}_j^*, \mathcal{I}) \right) \left(\prod_{i=1}^{j-1} \Phi(\mathcal{D}_i^*, \mathcal{I}) \right) \\ &\leq \sum_{i=1}^m (1 - Y(\mathcal{D}_i^*)) (1 - \Phi(\mathcal{D}_i, \mathcal{I})) . \end{aligned} \quad (4.115)$$

From the triangle inequality we have

$$\begin{aligned} \left| \Phi(\mathcal{D}_{m:1}, \mathcal{I}) - \prod_{i=1}^m \Phi(\mathcal{D}_i, \mathcal{I}) \right| &\leq \left| \Phi(\mathcal{D}_{m:1}, \mathcal{I}) - \Phi(\mathcal{D}_{m:1}^*, \mathcal{I}) \right| + \left| \Phi(\mathcal{D}_{m:1}^*, \mathcal{I}) - \prod_{i=1}^m \Phi(\mathcal{D}_i^*, \mathcal{I}) \right| \\ &\quad + \left| \prod_{i=1}^m \Phi(\mathcal{D}_i^*, \mathcal{I}) - \prod_{i=1}^m \Phi(\mathcal{D}_i, \mathcal{I}) \right| . \end{aligned} \quad (4.116)$$

Applying theorem 8 and eqs. (4.114) and (4.115) on the RHS yields eq. (4.58). \square

Proof of theorem 11

Proof. First, we derive an upper bound for $Y^2(\mathcal{A}_{m:1})$:

$$\begin{aligned} Y^2(\mathcal{A}_{m:1}) &\leq Y^2(\mathcal{A}_{m:1}^*) + (1 - Y(\mathcal{A}^{star_{m:1}}))^2 &&= \left(\frac{\|A_1^{m:1}\|_2^2}{d} \right)^2 + (1 - Y(\mathcal{A}_{m:1}^*))^2 && \text{(Theorem 7)} \\ &\leq \min_i \left(\frac{\|A_1^i\|_2^2}{d} \right)^2 + (1 - Y(\mathcal{A}_{m:1}^*))^2 && \text{(Lemma 6)} \\ &\leq \min_i Y^2(\mathcal{A}_i) + (1 - Y(\mathcal{A}_{m:1}^*))^2 . && \text{(Lemma 2)} \end{aligned}$$

Before taking the square root on each side, notice that for any $\epsilon \leq 0$, the non-catastrophic condition enforces that $\sqrt{Y^2(\mathcal{A}_i) + \epsilon} \leq Y(\mathcal{A}_i) + \epsilon/\sqrt{2}$. Indeed, since $Y(\mathcal{A}_i) > 1/\sqrt{2}$,

$$\begin{aligned} 1 - \sqrt{2}Y(\mathcal{A}_i) &< 0 \\ \Rightarrow \epsilon(1 - \sqrt{2}Y(\mathcal{A}_i) - \epsilon/2) &< 0 \\ \Rightarrow Y^2(\mathcal{A}_i) + \epsilon &< Y^2(\mathcal{A}_i) + \sqrt{2}\epsilon Y(\mathcal{A}_i) + \epsilon^2/2 \\ \Rightarrow \sqrt{Y^2(\mathcal{A}_i) + \epsilon} &< Y(\mathcal{A}_i) + \epsilon/\sqrt{2} . \end{aligned} \quad (4.117)$$

Hence,

$$Y(\mathcal{A}_{m:1}) \leq \min_i Y(\mathcal{A}_i) + (1 - Y(\mathcal{A}_{m:1}^*))^2/\sqrt{2} , \quad (4.118)$$

$$\min_i Y(\mathcal{A}_i) + (1 - Y^2(\mathcal{A}_{m:1}))^2/\sqrt{2} \quad (4.119)$$

which corresponds to the quasi-monotonicity statement. We then derive a lower bound on $Y(\mathcal{A}_{m:1})$:

$$\begin{aligned}
1 - Y(\mathcal{A}_{m:1}) &\leq 1 - Y(\mathcal{A}_{m:1}^*) && \text{(Theorem 7)} \\
&= 1 - \frac{\|A_1^{m:1}\|_2^2}{d} \\
&\leq \sum_{i=1}^m \left(1 - \frac{\|A_1^i\|_2^2}{d} \right) && \text{(Lemma 6)} \\
&\leq \sum_{i=1}^m \left(1 - \sqrt{Y^2(\mathcal{A}_i) - (1 - Y^2(\mathcal{A}_i))^2} \right). && \text{(Lemma 2)}
\end{aligned}$$

Direct computation suffices to show that for $Y(\mathcal{A}_i) \in [2^{-1/2}, 1]$, $\sqrt{Y^2(\mathcal{A}_i) - (1 - Y^2(\mathcal{A}_i))^2} \leq Y(\mathcal{A}_i) - (1 - Y^2(\mathcal{A}_i))^2$, hence

$$1 - Y(\mathcal{A}_{m:1}) \leq \sum_{i=1}^m (1 - Y(\mathcal{A}_i)) + (1 - Y^2(\mathcal{A}_i))^2, \quad (4.120)$$

which corresponds to the quasi-subadditivity property. To derive the approximate multiplicativity statement, let's factor the decoherent channels into their (left) polar decomposition $\mathcal{A}_i = \mathcal{D}_i \circ \mathcal{V}_i$. By relabeling $(\mathcal{V}_{i:1})^{-1} \circ \mathcal{D}_i \circ \mathcal{V}_{i:1} = \mathcal{D}'_i$, we have

$$Y(\mathcal{A}_{m:1}^*) = \sqrt{\Phi(\mathcal{D}'_1 \cdots \mathcal{D}'_m \mathcal{D}'_{m:1}, \mathcal{I})}. \quad (4.121)$$

From definition 7, we have that $Y(\mathcal{A}_i^*) - \Phi(\mathcal{D}'_i, \mathcal{I}) \leq \gamma_{\text{decoh}}^2 \left(1 - \sqrt{\Phi(\mathcal{D}'_i, \mathcal{I})} \right)^2$. We can use a telescopic expansion to get

$$\begin{aligned}
\prod_i Y(\mathcal{A}_i^*) - \prod_i \Phi(\mathcal{D}'_i, \mathcal{I}) &= \sum_{j=1}^m \left(\prod_{i=j+1}^m Y(\mathcal{A}_i^*) \right) (Y(\mathcal{A}_j^*) - \Phi(\mathcal{D}'_j, \mathcal{I})) \left(\prod_{i=1}^{j-1} \Phi(\mathcal{D}'_i, \mathcal{I}) \right) \\
&\hspace{15em} \text{(Telescopic sum)} \\
&\leq \gamma_{\text{decoh}}^2 \sum_{j=1}^m \left(\prod_{i=j+1}^m Y(\mathcal{A}_i^*) \right) \left(1 - \sqrt{\Phi(\mathcal{D}'_j, \mathcal{I})} \right)^2 \left(\prod_{i=1}^{j-1} \Phi(\mathcal{D}'_i, \mathcal{I}) \right) \\
&\hspace{15em} (4.122)
\end{aligned}$$

$$\begin{aligned}
&\leq \gamma_{\text{decoh}}^2 \sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}'_i, \mathcal{I})} \right)^2 \\
&= \gamma_{\text{decoh}}^2 \sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}'_i, \mathcal{I})} \right)^2. \hspace{10em} (4.123)
\end{aligned}$$

Using this, triangle inequality and eq. (4.112), we get

$$\begin{aligned}
\left| Y(\mathcal{A}_{m:1}^*) - \prod_i Y(\mathcal{A}_i^*) \right| &\leq \left| \sqrt{\Phi(\mathcal{D}'_1^* \cdots \mathcal{D}'_m^* \mathcal{D}'_{m:1}^*, \mathcal{I})} - \prod_i \Phi(\mathcal{D}'_i^*, \mathcal{I}) \right| \\
&\quad + \left| \prod_i \Phi(\mathcal{D}'_i^*, \mathcal{I}) - \prod_i Y(\mathcal{A}_i^*) \right| \quad (\text{Triangle ineq.}) \\
&\leq 4\gamma_{\text{decoh}}^2 \sum_{i=1}^m \sum_{j=i+1}^m (1 - \sqrt{\Phi(\mathcal{D}'_j^*, \mathcal{I})})(1 - \sqrt{\Phi(\mathcal{D}'_i^*, \mathcal{I})}) \\
&\quad + 2\gamma_{\text{decoh}}^2 \sum_i \left(1 - \sqrt{\Phi(\mathcal{D}'_i^*, \mathcal{I})} \right)^2 \quad (\text{Equations (4.112) and (4.123)}) \\
&= 2\gamma_{\text{decoh}}^2 \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}'_i^*, \mathcal{I})} \right) \right)^2 \quad (\text{Complete the square}) \\
&= 2\gamma_{\text{decoh}}^2 \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}'_i^*, \mathcal{I})} \right) \right)^2. \quad (4.124)
\end{aligned}$$

Notice that a usage of theorem 7 allows to translate $Y(\mathcal{A}_{m:1}^*)$ into $Y(\mathcal{A}_{m:1})$:

$$\begin{aligned}
(1 - Y(\mathcal{A}_{m:1}^*))^2 &\geq Y^2(\mathcal{A}_{m:1}) - Y^2(\mathcal{A}_{m:1}^*) \quad (\text{Theorem 7}) \\
&= (Y(\mathcal{A}_{m:1}) - Y(\mathcal{A}_{m:1}^*)) (Y(\mathcal{A}_{m:1}) + Y(\mathcal{A}_{m:1}^*)) \\
&> Y(\mathcal{A}_{m:1}) - Y(\mathcal{A}_{m:1}^*). \quad (\text{Non-catastrophic condition})
\end{aligned}$$

To remove the LK approximations from $\prod_i Y(\mathcal{A}_i^*)$, we use

$$\prod_i Y(\mathcal{A}_i) - \prod_i Y(\mathcal{A}_i^*) = \sum_{j=1}^m \prod_{i=j+1}^m Y(\mathcal{A}_i) (Y(\mathcal{A}_j) - Y(\mathcal{A}_j^*)) \prod_{i=1}^{j-1} Y(\mathcal{A}_i^*) \quad (4.125)$$

$$\leq \sum_{j=1}^m (Y(\mathcal{A}_j) - Y(\mathcal{A}_j^*)) \quad (4.126)$$

$$\leq \sum_{j=1}^m (1 - Y(\mathcal{A}_j^*))^2. \quad (4.127)$$

Using the triangle inequality and eqs. (4.123), (4.124) and (4.127) and theorem 7 yields

$$\begin{aligned}
\left| Y(\mathcal{A}_{m:1}) - \prod_i Y(\mathcal{A}_i) \right| &\leq (1 - Y(\mathcal{A}_{m:1}^*))^2 + \sum_{j=1}^m (1 - Y(\mathcal{A}_j^*))^2 \\
&\quad + \gamma_{\text{decoh}}^2 \sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}'_i^*, \mathcal{I})} \right)^2 \\
&\quad + 2\gamma_{\text{decoh}}^2 \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}'_i^*, \mathcal{I})} \right) \right)^2 \quad (4.128)
\end{aligned}$$

Invoking theorem 13 allows to naturally translates between $\Phi(\mathcal{D}_i^*, \mathcal{I})$ and $Y(\mathcal{A}_i)$, which completes the proof. \square

Proof of theorem 13

Proof. Let σ_i be the singular values of the LK operator of \mathcal{A} . The first part of the proof revolves around

$$\mathbb{E}(\sigma_i^2) \leq \mathbb{E}(\sigma_i) \leq \sqrt{\mathbb{E}(\sigma_i^2)}, \quad (4.129)$$

which implies that

$$Y^2(\mathcal{A}^*) \leq (\mathbb{E}(\sigma_i))^2 \leq Y(\mathcal{A}^*) . \quad (4.130)$$

First, let's demonstrate the lower bound eq. (4.61b):

$$\begin{aligned} \max_{V \in SU(d)} \Phi(\mathcal{V} \circ \mathcal{A}, \mathcal{U}) &\geq \max_{V \in SU(d)} \Phi(\mathcal{V} \circ \mathcal{A}^*, \mathcal{U}) && \text{(Theorem 8)} \\ &= (\mathbb{E}(\sigma_i))^2 && (4.131) \\ &\geq Y^2(\mathcal{A}^*) && \text{(Equation (4.130))} \\ &\geq Y^2(\mathcal{A}) - (1 - Y^2(\mathcal{A}))^2 . && \text{(Lemma 2)} \end{aligned}$$

Demonstrating the upper bound eq. (4.61a) follows the same reasoning:

$$\begin{aligned} \max_{V \in SU(d)} \Phi(\mathcal{V} \circ \mathcal{A}, \mathcal{U}) &\leq Y(\mathcal{A}^*) \max_{V \in SU(d)} \Phi(\mathcal{V} \circ \mathcal{A}^*, \mathcal{U}) + (1 - Y(\mathcal{A}^*)) + \frac{1}{2}(1 - Y(\mathcal{A}^*))^2 \\ &&& \text{(Equation (4.109))} \\ &= Y(\mathcal{A}^*) (\mathbb{E}(\sigma_i))^2 + (1 - Y(\mathcal{A}^*)) + \frac{1}{2}(1 - Y(\mathcal{A}^*))^2 \\ &\leq Y^2(\mathcal{A}^*) + (1 - Y(\mathcal{A}^*)) + \frac{1}{2}(1 - Y(\mathcal{A}^*))^2 && \text{(Equation (4.130))} \\ &= Y(\mathcal{A}^*) + \frac{3}{2}(1 - Y(\mathcal{A}^*))^2 \\ &\leq Y(\mathcal{A}) + \frac{3}{2}(1 - Y^2(\mathcal{A}))^2 . && \text{(Lemma 2 and eq. (4.97))} \end{aligned}$$

To tighten the lower bound at line 4.131, we may use the WSE decoherence constant:

$$\begin{aligned} (\mathbb{E}(\sigma_i))^2 &= Y(\mathcal{A}^*) - \gamma_{\text{decoh}}^2 (1 - \mathbb{E}(\sigma_i))^2 && \text{(Equability)} \\ &\geq Y(\mathcal{A}^*) - \gamma_{\text{decoh}}^2 (1 - Y(\mathcal{A}^*))^2 && \text{(Equation (4.130))} \\ &\geq Y(\mathcal{A}) - (1 - Y(\mathcal{A}^*))^2 - \gamma_{\text{decoh}}^2 (1 - Y(\mathcal{A}^*))^2 && \text{(Lemma 2)} \\ &\geq Y(\mathcal{A}) - (1 + \gamma_{\text{decoh}}^2)(1 - Y^2(\mathcal{A}))^2 , && ((Y^2(\mathcal{A}) \leq Y(\mathcal{A}^*))) \end{aligned}$$

which completes the proof. \square

Proof of theorem 10

Proof. First, we derive an upper bound for $\Phi(\mathcal{V} \circ \mathcal{D}_{m:1}, \mathcal{I})$:

$$\begin{aligned} \Phi(\mathcal{V} \circ \mathcal{D}_{m:1}, \mathcal{I}) &\leq \Phi(\mathcal{V} \circ \mathcal{D}_{m:1}^*, \mathcal{I}) \\ &\quad + \frac{1}{2} \left(\sum_{i=1}^m (1 - Y(\mathcal{D}_i^*)) \right)^2 + (1 - \Phi(\mathcal{V} \circ \mathcal{D}_{m:1}^*, \mathcal{I})) \sum_{i=1}^m (1 - Y(\mathcal{D}_i^*)) . \end{aligned} \quad (\text{Theorem 8})$$

Using lemmas 3 and 5, we get (let D_i^j be the i^{th} canonical Kraus operator of \mathcal{D}_j)

$$\begin{aligned} \Phi(\mathcal{V} \circ \mathcal{D}_{m:1}^*, \mathcal{I}) &\leq \left| \frac{\text{Tr}(V |D_1^m| \cdots |D_1^1|)}{d} \right|^2 \\ &\leq \min_i \left| \frac{\text{Tr} |D_1^i|}{d} \right|^2 \quad (\text{Lemma 5}) \\ &\leq \min_i \Phi(\mathcal{D}_i, \mathcal{I}), \quad (\text{Lemma 3}) \end{aligned}$$

which yields the quasi-monotonicity statement. Now, we derive a lower bound for $\Phi(\mathcal{V} \circ \mathcal{D}_{m:1}, \mathcal{I})$:

$$\begin{aligned} 1 - \sqrt{\Phi(\mathcal{V} \circ \mathcal{D}_{m:1}, \mathcal{I})} &\leq 1 - \sqrt{\Phi(\mathcal{V} \circ \mathcal{D}_{m:1}^*, \mathcal{I})} \quad (\text{Theorem 8}) \\ &= 1 - \left| \frac{\text{Tr}(V |D_1^m| \cdots |D_1^1|)}{d} \right| \quad (4.132) \end{aligned}$$

At this point, it seems tempting to use lemma 4, but recall that V is generally not Hermitian. However, we can get by as follows

$$\begin{aligned} 1 - \sqrt{\Phi(\mathcal{V} \circ \mathcal{D}_{m:1}, \mathcal{I})} &\leq 1 - \left| \text{Re} \left\{ \frac{\text{Tr}(V |D_1^m| \cdots |D_1^1|)}{d} \right\} \right| \\ &= 1 - \left| \frac{\text{Tr}(\text{Re}(V) |D_1^m| \cdots |D_1^1|)}{d} \right|, \quad (4.133) \end{aligned}$$

where $\text{Re}(V) := (V + V^\dagger)/2$ is Hermitian, which allows us to use lemma 4:

$$1 - \sqrt{\Phi(\mathcal{V} \circ \mathcal{D}_{m:1}, \mathcal{I})} \leq \left(1 - \left| \frac{\text{Tr} \text{Re}(V)}{d} \right| \right) + \sum_{i=1}^m \left(1 - \left| \frac{\text{Tr} |D_1^i|}{d} \right| \right) \quad (\text{Lemma 4})$$

WOLOG, we pick the global phase of V such that $\sqrt{\Phi(\mathcal{V}, \mathcal{I})} = \text{Tr} V / d \in \mathbb{R}_+$. From there we get

$$1 - \sqrt{\Phi(\mathcal{V} \circ \mathcal{D}_{m:1}, \mathcal{I})} \leq \left(1 - \sqrt{\Phi(\mathcal{V}, \mathcal{I})} \right) + \sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right). \quad (4.134)$$

To remove the square roots and the star, let's use $1 - x/2 - x^2/2 \leq \sqrt{1-x} \leq 1 - x/2$ for $x \in [0, 1]$ and lemma 3:

$$\begin{aligned}
1 - \Phi(\mathcal{V} \circ \mathcal{D}_{m:1}, \mathcal{I}) &\leq (1 - \Phi(\mathcal{V}, \mathcal{I})) + \sum_{i=1}^m (1 - \Phi(\mathcal{D}_i, \mathcal{I})) \\
&\quad + (1 - \Phi(\mathcal{V}, \mathcal{I}))^2 + \sum_{i=1}^m (1 - \Phi(\mathcal{D}_i^*, \mathcal{I}))^2 + \sum_{i=1}^m (1 - \Phi(\mathcal{D}_i, \mathcal{I}))(1 - Y^2(\mathcal{D}_i))
\end{aligned} \tag{4.135}$$

which corresponds to the quasi-subadditivity property. \square

Proof of theorem 14

Proof. Our goal is to bound

$$\left| \frac{\text{Tr}(V|D_1^m| \cdots |D_1^1|)}{d} \right|^2 = \left| \text{Re} \left\{ \frac{\text{Tr}(V|D_1^m| \cdots |D_1^1|)}{d} \right\} \right|^2 + \left| \text{Im} \left\{ \frac{\text{Tr}(V|D_1^m| \cdots |D_1^1|)}{d} \right\} \right|^2. \tag{4.136}$$

Proving theorem 14 is very similar to proving theorem 12, but the appended unitary V requires some extra care. Let's first bound the amplitude of the imaginary term. WOLOG, we pick the global phase of V such that $\sqrt{\Phi(\mathcal{V}, \mathcal{I})} = \text{Tr} V/d \in \mathbb{R}_+$.

$$\begin{aligned}
\left| \text{Im} \left\{ \frac{\text{Tr}(V|D_1^m| \cdots |D_1^1|)}{d} \right\} \right| &= \left| \text{Im} \left\{ \text{Tr} \left[\frac{(V - \text{Tr}(V)/d \mathbb{I})(|D_1^m| \cdots |D_1^1| - \text{Tr}(|D_1^m| \cdots |D_1^1|)/d \mathbb{I})}{d} \right] \right\} \right| \\
&\quad \text{(Adding real terms.)} \\
&\leq \left| \text{Tr} \left[\frac{(V - \text{Tr}(V)/d \mathbb{I})(|D_1^m| \cdots |D_1^1| - \text{Tr}(|D_1^m| \cdots |D_1^1|)/d \mathbb{I})}{d} \right] \right| \\
&\leq \frac{\|V - \text{Tr}(V)/d \mathbb{I}\|_2 \| |D_1^m| \cdots |D_1^1| - \text{Tr}(|D_1^m| \cdots |D_1^1|)/d \mathbb{I} \|_2}{\sqrt{d}} \\
&\quad \text{(Cauchy-Schwarz ineq.)} \\
&= \sqrt{1 - \Phi(\mathcal{V}, \mathcal{I})} \sqrt{Y(\mathcal{D}_{m:1}^*) - \Phi(\mathcal{D}_{m:1}^*, \mathcal{I})}. \tag{4.137}
\end{aligned}$$

We know from theorem 11 that $Y(\mathcal{D}_{m:1}^*) \approx \prod_i Y(\mathcal{D}_i^*)$. We also know from theorem 13 that $Y(\mathcal{D}_i^*) \approx \Phi(\mathcal{D}_i^*, \mathcal{I})$. From theorem 12 we know that $\prod_i \Phi(\mathcal{D}_i^*, \mathcal{I}) \approx \Phi(\mathcal{D}_{m:1}^*, \mathcal{I})$. By combining this information, we have that $Y(\mathcal{D}_{m:1}^*) \approx \Phi(\mathcal{D}_{m:1}^*, \mathcal{I})$. More precisely, by using

eqs. (4.114), (4.123) and (4.124), we get

$$\begin{aligned}
\left| \operatorname{Im} \left\{ \frac{\operatorname{Tr} (V|D_1^m| \cdots |D_1^1|)}{d} \right\} \right|^2 &\leq (1 - \Phi(\mathcal{V}, \mathcal{I})) \left[\gamma_{\text{decoh}}^2 \prod_{i=1}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \right)^2 \right. \\
&+ \frac{\gamma_{\text{decoh}}^4}{4} \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \right)^4 + \gamma_{\text{decoh}}^2 \sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right)^2 \\
&\left. + 2\gamma_{\text{decoh}}^2 \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \right)^2 \right], \tag{4.138}
\end{aligned}$$

meaning that the imaginary term is absolutely insignificant. To bound the real part of the trace, we mimic most of the proof technique used to prove theorem 12. Let's express the LK operators as $|D_1^i| = \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \mathbb{I}_d + \Delta_i$ and $V = \sqrt{\Phi(\mathcal{V}, \mathcal{I})} \mathbb{I}_d + \Delta_{m+1}$, and apply a first telescopic expansion:

$$\begin{aligned}
\operatorname{Re} \left\{ \operatorname{Tr} \left[\frac{V|D_1^m| \cdots |D_1^1|}{d} \right] \right\} &= \sqrt{\Phi(\mathcal{V}, \mathcal{I})} \prod_{i=1}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \\
&+ \sum_{j=1}^m \operatorname{Re} \left\{ \operatorname{Tr} \left[\frac{V|D_1^m| \cdots |D_1^{j+1}| \Delta_j}{d} \right] \right\} \prod_{i=1}^{j-1} \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \\
&\tag{Telescopic sum}
\end{aligned}$$

By applying the expansion again, and use $\operatorname{Tr} \Delta_i = 0$, we get:

$$\begin{aligned}
\operatorname{Re} \left\{ \operatorname{Tr} \left[\frac{V|D_1^m| \cdots |D_1^1|}{d} \right] \right\} &= \sqrt{\Phi(\mathcal{V}, \mathcal{I})} \prod_{i=1}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \\
&+ \sum_{j=1}^m \sum_{k=j+1}^m \operatorname{Re} \left\{ \frac{\operatorname{Tr} (V|D_1^m| \cdots |D_1^{k+1}| \Delta_k \Delta_j)}{d} \right\} \prod_{\substack{i=1 \\ i \neq j}}^{k-1} \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \\
&+ \sum_{j=1}^m \operatorname{Re} \left\{ \frac{\operatorname{Tr} (\Delta_{m+1} \Delta_j)}{d} \right\} \prod_{\substack{i=1 \\ i \neq j}}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})}. \\
&\tag{Telescopic sum, again}
\end{aligned}$$

After a simple application of the triangle inequality, we get

$$\begin{aligned}
\left| \operatorname{Re} \left\{ \operatorname{Tr} \left[\frac{V|D_1^m| \cdots |D_1^1|}{d} \right] \right\} - \sqrt{\Phi(\mathcal{V}, \mathcal{I})} \prod_{i=1}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right| &\leq \left| \sum_{j=1}^m \operatorname{Re} \left\{ \frac{\operatorname{Tr} (\Delta_{m+1} \Delta_j)}{d} \right\} \prod_{\substack{i=1 \\ i \neq j}}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right| \\
&+ \left| \sum_{j=1}^m \sum_{k=j+1}^m \operatorname{Re} \left\{ \frac{\operatorname{Tr} (V|D_1^m| \cdots |D_1^{k+1}| \Delta_k \Delta_j)}{d} \right\} \prod_{\substack{i=1 \\ i \neq j}}^{k-1} \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right| \\
&\tag{4.139}
\end{aligned}$$

The second term on the RHS is upper-bounded by the exact same technique as in theorem 12 (see the derivation of eq. (4.113)):

$$\left| \sum_{j=1}^m \sum_{k=j+1}^m \operatorname{Re} \left\{ \frac{\operatorname{Tr} \left(V |D_1^m| \cdots |D_1^{k+1}| \Delta_k \Delta_j \right)}{d} \right\} \prod_{\substack{i=1 \\ i \neq j}}^{k-1} \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right| \leq \frac{\gamma_{\text{decoh}}^2}{2} \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \right)^2. \quad (4.140)$$

The first term on the RHS of eq. (4.139) is bounded as follows:

$$\begin{aligned} \left| \sum_{j=1}^m \operatorname{Re} \left\{ \frac{\operatorname{Tr}(\Delta_{m+1} \Delta_j)}{d} \right\} \prod_{\substack{i=1 \\ i \neq j}}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right| &\leq \sum_{j=1}^m \left| \operatorname{Re} \left\{ \frac{\operatorname{Tr}(\Delta_{m+1} \Delta_j)}{d} \right\} \prod_{\substack{i=1 \\ i \neq j}}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right| \\ &\quad \text{(Triangle ineq.)} \\ &\leq \sum_{i=1}^m \left| \operatorname{Re} \left\{ \frac{\operatorname{Tr}(\Delta_{m+1} \Delta_i)}{d} \right\} \right| \quad (\sqrt{\Phi(\mathcal{A}_i^*, \mathcal{I})} \leq 1) \\ &\leq \sum_{i=1}^m \left| \operatorname{Tr} \left[\frac{\Delta_i (\Delta_{m+1} + \Delta_{m+1}^\dagger) / 2}{d} \right] \right| \\ &\quad (\Delta_i = \Delta_i^\dagger \text{ for } i \neq m+1.) \\ &\leq \sum_{i=1}^m \frac{\|\Delta_i\|_2}{\sqrt{d}} \frac{\|\operatorname{Re}(V) - \operatorname{Tr}(V)\mathbb{I}\|_2}{\sqrt{d}} \\ &\quad \text{(Cauchy-Schwarz ineq.)} \end{aligned} \quad (4.141)$$

This is where definition 7 (equability) is put to use. Recall that for $i \neq m+1$ we have $\frac{\|\Delta_i\|_2}{\sqrt{d}} = \gamma_{\text{decoh}}(\mathcal{D}_i) \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right)$ and that the WSE coherence constant is implicitly defined by $\frac{\|\operatorname{Re}(V) - \operatorname{Tr}(V)\mathbb{I}\|_2}{\sqrt{d}} = \gamma_{\text{coh}}(1 - \sqrt{\Phi(\mathcal{V}, \mathcal{I})})$, which means that

$$\left| \sum_{j=1}^m \operatorname{Re} \left\{ \frac{\operatorname{Tr}(\Delta_{m+1} \Delta_j)}{d} \right\} \prod_{\substack{i=1 \\ i \neq j}}^m \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right| \leq \gamma_{\text{decoh}} \gamma_{\text{coh}} \left(1 - \sqrt{\Phi(\mathcal{V}, \mathcal{I})} \right) \sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \quad \text{(Definition 7)}$$

Using $|a^2 - b^2| \leq |a - b||a + b|$, and reuniting the pieces, we get

$$\begin{aligned} \left| \Phi(\mathcal{V} \circ \mathcal{D}_{m:1}^*, \mathcal{I}) - \Phi(\mathcal{V}, \mathcal{I}) \prod_{i=1}^m \Phi(\mathcal{D}_i^*, \mathcal{I}) \right| &\leq 2\gamma_{\text{decoh}} \gamma_{\text{coh}} \left(1 - \sqrt{\Phi(\mathcal{V}, \mathcal{I})} \right) \sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \\ &\quad + \gamma_{\text{decoh}}^2 \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \right)^2 \\ &\quad + \left| \operatorname{Im} \left\{ \frac{\operatorname{Tr}(V |D_1^m| \cdots |D_1^1|)}{d} \right\} \right|^2. \end{aligned} \quad (4.142)$$

A straightforward application of eq. (4.115) and theorem 8 on the LHS (to get rid of the \star) yields eq. (4.64). \square

Proof of theorem 15

Proof. Let's factor the decoherent channels into their (left) polar decomposition $\mathcal{A}_i = \mathcal{D}_i \circ \mathcal{V}_i$. By relabeling $(\mathcal{V}_{i:1})^{-1} \circ \mathcal{D}_i \circ \mathcal{V}_{i:1} = \mathcal{D}'_i$ (notice that \mathcal{D}'_i are decoherent), we have

$$\mathcal{A}_{m:1} = \mathcal{V}_{m:1} \circ \mathcal{D}'_{m:1}. \quad (4.143)$$

First, let's find a lower bound on $\max_{W \in SU(d)} \Phi(\mathcal{W} \circ \mathcal{A}_{m:1}, \mathcal{U}_{m:1})$. A way to do this is to pick a wisely chosen argument for \mathcal{W} . Let's pick $\mathcal{W} = \mathcal{U}_{m:1} \circ (\mathcal{V}_{m:1})^\dagger$:

$$\max_{W \in SU(d)} \Phi(\mathcal{W} \circ \mathcal{A}_{m:1}, \mathcal{U}_{m:1}) \geq \Phi(\mathcal{U}_{m:1} \circ (\mathcal{V}_{m:1})^\dagger \circ \mathcal{A}_{m:1}, \mathcal{U}_{m:1}) \quad (4.144)$$

$$\geq \Phi(\mathcal{U}_{m:1} \circ (\mathcal{V}_{m:1})^\dagger \circ \mathcal{A}_{m:1}^*, \mathcal{U}_{m:1}) \quad (\text{Theorem 7})$$

$$= \Phi(\mathcal{D}'_{m:1}^*, \mathcal{I}) \quad (4.145)$$

$$\begin{aligned} &\geq \prod_{i=1}^m \Phi(\mathcal{D}'_i^*, \mathcal{I}) + \gamma_{\text{decoh}}^2 \prod_{i=1}^m \sqrt{\Phi(\mathcal{D}'_i^*, \mathcal{I})} \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}'_i^*, \mathcal{I})} \right) \right)^2 \\ &+ \frac{\gamma_{\text{decoh}}^4}{4} \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}'_i^*, \mathcal{I})} \right) \right)^4 \quad (\text{Equation (4.114)}) \end{aligned}$$

To bound $\prod_{i=1}^m \Phi(\mathcal{D}'_i^*, \mathcal{I})$, we express it as a sum of three terms:

$$\prod_{i=1}^m Y(\mathcal{A}_i) + \left(\prod_{i=1}^m Y(\mathcal{A}_i^*) - \prod_{i=1}^m Y(\mathcal{A}_i) \right) + \left(\prod_{i=1}^m \Phi(\mathcal{D}'_i^*, \mathcal{I}) - \prod_{i=1}^m Y(\mathcal{A}_i^*) \right). \quad (4.146)$$

To bound the second term, we used eq. (4.127). The third term of eq. (4.146) is bounded through eq. (4.123). Reuniting the pieces together, we get

$$\begin{aligned} \max_{W \in SU(d)} \Phi(\mathcal{W} \circ \mathcal{A}_{m:1}, \mathcal{U}_{m:1}) &\geq \prod_{i=1}^m Y(\mathcal{A}_i) - \gamma_{\text{decoh}}^2 \sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}'_i^*, \mathcal{I})} \right)^2 - \sum_{i=1}^m (1 - Y(\mathcal{A}_i^*))^2 \\ &- \gamma_{\text{decoh}}^2 \prod_{i=1}^m \sqrt{\Phi(\mathcal{D}'_i^*, \mathcal{I})} \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}'_i^*, \mathcal{I})} \right) \right)^2 \\ &- \frac{\gamma_{\text{decoh}}^4}{4} \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}'_i^*, \mathcal{I})} \right) \right)^4. \quad (4.147) \end{aligned}$$

With regards to the upper bound, we can first use theorem 8 to get

$$\begin{aligned} \max_{W \in SU(d)} \Phi(\mathcal{W} \circ \mathcal{A}_{m:1}, \mathcal{U}_{m:1}) &\leq \max_{W \in SU(d)} \left[\Phi(\mathcal{W} \circ \mathcal{A}_{m:1}^*, \mathcal{U}_{m:1}) + \frac{1}{2} \left(\sum_{i=1}^m (1 - Y(\mathcal{A}_i^*)) \right)^2 \right. \\ &\quad \left. + (1 - \Phi(\mathcal{W} \circ \mathcal{A}_{m:1}^*, \mathcal{U}_{m:1})) \sum_{i=1}^m (1 - Y(\mathcal{A}_i^*)) \right]. \end{aligned} \quad (4.148)$$

By using the flavored Von-Neumann inequality (lemma 5), followed by eq. (4.130), we get

$$\begin{aligned} \max_{W \in SU(d)} \Phi(\mathcal{W} \circ \mathcal{A}_{m:1}^*, \mathcal{U}_{m:1}) &\leq \left| \frac{\text{Tr} |D'_{m:1}|}{d} \right|^2 && \text{(Lemma 5)} \\ &\leq Y(\mathcal{D}'_{m:1}) && \text{(Equation (4.130))} \\ &\leq \prod_i Y(\mathcal{A}_i^*) + 2\gamma_{\text{decoh}}^2 \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \right)^2, && \text{(Equation (4.124))} \\ &\leq \prod_i Y(\mathcal{A}_i) + \sum_{i=1}^m (1 - Y(\mathcal{A}_i^*))^2 \\ &\quad + 2\gamma_{\text{decoh}}^2 \left(\sum_{i=1}^m \left(1 - \sqrt{\Phi(\mathcal{D}_i^*, \mathcal{I})} \right) \right)^2. && \text{(Equation (4.127))} \end{aligned}$$

Substituting this on the RHS of eq. (4.148) completes the proof. \square

4.10 Afterword

Based on the notion of equability, it is possible to reason more intuitively about the dimensional factors that appear when upper-bounding the diamond and spectral norm of a difference of channels.

4.10.1 A bound on the spectral radius of channel deviations

Without invoking any physical intuition, upper-bounding the spectral norm generally involves unruly dimensional factors. It turns out that these factors are essentially caused by pathological, usually unrealistic channels. In this section, I present refined general bounds that take into account pathological scenarios, but yield a simple expression in the typical case. To make sense of the results, an additional error parameter needs to be introduced. Indeed, not so surprisingly, ruling out extremal dephasers, which is equivalent to imposing the equability condition, still leaves some pathological scenarios which show up in bounding the spectral norm of channel deviations. Let A_j be canonical Kraus operators of a channel \mathcal{A} (A_1 being the LK operator). In general,

$$\sigma_{\max} \left(\sum_{j \neq 1} A_j A_j^\dagger \right) \leq \text{Tr} \sum_{j \neq 1} A_j A_j^\dagger = d(1 - \text{Tr} A_1 A_1^\dagger / d), \quad (4.149)$$

which is of order dr_{decoh} (notice the annoying dimensional factor). This inequality is saturated by channels with a Kraus decomposition of the form

$$\begin{aligned} A_1 &= |0\rangle\langle 0| + \sum_{i=1}^{d-1} \sqrt{1-\epsilon_i} |i\rangle\langle i|, \\ A_{i+2} &= \sqrt{\epsilon_i} |0\rangle\langle i| \text{ for } i \in \{1, 2, \dots, d-1\}, \end{aligned} \quad (4.150)$$

where ϵ_i are all on the same order. Physically speaking, all the state transitions – with comparable probabilities – point towards the same state. Of course, such error channel, which is labeled here as an *extremal relaxation*, is rather unrealistic for large dimensions: typically, certain transitions are much less likely than others. For instance, consider a n -qubit system for which the transition of a qubit from its excited state $|1\rangle$ to its ground state $|0\rangle$ happens with probability ϵ , by which I mean that the corresponding channel \mathcal{A} looks like a n -fold tensor product of qubit channels \mathcal{B} , individually defined as standard amplitude damping:

$$\begin{aligned} B_1 &= |0\rangle\langle 0| + \sqrt{1-\epsilon} |1\rangle\langle 1|, \\ B_2 &= \sqrt{\epsilon} |0\rangle\langle 1|. \end{aligned} \quad (4.151)$$

In this example, the transition from a computational state $|s\rangle$ to a computational state $|s'\rangle$ where s and s' have a Hamming distance of 1 happens with a probability on the order of ϵ . However, for a Hamming distance of 2, the probability drops to the order ϵ^2 . A quick calculation (let $(1 \pm \epsilon)^n \approx \pm n\epsilon$ for the sake of simplicity in this example) suffices to show that

$$\sigma_{\max} \left(\sum_{j \neq 1} A_j A_j^\dagger \right) = (1 + \epsilon)^n - 1 \approx n\epsilon, \quad (4.152)$$

and

$$\delta\Phi_{\text{decoh}}(\mathcal{A}) = 1 - \left(\frac{1 + \sqrt{1-\epsilon}}{2} \right)^{2n} \approx \frac{n\epsilon}{2}, \quad (4.153)$$

which yields

$$\sigma_{\max} \left(\sum_{j \neq 1} A_j A_j^\dagger \right) \approx 2\delta\Phi_{\text{decoh}}(\mathcal{A}), \quad (4.154)$$

which is typically much smaller than the general bound $2^n \delta\Phi_{\text{decoh}}(\mathcal{A})$.

This example motivates the following definition:

Definition 8: Relaxation bounding constant

Consider a non-catastrophic error channel $\mathcal{A} = \mathcal{U} \circ \mathcal{D}$ with canonical Kraus operators A_j (A_1 is the LK operator). The relaxation bounding constant d_{relax} is defined as

$$\sigma_{\max} \left(\sum_{j \neq 1} A_j A_j^\dagger \right) = d_{\text{relax}} \delta\Phi_{\text{decoh}}(\mathcal{A}). \quad (4.155)$$

It follows from our example that for equable channels, d_{relax} scales as the largest number of orthogonal states which all transition to a single state with comparable rates (and for which the rates are comparable to $\delta\Phi_{\text{decoh}}$).

With these new constants at hand, the following result is ready to be expressed:

Theorem 16: Bounding the spectral norm of channel deviations

Consider a non-catastrophic error channel $\mathcal{A} = \mathcal{D} \circ \mathcal{V}$ with target $\mathcal{U}_{\text{target}}$ and LK operator A_1 . Let $\delta\Phi_{\text{coh}} = \delta\Phi(\mathcal{V}, \mathcal{U}_{\text{target}})$, $\delta\Phi_{\text{decoh}} = \delta\Phi(\mathcal{D}, \mathcal{I})$. Then the spectral norm of $\mathcal{A} - \mathcal{U}_{\text{target}}$ is bounded as follows

$$\begin{aligned} \max_{M \in M_d(\mathbb{C})} \frac{\|(\mathcal{A} - \mathcal{U}_{\text{target}})[M]\|_2}{\|M\|_2} &\leq 2\sqrt{\Gamma_{\text{coh}}} \sqrt{\delta\Phi_{\text{coh}}} \\ &\quad + \sqrt{\Gamma_{\text{decoh}}} (\sqrt{\Gamma_{\text{decoh}}} + \sqrt{d_{\text{relax}}}) \delta\Phi_{\text{decoh}} + O(\delta\Phi_{\text{coh}}). \end{aligned} \quad (4.156)$$

Proof. Let A_i be canonical Kraus operators associated with \mathcal{A} and let A_1 be the LK operator. Let $U_{\text{target}} \in SU(d)$ be the targeted Kraus operator. The spectral norm of $\mathcal{A} - \mathcal{U}_{\text{target}}$ can be re-expressed as

$$\begin{aligned} \max_{M \in M_d(\mathbb{C})} \frac{\|(\mathcal{A} - \mathcal{U}_{\text{target}})[M]\|_2}{\|M\|_2} &= \max_{v \in \mathbb{C}^{d^2}} \frac{\|(\mathcal{A} - \mathcal{U}_{\text{target}})v\|_2}{\|v\|_2} \\ &= \max_{v \in \mathbb{C}^{d^2}} \frac{\|(A_1^* \otimes A_1 - U_{\text{target}}^* \otimes U_{\text{target}} + \sum_{i>1} A_i^* \otimes A_i)v\|_2}{\|v\|_2} \\ &= \max_{v, w \in \mathbb{C}^{d^2}} \frac{\left| \langle wv^\dagger, A_1^* \otimes A_1 - U_{\text{target}}^* \otimes U_{\text{target}} + \sum_{i>1} A_i^* \otimes A_i \rangle \right|}{\|v\|_2 \|w\|_2}, \end{aligned} \quad (4.157)$$

where on the last line I used $\|v\|_2 = \max_w \langle w, v \rangle / \|w\|_2$. Using the triangle inequality, it follows that

$$\begin{aligned} \max_{M \in M_d(\mathbb{C})} \frac{\|(\mathcal{A} - \mathcal{U}_{\text{target}})[M]\|_2}{\|M\|_2} &\leq \max_{v, w \in \mathbb{C}^{d^2}} \frac{\left| \langle wv^\dagger, A_1^* \otimes A_1 - U_{\text{target}}^* \otimes U_{\text{target}} \rangle \right|}{\|v\|_2 \|w\|_2} \\ &\quad + \max_{v, w \in \mathbb{C}^{d^2}} \sum_{i>1} \frac{\left| \langle wv^\dagger, A_i^* \otimes A_i \rangle \right|}{\|v\|_2 \|w\|_2}. \end{aligned} \quad (4.158)$$

The second term is bounded as follows:

$$\begin{aligned}
\sum_{i>1} \frac{|\langle wv^\dagger, A_i^* \otimes A_i \rangle|}{\|v\|_2 \|w\|_2} &\leq \sum_{i>1} \frac{\|\mathbb{I} \otimes A_i v\|_2 \|A_i^T \otimes \mathbb{I} w\|_2}{\|v\|_2 \|w\|_2} && \text{(Cauchy-Schwarz)} \\
&\leq \sqrt{\sum_{i>1} \frac{\|\mathbb{I} \otimes A_i v\|_2^2}{\|v\|_2^2}} \sqrt{\sum_{j>1} \frac{\|A_j^T \otimes \mathbb{I} w\|_2^2}{\|w\|_2^2}} && \text{(Cauchy-Schwarz)} \\
&\leq \sqrt{\sigma_{\max} \left(\sum_{i>1} A_i^\dagger A_i \right)} \sqrt{\sigma_{\max} \left(\sum_{j>1} A_j A_j^\dagger \right)} \\
&= \sqrt{\sigma_{\max} (\mathbb{I} - A_1^\dagger A_1)} \sqrt{d_{\text{relax}} \delta \Phi_{\text{decoh}}(\mathcal{A})} && \text{(Definition 8)}
\end{aligned}$$

The other factor is bounded as follows

$$\begin{aligned}
\sigma_{\max} (\mathbb{I} - A_1^\dagger A_1) &= 1 - \sigma_{\min}^2(A_1) \leq 2(1 - \sigma_{\min}(A_1)) && \text{(Diff. of squares)} \\
&\leq 2\Gamma_{\text{decoh}} \mathbb{E}[1 - \sigma_i(A_1)] && \text{(Definition 7)} \\
&= 2\Gamma_{\text{decoh}} (1 - \sqrt{\Phi_{\text{decoh}}(\mathcal{A}^*)}) && (4.159) \\
&= \Gamma_{\text{decoh}} \delta \Phi_{\text{decoh}}(\mathcal{A}) + O(\delta^2 \Phi_{\text{decoh}}(\mathcal{A})). && (4.160)
\end{aligned}$$

The first term on the RHS of eq. (4.158) is bounded by the spectral norm of the LK approximated error channel minus the identity:

$$\begin{aligned}
\frac{|\langle wv^\dagger, A_1^* \otimes A_1 - U_{\text{target}}^* \otimes U_{\text{target}} \rangle|}{\|v\|_2 \|w\|_2} &\leq \frac{\|(U_{\text{target}}^T A_1^* \otimes U_{\text{target}}^\dagger A_1 - \mathbb{I} \otimes \mathbb{I})v\|_2}{\|v\|_2} \\
&\leq \sigma_{\max} (U_{\text{target}}^T A_1^* \otimes U_{\text{target}}^\dagger A_1 - \mathbb{I} \otimes \mathbb{I}). && \text{(Cauchy-Schwarz)} \\
&&& (4.161)
\end{aligned}$$

Now, let $U_{\text{target}}^\dagger A_1 = V|A_1\rangle$, $|A_1\rangle = \mathbb{I} - \Delta_{\text{decoh}}$, and $V = \mathbb{I} - \Delta_{\text{coh}}$ (WOLOG, $\text{Tr } V \in \mathbb{R}_+$). By substituting this in eq. (4.161), it follows that

$$\begin{aligned}
\frac{|\langle wv^\dagger, A_1^* \otimes A_1 - U_{\text{target}}^* \otimes U_{\text{target}} \rangle|}{\|v\|_2 \|w\|_2} &\leq 2\sigma_{\max} (\Delta_{\text{coh}} + \Delta_{\text{decoh}} - \Delta_{\text{coh}} \Delta_{\text{decoh}}) \\
&\quad + \sigma_{\max}^2 (\Delta_{\text{coh}} + \Delta_{\text{decoh}} - \Delta_{\text{coh}} \Delta_{\text{decoh}}) && \text{(Trian. ineq.)} \\
&\leq 2\sigma_{\max} (\Delta_{\text{coh}}) + 2\sigma_{\max} (\Delta_{\text{decoh}}) + \sigma_{\max}^2 (\Delta_{\text{coh}}) \\
&\quad + 4\sigma_{\max} (\Delta_{\text{coh}}) \sigma_{\max} (\Delta_{\text{decoh}}) + \sigma_{\max}^2 (\Delta_{\text{decoh}}) \\
&\quad + 2\sigma_{\max}^2 (\Delta_{\text{coh}}) \sigma_{\max} (\Delta_{\text{decoh}}) + 2\sigma_{\max} (\Delta_{\text{coh}}) \sigma_{\max}^2 (\Delta_{\text{decoh}}) \\
&\quad + \sigma_{\max}^2 (\Delta_{\text{coh}}) \sigma_{\max}^2 (\Delta_{\text{decoh}}), && (4.162)
\end{aligned}$$

where on the last line, the triangle inequality and sub-multiplicativity of the spectral norm were used. From definition 7 (equable channels), it immediately follows that

$$\begin{aligned}\sigma_{\max}(\Delta_{\text{decoh}}) &= 1 - \sigma_{\min}(A_1) \\ &= \Gamma_{\text{decoh}} \mathbb{E}[1 - \sigma_i(A_1)]\end{aligned}\tag{Definition 7}$$

$$= \Gamma_{\text{decoh}} (1 - \sqrt{\Phi_{\text{decoh}}(\mathcal{A}^*)})\tag{4.163}$$

$$= \frac{\Gamma_{\text{decoh}}}{2} \delta\Phi_{\text{decoh}}(\mathcal{A}) + O(\delta^2\Phi_{\text{decoh}}(\mathcal{A})),\tag{4.164}$$

Similarly (let λ_j be the eigenvalues of V),

$$\begin{aligned}\sigma_{\max}(\Delta_{\text{coh}}) &= \max_j |1 - \lambda_j| \\ &= \max_j \sqrt{(1 - \text{Re}(\lambda_j))^2 + \text{Im}^2(\lambda_j)} \\ &= \max_j \sqrt{2} \sqrt{1 - \text{Re}(\lambda_j)} \\ &= \sqrt{2\Gamma_{\text{coh}}} \sqrt{1 - \mathbb{E}[\lambda_i]}\end{aligned}\tag{Definition 7}$$

$$= \sqrt{2\Gamma_{\text{coh}}} \sqrt{1 - \sqrt{\Phi_{\text{coh}}}}$$

$$= \sqrt{\Gamma_{\text{coh}}} \sqrt{\delta\Phi_{\text{coh}}} + O(\delta\Phi_{\text{coh}}).\tag{4.165}$$

Substituting eqs. (4.164) and (4.165) in eq. (4.162) yields

$$\frac{\left| \left\langle wv^\dagger, A_1^* \otimes A_1 - U_{\text{target}}^* \otimes U_{\text{target}} \right\rangle \right|}{\|v\|_2 \|w\|_2} \leq 2\sqrt{\Gamma_{\text{coh}}} \sqrt{\delta\Phi_{\text{coh}}} + \Gamma_{\text{decoh}} \delta\Phi_{\text{decoh}} + O(\delta\Phi_{\text{coh}}).\tag{4.166}$$

This completes the proof. \square

4.10.2 A bound on the diamond distance

A similar bounding technique can be applied to the diamond distance between channels \mathcal{A} and \mathcal{B} , defined as

$$\epsilon_{\diamond}(\mathcal{A}, \mathcal{B}) := \max_{\rho \in \mathcal{Q}_{d^2}} \frac{1}{2} \|(\mathcal{A} - \mathcal{B}) \otimes \mathcal{I}_d[\rho]\|_1.\tag{4.167}$$

Theorem 17: The diamond distance closely connects with the coherent infidelity

Consider a non-catastrophic error channel $\mathcal{A} = \mathcal{D} \circ \mathcal{V}$ with target $\mathcal{U}_{\text{target}}$ and LK operator A_1 . Let $\delta\Phi = \delta\Phi(\mathcal{A}, \mathcal{U}_{\text{target}})$, $\delta\Phi_{\text{coh}} = \delta\Phi(\mathcal{V}, \mathcal{U}_{\text{target}})$, $\delta\Phi_{\text{decoh}} = \delta\Phi(\mathcal{D}, \mathcal{I})$, $Y(\mathcal{A}) = Y$. Then the diamond distance between \mathcal{A} and $\mathcal{U}_{\text{target}}$ is bounded as follows

$$\sqrt{1 + Y^2 - 2\Phi} = \sqrt{\delta\Phi_{\text{coh}} + O(\delta^2\Phi)} \leq \epsilon_{\diamond}(\mathcal{A}, \mathcal{U}_{\text{target}}) \quad (4.168)$$

$$\epsilon_{\diamond}(\mathcal{A}, \mathcal{U}_{\text{target}}) \leq \sqrt{\Gamma_{\text{coh}}} \sqrt{\delta\Phi_{\text{coh}}} + 2\Gamma_{\text{coh}}\delta\Phi_{\text{coh}} + \Gamma_{\text{decoh}}\delta\Phi_{\text{decoh}} + O(\delta^2\Phi). \quad (4.169)$$

In the decoherence-limited regime, where $\delta\Phi_{\text{coh}} = \beta\delta\Phi_{\text{decoh}}^2$, the diamond distance can be bounded as

$$\delta\Phi \leq \epsilon_{\diamond}(\mathcal{A}, \mathcal{U}_{\text{target}}) \leq (\sqrt{\beta\Gamma_{\text{coh}}} + \Gamma_{\text{decoh}})\delta\Phi + O(\delta^2\Phi). \quad (4.170)$$

When errors are equable, the coherent infidelity essentially dictates the diamond distance:

$$\sqrt{\delta\Phi_{\text{coh}}} \lesssim \epsilon_{\diamond}(\mathcal{A}, \mathcal{U}_{\text{target}}) \lesssim \sqrt{\Gamma_{\text{coh}}} \sqrt{\delta\Phi_{\text{coh}}}. \quad (4.171)$$

The factor $\sqrt{\Gamma_{\text{coh}}}$ is only substantially large for coherently extremal channels, and stems from the fact that the diamond distance involves a maximization over inputs, while the average infidelity involves an average over inputs.

The diamond distance can only be substantially smaller than $O(\sqrt{\delta\Phi})$ in decoherence-limited scenarios, in which case it becomes of order $O(\delta\Phi)$. The leeway (see [WF14; Wal15]) in the relation between the diamond distance and the infidelity is mostly taken into account by purely unitary processes.

Proof. The lower bound comes directly from [WF14; Wal15; SWS15], up to a straightforward manipulation of dimensional factors.

The proof for the upper bound borrows many techniques from [Wal15], but also incorporates techniques from the previous section to trim down the dimensional factors.

From [Wat12] (Theorem 6), it follows that

$$\begin{aligned} 2\epsilon_{\diamond}(\mathcal{A}, \mathcal{U}_{\text{target}}) &= \sup_{\sigma, \rho \in \mathcal{Q}_d} \|(\sqrt{\rho} \otimes \mathbb{I}_d) \text{Choi}(\mathcal{A} - \mathcal{U}_{\text{target}})(\sqrt{\sigma} \otimes \mathbb{I}_d)\|_1 \\ &= \sup_{\sigma, \rho \in \mathcal{Q}_d} \|(\sqrt{\rho} \otimes \mathbb{I}_d) \left(\sum_k \text{col}(A_k) \text{col}^\dagger(A_k) - \text{col}(U_{\text{target}}) \text{col}^\dagger(U_{\text{target}}) \right) (\sqrt{\sigma} \otimes \mathbb{I}_d)\|_1 \\ &\hspace{15em} \text{(Equation (4.2))} \\ &= \sup_{\sigma, \rho \in \mathcal{Q}_d} \left\| \sum_k \text{col}(A_k \sqrt{\rho}^T) \text{col}^\dagger(A_k \sqrt{\sigma}^T) - \text{col}(U_{\text{target}} \sqrt{\rho}^T) \text{col}^\dagger(U_{\text{target}} \sqrt{\sigma}^T) \right\|_1 \\ &\hspace{15em} \text{(Equation (2.10))} \\ &= \sup_{\sigma, \rho \in \mathcal{Q}_d} \left\| \sum_k \text{col}(A_k \sqrt{\rho}) \text{col}^\dagger(A_k \sqrt{\sigma}) - \text{col}(U_{\text{target}} \sqrt{\rho}) \text{col}^\dagger(U_{\text{target}} \sqrt{\sigma}) \right\|_1 \\ &\hspace{15em} (\rho^T, \sigma^T \in \mathcal{Q}_d) \end{aligned}$$

From the triangle inequality:

$$2\epsilon_\diamond(\mathcal{A}, \mathcal{U}_{\text{target}}) \leq \sup_{\sigma, \rho \in \mathcal{Q}_d} \left\{ \left\| \text{col}(A_1 \sqrt{\rho}) \text{col}^\dagger(A_1 \sqrt{\sigma}) - \text{col}(U_{\text{target}} \sqrt{\rho}) \text{col}^\dagger(U_{\text{target}} \sqrt{\sigma}) \right\|_1 \right. \\ \left. + \left\| \sum_{k \neq 1} \text{col}(A_k \sqrt{\rho}) \text{col}^\dagger(A_k \sqrt{\sigma}) \right\|_1 \right\}. \quad (4.172a)$$

The $\left\| \sum_{k \neq 1} \text{col}(A_k \sqrt{\rho}) \text{col}^\dagger(A_k \sqrt{\sigma}) \right\|_1$ term can be upper-bounded by order $\delta\Phi_{\text{decoh}}$:

$$\begin{aligned} \left\| \sum_{k \neq 1} \text{col}(A_k \sqrt{\rho}) \text{col}^\dagger(A_k \sqrt{\sigma}) \right\|_1 &\leq \sum_{k \neq 1} \left\| \text{col}(A_k \sqrt{\rho}) \text{col}^\dagger(A_k \sqrt{\sigma}) \right\|_1 && \text{(Triangle ineq.)} \\ &= \sum_{k \neq 1} \left\| \|A_k \sqrt{\rho}\|_2 \|A_k \sqrt{\sigma}\|_2 \right\|_1 && (\|vw^\dagger\|_1 = \|v\|_2 \|w\|_2) \\ &\leq \sqrt{\sum_{j \neq 1} \left\| \|A_j \sqrt{\rho}\|_2^2 \right\|_2} \sqrt{\sum_{k \neq 1} \|A_k \sqrt{\sigma}\|_2^2} && \text{(Cauchy-Schwarz)} \\ &= \sqrt{\text{Tr}[(\mathbb{I} - A_1^\dagger A_1) \rho]} \sqrt{\text{Tr}[(\mathbb{I} - A_1^\dagger A_1) \sigma]} && \text{(TP condition)} \\ &= \Gamma_{\text{decoh}} \delta\Phi_{\text{decoh}} + O(\delta^2 \Phi_{\text{decoh}}). && \text{(Equation (4.160))} \end{aligned}$$

Hence, the order $\sqrt{\delta\Phi_{\text{coh}}}$ will arise from the LK approximation of \mathcal{A} . Let $A_1 = |A_1\rangle V$, $W = U_{\text{target}} V^\dagger$, $|A_1\rangle = \mathbb{I} - \Delta_{\text{decoh}}$ and $W = \mathbb{I} - \Delta_{\text{coh}}$. Then,

$$\begin{aligned} 2\epsilon_\diamond(\mathcal{A}^*, \mathcal{U}_{\text{target}}) &= \sup_{\sigma, \rho \in \mathcal{Q}_d} \left\| \text{col}(A_1 \sqrt{\rho}) \text{col}^\dagger(A_1 \sqrt{\sigma}) - \text{col}(U_{\text{target}} \sqrt{\rho}) \text{col}^\dagger(U_{\text{target}} \sqrt{\sigma}) \right\|_1 \\ &= \sup_{\sigma, \rho \in \mathcal{Q}_d} \left\| \text{col}(|A_1\rangle \sqrt{\rho}) \text{col}^\dagger(|A_1\rangle \sqrt{\sigma}) - \text{col}(W \sqrt{\rho}) \text{col}^\dagger(W \sqrt{\sigma}) \right\|_1 \\ &\leq \sup_{\sigma, \rho \in \mathcal{Q}_d} \left[\left\| \text{col}(|A_1\rangle \sqrt{\rho}) \text{col}^\dagger(|A_1\rangle \sqrt{\sigma}) - \text{col}(\sqrt{\rho}) \text{col}^\dagger(\sqrt{\sigma}) \right\|_1 \right. \\ &\quad \left. + \left\| \text{col}(\sqrt{\rho}) \text{col}^\dagger(\sqrt{\sigma}) - \text{col}(W \sqrt{\rho}) \text{col}^\dagger(W \sqrt{\sigma}) \right\|_1 \right] && \text{(Triangle ineq.)} \\ &= \sup_{\sigma, \rho \in \mathcal{Q}_d} \left[\left\| \text{col}(\Delta_{\text{decoh}} \sqrt{\rho}) \text{col}^\dagger(\Delta_{\text{decoh}} \sqrt{\sigma}) - \text{col}(\Delta_{\text{decoh}} \sqrt{\rho}) \text{col}^\dagger(\sqrt{\sigma}) - \text{col}(\sqrt{\rho}) \text{col}^\dagger(\Delta_{\text{decoh}} \sqrt{\sigma}) \right\|_1 \right. \\ &\quad \left. + \left\| \text{col}(\Delta_{\text{coh}} \sqrt{\rho}) \text{col}^\dagger(\Delta_{\text{coh}} \sqrt{\sigma}) - \text{col}(\Delta_{\text{coh}} \sqrt{\rho}) \text{col}^\dagger(\sqrt{\sigma}) - \text{col}(\sqrt{\rho}) \text{col}^\dagger(\Delta_{\text{coh}} \sqrt{\sigma}) \right\|_1 \right] \\ &\leq \sup_{\sigma, \rho \in \mathcal{Q}_d} \left[\left\| \Delta_{\text{decoh}} \sqrt{\rho} \right\|_2 \left\| \Delta_{\text{decoh}} \sqrt{\sigma} \right\|_2 + \left\| \Delta_{\text{decoh}} \sqrt{\rho} \right\|_2 \left\| \sqrt{\sigma} \right\|_2 + \left\| \sqrt{\rho} \right\|_2 \left\| \Delta_{\text{decoh}} \sqrt{\sigma} \right\|_2 \right. \\ &\quad \left. + \left\| \Delta_{\text{coh}} \sqrt{\rho} \right\|_2 \left\| \Delta_{\text{coh}} \sqrt{\sigma} \right\|_2 + \left\| \Delta_{\text{coh}} \sqrt{\rho} \right\|_2 \left\| \sqrt{\sigma} \right\|_2 + \left\| \sqrt{\rho} \right\|_2 \left\| \Delta_{\text{coh}} \sqrt{\sigma} \right\|_2 \right] \\ &\quad \text{(Triangle ineq. followed by } \|vw^\dagger\|_1 = \|v\|_2 \|w\|_2.) \\ &\leq 2\sigma_{\text{max}}(\Delta_{\text{coh}}) + 2\sigma_{\text{max}}(\Delta_{\text{decoh}}) + \sigma_{\text{max}}^2(\Delta_{\text{coh}}) + \sigma_{\text{max}}^2(\Delta_{\text{decoh}}). \quad (4.173) \end{aligned}$$

Using eqs. (4.164) and (4.165) yields

$$2\varepsilon_{\diamond}(\mathcal{A}^*, \mathcal{U}_{\text{target}}) \leq 2\sqrt{\Gamma_{\text{coh}}}\sqrt{\delta\Phi_{\text{coh}}} + \Gamma_{\text{decoh}}\delta\Phi_{\text{decoh}} + 4\Gamma_{\text{coh}}\delta\Phi_{\text{coh}} + O(\delta^2\Phi). \quad (4.174)$$

This completes the proof. \square

4.10.3 End of the first part

This chapter concludes the first part of this thesis. The introduced notions of decoherence and equability gave at least some semblance of order in the intricate set of quantum processes, especially given the existence of a unitary-decoherent polar factorization. The next part of this thesis focuses on process characterization methods. Similarly to this part, the chapters are based on articles for which my contribution was major. It is worth mentioning that the next articles were written prior to the work introduced in the present chapter. In particular, certain statements, phrased as conjectures or open problems, are in fact largely resolved by the work covered so far. As a result, the upcoming appended afterwords contain highly valuable new observations.

Part II

Quantum characterization through randomized benchmarking

Chapter 5

From randomized benchmarking experiments to gate-set circuit fidelity: how to interpret randomized benchmarking decay parameters

5.1 Foreword

The present chapter mainly consists in a literal transcription of [Car+18], for which my contribution was major. I made alterations to the original notation to ensure self-consistency within the thesis. For instance, I changed “incoherent” errors to “decoherence-limited” errors, a notion formalized more solidly in [CAE19] (which came after the following work).

5.2 Compendium

Randomized benchmarking (RB) protocols have become an essential tool for providing a meaningful partial characterization of experimental quantum operations. While the RB decay rate is known to enable estimates of the average fidelity of those operations under gate-independent Markovian noise, under gate-dependent noise this rate is more difficult to interpret rigorously. In this paper, we prove that single-qubit RB decay parameter p coincides with the decay parameter of the *gate-set circuit fidelity*, a novel figure of merit which characterizes the expected average fidelity over arbitrary circuits of operations from the gate-set. We also prove that, in the limit of high-fidelity single-qubit experiments, the possible alarming disconnect between the average gate fidelity and RB experimental results is simply explained by a basis mismatch between the gates and the state-preparation and measurement procedures, that is, to a unitary degree of freedom in labeling the Pauli matrices. Based on numerical evidence and physically motivated arguments, we conjecture that these results also hold for higher dimensions.

5.3 Introduction

The operational richness of quantum mechanics hints at an unprecedented computational power. However, this very richness carries over to a vast range of possible quantum error processes for which a full characterization is impractical for even a handful of quantum bits

(qubits). Randomized benchmarking (RB) experiments [EAZ05; Lev+07; Kni+08; Dan+09; MGE11; MGE12; CWE15; Cro+16] were introduced to provide a robust, efficient, scalable, SPAM-independent¹, partial characterization of specific sets of quantum operations of interest, referred to as gate-sets. Such experiments have been widely adopted across all platforms for quantum computing, eg. [Gae+12; Cór+13; Kel+14; Bar+14; Cas+16; Tak+16b; She+16; McK+16; McK+17b], and have become a critical tool for characterizing and improving the design and control of quantum bits (qubits).

Recently it has been shown that RB experiments on an arbitrarily large number of qubits will always produce an exponential decay under arbitrary Markovian error models (that is, where errors are represented as completely-positive maps). This ensures a well-defined theoretical characterization of these experiments and enables an important test for the presence of non-Markovian errors, in spite of the gauge freedom between the experimental quantities and a theoretical figure of merit such as the average gate fidelity [Pro+17a; Wal17; MPF18]. However, this theoretical advance still lacks a clear physical interpretation that rigorously connects the experimentally observed decay to a fidelity-based characterization of a physical set of gate-dependent errors. Linking an experimentally measured quantity to a physically meaningful figure of merit is not a mere intellectual satisfaction. It is crucial to ensure that a quantity measured in the context of a process characterization protocol indeed yields an outcome that assesses the quality of operations. What if a very noisy quantum device could yield a decent RB parameter? What if there exist scenarios where RB substantially underestimates the quality of a quantum device?

In this paper, we show that in the regime of high fidelity gates on single qubits, a simple physical interpretation of RB data does exist and allows a reliable characterization of quantum operations. Further we conjecture, based on numerical evidence, that such an interpretation extends to arbitrary dimensions. Consequently, this work provides the theoretical foundation behind a fundamental tool for identifying and eliminating errors through examining the results of RB experiments.

Consider a gate-set $\mathbf{G} = \{g\}$, where the elements g are in $U(d)$. The targeted realization of the gate-set is noted as $L(\mathbf{G})$, where $L : U(d) \rightarrow \text{CPTP}_d$ is defined as

$$L(g)[\rho] := g\rho g^\dagger, \quad (5.1)$$

$$L(\mathbf{G}) := \{L(g)|g \in \mathbf{G}\}, \quad (5.2)$$

where $\rho \in \mathcal{Q}_d(\mathbb{C})$. Similarly, the noisy implementation of the gate-set is denoted as $\Lambda(\mathbf{G})$, where $\Lambda : U(d) \rightarrow \text{CPTP}_d$ is a noise map. We denote a circuit composed of m elements by

$$\Lambda(g)_{m:1} := \Lambda(g_m) \cdots \Lambda(g_2)\Lambda(g_1). \quad (5.3)$$

For leakage-free RB experiments with arbitrarily gate-dependent (but still Markovian) errors, the average probability of an outcome μ after preparing a state ρ and applying a circuit of $m + 1$ operations that multiply to the identity is [Wal17; MPF18]

$$\mathbb{E}_{g_{m+1:1}} \langle \mu, \Lambda(g)_{m+1:1}[\rho] \rangle = Ap^m + B + \epsilon(m), \quad (5.4)$$

where $\langle M_1, M_2 \rangle := \text{Tr } M_1^\dagger M_2$ refers to the Hilbert-Schmidt inner product. On the right-hand side of eq. (5.4), A and B are independent of m (i.e., they only depend upon ρ , μ and $\Lambda(g)$)

¹SPAM stands for “State preparation and measurement”.

and $\epsilon(m)$ is a perturbative term that decays exponentially in m .

By design, RB gives some information about the error rate of motion-reversal (i.e., identity) circuits composed of gate-set elements. In this paper, we show how this information relates to general circuits. Consider the traditional notion of average fidelity for a noisy circuit $\tilde{\mathcal{C}}$ to a target unitary circuit \mathcal{C} ,

$$F(\tilde{\mathcal{C}}, \mathcal{C}) := \int \langle \tilde{\mathcal{C}}(\psi), \mathcal{C}(\psi) \rangle d\psi, \quad (5.5)$$

where the integral is taken uniformly over all pure states. Equation (5.5) corresponds to the definition of the usual notion of average gate fidelity, but is instead formulated in terms of “circuit”, which is to be understood as a sequence of elementary gates. We introduce this nuance to define a novel figure of merit, the *gate-set circuit fidelity*, which compares all possible sequences of m implemented operations from the physical gate-set realization $\Lambda(\mathbf{G})$ to their targets in $L(\mathbf{G})$,

Definition 9: Gate-set circuit fidelity

$$F(\Lambda(\mathbf{G}), L(\mathbf{G}), m) := \mathbb{E} [F(\Lambda(g)_{m:1}, L(g_{m:1}))], \quad (5.6)$$

$$= \frac{1}{|\mathbf{G}|^m} \sum_{g_i \in \mathbf{G}} F(\Lambda(g)_{m:1}, L(g_{m:1})), \quad (5.7)$$

where $g_{m:1}$ is a shorthand for $g_m \cdots g_1$. The case $m = 1$ yields the average fidelity of the gate-set realization $\Lambda(\mathbf{G})$ to $L(\mathbf{G})$. In general, the overall action of targeted circuits $L(g_{m:1})$ is reproduced by $\Lambda(g)_{m:1}$ with fidelity $F(\Lambda(\mathbf{G}), L(\mathbf{G}), m)$. Having access to the gate-set circuit fidelity enables going beyond quantifying the quality of gate-set elements as it also quantifies the quality of circuits based on those elements. In this paper, we prove that for all single-qubit gate-sets with fidelities close to 1 and for an appropriately chosen targeted gate-set realization $L(\mathbf{G})$, the gate-set circuit fidelity can be closely estimated via RB experiments, for all circuit lengths m , even in cases of highly gate-dependent noise models. This is possible because it turns out that $F(\Lambda(\mathbf{G}), L(\mathbf{G}), m)$ essentially behaves like an exponential decay in m , uniquely parameterized by the RB decay constant p . The robust inclusion of gate-dependence is a major step forward since it encompasses very realistic noise models. We conjecture this result to hold for higher dimensions, based on numerical evidences and physically motivated arguments. Notice that the gate-set circuit fidelity quantifies the expected fidelity of *all* circuits (built from gate-set elements), and not only motion-reversal ones. This is an important observation to keep in mind because although RB experiments intrinsically revolve around motion-reversal circuits, the figure of merit that it yields isn't limited to such paradigm. Quantifying the quality of all circuits is much more useful than quantifying identity ones.

5.4 The dynamics of the gate-set circuit fidelity

It follows from the RB literature [EAZ05; MGE11] that for gate-independent noise models of the form $\Lambda(g) = \mathcal{E} \circ L(g)$ or $\Lambda(g) = L(g) \circ \mathcal{E}$, where $\mathcal{E} \in \text{CPTP}_d$ is a fixed error, given a

2-design gate-set \mathbf{G} , the gate-set circuit fidelity behaves exactly as

$$F(\Lambda(\mathbf{G}), L(\mathbf{G}), m) = \frac{1}{d} + \frac{d-1}{d} p^m, \quad (5.8)$$

where p is estimated through standard RB by fitting to eq. (5.4) with $\epsilon(m) = 0$ and d is the dimension of the system. The relationship between the survival probability decay curve and the decay in eq. (5.8) shouldn't be surprising. Indeed, consider a RB experiment with a noise model of the form $\Lambda(g_i) = \mathcal{E} \circ L(g_i)$ and a perfect inversion step $L(g_{m+1})$ and perfect SPAM. In such case, the gate-set circuit fidelity and the survival probability exactly coincide. A less trivial matter is to show the link between the RB decay parameter and eq. (5.8) for gate-dependent leakage-free noise models for which the choice of targeted gate-set is to be treated more carefully. In fact, as we will show, a poor choice of targeted gate-set can lead to a strong violation of eq. (5.8) in the sense that $\delta F(\Lambda(\mathbf{G}), L(\mathbf{G}), m) \equiv 1 - F(\Lambda(\mathbf{G}), L(\mathbf{G}), m)$ can relatively differ from $1 - (\frac{1}{d} + \frac{d-1}{d} p^m)$ by multiple orders of magnitude. An appropriate choice of targeted gate-set will essentially restore the decay relation shown in eq. (5.8).

Equation (5.8) generalizes to

$$F(\Lambda(\mathbf{G}), L(\mathbf{G}), m) = \frac{1}{d} + \frac{d-1}{d} f_{\text{trls}}(\Lambda(\mathbf{G}), L(\mathbf{G}), m), \quad (5.9)$$

where the fidelity on the traceless hyperplane is similar to the gate-set circuit fidelity, but is averaged over the traceless part of the pure states, $\psi_{\text{trls}} = \psi - \mathbb{I}/d$:

$$f_{\text{trls}}(\Lambda(\mathbf{G}), L(\mathbf{G}), m) := \frac{\mathbb{E} \left(\int \langle \Lambda(g)_{m:1}[\psi_{\text{trls}}], L(g_{m:1})[\psi_{\text{trls}}] \rangle d\psi \right)}{\int \langle \psi_{\text{trls}}, \psi_{\text{trls}} \rangle d\psi}. \quad (5.10)$$

The integrand in the numerator of the right-hand side of eq. (5.10) can be visualized as the fidelity restricted on the Bloch space, comparing the ideally mapped Bloch vectors $\psi_{\text{trls}} \rightarrow L(g_{m:1})(\psi_{\text{trls}})$ to their noisy analog $\Lambda(g)_{m:1}[\psi_{\text{trls}}]$. Equation (5.9) is quickly obtained by realizing that the symmetric integral over the Bloch space $\int \psi_{\text{trls}} d\psi = 0$.

Under gate-dependent noise, $1 - f_{\text{trls}}(\Lambda(\mathbf{G}), L(\mathbf{G}), 1)$ could relatively differ from $1 - p$ by several orders of magnitude [Pro+17a; QK18]. Such discrepancy was seen as a serious concern: the observed RB decay seemingly fails in characterizing the quality of quantum operations. To see the possible immense disconnect between p and $f_{\text{trls}}(\Lambda(\mathbf{G}), L(\mathbf{G}), 1)$, consider the canonical example where single-qubit gates are perfectly implemented, but differ from the targets $\mathcal{G} \in L(\mathbf{G})$ by a labeling of the Pauli axes:

$$\Lambda(X) = L(Y), \quad (5.11a)$$

$$\Lambda(Y) = L(Z), \quad (5.11b)$$

$$\Lambda(Z) = L(X). \quad (5.11c)$$

This noise model would lead to an absence of decay in the survival probability, that is $p = 1$. Indeed, motion-reversal circuits are perfectly implementing the identity gate, regardless of the length of the circuit. A quick calculation results in $f_{\text{trls}}(\Lambda(\mathbf{G}), L(\mathbf{G}), m) = 0$, which demonstrates a difference in orders of magnitude $|\log(1 - p) - \log(1 - f_{\text{trls}}(\Lambda(\mathbf{G}), L(\mathbf{G}), 1))|$ that tends to infinity as $p \rightarrow 1$. The RB experiment indicates no operational error while the average gate fidelity indicates 1/2. Does the outcome of RB massively underestimate the

error? Notice that since the implementation error is a permutation of labels, there is actually no observable error in the device. The alarmingly low value of gate-set circuit fidelity of $\Lambda(\mathbf{G})$ to $L(\mathbf{G})$ is simply a consequence of a poor choice of targeted gate-set realization.

As a more involved example, let the noise model be $\Lambda(g) = \mathcal{U}L(g)\mathcal{U}^\dagger$ for any non-identity unitary channel $\mathcal{U} = L(U)$ and let the set of targeted operations be $L(\mathbf{G})$ (this includes our previous mislabeling example as a special scenario). In such cases $f_{\text{trls}}(\Lambda(\mathbf{G}), L(\mathbf{G}), 1)$ can take any value in the interval $[0, 1)$, depending on the choice of \mathcal{U} . However, using the same argument as in the previous example, the survival probability is not subject to a decay ($p = 1$), showing once again how the decay parameter could arbitrarily differ from a poorly defined average gate fidelity. This apparent disconnect arises due to a *basis mismatch* between the bases in which the noisy gate-set and the targeted gate-set are defined. A reconciliation of the RB observations with a gate-set circuit fidelity is obtained by changing the set of targets to $L(UGU^\dagger)$ since $f_{\text{trls}}(\Lambda(\mathbf{G}), L(UGU^\dagger), 1) = 1$. One might argue that implementing $\Lambda(\mathbf{G}) = \mathcal{U}L(\mathbf{G})\mathcal{U}^\dagger$ instead of the target realization $L(\mathbf{G})$ should raise an operational error. Not necessarily: consider a circuit uniquely constructed from operations $\Lambda(g_i) \in \Lambda(\mathbf{G})$. According to Born's rule, the probability of measuring the outcome i associated with the positive operator μ_i after performing the circuit on a state ρ is:

$$\begin{aligned}
p_i &= \langle \mu_i, \Lambda(g)_{m:1}[\rho] \rangle \\
&= \langle \mu_i, \mathcal{U}L(g_m)\mathcal{U}^\dagger \cdots \mathcal{U}L(g_2)\mathcal{U}^\dagger \mathcal{U}L(g_1)\mathcal{U}^\dagger[\rho] \rangle \\
&= \langle \mu_i, \mathcal{U}L(g_{m:1})\mathcal{U}^\dagger[\rho] \rangle \\
&= \langle \mu'_i, L(g_{m:1})[\rho'] \rangle,
\end{aligned} \tag{5.12}$$

where $\rho' = \mathcal{U}^\dagger[\rho]$, $\mu'_i = \mathcal{U}^\dagger[\mu_i]$. That is, the error can be interpreted as part of SPAM procedures rather than from operations. Since the unitary transformation can be pushed to either SPAM procedures or coherent manipulations, it should be seen as a mismatch between them. Indeed, the physical unitary conjugation $\Lambda(\mathbf{G}) = \mathcal{U}L(\mathbf{G})\mathcal{U}^\dagger$ doesn't affect the *internal action* of operations, but exclusively the connection between operations and SPAM procedures. Changing the targeted gate-set realization $L(\mathbf{G})$ to $L(UGU^\dagger)$ is allowed by the degree of freedom in labeling what is the basis for SPAM procedures and what is the reference basis for processes.

In section 5.7.1, we show how exactly the disconnect between p^m and $f_{\text{trls}}(\Lambda(\mathbf{G}), L(UGU^\dagger), m)$ depends on the choice of targeted gate-set realization $L(UGU^\dagger)$. That is, we provide an expression of the form

$$f_{\text{trls}}(\Lambda(\mathbf{G}), L(UGU^\dagger), m) = C(U)p^m + D(m, U), \tag{5.13}$$

where $U \in U(d)$ and corresponds to a physical change of reference (see theorem 20). A first interesting observation is that $D(m, U)$ is typically negligible or becomes rapidly negligible as it is also exponentially suppressed in m^2 . This means that the relative variation in f_{trls} as

²Since $D(1, U)$ is typically close to 0, the exponential suppression is quite effective compared to $p^m \approx 1 - m(1 - p)$ which is essentially linear for small m .

the circuit grows in length,

$$\frac{f_{\text{trls}}(\Lambda(\mathbf{G}), L(\mathbf{UGU}^\dagger), m+1)}{f_{\text{trls}}(\Lambda(\mathbf{G}), L(\mathbf{UGU}^\dagger), m)} = p + \epsilon(m, U), \quad (5.14)$$

depends weakly on the choice of targeted gate-set. More precisely, $\epsilon(m, U)$ is composed of two factors: the first one decays exponentially in m and is at most of order $(1-p)^{m/2}$, while the second carries the dependence in U ; the existence of a specific choice of U such that this last factor becomes at most of order $(1-p)^{3/2}$ is proven in the single-qubit case (section 5.7.2), and conjectured to hold in general. The explicit behaviour of $\epsilon(m, U)$ given a numerically simulated gate-dependent noise model is illustrated in fig. 5.1.

Consequently, the gate-set circuit fidelity can be updated with a good approximation through the recursion relation

$$F(\Lambda(\mathbf{G}), L(\mathbf{UGU}^\dagger), m+1) \approx \frac{1}{d} + p \left(F(\Lambda(\mathbf{G}), L(\mathbf{UGU}^\dagger), m) - \frac{1}{d} \right). \quad (5.15)$$

Roughly speaking, this means that the choice of basis in which are expressed the targets in $L(\mathbf{UGU}^\dagger)$ is not highly significant when it comes to updating the gate-set circuit fidelity as the circuit grows in depth. The RB decay rate p enables the decrease in fidelity due to adding a gate to a circuit to be predicted.

However, to provide insight on the total value of the gate-set circuit fidelity given a circuit's length m , we need a stronger relation between the RB estimate of p and the gate-set circuit fidelity. Fortunately, the basis freedom in the choice of targeted gate-set can be fixed in a way that allows us to estimate the total change in gate-set circuit fidelity for arbitrary circuit's lengths.

In section 5.7.2, we prove that the potentially large disconnect between p and $f_{\text{trls}}(\Lambda(\mathbf{G}), L(\mathbf{UGU}^\dagger), 1)$ under general gate-dependent noise is almost completely accounted for by a basis mismatch which, as we argued earlier, doesn't exactly correspond to a process error since unitary conjugation does not affect the internal dynamics of operations.

Theorem 18: The RB decay connects with a (physical) circuit fidelity

Consider a 2-design gate-set \mathbf{G} . For any single-qubit noisy gate-set physical realization $\Lambda(\mathbf{G})$ perturbed from $L(\mathbf{G})$, there exists a targeted gate-set realization $L(\mathbf{UGU}^\dagger)$, where $U \in U(d)$ corresponds to a physical unitary, such that

$$F(\Lambda(\mathbf{G}), L(\mathbf{UGU}^\dagger), m) = \frac{1}{d} + \frac{d-1}{d} p^m + O((1-p)^2). \quad (5.16)$$

In fact, we conjecture this result to hold for any dimension, or at least for most realistic gate-dependent noise models. To grasp the physical reasoning behind this, we refer to the end of section 5.7.2, as it rests on some prior technical analysis. The extension of theorem 18 to 2-qubit systems is supported by numerical evidences (see sections 5.7.1 to 5.7.2).

The unitary freedom appearing in the gate-set circuit fidelity means that there exists an infinite amount of fidelity-based figures of merit describing noisy circuits, one for each infinitely many targeted gate-set realization $L(\mathbf{UGU}^\dagger)$. Of course, there exist choices of targeted operations that yield in gate-set circuit fidelities that differ from eq. (5.16) (see

[Pro+17a; QK18]); the example shown in eqs. (5.11a) to (5.11c) is an elementary instance thereof. Theorem 18 simply states that there exists a natural choice of gate-set realization $L(UGU^\dagger)$ that allows connecting the outcome of an RB experiment to a gate-set circuit fidelity. The choice of basis is like taking the perspective of the gates rather than the perspective of SPAM procedures (as is implicitly done when defining gates relative to the energy eigenbasis of the system). In this picture, the gate-set circuit fidelity describes the accuracy of the internal behaviour of operations as they act in concert.

To reformulate the result, the family of circuits $\Lambda(g)_{m:1}$ built from a composition of m noisy operations $\Lambda(g_i) \in \Lambda(\mathbf{G})$ mimics the family of circuit realizations $L(Ug_{m:1}U^\dagger)$ with fidelity $\frac{1}{d} + \frac{d-1}{d}p^m$. In the paradigm where the initially targeted realizations $L(g_i) \in L(\mathbf{G})$ are defined with respect to SPAM procedures, U captures the misalignment between the basis in which the operations $\Lambda(g_i) \in \Lambda(\mathbf{G})$ are defined and the basis defined by SPAM procedures. This goes farther: consider an additional 2-design gate-set \mathbf{H} , for which the targeted operational realizations $L(h) \in L(\mathbf{H})$ are also defined respect to SPAM procedures. From theorem 18, there exists a physical unitary V for which $\Lambda_{\mathbf{H}}(h)_{m:1}$ overlaps with the action of $L(Vh_{m:1}V^\dagger)$ with fidelity $\frac{1}{d} + \frac{d-1}{d}q^m$ (where q is estimated through RB). $U^\dagger V$ captures the basis mismatch between the gate-sets \mathbf{G} and \mathbf{H} . Such a non-trivial mismatch could easily be imagined if, for instance, the physical realization $\Lambda_{\mathbf{H}}$ of gates belonging to \mathbf{H} were obtained through a different physical process than the one modeled by Λ , or calibrated with regards to alternate points of reference.

5.5 Finding the appropriate set of targeted gate realizations for specific noise models

We now discuss how the appropriate unitary conjugation on the initial targeted gate-set can be calculated for specific noise models, whether from numerical simulations, analytic approximations, or tomographic reconstructions. As shown in theorem 20 and eq. (5.13), the total change of gate-set circuit fidelity depends on the physical basis in which the target gate-set realization is expressed. In the single-qubit case, we showed the existence of a physical basis \mathcal{U} that reconciles $f_{\text{trls}}(\Lambda(\mathbf{G}), L(UGU^\dagger), m)$ with p^m through theorem 18. One might suspect that the unitary U can be found through the maximization of the gate-set fidelity:

$$U = \underset{V \in U(d)}{\operatorname{argmax}} F(\Lambda(\mathbf{G}), L(V\mathbf{G}V^\dagger), 1), \quad (5.17)$$

and indeed this would handle noise models of the form $\Lambda(\mathbf{G}) = \mathcal{U}\mathcal{E}L(\mathbf{G})\mathcal{U}^\dagger$, as

$$p = f_{\text{trls}}(\Lambda(\mathbf{G}), L(UGU^\dagger), 1) \geq f_{\text{trls}}(\Lambda(\mathbf{G}), L(\mathbf{G}), 1).$$

However, this hypothesis fails for simple noise models of the form $\Lambda(\mathbf{G}) = \mathcal{U}^\dagger\mathcal{E}L(\mathbf{G})\mathcal{U}$, where

$$p = f_{\text{trls}}(\Lambda(\mathbf{G}), L(UGU^\dagger), 1) \leq f_{\text{trls}}(\Lambda(\mathbf{G}), L(\mathbf{G}), 1).$$

Those last two examples show that p can be greater or lesser than $f_{\text{trls}}(\Lambda(\mathbf{G}), L(\mathbf{G}), 1)$, depending on the noise model. More examples are derived in [Pro+17a; QK18]. This particular case study is informative as these two last noise models share something in common: there

exists a choice of unitary U that cancels the noisy map on the right of the noisy gate-set realization. Although such exact cancellation is not always possible, we now show that a close approximation is sufficient. Consider the slightly more general noise model of the form $\Lambda(\mathbf{G}) = \mathcal{E}_L L(\mathbf{G}) \mathcal{E}_R$, where we allow fixed but arbitrary error maps to the left and the right of a target gate-set realization. It can be shown while staying under the scope of the original analysis provided in [MGE11; MGE12] that $p^m = f_{\text{trls}}(\mathcal{E}_R \mathcal{E}_L L(\mathbf{G}), L(\mathbf{G}), m)$, since $\mathcal{E}_R \mathcal{E}_L$ is the effective error map between two otherwise perfect implementations of the gate-set elements. In the single-qubit case (and for many, if not all physically motivated higher dimensional noise models) there exists a unitary operation $U \in U(d)$ such that

$$F(\mathcal{E}_R \mathcal{E}_L, \mathcal{I}) = F(\mathcal{E}_L L(\mathbf{G}) \mathcal{E}_R, L(UGU^\dagger), 1) + O((1-p)^2), \quad (5.18)$$

(see section 5.7.2). That is, the fidelity of the effective error channel occurring between two noisy gate implementations can be seen as the gate-set circuit fidelity between a noisy gate-set realization and an appropriately targeted realization. A choice of such physical unitary is

$$U = \operatorname{argmax}_{V \in U(d)} F(\mathcal{E}_R V, \mathcal{I}), \quad (5.19)$$

which essentially cancels the unitary part of \mathcal{E}_R ³. Another way to see this is that the unitary freedom allows us to re-express the errors $\mathcal{E}_L, \mathcal{E}_R$ as

$$\begin{aligned} \mathcal{E}_L &\rightarrow \mathcal{U}^\dagger \mathcal{E}_L \\ \mathcal{E}_R &\rightarrow \mathcal{E}_R \mathcal{U}. \end{aligned}$$

We can then choose the unitary that depletes $\mathcal{E}_R \mathcal{U}$ from any coherent component. Intuitively, re-expressing the error on one side to make it decoherence-limited prevents any type of unitary conjugation of the form $\Lambda(\mathbf{G}) = \mathcal{U} \mathcal{E} L(\mathbf{G}) \mathcal{U}^\dagger$.

For more general gate-dependent noise models, the idea remains more or less the same. As shown in section 5.7.2, the right error \mathcal{E}_R is replaced by its generalization, the 4th order right error $\mathcal{E}_R^{(4)} = \mathbb{E}_{g_i \in \mathbf{G}} \left[L(g_{4:1})^\dagger \Lambda(g)_{4:1} \right]$ (eq. (5.41a)). From there, we find:

Theorem 19: Finding the appropriate targeted gate-set

A proper choice of physical basis for which eq. (5.16) applies is

$$U = \operatorname{argmax}_{V \in U(2)} F \left(\mathbb{E}_{g_i \in \mathbf{G}} \left[L(g_{4:1})^\dagger \Lambda(g)_{4:1} \right] V, \mathcal{I} \right), \quad (5.20)$$

$U = L(U)$ cancels the unitary part of the 4th order right error.

This provides a means to guide the search of the appropriate targeted realization $L(UGU^\dagger)$ given a numerical noise model $\Lambda(\mathbf{G})$. Indeed, the 4th order right error is easily

³Of course, $\operatorname{argmax}_{V \in U(d)} F(V^\dagger \mathcal{E}_L, \mathcal{I})$ would also fulfill eq. (5.18).

found, either by direct computation of the average $\mathbb{E}_{g_i \in \mathbf{G}} \left[L(g_{4:1})^\dagger \Lambda(g)_{4:1} \right]$, or more efficiently by solving the eigensystem defined in eq. (5.28a). For higher dimensions, the optimization defined in eq. (5.20) can be solved via a gradient ascent parametrized over the $d^2 - 1$ degrees of freedom of $SU(d)$ (although the result of theorem 18 is only conjectured for $d > 2$).

In the single-qubit case, the optimization procedure can be replaced by an analytical search. Given the process matrix $\mathcal{E}_R^{(4)}$ of the 4th order right error, it suffices to find the polar decomposition of its 3×3 submatrix acting on the Bloch vectors: $\mathcal{E}_R^{(4)} \Pi_{\text{trls}} = \mathcal{D}_{\text{trls}} \mathcal{V}_{\text{trls}}$. The unitary factor \mathcal{V} corresponds to U^\dagger , while the positive factor \mathcal{D} captures a decoherence-limited process (rigorously defined in eq. (5.44)).

With this at hand, we performed numerically simulated RB experiments under gate-dependent noise models. Each of the 24 Cliffords was constructed by a sequence of X and Y pulses, $G_x = P(\sigma_x, \pi/2)$ and $G_y = P(\sigma_y, \pi/2)$, where

$$P(H, \theta) := e^{i\theta H/2}. \quad (5.21)$$

The 2-qubit Cliffords were obtained through the construction shown in [Bar+14; Cór+13], where the 11520 gates are composed of single-qubit Clifford and CZ gates. The implementation of the 2-qubit entangling operation was consistently performed with an over-rotation: $\Lambda(\text{CZ}) = L\left(P(\sigma_z^1 \sigma_z^2 - \sigma_z^1 - \sigma_z^2, \pi/2 + 10^{-1})\right)$. In fig. 5.2, the single-qubit gate generators are modeled with a slight over-rotation: $\Lambda(X) = L\left(P(\sigma_x, \pi/2 + 10^{-1})\right)$ and $\Lambda(Y) = L\left(P(\sigma_y, \pi/2 + 10^{-1})\right)$. This model exemplifies the failure of the maximization hypothesis proposed in eq. (5.17). In figs. 5.1 and 5.3, the single-qubit gate generators are followed by a short Z pulse, $\Lambda(X) = L\left(P(\sigma_z, \theta_z)\right)L(X)$ and $\Lambda(Y) = L\left(P(\sigma_z, \theta_z)\right)L(Y)$, which reproduces the toy model used in [Pro+17a].

5.6 Conclusion

RB experiments estimate the survival probability decay parameter p of motion-reversal circuits constituted of operations from a noisy gate-set realization $\Lambda(\mathbf{G})$ of increasing length (see eq. (5.4)). While motion-reversal is intrinsic to the experimental RB procedure, the estimated decay constant p can be interpreted beyond this paradigm. In this paper we have shown that, in a physically relevant limit, the very same parameter determines an interesting figure of merit, namely the gate-set circuit fidelity (defined in eq. (5.6)): as a random operation from $\Lambda(\mathbf{G})$ is introduced to a random circuit constructed from elements in $\Lambda(\mathbf{G})$, p captures the expected relative change in the gate-set circuit fidelity through eq. (5.15).

It is also possible to characterize the full evolution of gate-set circuit fidelity as a function of the circuit length. In this paper, we have also demonstrated that given a single-qubit noisy gate-set realization $\Lambda(\mathbf{G})$ perturbed from $L(\mathbf{G})$, there exists an alternate set of target gate realizations obtained through a physical basis change $L(UGU^\dagger)$ such that the gate-set circuit fidelity takes the simple form given in eq. (5.16). This gives a rigorous underpinning to previous work that has assumed that the experimental RB decay parameter robustly determines a relevant average gate fidelity (eq. (5.5)) for experimental control under generic gate-dependent scenarios. We conjecture a similar result to hold for higher dimensions and provide numerical evidence and physically motivated arguments to support this conjecture.

Given any specific numerical noise model $\Lambda(\mathbf{G})$ perturbed from $L(\mathbf{G})$, we showed how to obtain a physical unitary \mathcal{U} for which eq. (5.16) holds. The procedure can be seen as a fidelity maximization of the 4th order right error acting on the gate-set through a unitary correction (see theorem 19).

The introduction of such a physical basis adjustment is natural because it has no effect on how errors accumulate as a function of the sequence length. Rather, it only reflects a basis mismatch to the experimental SPAM procedures. This is in principle detectable by RB experiments but in practice not part of the goals of such diagnostic experiments. In particular, differences in the (independent) basis adjustments required for distinct gate-sets will not robustly appear in any RB-type characterization of the individual gate-sets, but will be detected when comparing RB experiments performed over distinct gate-sets (e.g. comparing independent single-qubit RB on two qubits - which has no two-qubit entangling gate - with standard two-qubit RB). We leave the problem of characterizing relative basis mismatch between independent gate-sets as a subject for further work.

5.7 Supplementary material

5.7.1 An expression for the total change in the gate-set circuit fidelity

In this section, we extend the standard RB analysis under gate-dependent noise provided in [Wal17; MPF18] in order to prove the claim from eq. (5.14) that standard RB returns the relative variation of the gate-set circuit fidelity.

In the following, a linear map $\mathcal{A} : M_d(\mathbb{C}) \rightarrow M_d(\mathbb{C})$ and its Liouville representation $A \in M_{d^2}(\mathbb{C})$ are represented by the same symbol. Context will suffice to differentiate them. Let $\Pi_{\text{trls}}(\rho) = \rho - \mathbb{I}_d \text{Tr} \rho / d$ be the projector onto the traceless hyperplane. We denote the Frobenius norm, which is defined by the Hilbert-Schmidt inner product, as $\|\cdot\|_2$. For instance, in the qubit case $\|\Pi_{\text{trls}}\|_2^2 = 3$. We denote the induced 2-norm as $\|\cdot\|_\infty$, which corresponds to the maximal singular value (and to the ∞ -Schatten norm). Let e_j be the canonical unit vectors, $A = \sum_{j,k} a_{j,k} e_j e_k^T$, and

$$\text{vec}(A) = \sum_{j,k} a_{j,k} e_k \otimes e_j. \quad (5.22)$$

Using the identity

$$\text{vec}(ABC) = (C^T \otimes A) \text{vec}(B), \quad (5.23)$$

we have

$$\begin{aligned} f_{\text{trls}}(\Lambda(\mathbf{G}), L(\mathbf{G}), m) &= \mathbb{E}_{g_i \in \mathbf{G}} \left(\frac{\langle \Lambda(g)_{m:1} \Pi_{\text{trls}}, L(g_{m:1}) \Pi_{\text{trls}} \rangle}{\|\Pi_{\text{trls}}\|_2^2} \right) \\ &= \frac{\text{vec}^\dagger(\Pi_{\text{trls}}) \mathcal{T}^m \text{vec}(\Pi_{\text{trls}})}{\|\Pi_{\text{trls}}\|_2 \|\Pi_{\text{trls}}\|_2} \end{aligned} \quad (5.24)$$

where the twirling superchannel $\mathcal{T} : M_{d^2}(\mathbb{C}) \rightarrow M_{d^2}(\mathbb{C})$ is defined as

$$\mathcal{T} := \mathbb{E}_{g \in \mathbf{G}} [L_{\text{trls}}(g) \otimes \Lambda(g)] \quad (5.25)$$

and $L_{\text{trls}}(g) := \Pi_{\text{trls}}L(g)\Pi_{\text{trls}}$ (see [CW15; Pro+17a; Wal17]). Changing the gate-set $L(\mathbf{G})$ to $L(U\mathbf{G}U^\dagger)$ for some physical unitary $U \in U(d)$ leaves Π_{trls} invariant: $\Pi_{\text{trls}} = \mathcal{U}\Pi_{\text{trls}}\mathcal{U}^\dagger$. Therefore

$$f_{\text{trls}}(\Lambda(\mathbf{G}), L(U\mathbf{G}U^\dagger), m) = \frac{\text{vec}^\dagger(\mathcal{U}\Pi_{\text{trls}})}{\|\Pi_{\text{trls}}\|_2} \mathcal{T}^m \frac{\text{vec}(\mathcal{U}\Pi_{\text{trls}})}{\|\Pi_{\text{trls}}\|_2}. \quad (5.26)$$

The spectrum of \mathcal{T} is unchanged under the basis change $L(g) \rightarrow L(UgU^\dagger)$. Moreover, its most important eigenvectors are as follows:

Lemma 7

Let p be the highest eigenvalue of \mathcal{T} and

$$\mathcal{A}_m := p^{-m} \mathbb{E}_{g_i \in \mathbf{G}} \left[(L_{\text{trls}}(g_{m:1}))^\dagger \Lambda(g)_{m:1} \right], \quad (5.27a)$$

$$\mathcal{B}_m := p^{-m} \mathbb{E} \left[\Lambda(g)_{m:1} (L_{\text{trls}}(g_{m:1}))^\dagger \right]. \quad (5.27b)$$

Then we have

$$\text{vec}^\dagger(\mathcal{A}_\infty^T) \mathcal{T} = p \text{vec}^\dagger(\mathcal{A}_\infty^T), \quad (5.28a)$$

$$\mathcal{T} \text{vec}(\mathcal{B}_\infty) = p \text{vec}(\mathcal{B}_\infty). \quad (5.28b)$$

Proof. By eq. (5.23),

$$\text{vec}(\mathcal{B}_m) = p^{-m} \mathbb{E}_{g_i \in \mathbf{G}} ((L^*(g_{m:1})_{\text{trls}} \otimes \Lambda(g)_{m:1}) \text{vec}(\Pi_{\text{trls}})). \quad (5.29)$$

As the Liouville representation is real-valued (WOLOG) and the g_i are independently averaged,

$$\text{vec}(\mathcal{B}_m) = (\mathcal{T}/p)^m \text{vec}(\Pi_{\text{trls}}). \quad (5.30)$$

Since the noisy gate-set realization $\Lambda(\mathbf{G})$ is a small perturbation from $L(\mathbf{G})$, the spectrum of \mathcal{T} will be slightly perturbed from $\{1, 0, 0, \dots\}$. Therefore $(\mathcal{T}/p)^m$ approaches a rank-1 projector as m increases and so $\text{vec}(\mathcal{B}_\infty)$ is a +1-eigenvector of \mathcal{T}/p .

The same argument applies to \mathcal{A}_∞^T . \square

Lemma 7 allows us to write

$$\mathcal{T} = p \frac{\text{vec}(\mathcal{B}_\infty) \text{vec}^\dagger(\mathcal{A}_\infty^T)}{\langle \mathcal{A}_\infty^T, \mathcal{B}_\infty \rangle} + \Delta, \quad (5.31)$$

with $\Delta \text{vec}(\mathcal{B}_\infty) = \text{vec}^\dagger(\mathcal{A}_\infty^T) \Delta = 0$. In eq. (5.26), we can expand the vectors as

$$\frac{\text{vec}^\dagger(\mathcal{U}\Pi_{\text{trls}})}{\|\Pi_{\text{trls}}\|_2} = a(U) \frac{\text{vec}^\dagger(\mathcal{A}_\infty^T)}{\|\mathcal{A}_\infty^T\|_2} + \sqrt{1 - a^2(U)} w^\dagger(U) \quad (5.32a)$$

$$\frac{\text{vec}(\mathcal{U}\Pi_{\text{trls}})}{\|\Pi_{\text{trls}}\|_2} = b(U) \frac{\text{vec}(\mathcal{B}_\infty)}{\|\mathcal{B}_\infty\|_2} + \sqrt{1 - b^2(U)} v(U) \quad (5.32b)$$

where

$$a(U) := \frac{\langle \mathcal{A}_\infty^T, \mathcal{U} \rangle}{\|\Pi_{\text{trls}}\|_2^2} \left(\frac{\|\mathcal{A}_\infty\|_2^2}{\|\Pi_{\text{trls}}\|_2^2} \right)^{-1/2}, \quad (5.33)$$

$$b(U) := \frac{\langle \mathcal{U}, \mathcal{B}_\infty \rangle}{\|\Pi_{\text{trls}}\|_2^2} \left(\frac{\|\mathcal{B}_\infty\|_2^2}{\|\Pi_{\text{trls}}\|_2^2} \right)^{-1/2}. \quad (5.34)$$

and $v(U)$, $w(U)$ are implicitly defined unit vectors. Using this expansion together with eq. (5.31) in eq. (5.26) yields the following result:

Theorem 20: Total gate-set circuit fidelity

The gate-set circuit fidelity obeys

$$F(\Lambda(\mathbf{G}), L(UGU^\dagger), m) = \frac{1}{d} + \frac{d-1}{d} (C(U)p^m + D(m, U)), \quad (5.35)$$

where

$$\begin{aligned} C(U) &:= \frac{\langle \mathcal{A}_\infty^T, \mathcal{U} \rangle \langle \mathcal{U}, \mathcal{B}_\infty \rangle}{\|\Pi_{\text{trls}}\|_2^2 \|\Pi_{\text{trls}}\|_2^2} \left(\frac{\langle \mathcal{A}_\infty^T, \mathcal{B}_\infty \rangle}{\|\Pi_{\text{trls}}\|_2^2} \right)^{-1} \\ &= \frac{\langle \Pi_{\text{trls}}, \mathcal{A}_\infty \mathcal{U} \rangle \langle \Pi_{\text{trls}}, \mathcal{U}^\dagger \mathcal{B}_\infty \rangle}{\|\Pi_{\text{trls}}\|_2^2 \|\Pi_{\text{trls}}\|_2^2} \frac{\|\Pi_{\text{trls}}\|_2^2}{\langle \Pi_{\text{trls}}, \mathcal{A}_\infty \mathcal{B}_\infty \rangle} \end{aligned} \quad (5.36a)$$

$$D(m, U) := \sqrt{1 - a^2(U)} \sqrt{1 - b^2(U)} w(U)^\dagger \Delta^m v(U). \quad (5.36b)$$

In [Pro+17a; Wal17; MPF18] it is shown that standard RB provides an estimate of p . Notice that p is independent of the basis appearing in the targeted gate-set realization, $L(UGU^\dagger)$.

From eq. (5.35), it is straightforward to show that

$$\begin{aligned} \epsilon(m, U) &:= \frac{f_{\text{trls}}(\Lambda(\mathbf{G}), L(UGU^\dagger), m+1)}{f_{\text{trls}}(\Lambda(\mathbf{G}), L(UGU^\dagger), m)} - p = \\ &= \sqrt{1 - a^2(U)} \sqrt{1 - b^2(U)} \frac{w(U)^\dagger \Delta^m (\Delta - p \Pi_{\text{trls}}) v(U)}{f_{\text{trls}}(\Lambda(\mathbf{G}), L(UGU^\dagger), m)}, \end{aligned} \quad (5.37)$$

which is exponentially suppressed. We show in the next section that the eigenvalues of Δ are at most of order $\sqrt{1-p}$, which ensures a very fast decay, as shown in fig. 5.1. Equation (5.14) is in fact a reformulation of eq. (5.37).

5.7.2 Varying the ideal gate-set of comparison

In this section, we prove theorem 18 by determining how the basis change in the target gate-set representation $L(UGU^\dagger)$ affects the coefficients in eq. (5.35).

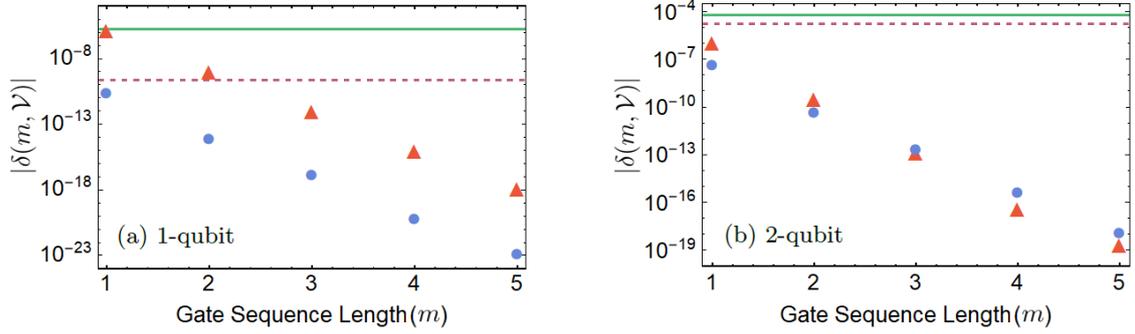


FIGURE 5.1: Absolute value of the deviation $\delta(m, \mathcal{V}) \equiv \epsilon(m, V)$ (the label of the y -axis corresponds to the original notation appearing in [Car+18]), described in eq. (5.14) (also see eq. (5.37)), as function of circuit length m with noise model generated by $\Lambda(X) = L(P(\sigma_z, 10^{-1}))L(X)$ and $\Lambda(Y) = L(P(\sigma_z, 10^{-1}))L(Y)$, $\Lambda(CZ) = L(P(\sigma_z^1 \sigma_z^2 - \sigma_z^1 - \sigma_z^2, \pi/2 + 10^{-1}))$ (see eq. (5.21)). The red triangles are obtained with the choice of basis $V = I$, while the blue circles are obtained with the choice $V = U$ where U is found through eq. (5.20). The purple horizontal dashed line corresponds to $(1 - p)^2$, while the full green line corresponds to $(1 - F(\Lambda(\mathbf{G}), L(\mathbf{G}), 1))^2$. For both ideal gate-sets $L(\mathbf{G})$ and $L(U\mathbf{G}U^\dagger)$, the deviation becomes quickly negligible as the sequence length increases. In fact, in the case $V = U$ (blue circles), the deviation is always below $(1 - p)^2$.

Let $L(\mathbf{G})$ be a target gate-set realization defined with respect to the SPAM procedures. We can write the elements of a noisy gate-set as

$$\Lambda(g) = L(g) + \delta(g)L(g), \quad (5.38)$$

so that the perturbations $\delta(g) \in M_{d^2}(\mathbb{C})$ both capture the errors in the noisy gate and the mismatch with the targeted reference basis. Under gate-independent noise with no basis mismatch, i.e. $\Lambda(g) = \mathcal{E}L(g)$, the error can be expressed as $\mathcal{E} = \mathcal{I} + \delta(g)$. The infidelity of \mathcal{E} to the identity is defined as $\delta F(\mathcal{E}, \mathcal{I}) := 1 - F(\mathcal{E}, \mathcal{I})$. Notice that $\delta F(\mathcal{E}, \mathcal{I}) = F(\delta(g), \mathcal{I})$. A basis mismatch will change the infidelity of the perturbations roughly to $\delta F(\mathcal{U}\mathcal{E}, \mathcal{I}) + \delta F(\mathcal{U}^\dagger, \mathcal{I})$ for some unitary channel \mathcal{U} , which will typically differ substantially from the fidelity inferred from the associated RB experiment.

Experimentally, such basis mismatches will be relatively small as operations will be somewhat consistent with SPAM procedures. Under this assumption, we now show that there exists an alternate perturbative expansion,

$$\Lambda(g) = L(UgU^\dagger) + \delta(UgU^\dagger)L(UgU^\dagger), \quad (5.39)$$

for which $\mathbb{E}_{g \in \mathbf{G}} F(\delta(UgU^\dagger), \mathcal{I})$ is in line with the data resulting from an RB experiment.

In section 5.7.1, we showed that $(\mathcal{T}/p)^n$ converges to a rank-1 projector. We now quantify the rate of convergence. Recall that \mathcal{T} is perturbed from a rank-1 projector with spectrum

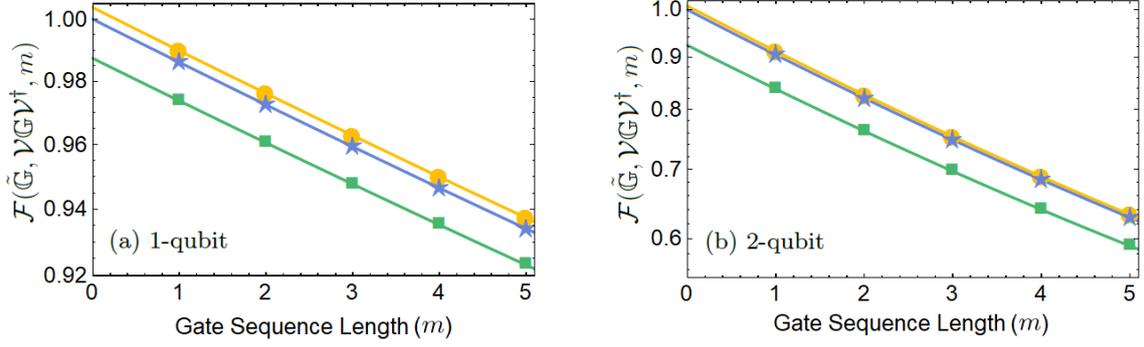


FIGURE 5.2: gate-set circuit fidelity $F(\Lambda(\mathbf{G}), L(\mathbf{V}\mathbf{G}\mathbf{V}^\dagger), m \equiv \mathcal{F}(\tilde{\mathbf{G}}, \mathcal{V}\mathbf{G}\mathcal{V}^\dagger, m)$ (the label of the y -axis corresponds to the original notation appearing in [Car+18]) as function of circuit length m with noise model generated by $\Lambda(X) = L(P(\sigma_z, 10^{-1}))L(X)$ and $\Lambda(Y) = L(P(\sigma_z, 10^{-1}))L(Y)$, $\Lambda(CZ) = L(P(\sigma_z^1\sigma_z^2 - \sigma_z^1 - \sigma_z^2, \pi/2 + 10^{-1}))$ (see eq. (5.21)). The different colors portray choices of basis; the yellow circles $V = I$, the blue stars $V = U$ where U is found through eq. (5.20), and the green squares $V = U^2$. Here the lines correspond to the fit for sequence lengths of $m=5$ to 10 . The choice $V = U$ produces the evolution prescribed by theorem 18, which through extrapolation has an intercept of 1.

$\{1, 0, 0, \dots\}$. Hence, by the Bauer-Fike theorem [BF60], for any eigenvalue $\lambda \neq p$ of \mathcal{T} ,

$$\begin{aligned}
|\lambda - 0| &\leq \|\mathbb{E}_{g \in \mathbf{G}} [L_{\text{trls}}(g) \otimes \delta(g)L(g)]\|_\infty && \text{(Bauer-Fike)} \\
&\leq \mathbb{E}_{g \in \mathbf{G}} \|[L_{\text{trls}}(g) \otimes \delta(g)L(g)]\|_\infty && \text{(triangle ineq.)} \\
&= \mathbb{E}_{g \in \mathbf{G}} \|\delta(g)\|_\infty && \text{(Unitary invariance)} \\
&\leq O\left(\mathbb{E}_{g \in \mathbf{G}} \sqrt{F(\delta(g), \mathcal{I})}\right) && \text{([Wal15])} \\
&\leq O\left(\sqrt{\mathbb{E}_{g \in \mathbf{G}} F(\delta(g), \mathcal{I})}\right) && \text{(concavity)}
\end{aligned}$$

This spectral profile implies that $(\mathcal{T}/p)^n$ converges quickly to a rank-1 operator since the eigenvalues close to zero are exponentially suppressed.

Hence, we can approximate the asymptotic eigen-operators defined in eqs. (5.27a) and (5.27b) as:

$$\mathcal{A}_\infty = \mathcal{A}_4 + O\left([\mathbb{E}_{g \in \mathbf{G}} F(\delta(g), \mathcal{I})]^2\right), \quad (5.40a)$$

$$\mathcal{B}_\infty = \mathcal{B}_4 + O\left([\mathbb{E}_{g \in \mathbf{G}} F(\delta(g), \mathcal{I})]^2\right). \quad (5.40b)$$

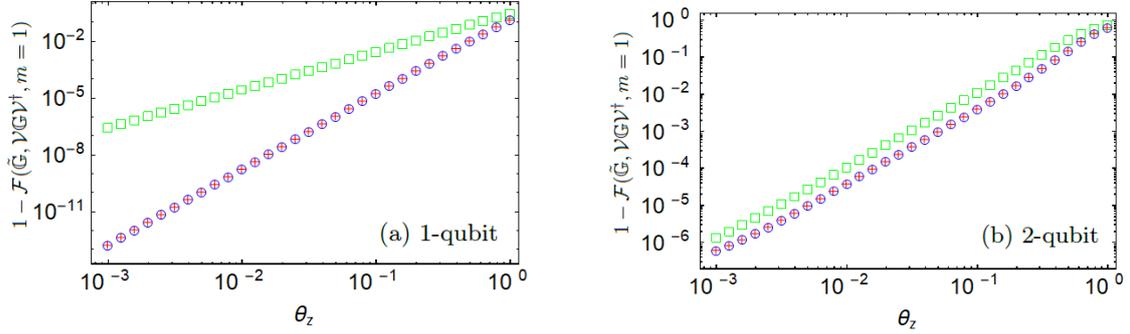


FIGURE 5.3: $1 - \mathcal{F}(\tilde{\mathbf{G}}, \mathcal{V}L(\mathbf{G})\mathcal{V}^\dagger, m = 1) \equiv \delta F(\Lambda(\mathbf{G}), L(V\mathbf{G}V^\dagger), m = 1)$ (the label of the y -axis corresponds to the original notation appearing in [Car+18]) as function of circuit length m with noise model generated by $\Lambda(X) = L(P(\sigma_z, 10^{-1}))L(X)$ and $\Lambda(Y) = L(P(\sigma_z, 10^{-1}))L(Y)$, $\Lambda(CZ) = L(P(\sigma_z^1\sigma_z^2 - \sigma_z^1 - \sigma_z^2, \pi/2 + 10^{-1}))$ (see eq. (5.21)), with $\mathcal{V} = \mathcal{I}$ (green squares) and $\mathcal{V} = \mathcal{U}$ (blue circles) where \mathcal{U} is found through eq. (5.20). The red crosses correspond to $(1 - p)/2$ obtained through RB experiments.

In the simple noise model $\mathcal{E}_L L(\mathbf{G}) \mathcal{E}_R$, $\mathcal{A}_\infty \propto \Pi_{\text{trls}} \mathcal{E}_R$ and $\mathcal{B}_\infty \propto \mathcal{E}_L \Pi_{\text{trls}}$. To pursue the analogy, we denote the m^{th} order right and left errors as

$$\mathcal{E}_R^{(m)} = \mathbb{E}_{g \in \mathbf{G}} \left[(L(g_{m:1}))^\dagger \Lambda(g)_{m:1} \right], \quad (5.41a)$$

$$\mathcal{E}_L^{(m)} = \mathbb{E}_{g \in \mathbf{G}} \left[\Lambda(g)_{m:1} (L(g_{m:1}))^\dagger \right]. \quad (5.41b)$$

Combining eq. (5.41) and eq. (5.40), we get

$$\mathcal{A}_\infty \propto \Pi_{\text{trls}} \mathcal{E}_R^{(4)} + O\left([\mathbb{E}_{g \in \mathbf{G}} F(\delta(g), \mathcal{I})]^2\right), \quad (5.42a)$$

$$\mathcal{B}_\infty \propto \mathcal{E}_L^{(4)} \Pi_{\text{trls}} + O\left([\mathbb{E}_{g \in \mathbf{G}} F(\delta(g), \mathcal{I})]^2\right). \quad (5.42b)$$

The structure of single-qubit error channels allows us to pursue a deeper analysis. It follows from the channel analysis provided in [RSW02] that, for high-fidelity qubit-channels, the 3×3 submatrix acting on the traceless hyperplane can always be decomposed as

$$\mathcal{E} \Pi_{\text{trls}} = \mathcal{D} \mathcal{V} \Pi_{\text{trls}} \quad (5.43)$$

where \mathcal{V} is a physical unitary, and \mathcal{D} is a decoherence-limited process. Here we label a channel \mathcal{D} as “decoherence-limited” if

$$\frac{\langle \Pi_{\text{trls}}, \mathcal{D} \rangle}{\|\Pi_{\text{trls}}\|_2^2} = \frac{\|\mathcal{D} \Pi_{\text{trls}}\|_2}{\|\Pi_{\text{trls}}\|_2} + O(\delta^2 F(\mathcal{D}, \mathcal{I})). \quad (5.44)$$

Decoherence-limited channels have the additional property that, given an error channel \mathcal{E} [CWE16],

$$\frac{\langle \Pi_{\text{trls}}, \mathcal{D}\mathcal{E} \rangle}{\|\Pi_{\text{trls}}\|_2^2} = \frac{\langle \Pi_{\text{trls}}, \mathcal{D} \rangle}{\|\Pi_{\text{trls}}\|_2^2} \frac{\langle \Pi_{\text{trls}}, \mathcal{E} \rangle}{\|\Pi_{\text{trls}}\|_2^2} + O(\delta^2 F(\mathcal{D}\mathcal{E}, \mathcal{I})) . \quad (5.45)$$

Expressing the 4th order right error $\mathcal{E}_R^{(4)}$ as

$$\mathcal{E}_R^{(4)} \Pi_{\text{trls}} = \mathcal{D}\mathcal{V}\Pi_{\text{trls}} . \quad (5.46)$$

allows us to maximally correct it through a physical unitary:

$$F(\mathcal{E}_R^{(4)} \mathcal{V}^\dagger, \mathcal{I}) = \max_{\mathcal{U}} F(\mathcal{E}_R^{(4)} \mathcal{U}, \mathcal{I}) \geq F(\mathcal{E}_R^{(4)}, \mathcal{I}) . \quad (5.47)$$

Using the property expressed in eq. (5.45), we get:

$$\begin{aligned} \frac{\langle \Pi_{\text{trls}}, \mathcal{E}_R^{(4)} \mathcal{V}^\dagger \mathcal{V} \mathcal{E}_L^{(4)} \rangle}{\|\Pi_{\text{trls}}\|_2} &= \frac{\langle \Pi_{\text{trls}}, \mathcal{E}_R^{(4)} \mathcal{V}^\dagger \rangle}{\|\Pi_{\text{trls}}\|_2} \frac{\langle \Pi_{\text{trls}}, \mathcal{V} \mathcal{E}_L^{(4)} \rangle}{\|\Pi_{\text{trls}}\|_2} \\ &+ O\left([\mathbb{E}_{g \in \mathbf{G}} F(\delta(g), \mathcal{I})]^2\right) . \end{aligned} \quad (5.48)$$

Looking back at theorem 20 and using eqs. (5.42a), (5.42b) and (5.48) results in

$$C(V^\dagger) = 1 + O\left([\mathbb{E}_{g \in \mathbf{G}} F(\delta(g), \mathcal{I})]^2\right) . \quad (5.49)$$

Since both \mathcal{V} and $\mathcal{E}_L^{(4)}$ have at most infidelity of order $\mathbb{E}_{g \in \mathbf{G}} F(\delta(g), \mathcal{I})$, it follows that the composition $\mathcal{V} \mathcal{E}_L^{(4)}$ must also have an infidelity of order $\mathbb{E}_{g \in \mathbf{G}} F(\delta(g), \mathcal{I})$, which guarantees

$$\sqrt{1 - b^2(V^\dagger)} = O\left(\sqrt{\mathbb{E}_{g \in \mathbf{G}} F(\delta(g), \mathcal{I})}\right) , \quad (5.50)$$

while decoherence limitations guarantee

$$\sqrt{1 - a^2(V^\dagger)} = O\left(\mathbb{E}_{g \in \mathbf{G}} F(\delta(g), \mathcal{I})\right) . \quad (5.51)$$

Using

$$|w(V^\dagger)^\dagger \Delta v(V^\dagger)| \leq \mathbb{E} \|\delta(g)\|_\infty \leq O\left(\sqrt{\mathbb{E}_{g \in \mathbf{G}} F(\delta(g), \mathcal{I})}\right) \quad (5.52)$$

in eq. (5.36b), we find

$$D(1, V^\dagger) = O\left([\mathbb{E}_{g \in \mathbf{G}} F(\delta(g), \mathcal{I})]^2\right) , \quad (5.53)$$

which, together with eqs. (5.35) and (5.49) leads to

$$f_{\text{trls}}(\Lambda(\mathbf{G}), L(V^\dagger \mathbf{G} V), m) = p^m + O\left([\mathbb{E}_{g \in \mathbf{G}} F(\delta(g), \mathcal{I})]^2\right) . \quad (5.54)$$

This expression allows us to pick a better perturbative expansion than eq. (5.38). Indeed, choosing

$$\Lambda(g) = L(V^\dagger g V) + \delta(V^\dagger g V)L(V^\dagger g V), \quad (5.55)$$

ensures that the noisy operations $\mathcal{I} + \delta(V^\dagger g V)$ have a gate-set circuit infidelity which is more in line with the RB data:

$$\mathbb{E}_{g \in \mathbf{G}} F(\delta(V^\dagger g V), \mathcal{I}) = \frac{d-1}{d}(1-p) + O\left([\mathbb{E}_{g \in \mathbf{G}} F(\delta(g), \mathcal{I})]^2\right). \quad (5.56)$$

Iterating the analysis leads to

$$f_{\text{trls}}(\Lambda(\mathbf{G}), L(V^\dagger \mathbf{G} V), m) = p^m + O((1-p)^2). \quad (5.57)$$

This completes the demonstration of theorem 18.

Our current proof technique relies on the structure of single-qubit channels. For higher dimensions, we conjecture that an analog of theorem 18 holds, although the scaling with the dimension is unclear.

Conjecture 1: Polar decomposition for high-fidelity quantum channels

If the fidelity of $\mathcal{E}_R^{(4)}$ is high, then \exists a physical unitary \mathcal{V}^\dagger s.t. $\mathcal{E}_R^{(4)}\mathcal{V}^\dagger$ is decoherence-limited.

As we now show constructively, conjecture 1 holds for physically motivated noise models composed of generalized dephasing, amplitude damping, and unitary processes. Under such noise models,

$$\mathcal{E}_R^{(4)} = \mathcal{U}_T \mathcal{D}_T \cdots \mathcal{U}_2 \mathcal{D}_2 \mathcal{U}_1 \mathcal{D}_1 \quad (5.58)$$

for some unitaries \mathcal{U}_i and decoherence-limited channels \mathcal{D}_i .

The channel $\mathcal{U} \mathcal{D} \mathcal{U}^\dagger$ is decoherence-limited for any physical unitary \mathcal{U} , and the composition of decoherence-limited channels is also decoherence-limited, so eq. (5.58) can be rewritten as $\mathcal{E}_R^{(4)} = \mathcal{D} \mathcal{V}$, where \mathcal{D} and \mathcal{V} are decoherence-limited and unitary respectively:

$$\mathcal{D} = (\mathcal{U}_T \mathcal{D}_T \mathcal{U}_T^\dagger) \cdots (\mathcal{U}_{T:1} \mathcal{D}_1 \mathcal{U}_{T:1}^\dagger) \quad (5.59)$$

$$\mathcal{V} = \mathcal{U}_{T:1}. \quad (5.60)$$

Acknowledgments

The authors acknowledge helpful discussions with Timothy J. Proctor and Hui Khoon Ng. This research was supported by the U.S. Army Research Office through grant W911NF-14-1-0103. This research was undertaken thanks in part to funding from TQT, CIFAR, the Government of Ontario, and the Government of Canada through CFREE, NSERC and Industry Canada.

5.8 Afterword

5.8.1 Extending the results to all dimensions

Extending theorem 18 for d -dimensional systems follows from the work developed in chapter 4.

Theorem 21: The RB decay connects with a (physical) circuit fidelity

Let \mathbf{G} be a 2-design gate-set. Consider noisy gate-set physical realization $\Lambda(\mathbf{G})$ perturbed from $L(\mathbf{G})$, for which errors are equable in the strong sense and for which the relaxation bounding constant d_{relax} is small enough so that when invoking theorem 16,

$$\|\Lambda(g) - L(g)\|_{\infty} \leq O\left(\sqrt{\delta F(\Lambda(g), L(g))}\right). \quad (5.61)$$

Then, there exists a targeted gate-set realization $L(UGU^{\dagger})$, where $U \in U(d)$ correspond to a physical unitary, such that

$$F(\Lambda(\mathbf{G}), L(UGU^{\dagger}), m) = \frac{1}{d} + \frac{d-1}{d} p^m + O((1-p)^2), \quad (5.62)$$

where the decay constant p can be efficiently estimated via randomized benchmarking.

Notice that according to theorem 16, to violate eq. (5.61) while ensuring SSE errors, d_{relax} would have to be larger than $O((\delta F)^{-1})$, which is a very unrealistic error scenario. In the instance where the infidelity is on the order of 10^{-2} , d_{relax} would have to be significantly larger than a hundred. That is, violating eq. (5.61) would necessitate a process for which more than a hundred of orthogonal states would have comparable rates of transition to the same state (and for which the rates would be on the order of δF).

In other words, the condition stated by eq. (5.61) is ensured by ruling out extremal dephasers, extremal unitaries, and extremal transitions. Let me now proceed to the proof of theorem 21.

Proof. The extension of theorem 18 essentially follows from two results:

- i. The spectral norm of realistic channel deviations scales at most as $O(\sqrt{\delta F})$.
- ii. There exists a polar decomposition for (non-catastrophic) quantum channels of arbitrary dimensions.

Those two results are actually directly shown in chapter 4 (theorems 9 and 16 respectively), but the road between them and the generalization of theorem 18 might not be immediately clear to the reader.

Let me split the demonstration in two. First, since by assumption the spectral norm deviations scales at most as $O(\sqrt{\delta F})$, the m -fold noisy twirl converges very quickly to a fixed form. In other words, by Bauer-Fike theorem (or via first-order perturbation theory) the eigenvalues perturbed from 0 of the matrix

$$\mathcal{T} = \mathbb{E}_{g \in \mathbf{G}} [L_{\text{trls}}(g) \otimes \Lambda(g)] \quad (5.63)$$

are at most of order $\sqrt{\mathbb{E}_{g \in \mathbf{G}}[\delta F(\Lambda(g), L(g))]}$. In particular, that means that \mathcal{T}^4/p^4 is, up to order $(\mathbb{E}_{g \in \mathbf{G}}[\delta F(\Lambda(g), L(g))])^2$, a rank-1 projector. This simple realization allows to reformulate the statement of theorem 20 in a much more digestible way. That is, the gate-set circuit fidelity can be expressed as

$$F(\Lambda(\mathbf{G}), L(UGU^\dagger), m) = \frac{1}{d} + \frac{d-1}{d} (C(U)p^m + D(m, U)) , \quad (5.64)$$

where

$$C(U) = \frac{f_{\text{trls}}(\mathcal{E}_R^{(4)}U)f_{\text{trls}}(U^\dagger\mathcal{E}_L^{(4)})}{f_{\text{trls}}((\mathcal{E}_R^{(4)}U)(U^\dagger\mathcal{E}_L^{(4)}))} + O\left(\left(\mathbb{E}_{g \in \mathbf{G}}[\delta F(\Lambda(g), L(UGU^\dagger))]\right)^2\right) , \quad (5.65a)$$

$$|D(1, U)| \leq \sqrt{1 - \frac{u(\mathcal{E}_R^{(4)}U)}{f_{\text{trls}}^2(\mathcal{E}_R^{(4)}U)}} \sqrt{1 - \frac{u(U^\dagger\mathcal{E}_L^{(4)})}{f_{\text{trls}}^2(U^\dagger\mathcal{E}_L^{(4)})}} \sqrt{\mathbb{E}_{g \in \mathbf{G}}[\delta F(\Lambda(g), L(UGU^\dagger))]} . \quad (5.65b)$$

This ends the first part of the demonstration. The second part makes use of the polar decomposition of non-catastrophic quantum channels.

By picking U such that either $\mathcal{E}_R^{(4)}U$ or $U^\dagger\mathcal{E}_L^{(4)}$ is decoherent, the quasi-multiplicativity of the fidelity on the traceless hyperplane is ensured:

$$f_{\text{trls}}((\mathcal{E}_R^{(4)}U)(U^\dagger\mathcal{E}_L^{(4)})) = f_{\text{trls}}(\mathcal{E}_R^{(4)}U)f_{\text{trls}}(U^\dagger\mathcal{E}_L^{(4)}) + O\left(\left(\mathbb{E}_{g \in \mathbf{G}}[\delta F(\Lambda(g), L(UGU^\dagger))]\right)^2\right) , \quad (5.66)$$

which in turns implies that

$$C(U) = 1 + O\left(\left(\mathbb{E}_{g \in \mathbf{G}}[\delta F(\Lambda(g), L(UGU^\dagger))]\right)^2\right) . \quad (5.67)$$

The one-to-one correspondence between the unitarity and the fidelity of decoherent WSE channels implies

$$D(1, U) \leq O\left(\left(\mathbb{E}_{g \in \mathbf{G}}[\delta F(\Lambda(g), L(UGU^\dagger))]\right)^2\right) . \quad (5.68)$$

□

To phrase theorem 21 in less technical terms, the action of noisy 2-designs implicitly defines an operational reference basis for which the circuit fidelity takes a simple form:

$$F(\Lambda(\mathbf{G}), L(UGU^\dagger), m) \approx \frac{1}{d} + \frac{d-1}{d} p^m . \quad (5.69)$$

The decay constant can be efficiently estimated via a standard RB experiment, even in the advent of gate-dependent errors. In particular, for $m = 1$, this means that the outcome of a standard RB experiment really does directly connect with an average gate fidelity over the gate-set physical realization $\Lambda(\mathbf{G})$.

5.8.2 Comparing data from different reference bases

The derivation of theorem 21 brings an important subject of awareness (which is shortly discussed earlier in this chapter): experimental results arising from different gate-sets or from similar gate-sets subject to different error models may yield metrics implicitly defined through the scope of different targeted realizations. This doesn't mean that the results of two different RB experiments performed on different hardware cannot be compared: the obtained fidelities both translate how well certain operations can be performed between themselves. However, by design, an RB experiment performed over a gate-set \mathbf{G} doesn't provide much information about how well the physical realizations $\Lambda(g) \in \Lambda(\mathbf{G})$ connect with other sets of operations, unless those other sets share elements in common with $\Lambda(\mathbf{G})$. This yields an immediate question: what sufficient set of operations should two gate-sets share so that they implicitly define the same (or a very close) reference? This question becomes relevant when considering protocols such as iterated interleaved RB or simultaneous RB [Gam+12; Mag+12; She+16], where the results of different RB experiments are compared to infer some information about specific gates or about the level of cross-talk.

More generally, in the presence of gate-dependent errors, accounting for variations of the implicitly defined comparative gate-set realization $L(\mathbf{G})$ should take part – either in the data analysis or in the experimental design itself – of any sensitive comparison scheme; without such precaution, the interpretation of data could be compromised by gate-dependent effects that significantly alter the implicitly defined targeted realization mapping from one experiment to the next.

Chapter 6

Characterizing Universal Gate Sets via Dihedral Benchmarking

6.1 Foreword

The present chapter mainly consists in a literal transcription of [CWE15], for which my contribution was major.

6.2 Compendium

We describe a practical experimental protocol for robustly characterizing the error rates of non-Clifford gates associated with dihedral groups, including small single-qubit rotations. Our dihedral benchmarking protocol is a generalization of randomized benchmarking that relaxes the usual unitary 2-design condition. Combining this protocol with existing randomized benchmarking schemes enables practical universal gate sets for quantum information processing to be characterized in a way that is robust against state-preparation and measurement errors. In particular, our protocol enables direct benchmarking of the $\pi/8$ gate even under the gate-dependent error model that is expected in leading approaches to fault-tolerant quantum computation.

6.3 Introduction

A universal quantum computer is a device allowing for the implementation of arbitrary unitary transformations. As with any scenario involving control, a practical quantum computation will inevitably have errors. While the complexity of quantum dynamics is what enables the unique capabilities of quantum computation, including important applications such as quantum simulation and Shor’s factoring algorithm, that same complexity poses a unique challenge to efficiently characterizing the errors. One approach is quantum process tomography [CN97; PCZ97], which completely characterizes the errors on arbitrary quantum gates but requires resources that scale exponentially in the number of qubits. Moreover, quantum process tomography is sensitive to state-preparation and measurement (SPAM) errors [Mer+13].

Randomized benchmarking [EAZ05; Kni+08; MGE11; MGE12] using a unitary 2-design [Dan+09], such as the Clifford group, overcomes both of these limitations by providing an estimate of the error rate per gate averaged over the 2-design. More specifically, it is a method

for efficiently estimating the average fidelity

$$F(\mathcal{E}) := \int d\psi \langle \psi | \mathcal{E}(\psi) | \psi \rangle \quad (6.1)$$

of a noise map \mathcal{E} associated with any group of quantum operations forming a unitary 2-design in a way that is robust against SPAM errors. This partial information is useful in practice as it provides an efficient means of tuning-up experimental performance, and, moreover, provides a bound on the threshold error rate required for fault tolerant quantum computing [Got10] that becomes tight when the noise is stochastic [MGE12; WF14; SWS15; Puz+14; Mag+13; GB15].

An important limitation of existing randomized benchmarking methods is that they are only efficient in the number of qubits [EAZ05; Dan+09; MGE11] for non-universal sets of gates such as the Clifford group. While Clifford gates play an important role in many fault-tolerant approaches to quantum computation [Got10], one still needs a means of benchmarking an additional non-Clifford gate required for universality. One approach is to separately benchmark distinct unitary 2-designs [Bar+14]. While this approach is relatively straightforward for characterizing gates at the physical level, it is unclear how to apply this approach in the context of leading fault-tolerant proposals wherein particular non-Clifford operations required for universality, such as the $\pi/8$ gate, are implemented via magic state distillation and gate injection [BK05; MEK12], which is a complex procedure that will be subject to dramatically different error rates than those of the (physical or logical) Clifford gates. Alternatively, randomized benchmarking tomography [Kim+14] can be employed, although the fast decay curves can have a large uncertainty due to fitting an exponential to a small number of significant data points.

In the present paper, we describe a protocol for benchmarking the average fidelity of a group of operations corresponding to the dihedral group which does not satisfy the usual 2-design constraint for randomized benchmarking. However, we show that the dihedral benchmarking protocol still allows the average fidelity to be estimated while retaining many of the benefits of standard randomized benchmarking. Furthermore, by combining our dihedral benchmarking protocol with both standard [MGE11] and interleaved randomized benchmarking [Mag+12], we give an explicit method for characterizing the average fidelity of the $\pi/8$ gate directly. This is of particular interest because the $\pi/8$ gate combined with the generators of the Clifford group (e.g., the CNOT, the Hadamard and the Pauli gates) provides a standard gate set for generating universal quantum computation. Moreover our protocol enables characterization of non-Clifford gates associated with small angle rotations, which are of interest to achieve more efficient fault-tolerant circuits [LC13; For+15; DCP15]. Furthermore, the dihedral benchmarking protocol can be implemented either at the physical or logical level, but will find its greatest impact in the latter case, which is relevant to fault-tolerant quantum computation via magic-state injection. In that setting, the quality of a logical $\pi/8$ gate implemented via gate-injection will depend in a complex way on the quality of the input (distilled) magic state, the errors on the physical stabilizer operations, and the errors in the stabilizer measurements, all of which are required for the injection routine. Applying our protocol at the logical level provides a direct means of benchmarking the logical error rate of the injected $\pi/8$ gate, which may be dramatically different from the error rates achieved for the logical Clifford operations under the fault-tolerant encoding, without assessing the performance of the individual components. Remarkably, dihedral benchmarking

overcomes the key assumption of ‘weak gate-dependence’ of the noise that limits previous benchmarking protocols. Specifically, the protocol is robust in the important setting when the error on the non-Clifford gate, such as the $\pi/8$ gate, is substantially different from the error on the Clifford operations. As noted above, this is the expected scenario in leading approaches to fault-tolerant quantum computation.

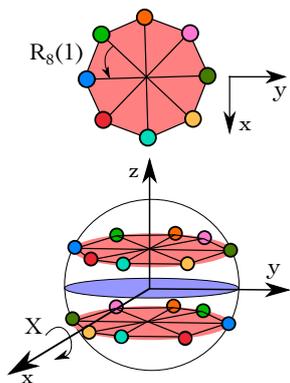


FIGURE 6.1: The orbit, under the action of the dihedral group \mathbf{D}_8 , of an input state located at a 45° degree latitude on the Bloch sphere. $R_8(z)$ are the rotations of the octagon, while X is a reflection (or a rotation in 3 dimensions, with the rotation axis parallel to the octagon’s surface). The $\pi/8$ gate corresponds to the smallest rotation $R_8(1)$.

6.4 Characterizing single-qubit unitary groups

We now outline a protocol that yields the average gate fidelity of the experimental implementation of a single-qubit unitary group of the form

$$\mathbf{D}_j = \langle R_j(1), X \rangle, \quad (6.2)$$

where $\langle \dots \rangle$ denotes the group generated by the arguments, j is a positive integer (or an arbitrary real number), and

$$R_j(z) := e^{\pi izZ/j} = \cos(\pi z/j)\mathbb{I} + i \sin(\pi z/j)Z. \quad (6.3)$$

Up to an overall sign, \mathbf{D}_j is a representation of the dihedral group of order $2j$, with $XR_j(z) = R_j(-z)X$, which is not a unitary 2-design and includes gates producing arbitrarily small rotations as j increases. Note that the choice of rotation axis is arbitrary, and that any single-qubit gate can be written as $R_j(1)$ relative to some axis. Consequently, our protocol will allow any single-qubit gate to be benchmarked. The Bloch sphere representation of \mathbf{D}_8 acting on a qubit state is illustrated in fig. 6.1. This group contains the so-called $\pi/8$ gate, which corresponds to the $R_8(1)$ rotation.

The dihedral benchmarking protocol for a fixed integer j is as follows. (Note that j can also be a real number, in which case the sums below are replaced by integrals.)

1. Choose two strings of length m , $\mathbf{z} = (z_1, \dots, z_m) \in \mathbb{Z}_j^m$ and $\mathbf{x} = (x_1, \dots, x_m) \in \mathbb{Z}_2^m$, independently and uniformly at random.
2. Prepare a system in an arbitrary initial state ρ .¹

¹The constants A and B appearing in eqns. 6.5 and 6.6 depend on state preparation, as shown in eqns. 6.26–6.28. These constants may be maximized by choosing an appropriate state preparation (and the corresponding measurement). In particular, optimal states for eqn. 6.5 and eqn. 6.6 are $|0\rangle\langle 0|$ and $|+\rangle\langle +|$ respectively.

3. At each time step $t = 1, \dots, m$, apply $R_j(z_t)X^{x_t}$.
4. Apply the recovery gate, defined as

$$g_{m+1} := X^{b_1}Z^{b_2} \prod_{t=1}^m [R_j(z_t)X^{x_t}]^\dagger,$$

where $b_1, b_2 \in \mathbb{Z}_2$ are fixed by considerations below.

5. Perform a POVM $\{\mu, \mathbb{I} - \mu\} \rightarrow \{+1, -1\}$ for some $\mu \approx \rho$, to estimate the probability $\Pr(+1|m, x, z, b_1, b_2)$ of outcome $+1$.
6. Repeat steps 1-5 k times, where k is fixed by the requirement to estimate the average recovery probability

$$p(m, b_1, b_2) := |2j|^{-(m+1)} \sum_{x,z} \Pr(+1|m, x, z, b_1, b_2)$$

to a desired precision (see [MGE12; GFC14; WF14] for details on the required sampling complexity).

For $b_1 = b_2 = 0$, the average recovery probability is

$$p(m, 0, 0) = A_0 + A_1 \lambda_1^m + A_2 \lambda_2^m, \quad (6.4)$$

where A_i are constants absorbing SPAM factors. Fitting two exponentials is generally more difficult than fitting a single exponential, so we propose instead fitting to

$$\begin{aligned} p(m, 0, 0) + p(m, 0, 1) \\ - p(m, 1, 0) - p(m, 1, 1) = 4A_2 \lambda_2^m \end{aligned} \quad (6.5)$$

and

$$p(m, 0, 0) - p(m, 0, 1) = 2A_1 \lambda_1^m. \quad (6.6)$$

As we will show below, the average gate fidelity is related to the fit parameters λ_1 and λ_2 by

$$F(\mathcal{E}_{\mathbf{D}_j}) = \frac{1}{2} + \frac{1}{6}(\lambda_2 + 2\lambda_1), \quad (6.7)$$

where $\mathcal{E}_{\mathbf{D}_j}$ is the noise over \mathbf{D}_j and we assume that the noise is completely positive and trace-preserving and is also gate and time-independent (though perturbative approaches to relax these assumptions can be considered [MGE12; WF14])

6.5 Characterizing the $\pi/8$ gate

The $\pi/8$ gate, or $R_8(1)$ in the notation of eq. (6.3), is important in many implementations because it is used to supplement the Clifford gates to achieve universal quantum computation. In leading approaches to fault-tolerant error-correction, the $\pi/8$ gate is physically realized via magic-state injection [BK05], in which magic states are acted upon by Clifford transformation and post-selected stabilizer measurements. Because the physical (logical) Clifford

gates are applied directly (transversally) whereas the $\pi/8$ gate is implemented through the above method, the error on the $\pi/8$ gate may be substantially different and requires separate characterization. While the quality of the injected gate can be assessed by measuring the quality of the input and output magic states as well as benchmarking the required stabilizer operations, here we provide a direct method to estimate the average gate fidelity of the $\pi/8$ gate.

The $\pi/8$ gate is contained in \mathbf{D}_8 (see eqn. 6.2), which contains \mathbf{D}_4 as a subgroup. One approach to characterizing the $\pi/8$ gate is to benchmark \mathbf{D}_4 and \mathbf{D}_8 separately. If the average fidelity over \mathbf{D}_8 and \mathbf{D}_4 are similar, this is an indication that the $\pi/8$ gate has similar average fidelity as the Clifford group. However, typically this will not hold for the reasons stated above, in which case we suggest the following protocol. First benchmark \mathbf{D}_4 as per the above protocol. Then adapt interleaved randomized benchmarking [Gam+12] to the above protocol by replacing steps 3 and 4 (with $j = 4$) with the two following steps:

3'. At each time step $t = 1, \dots, m$, apply $R_8(1)R_4(z_t)X^{x_t}$.

4'. Apply the recovery gate, defined as

$$g_{m+1} := X^{b_1} Z^{b_2} \prod_{t=1}^m [R_8(1)R_4(z_t)X^{x_t}]^\dagger.$$

We require the sequence length to be even to ensure that the recovery gate is in \mathbf{D}_4 , which follows from the commutation relation $XR_j(z) = R_j(-z)X$. For $b_1 = b_2 = 0$, the average recovery probability is similar to eqn. 6.4, but with different decay parameters:

$$p(m, 0, 0) = A'_0 + A'_1 \lambda_1'^m + A'_2 \lambda_2'^m + . \quad (6.8)$$

As in the previous section, the curves

$$\begin{aligned} p(m, 0, 0) + p(m, 0, 1) \\ - p(m, 1, 0) - p(m, 1, 1) = 4A'_2 \lambda_2'^m \end{aligned} \quad (6.9)$$

and

$$p(m, 0, 0) - p(m, 0, 1) = 2A'_1 \lambda_1'^m \quad (6.10)$$

can be fitted instead to extract the decay parameters. The average gate fidelity of the composed noise channel $\mathcal{E}_{\pi/8} \circ \mathcal{E}_{\mathbf{D}_4}$ (where $\mathcal{E}_{\pi/8}$ is the noise on the $\pi/8$ gate and \circ denotes channel composition) is related to the fit parameters λ_1' and λ_2' by

$$F(\mathcal{E}_{\pi/8} \circ \mathcal{E}_{\mathbf{D}_4}) = \frac{1}{2} + \frac{1}{6}(\lambda_2' + 2\lambda_1'). \quad (6.11)$$

The average gate fidelity of the $\pi/8$ gate, $F(\mathcal{E}_{\pi/8})$ (as opposed to the fidelity of the composite noise channel), can then be estimated from the relation [Kim+14; CWE16]

$$\begin{aligned} & |\Phi(\mathcal{E}_{\pi/8}) - \Phi(\mathcal{E}_{\mathbf{D}_4})\Phi(\mathcal{E}_{\pi/8} \circ \mathcal{E}_{\mathbf{D}_4}) - (1 - \Phi(\mathcal{E}_{\mathbf{D}_4}))(1 - \Phi(\mathcal{E}_{\pi/8} \circ \mathcal{E}_{\mathbf{D}_4}))| \\ & \leq 2\sqrt{\Phi(\mathcal{E}_{\mathbf{D}_4})\Phi(\mathcal{E}_{\pi/8} \circ \mathcal{E}_{\mathbf{D}_4})(1 - \Phi(\mathcal{E}_{\mathbf{D}_4}))(1 - \Phi(\mathcal{E}_{\pi/8} \circ \mathcal{E}_{\mathbf{D}_4}))}, \end{aligned} \quad (6.12)$$

where

$$\Phi(\mathcal{E}) := \frac{3}{2}F(\mathcal{E}) - \frac{1}{2} \quad (6.13)$$

is the process fidelity. This bound is loose in general but tight when the Clifford gates in \mathbf{D}_4 have much higher fidelity than the $\pi/8$ gate (which is the regime of interest when optimizing the overhead and fidelity of the distillation and injection routines) [Fow+12].

Smaller rotations can also be characterized in a similar fashion. The average fidelity of a small rotation $R_J(1)$ is estimated by implementing the same scheme, replacing $\pi/8$ with $R_J(1)$, \mathbf{D}_4 with \mathbf{D}_j such that $2N \cdot j = J$ for any fixed choice of $N \in \mathbb{N}$, and by restricting the sequence lengths to be multiples of $2N$.

6.6 Analysis

We now derive the formula for the decay curves expressed in eqns. 6.4–6.6, 6.8–6.10 together with the average fidelity eqns. 6.7 and 6.11. For convenience, we will use the Pauli-Liouville representation of channels in which channel composition corresponds to matrix multiplication (see, e.g., Ref. [WF14] for details). The Pauli-Liouville representation of an abstract channel \mathcal{E} , which we denote with the same letter \mathcal{E} (the context suffices to differentiate the matrix from the abstract channel), is the matrix of trace inner products between Pauli matrices P_j and their images $\mathcal{E}(P_k)$ under \mathcal{E} ,

$$\mathcal{E}_{jk} = \text{Tr}(P_j \mathcal{E}(P_k)). \quad (6.14)$$

We assume that the experimental noise is completely positive and trace-preserving (CPTP) and is also gate and time-independent (though perturbative approaches to relax these assumptions can be considered [MGE12; WF14]), so that we can represent the experimental implementation of $R_J(1)R_j(z)X^x$ as

$$\Lambda_J(R_J(1)) \circ \Lambda_{\mathbf{D}_j}(R_j(z)X^x), \quad (6.15)$$

where $\Lambda_J, \Lambda_{\mathbf{D}_j}$ are maps from $U(2)$ to noisy operations in CPTP_2 :

$$\Lambda_J(R_J(1)) = L(R_J(1)) \circ \mathcal{E}_{R_J(1)} \quad (6.16a)$$

$$\Lambda_{\mathbf{D}_j}(R_j(z)X^x) = \mathcal{E}_{\mathbf{D}_j} \circ L(R_j(z)X^x), \quad (6.16b)$$

and $L : U(d) \rightarrow \text{CPTP}_2$ is implicitly defined as

$$L(U)[\rho] = U\rho U^\dagger. \quad (6.17)$$

$\mathcal{E}_{R_J(1)}, \mathcal{E}_{\mathbf{D}_j} \in \text{CPTP}_2$ are quantum error channels.

The standard \mathbf{D}_j benchmarking protocol can be obtained by setting $J = 1$ and $\mathcal{E}_{R_J(1)} = \mathbb{I}_4$, while the interleaved case corresponds to $J = 2N \cdot j$.

For $j > 2$, the Pauli-Liouville representation of $L(R_j(z)X^x)$ is block diagonal with three

blocks, where the blocks corresponds to three inequivalent irreducible representations (irreps) of the dihedral group, namely,

$$\begin{aligned}
\phi_0(R_j(z)X^x) &= 1 && \text{(trivial representation)} \\
\phi_1(R_j(z)X^x) &= \begin{pmatrix} \cos(2\pi z/j) & (-1)^{x+1} \sin(2\pi z/j) \\ \sin(2\pi z/j) & (-1)^x \cos(2\pi z/j) \end{pmatrix} && \text{(faithful representation)} \\
\phi_2(R_j(z)X^x) &= (-1)^x. && \text{(parity representation)}
\end{aligned}$$

This is easily seen by looking at the action of \mathbf{D}_j on the Bloch sphere (see fig. 6.1). The trivial representation emerges from the unitality and trace-preserving properties of unitary operations, which map any Bloch shell of constant radius to itself, including the center point. The parity representation encodes the fact that the $\pm Z$ poles of the Bloch sphere are invariant under conjugation by $R_j(z)$ and swapped under conjugation by X . The two-dimensional representation encodes the action of $R_j(z)X^x$ on the XY -plane of the Bloch sphere.

The *twirl* of a channel \mathcal{E} over a group \mathbf{G} is defined as

$$\mathcal{E}^{\mathbf{G}} = |\mathbf{G}|^{-1} \sum_{g \in \mathbf{G}} L(g^{-1}) \mathcal{E} L(g). \quad (6.18)$$

As a consequence of Schur's lemmas (see the supplementary information of Ref. [Gam+12]), the twirl of any channel over \mathbf{D}_j is

$$\mathcal{E}^{\mathbf{D}_j} = (\mathcal{E}_{R_j(1)} \mathcal{E}_{\mathbf{D}_j})^{\mathbf{D}_j} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda'_1 & 0 & 0 \\ 0 & 0 & \lambda'_1 & 0 \\ 0 & 0 & 0 & \lambda'_2 \end{pmatrix} = \phi_0(\mathbb{I}) \oplus \lambda'_1 \phi_1(\mathbb{I}) \oplus \lambda'_2 \phi_2(\mathbb{I}) \quad (6.19)$$

for $j > 2$, where $\lambda'_2 := \mathcal{E}_{33}$ and $\lambda'_1 := \frac{\mathcal{E}_{11} + \mathcal{E}_{22}}{2}$ and the diagonal blocks correspond to the three inequivalent irreps of \mathbf{D}_j . Defining

$$\begin{aligned}
\mathcal{A}_m &:= (2j)^{-m} \sum_{x,z} \prod_{i=m}^1 L(R_j(1)) \mathcal{E}_{R_j(1)} \mathcal{E}_{\mathbf{D}_j} L(R_j(z_i)X^{x_i}) \\
\mathcal{B}_m &:= (2j)^{-m} \sum_{x,z} \prod_{i=1}^m L(X^{x_i} R_j(z_i)^\dagger R_j(1)^\dagger),
\end{aligned} \quad (6.20)$$

the average over all sequences of length m can be expressed as the effective channel

$$\mathcal{C} = \mathcal{E}_{\mathbf{D}_j} L(X^{b_1} Z^{b_2}) \mathcal{B}_m \mathcal{A}_m. \quad (6.21)$$

But $\mathcal{B}_m \mathcal{A}_m$ can be re-expressed as

$$\begin{aligned}
\mathcal{B}_m \mathcal{A}_m &= \mathcal{B}_{m-1} (\mathcal{E}_{R_j(1)} \mathcal{E}_{\mathbf{D}_j})^{\mathbf{D}_j} \mathcal{A}_{m-1} \\
&= (\mathcal{E}_{R_j(1)} \mathcal{E}_{\mathbf{D}_j})^{\mathbf{D}_j} \mathcal{B}_{m-1} \mathcal{A}_{m-1} \\
&= \prod_{j=1}^m (\mathcal{E}_{R_j(1)} \mathcal{E}_{\mathbf{D}_j})^{\mathbf{D}_j},
\end{aligned} \quad (6.22)$$

where the second line follows from the fact that $\mathcal{E}^{\mathbf{D}_j}$ is proportional to the identity in each of the blocks.

With these definitions, the average fidelity is [Nie02; Kim+14]

$$\begin{aligned} F(\mathcal{E}) &= \frac{1}{2} + \frac{1}{6}(\mathcal{E}_{11} + \mathcal{E}_{22} + \mathcal{E}_{33}) \\ &= \frac{1}{2} + \frac{1}{6}(\lambda'_2 + 2\lambda'_1) \end{aligned} \quad (6.23)$$

as in eqn. 6.7 and 6.11. Using eqn. 7.14, the effective channel \mathcal{C} from eqn. 6.21 can readily be expressed as

$$\mathcal{C} = \mathcal{E}_{\mathbf{D}_j} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & (-1)^{b_2} \lambda_1'^m & 0 & 0 \\ 0 & 0 & (-1)^{b_1+b_2} \lambda_1'^m & 0 \\ 0 & 0 & 0 & (-1)^{b_1} \lambda_2'^m \end{pmatrix}. \quad (6.24)$$

Therefore the recovery probability is

$$\begin{aligned} p(m, b_1, b_2) &= \text{Tr}(\mu \mathcal{C}(\rho)) \\ &= (-1)^{b_1} A_2 \lambda_2'^m + \left((-1)^{b_1+b_2} B_1 + (-1)^{b_2} B_2 \right) \lambda_1'^m + A_0, \end{aligned} \quad (6.25)$$

where

$$A_2 := 2^{-1} \cdot \text{Tr}(\mu \cdot \mathcal{E}_{\mathbf{D}_j}(Z)) \cdot \text{Tr}(\rho Z), \quad (6.26)$$

$$B_1 := 2^{-1} \cdot \text{Tr}(\mu \cdot \mathcal{E}_{\mathbf{D}_j}(Y)) \cdot \text{Tr}(\rho Y), \quad (6.27)$$

$$B_2 := 2^{-1} \cdot \text{Tr}(\mu \cdot \mathcal{E}_{\mathbf{D}_j}(X)) \cdot \text{Tr}(\rho X), \quad (6.28)$$

$$A_0 := 2^{-1} \cdot \text{Tr}(\mu \cdot \mathcal{E}_{\mathbf{D}_j}(\mathbb{I})). \quad (6.29)$$

Eqns. 6.4–6.6, 6.8–6.10 then follow from appropriate choices of b_1, b_2 and simple algebra.

6.7 Numerical simulation

Although the previous analysis is derived for gate- and time-independent noise, the randomized benchmarking protocol is both theoretically and practically robust to some level of gate-dependent noise [MGE12; WF14]. We now illustrate through numerical simulations that this robustness holds for the dihedral benchmarking protocol, particularly in the regime where the noise is strongly gate-dependent (as expected when the gates are implemented using different methods, namely, direct unitaries and magic state injection).

For our simulations, each operation within the dihedral group \mathbf{D}_8 is generated by composing two gates; the first from \mathbf{D}_4 and the second is either identity or the $\pi/8$ gate. The error associated with the first gate is a simple depolarizing channel with an average fidelity of 0.9975. For the second gate, the error arises only after the $\pi/8$ gate, and corresponds to an over-rotation with an average fidelity of 0.99. The total average fidelity over \mathbf{D}_8 is 0.9925. Fig. 6.2 shows the two decay curves described by eqns. 6.5 and 6.6. Weighted non-linear

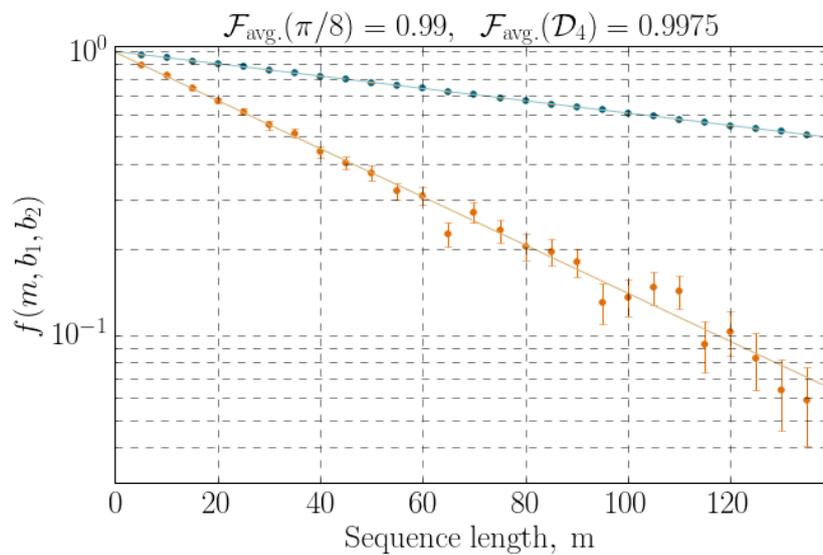


FIGURE 6.2: (Color online) Decay curves corresponding to eqns. 6.5 and 6.6 for a standard randomized benchmarking simulation with $A_2 \approx \frac{1}{4}$ and $A_1 \approx \frac{1}{2}$ respectively. The shallow (blue) and steep (orange) lines correspond to eqn. 6.5 and eqn. 6.6 respectively. Each data point is obtained after averaging 500 sequences of fixed length. A weighted non-linear regression (performed using the `scipy` Python package) gives an estimate of 0.9924(1) for the average fidelity over \mathcal{D}_8 , compared to the analytic value 0.9925. See *Numerical simulation* for details.

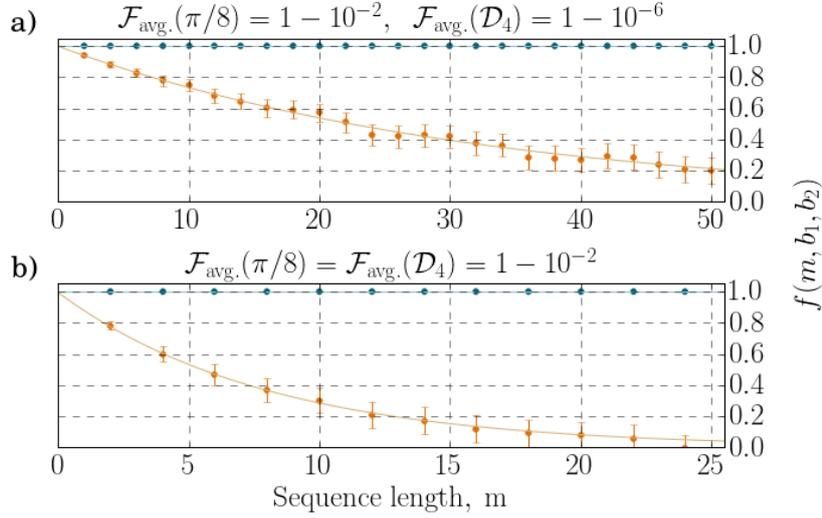


FIGURE 6.3: Decay curves corresponding to eqns. 6.5 and 6.6 for an interleaved randomized benchmarking simulation with $A_2 \approx \frac{1}{4}$ and $A_1 \approx \frac{1}{2}$ respectively. The shallow (blue) and steep (orange) lines correspond to eqn. 6.5 and eqn. 6.6 respectively. Each data point is obtained after averaging 500 sequences of fixed length. Figure a) corresponds to high fidelity Clifford operations and a relatively noisy $\pi/8$ gate. Figure b) corresponds to errors of the same magnitude on the Clifford operations and the $\pi/8$ gate. See *Numerical simulation* for details.

regressions give an estimate of 0.9924(1) for the average fidelity, which is consistent with the analytic value. We also simulate the interleaved randomized benchmarking protocol in two different regimes (see fig. 6.3). The first regime (fig. 6.3a) corresponds to over-rotation errors that are small for the Clifford operations, with average fidelity $1 - 10^{-6}$, but large for the $\pi/8$ gate, with average fidelity $1 - 10^{-2}$. The estimate of the fidelity of the $\pi/8$ gate via our protocol, 0.9901(2), is extremely precise in this regime. The second regime (fig. 6.3b) corresponds to a similar over-rotation with average fidelity 0.99 both for the Clifford group and the $\pi/8$ gate. In this case the estimated value of $F(\mathcal{E}_{\pi/8})$ is 0.980 and the bound from eqn. 6.12 only guarantees $F(\mathcal{E}_{\pi/8})$ to lie the interval $[0.958, 1.000]$. The rather loose bound in this regime is an open problem for interleaved randomized benchmarking and is not specific to the current protocol.

6.8 Conclusion

We have provided a protocol that extracts the average fidelity of the error arising over a group of single-qubit operations corresponding to the dihedral group. While we have explicitly assumed that the rotation axis is the z axis, this is an arbitrary choice. Since any single-qubit unitary can be written as a rotation about some axis on the Bloch sphere, our protocol can be used to characterize any single-qubit gate.

Of particular importance are \mathbf{D}_8 and \mathbf{D}_4 , which enable efficient and precise benchmarking of the $\pi/8$ gate that plays a unique and important role in leading proposals for fault-tolerant quantum computation. We have illustrated numerically that the fidelity of the $\pi/8$ gate can

be estimated using an interleaved version of our protocol. This estimate is precise when the quality of Clifford gates is significantly greater than the $\pi/8$ gate, which is a regime relevant to near-term small-scale demonstrations of universal fault-tolerant quantum computation where Clifford operations are performed transversally while the quality of the $\pi/8$ gate is limited by the relatively high cost of magic state distillation.

Acknowledgments— This research was supported by the U.S. Army Research Office through grant W911NF-14-1-0103, CIFAR, the Government of Ontario, and the Government of Canada through NSERC and Industry Canada.

6.9 Afterword

6.9.1 Group-based randomized benchmarking

Dihedral benchmarking was the first RB protocol that suggested using a family of gate-sets that do not form a 2-design, but that certainly exhibit a group structure. The idea can be naturally extended to other groups [Cro+16; Hel+18; SH18; OWE18; BE18; Erh+19]. Essentially, by performing a twirl over a group structure, one can immediately invoke Schur’s orthogonality relations to simplify the structure of a given error channel (for a group 2-design, errors are transformed into simple depolarizing channels). The simplified channel is then typically repeated, and a signal is extracted (the average recovery probability). Since the repeated simplified channel has few parameters, it is generally possible to tie the signal to a fitting model, and estimate desired error parameters.

Consider the simpler case of a randomizing gate-set \mathbf{G} that forms a group (here \mathbf{G} is a subgroup of $U(d)$) and for which the irreps composing the Liouville representation of ideal operations remain inequivalent. The Liouville matrix of a group element $g \in \mathbf{G}$ is equivalent to a direct sum of inequivalent irreps ϕ_i ,

$$L(g) = \mathcal{B} \left(\bigoplus_i \phi_i(g) \right) \mathcal{B}^{-1}, \quad (6.30)$$

where \mathcal{B} is some invertible $d^2 \times d^2$ matrix. It follows immediately from Schur’s orthogonality relations that a twirl of a error matrix \mathcal{E} over the group \mathbf{G} would yield:

$$\mathcal{E}^{\mathbf{G}} = \frac{1}{|\mathbf{G}|} \sum_{g \in \mathbf{G}} L(g) \mathcal{E} L(g^{-1}) = \mathcal{B} \left(\bigoplus_i \lambda_i \phi_i(e) \right) \mathcal{B}^{-1}, \quad (6.31)$$

where

$$\lambda_i = \frac{\langle \mathcal{B} (0 \oplus \dots \oplus 0 \oplus \phi_i(e) \oplus 0 \oplus \dots \oplus 0) \mathcal{B}^{-1}, \mathcal{E} \rangle}{\text{Tr } \phi_i(e)}. \quad (6.32)$$

By repeating the twirled channel many times, as depicted in the gate-independent analysis of group-based RB protocols, and by choosing specific state preparation and measurement procedures, one typically get signal which consists in a sum of exponentials:

$$\langle \mu | (\mathcal{E}^{\mathbf{G}})^m | \rho \rangle = \sum_i A_i(\rho, \mu) \lambda_i^m. \quad (6.33)$$

The model can get a little more intricate, but not overbearingly so, when the Liouville representation of the randomizing group contains equivalent irreps [OWE18]. The process fidelity of \mathcal{E} to the identity is retrieved via

$$\Phi(\mathcal{E}, \mathcal{I}) = \sum_i \lambda_i \frac{\text{Tr} \phi_i(e)}{\text{Tr} L(e)}. \quad (6.34)$$

This means that one typically want to either estimate all the λ_i s, or sample from them with weights $w_i = \text{Tr} \phi_i(e)/d^2$. Of course, one might be interested in the λ_i s themselves for the sake of noise reconstruction techniques.

To estimate the λ_i s, the strategy is typically to vary over different SPAM procedures (i.e. (ρ, μ) pairs) in order to put emphasis on different decays or different linear combinations of decays. Recall that in dihedral benchmarking, the recovery operation is the sequence inverse supplemented by a Pauli of the form $X^{b_1} Z^{b_2}$. This extra (compiled) Pauli in the last step can be seen as taking part of the measuring procedure, that is taking μ to $\mu(b_1, b_2) = X^{b_1} Z^{b_2} \mu Z^{b_2} X^{b_1}$. After gathering enough data points from eq. (6.33), by changing the sequence length m and the SPAM procedures (ρ, μ) , $\Phi(\mathcal{E}, \mathcal{I})$ is estimated. For instance, in dihedral benchmarking, multiple types of SPAM procedures are performed and specific linear combinations of the respective recovery probabilities are considered in order to isolate the decays. That is, the fitting model eq. (6.33) is effectively broken up in simpler fitting models of the form

$$f_i(m) = A_i' \lambda_i^m. \quad (6.35)$$

Fitting a single exponential is typically much easier and stable than fitting a sum of exponentials for which the individual decays are close to each other. The trick of isolating the decays that appears in dihedral benchmarking can be generalized for other groups by simply appealing to character orthogonality relations [Hel+18].

6.9.2 Generator-based randomized benchmarking

Group-based RB borrows most of its tools from the representation theory of groups, and in particular from Schur's orthogonality relations (including character theory [Hel+18]). While the algebraic richness of group structures guarantees important simplifications of error accumulation in random circuits, it also typically demands more and more elaborate circuit constructions as the system dimension grows in size². This in turns complexifies the implementation of RB experiments, as their realization requires the design of efficient gate samplers and gate compilers. This is not necessarily a fundamental problem. For instance, there exist efficient recipes for constructing Clifford gate compilers and samplers. Moreover, designing gate compilers, despite being demanding, should be seen as an essential part in the realization of a computing device. The more fundamental reason to consider algebraically lighter randomizing schemes is twofold:

²This of course depends on the group and on the set of primitive operations. In some architectures, the n -qubit Pauli group might be directly obtained from native operations.

- i. By design, group operations have lower fidelities than the primitive operations that generate them. In most proposed architectures, for which the qubit connectivity is limited, the discrepancy between those fidelities could grow considerably with the system's dimension. Since circuits composed of m random group elements will see their fidelity drop more rapidly than sequences of m primitive operations, one can imagine a regime where the signal given by group-based RB, eq. (6.33), would be very faint in comparison to the signal coming from a generator-based random sequence. Designing RB experiments based on weaker algebraic structures would allow the possibility to characterize devices in regimes where more traditional group-based RB would be useless.
- ii. The connection between primitive operations and the result of group-based RB experiments could become fainter as circuit constructions for generating group elements gain in depth. Interleaving primitive operations with group elements would yield single-gate fidelity estimates subject to the bounds derived in chapter 3, which would render the estimate completely meaningless as the infidelity of the randomizing set gets larger.

These concerns, which are only valid in specific (but relevant) regimes, could be alleviated by simplifying the structure of the randomizing gate-set [Pro+18; SH18; Erh+19]. Notice that the work of [Pro+18; SH18] suggest a generator-based approach, which differs from the scalable group-based technique known as cycle benchmarking, developed in [Erh+19]. Despite invoking group structures, cycle benchmarking is not vulnerable to the two previous scalability concerns since the randomizing group (the n -fold Pauli group) connects directly with a set of generators, and doesn't require intricate constructions as the system size increases. In the next chapter, I will introduce some original elements of theory regarding generator-based RB under gate-dependent effects.

Chapter 7

Generator-based randomized benchmarking

7.1 Foreword

This chapter contains two related projects regarding generator-based RB protocols.

The first one, in which I heavily contributed, concerns the analysis of a well-known RB protocol implemented in 2008 that was using a randomizing gate-set that was neither a 2-design nor a group [Boo+19]. The protocol can in fact be cast as a simple case of generator-based RB experiment. Notice here that the term “generator” refers to the generators of a group, and doesn’t need to relate to native gates (although primitive gates form a generating set). The simplicity of the analysis provided in [Boo+19] is a good stepping stone to reason about more advanced generator-based RB protocols.

The second (multipart) project concerns a generator-based RB protocol labeled as “direct randomized benchmarking” (DRB) [Pro+18]. The protocol itself, as well as some analysis techniques were mostly lead by Dr. Timothy J. Proctor. However, I did take part in reasoning about “sequence-asymptotic 2-designs” (a notion that differs from “asymptotic 2-designs” introduced by of Gross *et al.* [GAE07]), and about the analysis of the DRB signal under general realistic gate-dependent effects. Hence, in this chapter, I will briefly introduce the protocol and focus mainly on its signal analysis without covering the statistical considerations regarding the fitting procedures.

7.2 Randomized Benchmarking under Different Gate-sets

The present section consists in a literal transcription of the first (theory oriented) sections of [Boo+19], for which my contribution was major. I only left the discussion regarding the gate-independent analysis of “NIST RB”.

7.2.1 Compendium

We provide a comprehensive analysis of the differences between two important standards for randomized benchmarking (RB): the Clifford-group RB protocol proposed originally in Emerson et al (2005) and Dankert et al (2006), and a variant of that RB protocol proposed later by the NIST group in Knill et al, PRA (2008). Our analysis provides an important first step towards developing definitive standards for benchmarking quantum gates and a more

rigorous theoretical underpinning for the NIST protocol and other RB protocols lacking a group-structure.

7.2.2 Introduction

Clifford-group randomized benchmarking (RB) [EAZ05; Dan+06; Dan+09] has become the *de facto* standard tool for assessing and optimizing the quantum control required for quantum computing systems by estimating error rates associated with sets of elementary gates operations. It has been known for some time that this protocol leads to an invariant exponential decay [EAZ05; Dan+06; Lev+07; Dan+09] because it is equivalent to a sequence of twirls [Lev+07] with unitary-two designs [Dan+06; Dan+09].

More recently, the robustness of the Clifford-group RB protocol has been supported by a rigorous theoretical framework, including proofs that an exponential fidelity decay will be observed under very broad experimental conditions, including essentially arbitrary state preparation and measurement errors [MGE11; MGE12] and gate-dependent errors [Wal17], as well as proofs that the observed error rate relates directly to a well-defined notion of gate-fidelity [Wal17; Pro+17a; Car+18], which fully overcome recent concerns about relating measured RB error rates to a meaningful concept of gate-fidelity under gate-dependent errors [Pro+17a].

While a wide-variety of group-based generalizations of RB have been proposed in recent years, *e.g.* [Bar+14; WF14; CWE15; WBE15; Cro+16; WBE16; Wal17; OWE18; SH18; Hel+18], in this Letter we focus on clarifying the physical relevance of a standing conflation in the literature between the now standard Clifford-group RB protocol proposed in [EAZ05; Dan+06; Dan+09] and an alternate version of RB proposed later by NIST [Kni+08]. As described below, these are distinct protocols that measure distinct properties of the error model and thus can produce different error rate estimates under the same, realistic experimental conditions. Moreover, because the NIST protocol does not admit a closed-group or unitary two-design structure, the rigorous theoretical framework justifying Clifford-group RB does not trivially extend to support the physical interpretation and robustness of NIST RB.

In this Letter we identify the operationally-relevant differences between the Clifford-group RB protocol and the NIST version of RB which clarifies how they can lead to very different error rate estimates given the same error model (as defined in terms of the elementary control pulses). We then provide the first rigorous proof that the NIST RB protocol does indeed produce an exponential decay under gate-independent error models. This is an important step toward developing a theoretical justification for the NIST protocol and other RB protocols that do not admit a group-structure in the case of gate-dependent errors and the ultimate goal of a theoretical framework within which error reconstruction under RB protocols with different gate sets can be extracted in a unified and consistent manner. Our analysis is thus also essential for comparing cross-platform benchmarking methods and standards for quantum computing.

7.2.3 Background and Motivation

The original proposal for randomized benchmarking from Emerson *et al.* [EAZ05] considered implementing long sequences of quantum gates drawn *uniformly* at random from the group $SU(d)$ for any quantum systems with Hilbert space dimension d .

That work proved that the measured fidelity would follow an exponential decay with a decay rate that is fixed uniquely by the error model, that is, the measured decay rate would not depend on the choice of initial state or the specific random quantum gate sequences.

This protocol suffered from two limitations: the random gates were drawn from a continuous set, which is impractical even for $d = 2$, and the protocol would not be efficient for large systems because a typical random element of $U(d)$ requires exponentially long gate sequences under increasing numbers of qubits. Additionally, in that limit the inversion gate may not be computed efficiently.

However, practical and efficient solutions to both of these problems were proposed in Dankert *et al.* [Dan+06; Dan+09] in 2006, which proved and observed that drawing gates uniformly at random from the Clifford group would lead to the same exponential decay rate as computed in the protocol proposed earlier in Emerson *et al.* [EAZ05], which follows from the unitary 2-design property of the Clifford group.

This connection is made more explicit through the observation that a random sequence of gates drawn from any group is equivalent to an independent sequence of twirls under that group, as shown explicitly in [Lev+07] and had been conjectured earlier in [EAZ05].

Collectively these papers define what is now known as Clifford-group RB, an efficient and practical method for assessing error rates for quantum processors on arbitrarily large numbers of qubits, summarized here as Protocol 1. This Clifford-group RB protocol has become a *de facto* standard for benchmarking and optimizing gate performance and has been implemented by a large number of groups across various hardware platforms to characterize single- and multi-qubit gate operations, see, *e.g.*, Refs [Muh+15; Kel+14; Bar+14; McK+17b; She+16; Xia+15; Tar].

The theoretical underpinnings of the standard protocol were clarified and further developed by Magesan *et al.* [MGE11; MGE12], which showed that the exponential decay rate was robust to state preparation and measurement errors (SPAM), and by Wallman [Wal17] and Dugas *et al.* [Car+18], which showed that the exponential decay rate was meaningfully related to a gate-fidelity in spite of the gauge freedom highlighted by Proctor *et al.* [Pro+17a] that occurs in the usual definition of the average gate-fidelity.

Additionally, the work of Wallman [Wal17] established that the RB error rate is robust to very large variations in the error model over the gate set (known as gate-dependent error models) and thus established that RB can also be an effective tool for diagnosing non-Markovian errors. This follows from the fact that only non-Markovian errors (including what are sometimes called time-dependent Markovian errors) can produce a statistically significant deviation from an exponential decay under a Clifford-group RB experiment.

A different version of the 2005 Emerson *et al.* [EAZ05] protocol was proposed by Knill *et al.* [Kni+08] in 2008 and implemented in the NIST ion trap. This proposal involved the same kind of motion reversal experiment proposed in Emerson *et al.* [EAZ05] but selects a random sequences of gates drawn from a *non-uniform* sampling of the single-qubit Cliffords, defined as "Pauli-randomized $\pi/2$ gates". The precise recipe for this protocol is summarized as Protocol 2. The NIST version of the randomized benchmarking protocol continues to be implemented mainly in ion traps [Bro+11; Har+14]. We note that in contrast to the earlier Clifford-group RB protocol which is defined for single- and multi-qubit gate operations, the NIST version of RB is defined only for single-qubit gate operations.

In this Letter, we prove that the measured fidelity under the NIST protocol will follow an exponential decay, which has never been established for this protocol, and relate the decay

rate to the intrinsic properties of the error model, demonstrating how it differs from the properties measured by Clifford-group RB. This analysis also provides first step towards developing a self-consistent theoretical framework for interpreting and relating the results of the large and growing family of RB-style protocols, which all share the structure of applying random sequences of gates and differ mainly through the choice of which random gate-sets [CWE15; Cro+16; WBE16; CW15; WBE15; Gam+12; Bar+14; Pro+18; Kim+14; Eme+07; MGE12; She+16; Wal17; WG18; Xia04].

Finally, an additional motivation for the present work comes from the recent conceptual development [WE16] establishing how accurately RB error estimation methods can inform the design and ‘in vivo’ performance of large-scale quantum computations. This development overcomes a standing criticism of RB protocols that the very nature of a randomization protocol limits these protocols to detect only the stochastic component of coherent errors - and hence that RB-type protocols are not able to capture the full impact of these errors. Coherent errors are those that typically arise from imperfect quantum control due to residual mis-calibrations ¹ and pose a major challenge for reliable quantum computation. However, this perceived limitation has become a strength of RB protocols thanks to the concept of randomized compiling [WE16].

Randomized compiling is an important generalization and improvement to the concept of Pauli-Frame Randomization (PFR) proposed earlier in [Kni05] that does not require any overhead for the randomization and works for universal gate sets ². When implementing a quantum algorithm via randomized compiling, the only performance limiting component of a coherent error is precisely the stochastic component that is detected via RB protocols. In summary, a precise and accurate understanding of RB error estimates is highly relevant because RB detects precisely the component of the error that determines the ‘in vivo’ performance of the gate operations within a large-scale circuit performed via randomized compiling.

¹Note that cross-talk is a non-trivial coherent error that results from control errors affecting distant qubits.

²In particular, relative to PFR, randomized compiling (i) does not add additional overhead to each clock cycle, which it achieves by ‘compiling in’ the randomizing gates, (ii) works for universal gate sets, and (iii) rigorously characterizes how close the effective error model is to a purely stochastic error model under errors gate-dependent errors

7.2.4 Results

Standard RB vs NIST RB

The standard RB protocol (SRB) [EAZ05; Dan+06; Dan+09] is summarized in protocol 1.

Protocol 1: Standard Clifford-group RB, as described in [EAZ05; Dan+06; Dan+09].

1. Sample a set of m gates $\{g_1, \dots, g_m\}$ picked independently and uniformly at random from the Clifford group \mathbf{C} defined in eq. (7.3);
2. Determine the recovery gate g_{m+1} (see text below);
3. Prepare a state $\rho \approx |0\rangle\langle 0|$;
4. Perform the sampled gates from step 1, followed by the recovery gate g_{m+1} determined in step 2:
 $\Lambda(g)_{m+1:1} = \Lambda(g_{m+1}) \circ \dots \circ \Lambda(g_1)$;
5. Measure a POVM $\{\mu, \mathbb{I} - \mu\}$, where the first observable is $\mu \approx L(g_{m+1:1})[|0\rangle\langle 0|]$, and respective outcome labels are {"recovery", "non-recovery"};
6. Repeat steps 3–5 a number times to estimate the probability of observing the "recovery" event $\Pr(\text{"recovery"}|\{g_i\}, m) = \text{Tr } \mu \Lambda(g)_{m+1:1}[\rho]$;
7. Repeat steps 1–6 for s different sets of m randomly sampled gates $\{g_i\}$;
8. Repeat steps for 1–7 for different values of m of random gates.
9. Fit the estimated recovery probabilities to the decay model

$$A_{\mathbf{C}} p_{\mathbf{C}}^m + B_{\mathbf{C}} ; \tag{7.1}$$

10. Estimate the Clifford gate-set infidelity through

$$\delta F(\Lambda(\mathbf{C}), L(\mathbf{C})) = (1 - p_{\mathbf{C}})/2. \tag{7.2}$$

The recovery operations mentioned in step 2 is usually an inversion gate, where $g_{m+1} = (g_{m:1})^{-1}$, in which case the recovery observable simply corresponds to the initial state: $\mu \approx |0\rangle\langle 0|$. However, performing the inverse only up to a random bit flip, i.e $g_{m+1} = X_{\pi}^b (g_{m:1})^{-1}$, leads to a simpler decay model with less free parameters since it fixes $B_{\mathbf{C}}$ [Har+19]. Of course in this case one has to keep track of the bit flip, that is $\mu \approx |b\rangle\langle b|$. Such a randomized recovery operation was proposed originally in [Kni+08].

SRB is typically implemented using the Clifford group \mathbf{C} as a randomizing gate-set, as specified in the first step of protocol 1, but the derivation of the decay model shown in eq. (7.1) holds for any unitary 2-design [Dan+06; Dan+09]. The Clifford group is defined as follows. First consider the pulses along any Cartesian axis system

$$X_{\theta} := e^{-i\theta/2 \sigma_X}, Y_{\theta} := e^{-i\theta/2 \sigma_Y}, Z_{\theta} := e^{-i\theta/2 \sigma_Z},$$

where σ_i denote the unitary Pauli matrices. The Pauli group \mathbf{P} is defined in terms of the identity operation and 3 elementary π pulses: $\mathbf{P} := \{\mathbb{I}, X_\pi, Y_\pi, Z_\pi\}$.

The Clifford group \mathbf{C} is defined as the normalizer of the Pauli group and can be obtained from the Pauli group composed with the coset $\mathbf{S} := \{\mathbb{I}, X_{\pi/2}, Y_{\pi/2}, Z_{\pi/2}, Z_{\pi/2}X_{\pi/2}, X_{-\pi/2}Z_{-\pi/2}\}$:

$$\mathbf{C} := \mathbf{S} \cdot \mathbf{P} = \{S \cdot P \mid S \in \mathbf{S}, P \in \mathbf{P}\}. \quad (7.3)$$

Some other experimental groups performed RB using alternate 2-design gate-sets in step 1 of protocol 1 [Bar+14]. Amongst the set of possible unitary 2-designs, it is worth mentioning those following subsets of \mathbf{C} . Consider the cyclic group $\mathbf{T} := \{\mathbb{I}, Z_{\pi/2}X_{\pi/2}, X_{-\pi/2}Z_{-\pi/2}\}$, then the following sets both form 2-designs of order 12:

$$\mathbf{C}_{12} := \mathbf{T} \cdot \mathbf{P} = \{T \cdot P \mid T \in \mathbf{T}, P \in \mathbf{P}\}, \quad (7.4)$$

$$\sqrt{\mathbf{Z}}\mathbf{C}_{12} := Z_{\pi/2} \cdot \mathbf{C}_{12} = \{Z_{\pi/2} \cdot C \mid C \in \mathbf{C}_{12}\}, \quad (7.5)$$

with $\mathbf{C}_{12} \cup \sqrt{\mathbf{Z}}\mathbf{C}_{12} = \mathbf{C}$. Obviously, the decay parameters as well as the infidelity depend on the randomizing gate-set (hence the indices).

The validity of the decay model and the connection between the decay parameter and the gate-set infidelity have been demonstrated in the case of gate-independent Markovian noise scenarios in [EAZ05]. The proofs of eq. (7.1) and eq. (7.2) have been generalized to encompass gate-dependent noise scenarios in [Wal17; MPF18] and [Car+18] respectively³.

Although the proof techniques can get mathematically heavy, their essence remains simple: the algebraic richness of 2-designs prevents errors to accumulate in an unpredictable way as the circuit grows in length. As we show with more care in the next section, the random sampling over the gate-set tailors the effective errors at each cycle in a depolarizing channel for which the evolution is parameterized by a single real number p . The errors are stripped out of all their properties except one, which turns out to be in one-to-one correspondence with their average infidelity. By modifying the sampled circuits lengths, we can estimate the parameter p and retrieve the infidelity.

While unitary 2-designs are provably effective randomizing gate-sets, leading to the model portrayed in eq. (7.1), some algebraically weaker gate-sets have indicated a similar exponential decaying behaviour.

The gate-set \mathbf{N} used in NIST RB [Kni+08] is a composition of a set $\mathbf{Q} := \{X_{\pm\pi/2}, Y_{\pm\pi/2}\}$, consisting of $\pi/2$ pulses in the xy-plane, with the Pauli operators:

$$\mathbf{N} := \mathbf{Q} \cdot \mathbf{P} = \{Q \cdot P \mid Q \in \mathbf{Q}, P \in \mathbf{P}\}. \quad (7.6)$$

\mathbf{N} has order 8, and although it contains all its inverse elements (that is $\forall g \in \mathbf{N}, \exists g^{-1} \in \mathbf{N}$ s.t. $g \cdot g^{-1} = \mathbb{I}$), it is not closed under multiplication. It does not form a group, nor a 2-design; however, the closure $\langle \mathbf{N} \rangle$ forms the Clifford group \mathbf{C} .

RB sequences can be seen as Markov chains [MEK12], where the elements of the chain are the aggregate circuits, that is $C_1 = g_1, C_2 = g_2g_1, C_m = g_{m:1}$. Indeed, the probability distribution on circuits $C_m = g_{m:1}$ simply depends on the circuit C_{m-1} and on the probability distribution of the random gate applied at step m . In standard RB, C_i is always uniformly

³In gate-dependent noise scenarios, the connection between the RB decay parameter and the gate-set infidelity remains a (strongly supported) conjecture for $d > 2$.

distributed over the Clifford group. In NIST RB, C_{2n} (or C_{2n+1}) converges to a uniform distribution over C_{12} (or $\sqrt{Z}C_{12}$), as shown in fig. 7.1.

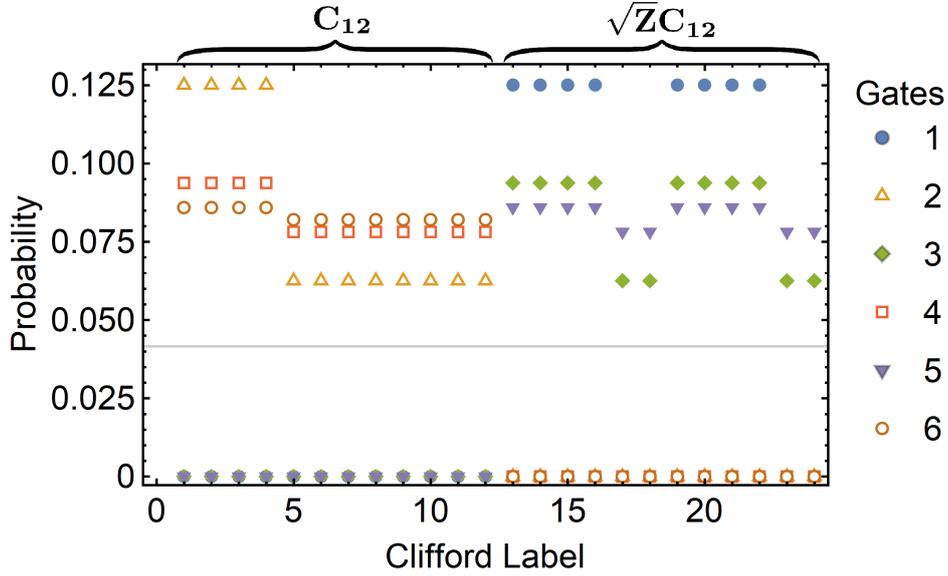


FIGURE 7.1: Probability distribution over the Clifford gates C (labelled as in [Bar+14]) after m gates (i.e clock cycles) of NIST RB drawn uniformly at random from $N \subset C$. This leads to a *non-uniform* sampling over the Cliffords that varies as m increases. Asymptotically, for a sequence of even (or odd) length, the probability distribution tends toward a uniform distribution over C_{12} (or $\sqrt{Z}C_{12}$). The grey line indicates an equal probability over the full 24 Clifford group C .

While this approach to RB has been useful for estimating error rates [Kni+08; Bro+11; Har+14], in the absence of a unitary 2-design structure, it is not clear how to relate the measured probabilities from protocol 2 to the usual decay predicted under SRB, or to any infidelity for that matter. In this paper we provide a concrete analysis of the outcome of protocol 2, which yields a justification and interpretation for the decay model eqs. (7.7) and (7.8).

It is important to emphasize that NIST RB now falls into a family of RB protocols defined as “direct RB” [Pro+18]. The analysis below gives a concrete instance of direct RB that both justifies and interprets past experiments and gives an insightful example of the main idea behind direct RB.

Protocol 2: NIST RB, as described in [Kni+08].

1. Sample a set of m gates $\{g_1, \dots, g_m\}$ picked independently and uniformly at random from the NIST gate-set \mathbf{N} defined in eq. (7.6);
- 2–8. Idem as in protocol 1.
9. Fit the estimated recovery probabilities to the decay model

$$A_{\mathbf{N}} p_{\mathbf{N}}^m + B_{\mathbf{N}} . \quad (7.7)$$

10. Estimate the NIST gate-set infidelity through

$$\delta F(\Lambda(\mathbf{N}), L(\mathbf{N})) = (1 - p_{\mathbf{N}})/2 . \quad (7.8)$$

Theoretical Analysis of NIST RB under a gate-independent error model

The goal of this section is to provide the key insight behind the mechanics of NIST RB. To lighten up the mathematical machinery, we assume a gate-independent error model, where the noisy gates are followed by an error \mathcal{E} :

$$\Lambda(g) = \mathcal{E} \circ L(g) , \quad (7.9)$$

where

$$L(g)[\rho] = g\rho g^\dagger . \quad (7.10)$$

In such model, the gate-set infidelities $\delta F(\Lambda(\mathbf{C}), L(\mathbf{C}))$ and $\delta F(\Lambda(\mathbf{N}), L(\mathbf{N}))$ are *de facto* equal to the infidelity of the error $\delta F(\mathcal{E}, \mathcal{I})$. We proceed in showing that $\delta F(\mathcal{E}, \mathcal{I})$ can be estimated by both protocols 1 and 2.

The recovery probabilities look like

$$\text{Tr} \left[\mu \mathcal{E} L \left(\underbrace{X_\pi^b g_{m:1}^{-1}}_{g_{m+1}} \right) \mathcal{E} L(g_m) \cdots L(g_2) \mathcal{E} L(g_1) [\rho] \right] . \quad (7.11)$$

Shoving the last error \mathcal{E} as well as the random bit flip X_π^b in the measurement procedure (that is, $\mu \rightarrow L(X_\pi^b) \mathcal{E}^\dagger[\mu]$), leaves us with the random sequence which is at the heart of both NIST RB and SRB protocols:

$$S(\{g_i\}) = L(g_{m:1}^{-1}) \mathcal{E} L(g_m) \cdots \mathcal{E} L(g_2) \mathcal{E} L(g_1) . \quad (7.12)$$

In SRB, the next step in the analysis consists in redefining the gates as $g_i = g'_i g'_{i-1}^{-1}$ (with $g_1 = g'_1$), where both g_i and g'_i are picked uniformly at random from the randomizing set. Such a relabeling is possible because the randomizing gate-set is usually a group. Averaging

over all sequences yields

$$\begin{aligned}\mathbb{E}_{g'_i \in \mathbf{C}} S(\{g_i\}) &= \mathbb{E}_{\{g'_i\}} L(g'_m)^{-1} \mathcal{E} L(g'_m) \cdots L(g'_2)^{-1} \mathcal{E} L(g'_2) L(g'_1)^{-1} \mathcal{E} L(g'_1) \\ &= (\mathcal{E}^{\mathbf{C}})^m,\end{aligned}\quad (7.13)$$

where

$$\mathcal{E}^{\mathbf{C}} := |\mathbf{C}|^{-1} \sum_{g \in \mathbf{C}} L(g)^{-1} \mathcal{E} L(g) \quad (7.14)$$

is referred to as the twirl of the error \mathcal{E} over the gate-set \mathbf{C} . If \mathbf{C} is a 2-design, then the twirled channel $\mathcal{E}^{\mathbf{C}}$ is reduced to a depolarizing channel. To mathematically concretize the description of a channel \mathcal{E} , we resort to the 4×4 Pauli-Liouville representation, which is defined as

$$\mathcal{E}_{ij} := \frac{1}{2} \text{Tr} B_j^\dagger \mathcal{E}(B_i) \quad (7.15)$$

where $B_0 = \mathbb{I}$, $B_1 = \hat{\sigma}_x$, $B_2 = \hat{\sigma}_y$, $B_3 = \hat{\sigma}_z$. In such representation, the depolarizing channel $\mathcal{E}^{\mathbf{C}}$ is expressed as a diagonal matrix $\text{diag}(1, p_{\mathbf{C}}, p_{\mathbf{C}}, p_{\mathbf{C}})$, where $p_{\mathbf{C}}$ is a real number close to 1:

$$p_{\mathbf{C}} = \frac{\mathcal{E}_{11} + \mathcal{E}_{22} + \mathcal{E}_{33}}{3}. \quad (7.16)$$

The averaged core sequence hence evolves as

$$\mathbb{E}_{g_i \in \mathbf{C}} S(\{g_i\}) = (\mathcal{E}^{\mathbf{C}})^m = \text{diag}(1, p_{\mathbf{C}}^m, p_{\mathbf{C}}^m, p_{\mathbf{C}}^m). \quad (7.17)$$

Deriving eq. (7.1) is then simply a matter of incorporating SPAM procedures in the evaluation of the recovery probabilities. Straightforward algebra links the infidelity of \mathcal{E} with its diagonal Liouville matrix elements through

$$\delta F(\mathcal{E}, \mathbb{I}) = \frac{1}{2} - \frac{\mathcal{E}_{11} + \mathcal{E}_{22} + \mathcal{E}_{33}}{6}. \quad (7.18)$$

The relation between the decay constant $p_{\mathbf{C}}$ and the gate-set infidelity $r_{\mathbf{C}} = r(\mathcal{E}, \mathbb{I})$ results from combining eq. (7.16) and eq. (7.18).

The relabeling trick resulting in a m -composite depolarizing channel is not possible in NIST RB: \mathbf{N} is neither a group nor a 2-design. However, although \mathbf{N} has a weaker algebraic structure, it is not completely devoid of interesting properties. Indeed, every element of \mathbf{N} can be written as $P_{\text{left}} \cdot Q \cdot P_{\text{right}}$, where $P_{\text{left}}, P_{\text{right}} \in \mathbf{P}$ and $Q \in \mathbf{Q}$. Using this, we can relabel every gate g_i as

$$g_1 = P_1 Q_1, \quad (7.19a)$$

$$g_i = P_i Q_i P_{i-1}^{-1} \quad (i = 2, \dots, m), \quad (7.19b)$$

$$g_{m:1}^{-1} = Q_{m:1}^{-1} P_m^{-1} \quad (7.19c)$$

where P_i is chosen UAR from the Pauli group \mathbf{P} , and Q_i are chosen UAR from \mathbf{Q} . Using such a manipulation and randomizing over the Paulis transform the core sequence into

$$\mathbb{E}_{P_i \in \mathbf{P}} S(\{g_i\}) = L \left(Q_{m:1}^{-1} \right) \mathcal{E}^{\mathbf{P}L(Q_m)} \cdots \mathcal{E}^{\mathbf{P}L(Q_2)} \mathcal{E}^{\mathbf{P}L(Q_1)}, \quad (7.20)$$

where $\mathcal{E}^{\mathbf{P}}$ is the error channel twirled over the Pauli group. In the Pauli-Liouville picture, the Pauli group has 4 inequivalent irreps; the twirled channel is diagonal:

$$\mathcal{E}^{\mathbf{P}} = \text{diag}(1, x, y, z), \quad (7.21)$$

where $x = \mathcal{E}_{11}$, $y = \mathcal{E}_{22}$, $z = \mathcal{E}_{33}$. The relabeling method still can't be used with the Q_i 's, but the simplification of the noise channel \mathcal{E} through the Pauli twirl unveils a recursive approach. Consider the $m = 1$ case:

$$\mathbb{E}_{g_1 \in \mathbf{N}} S(\{g_1\}) = \mathbb{E}_{Q_1 \in \mathbf{Q}} L \left(Q_1^{-1} \right) \mathcal{E}^{\mathbf{P}L(Q_1)} = \mathcal{E}^{\mathbf{N}}, \quad (7.22)$$

where the twirl over the NIST gate-set results in

$$\mathcal{E}^{\mathbf{N}} = \text{diag} \left(1, \frac{x+z}{2}, \frac{y+z}{2}, \frac{x+y}{2} \right). \quad (7.23)$$

The $m = 2$ case suggests a recursion relation:

$$\mathbb{E}_{g_i \in \mathbf{N}} S(\{g_i\}) = (\mathcal{E}^{\mathbf{N}} \mathcal{E}^{\mathbf{P}})^{\mathbf{N}}, \quad (7.24a)$$

$$(\mathcal{E}^{\mathbf{N}} \mathcal{E}^{\mathbf{P}})^{\mathbf{N}} = \text{diag}(1, x_2, y_2, z_2), \quad (7.24b)$$

where

$$x_2 = \frac{x \frac{(x+z)}{2} + z \frac{(x+y)}{2}}{2}, \quad (7.25a)$$

$$y_2 = \frac{y \frac{(y+z)}{2} + z \frac{(x+y)}{2}}{2}, \quad (7.25b)$$

$$z_2 = \frac{x \frac{(x+z)}{2} + y \frac{(y+z)}{2}}{2}. \quad (7.25c)$$

Indeed, the general case can be expressed as

$$\mathbb{E}_{g_i \in \mathbf{N}} S(\{g_i\}) = \left(\left((\mathcal{E}^{\mathbf{N}} \mathcal{E}^{\mathbf{P}})^{\mathbf{N}} \mathcal{E}^{\mathbf{P}} \right)^{\mathbf{N}} \mathcal{E}^{\mathbf{P}} \cdots \right)^{\mathbf{N}}, \quad (7.26a)$$

$$\left(\left((\mathcal{E}^{\mathbf{N}} \mathcal{E}^{\mathbf{P}})^{\mathbf{N}} \mathcal{E}^{\mathbf{P}} \right)^{\mathbf{N}} \mathcal{E}^{\mathbf{P}} \cdots \right)^{\mathbf{N}} = \text{diag}(1, x_m, y_m, z_m), \quad (7.26b)$$

where the recursion relation can be stated as

$$x_m = \frac{x \cdot x_{m-1} + z \cdot z_{m-1}}{2}, \quad (7.27a)$$

$$y_m = \frac{y \cdot y_{m-1} + z \cdot z_{m-1}}{2}, \quad (7.27b)$$

$$z_m = \frac{x \cdot x_{m-1} + y \cdot y_{m-1}}{2}. \quad (7.27c)$$

Using basic linear algebra, this system of recursive equations can be expressed as

$$\begin{bmatrix} x_m \\ y_m \\ z_m \end{bmatrix} = M \begin{bmatrix} x_{m-1} \\ y_{m-1} \\ z_{m-1} \end{bmatrix} = M^m \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad (7.28)$$

where

$$M = \frac{1}{2} \begin{bmatrix} x & 0 & z \\ 0 & y & z \\ x & y & 0 \end{bmatrix}. \quad (7.29)$$

x, y and z differ from 1 by at most order $\delta F(\mathcal{E}, \mathbb{I})$. Hence, up to the second order in the infidelity, M has the following spectrum:

$$\lambda_1 \approx \frac{x + y + z}{3} = p_C, \quad (7.30a)$$

$$\lambda_2 \approx \frac{x + y}{4}, \quad (7.30b)$$

$$\lambda_3 \approx -\frac{x + y + 4z}{12}. \quad (7.30c)$$

Since $\lambda_1 \approx 1$, $\lambda_2 \approx 1/2$ and $\lambda_3 \approx -1/2$, M^m converges very quickly to a rank-1 operator as m increases. This means that for m large enough so that $1/2^m$ becomes negligible, x_m, y_m, z_m are proportional to λ_1^m :

$$\mathbb{E}_{g_i \in \mathbf{N}} S(\{g_i\}) \approx \text{diag}(1, c_1 \lambda_1^m, c_2 \lambda_1^m, c_3 \lambda_1^m), \quad (7.31)$$

where c_i are proportionality constants. Equation (7.7) is obtained by incorporating the SPAM procedures in evaluating the recovery probabilities, and by relabeling λ_1 as p_N . Finally, the relation between the decay p_N and the gate-set infidelity $\delta F(\Lambda(\mathbf{N}), L(\mathbf{N})) = \delta F(\mathcal{E}, \mathbb{I})$ is retrieved via eq. (7.30a):

$$\delta F(\Lambda(\mathbf{N}), L(\mathbf{N})) = (1 - p_N)/2 + O(\delta^2 F), \quad (7.32)$$

which essentially states that the NIST RB decay parameter p_N provides a very good estimates of the gate-set infidelity $\delta F(\Lambda(\mathbf{N}), L(\mathbf{N}))$ through eq. (7.8).

With this analysis behind us, let's compare the internal mechanics of protocols 1 and 2. First of all, both protocols make use of randomizing gate-sets, \mathbf{C} and \mathbf{N} respectively. In both cases, the randomization tailors the error dynamics such that the average core sequence $\mathbb{E}_{g_i \in \mathbf{G}} S(\{g_i\})$ evolves with respect to a single decay parameter, as show in eqs. (7.17)

and (7.31).

An interesting difference here is that the Clifford randomization simplifies the error into a 1-parameter depolarizing channel at each time step, while the NIST randomization doesn't, as shown in eq. (7.23). In the latter case, certain error components remain "imperfectly shuffled" after a few random gates, leaving space for a multi-parameterized noise evolution portrayed by eqs. (7.28) and (7.29). However, as the random sequence gets longer, the evolution quickly converges to a 1-parameter decay. The fact that this decay relates to the infidelity shouldn't be surprising, since $\text{diag}(0, 1, 1, 1)$ is a channel component that commutes with every unitary (so is "immune" to twirling). Given an error \mathcal{E} , its corresponding coefficient is $(\mathcal{E}_{11} + \mathcal{E}_{22} + \mathcal{E}_{33})/3$, which is in one-to-one correspondence with the infidelity $\delta F(\mathcal{E}, \mathbb{I})$ via eq. (7.18).

7.3 Direct randomized benchmarking

Section 7.3.1 consists of a quick summary of the protocol derived mostly by Dr. Timothy J. Proctor in [Pro+18]. The subsequent subsections, however, consists of some analytic tools regarding the protocol in question, for which my contribution was major.

7.3.1 The protocol

The theory behind direct randomized benchmarking allows to benchmark a wide family of gate-sets, but for the sake of simplicity, consider the simple gate-set \mathbf{H} which consists in operations that are native to the device, that can be operated during the same cycle, and that ultimately generate the n -qubit Clifford group \mathbf{C} . Such gate-set will be referred to as a native 1-cycle generating set of \mathbf{C} . A three qubit example could be, if the device allows it,

$$\mathbf{H} = \{\text{CNOT}_{1,2} \otimes h_3, h_1 \otimes \text{CNOT}_{2,3}, h_1 \otimes h_2 \otimes h_3 | h_i \in \{H, \sqrt{Z}, X\} \subset U(d)\}. \quad (7.33)$$

Consider a native 1-cycle generating set \mathbf{H} of the Clifford group \mathbf{C} , consider a probability distribution Ω with full support over elements of \mathbf{H} , and consider a Markovian error map $\Lambda : U(d) \rightarrow \text{CPTP}_d$. Then consider the following protocol:

Protocol 3: Direct randomized benchmarking.

1. Sample a Clifford gate g_0 uniformly from the Clifford group \mathbf{C} ; sample a set of m gates $\{g_1, \dots, g_m\}$ picked independently from the distribution Ω over the generating set \mathbf{H} ;
2. Determine the recovery gate g_{m+1} (see text below);
3. Prepare a state ρ ;
4. Perform the sampled gates (starting with the implementation of g) from step 1, followed by the recovery gate g_{m+1} determined in step 2:
 $\Lambda(g)_{m+1:0} = \Lambda(g_{m+1}) \circ \dots \circ \Lambda(g_0)$;
5. Measure a POVM $\{\mu, \mathbb{I} - \mu\}$, where the first observable is $\mu \approx L(g_{m+1:0})(\rho)$, and respective outcome labels are {"recovery", "non-recovery"};
6. Repeat steps 3–5 a number of times to estimate the probability of observing the "recovery" event $\Pr(\text{"recovery"} | \{g_i\}, m) = \text{Tr } \mu \Lambda(g)_{m+1:0}[\rho]$;
7. Repeat steps 1–6 for s different sets of m randomly sampled gates $\{g_i\}$;
8. Repeat steps 1–7 for different values of m of random gates.
9. Fit the estimated recovery probabilities to the decay model

$$A_0(\rho, \mu) + A_1(\rho, \mu) p^m ; \quad (7.34)$$

10. Estimate the native 1-cycle gate-set weighted infidelity through

$$\sum_{h \in H} \Omega(h) \delta F(\Lambda(h), \mathcal{B}L(h)\mathcal{B}^{-1}) = \frac{d-1}{d}(1-p), \quad (7.35)$$

where $\mathcal{B} \in M_{d^2}(\mathbb{C})$ implicitly defines a specific gauge.

Similarly to protocol 1, choosing $\rho \approx |0\rangle\langle 0|^{\otimes n}$ and the recovery gate to be the sequence inverse up to random bit flips, i.e. sample g_{m+1} from $X^{b_1} \otimes \dots \otimes X^{b_n} (g_{m:0})^{-1}$ where $b_i \sim \text{Bernouilli}(1/2)$, allows to effectively remove the constant $A_0(\rho, \mu)$ from the fitting model [Har+19].

The fitting model suggested at step 9 is an approximation of the exact signal, which is of the form

$$A_0(\rho, \mu) + A_1(\rho, \mu) p^m + \sum_{i>1} C_i(\rho, \mu) \lambda_i^m, \quad (7.36)$$

where $|\lambda_i|^m$ are expected to decay to 0 much faster than p^m . If $|\lambda_i|$ are on the same order as p , direct RB may fail to yield a reliable signal. In such case, the algebraic structure needs to be enriched, either by modifying the distribution Ω , or by adding elements (and potentially more cycles) to the randomizing gate-set.

Instead of being approximated, the fitting model could also take the additional terms

into account, via techniques such as the Padé-Laplace method [YC87]. Recall that the appearance of additional fast-decaying constants was observed in the simple analysis of protocol 2 (NIST RB). In particular, it was shown that those supplementary decays were roughly upper-bounded in absolute value by $1/2$.

The consideration of native 1-cycle gate-set as randomizing sets is mostly for the sake of simplicity. Choosing a gate-set for which the elements generate a 2-design is sufficient. This generalization includes protocol 2.

Once more analysis tools will be at disposal, I will include a short discussion regarding the gauge transformation $L(h) \rightarrow \mathcal{B}L(h)\mathcal{B}^{-1}$.

7.3.2 Generalizing twirls

In [SH18], França and Hashagen develop tools for treating twirls over generators. They successfully demonstrate a decay model for some set of generators (closed under inversion), given a gate-independent approximately covariant error model. Here I propose a different method of analysis⁴, which allows to treat a wider variety of gate-sets (including native 1-cycle generating gate-sets), as well as general gate-dependent error models. I will focus on generators of the Clifford group for the sake of simplicity, but the idea should generalize naturally to finite groups.

Let $\phi_0, \phi_1, \phi_2, \dots, \phi_{N-1}$ represent the isomorphism classes of irreducible representations of the n -qubit Clifford group \mathbf{C} (let the unitary matrices ϕ_i be of dimension $d_i \times d_i$). Moreover, let $\phi_0(g) = 1$ and $\phi_1(g) = \phi_{\text{Bloch}}(g)$ be the trivial representation and the (irreducible) representation of the group when acting on Bloch vectors. The process matrix is equivalent to $L(g) = \phi_0(g) \oplus \phi_1(g)$. Notice that ϕ_{Bloch} is irreducible iff the group forms a 2-design.

Consider the direct sum of all inequivalent irreps

$$R(g) := \bigoplus_{i=0}^{N-1} \phi_i(g). \quad (7.37)$$

This special $d_R \times d_R$ unitary representation is the heart of the following generalization of twirls:

⁴This method of analysis was co-developed with Dr. Timothy J. Proctor.

Definition 10: Twirls

Consider a generating set \mathbf{H} of the n -qubit Clifford group \mathbf{C} (possibly \mathbf{C} itself), a probability distribution Ω with full support over elements of \mathbf{H} , and a Markovian error model $\Lambda : U(d) \rightarrow M_{d^2}(\mathbb{C})$. Let the $d^2 \times d^2$ process matrix of ideal operations be expressed as $L(g) = \phi_0(g) \oplus \phi_1(g) = 1 \oplus \phi_{\text{Bloch}}(g)$. Let the $d_R \times d_R$ matrix $R(g)$ be defined as in eq. (7.37).

The noiseless twirl $\mathcal{T}_{\Omega,L} : M_{d^2 \times d_R} \rightarrow M_{d^2 \times d_R}$ is a linear operation defined as

$$\mathcal{T}_{\Omega,L}[\mathcal{A}] = \sum_g \Omega(g) L(g) \mathcal{A} R^\dagger(g), \quad (7.38)$$

where \mathcal{A} is a $d^2 \times d_R$ matrix. The noisy twirl is similarly defined as

$$\mathcal{T}_{\Omega,\Lambda}[\mathcal{A}] = \sum_g \Omega(g) \Lambda(g) \mathcal{A} R^\dagger(g). \quad (7.39)$$

If elements of $M_{d^2 \times d_R}$ are column-vectorized, twirls can be expressed as a matrices:

$$\mathcal{T}_{\Omega,L} = \sum_g \Omega(g) R^*(g) \otimes L(g) \quad (7.40)$$

$$\mathcal{T}_{\Omega,\Lambda} = \sum_g \Omega(g) R^*(g) \otimes \Lambda(g). \quad (7.41)$$

Notice the slight abuse of notation; just as for superoperators, the usage of the matrix form will be distinguished from the abstract form depending on the context.

This definition might seem abstruse at a first glance: the domain of twirls doesn't correspond to $d^2 \times d^2$ matrices (which include superoperators), and the definition relies on the representation $R(g)$, for which the dimension is much larger than d^2 (Notice that it contains $L(g)$ as a block.). However, as I will show later, it is precisely this domain extension that will allow to treat gate-dependent error models.

Before pursuing with analyzing generator-based RB, consider the simpler case where twirls defined as in definition 10 are carried uniformly over the Clifford group.

The Clifford twirl

Consider the noiseless twirl matrix $\mathcal{T}_{\Pi,L}$ where the group of interest is the Clifford group \mathbf{C} , and Π is the uniform distribution. From definition 10, the twirl matrix is expressed as:

$$\mathcal{T}_{\Pi,L} = \frac{1}{|\mathbf{C}|} \sum_{g \in \mathbf{C}} R^*(g) \otimes L(g) \quad (7.42)$$

$$= \frac{1}{|\mathbf{C}|} \sum_{g \in \mathbf{C}} \left(\bigoplus_{i=0}^{i=N-1} \phi_i^*(g) \right) \otimes \left(\bigoplus_{i=0}^{i=1} \phi_i(g) \right). \quad (7.43)$$

Now, consider a $d^2 \times d_R$ matrix \mathcal{A} . This matrix can be decomposed in distinct sectors:

$$\mathcal{A} = \left[\begin{array}{c|c|c|c} \mathcal{A}^{(0,0)} & \mathcal{A}^{(0,1)} & \dots & \mathcal{A}^{(0,N-1)} \\ \hline \mathcal{A}^{(1,0)} & \mathcal{A}^{(1,1)} & \dots & \mathcal{A}^{(1,N-1)} \end{array} \right], \quad (7.44)$$

where $\mathcal{A}^{(0,i)}$ are $1 \times d_i$ matrices, and $\mathcal{A}^{(1,i)}$ are $(d^2 - 1) \times d_i$ matrices. From simple vectorization identities, it follows that

$$\mathcal{T}_{\Pi,L}[\mathcal{A}] = \frac{1}{|\mathbf{C}|} \sum_{g \in \mathbf{C}} \left[\begin{array}{c|c|c|c} \phi_0(g)\mathcal{A}^{(0,0)}\phi_0^\dagger(g) & \phi_0(g)\mathcal{A}^{(0,1)}\phi_1^\dagger(g) & \dots & \phi_0(g)\mathcal{A}^{(0,N-1)}\phi_N^\dagger(g) \\ \hline \phi_1(g)\mathcal{A}^{(1,0)}\phi_0^\dagger(g) & \phi_1(g)\mathcal{A}^{(1,1)}\phi_1^\dagger(g) & \dots & \phi_1(g)\mathcal{A}^{(1,N-1)}\phi_N^\dagger(g) \end{array} \right]. \quad (7.45)$$

It follows directly from Schur's lemma (see [Art, p.308, Theorem 10.7.6]) that only 2 sectors do not vanish:

$$\mathcal{T}_{\Pi,L}[\mathcal{A}] = \left[\begin{array}{c|c|c|c|c} \mathcal{A}^{(0,0)} & & 0 & & 0 \\ \hline & & \phi_{\text{Bloch}}(e) \cdot \frac{\text{Tr } \mathcal{A}^{(1,1)}}{\text{Tr } \phi_{\text{Bloch}}(e)} & & 0 \\ \hline 0 & & 0 & \dots & 0 \\ \hline & & & \dots & 0 \\ \hline 0 & & & & 0 \end{array} \right]. \quad (7.46)$$

In other words, $\mathcal{T}_{\Pi,L}$ is a rank-2 projector. This can be seen again by rewriting the twirl operator as (recall that $L(g) = \phi_0(g) \oplus \phi_1(g)$)

$$\mathcal{T}_{\Pi,L} = \underbrace{\frac{1}{|\mathbf{C}|} \sum_g \phi_0^*(g) \otimes L(g)}_{\text{Rank 1 proj.}} \oplus \underbrace{\frac{1}{|\mathbf{C}|} \sum_g \phi_1^*(g) \otimes L(g)}_{\text{Rank 1 proj.}} \oplus_{i=2}^{i=N-1} \underbrace{\frac{1}{|\mathbf{C}|} \sum_g \phi_i^*(g) \otimes L(g)}_0. \quad (7.47)$$

From Weyl's inequalities, the noisy twirl $\mathcal{T}_{\Pi,\Lambda}$ only deviates slightly from a rank-2 projector:

$$\mathcal{T}_{\Pi,\Lambda} = \underbrace{\frac{1}{|\mathbf{C}|} \sum_g \phi_0^*(g) \otimes \Lambda(g)}_{\text{Close to a rank 1 proj.}} \oplus \underbrace{\frac{1}{|\mathbf{C}|} \sum_g \phi_1^*(g) \otimes \Lambda(g)}_{\text{Close to a rank 1 proj.}} \oplus_{i=2}^{i=N-1} \underbrace{\frac{1}{|\mathbf{C}|} \sum_g \phi_i^*(g) \otimes \Lambda(g)}_{\text{Close to 0}}. \quad (7.48)$$

How can one reason about this matrix? Algebraically speaking, implementing quantum operations is an attempt to recreate, through physical processes, the action of $L(g) = \phi_0(g) \oplus \phi_1(g)$ on quantum states. For any group 2-design, ϕ_{Bloch} is an irrep. A finite group (such as the Clifford group) has a finite amount of inequivalent irreps (represented here as ϕ_i). An important result in representation theory, known as Schur's lemma, states that those irreps induce intrinsically different actions. In particular, irreps must obey strong inner product laws, known as "Schur orthogonality relations" of irrep entries [Ser, Section 2.2, corollaries 2 and 3.]:

$$\sum_{g \in \mathbf{C}} [\phi_i(g)]_{jk}^* [\phi_{i'}(g)]_{j'k'} = \delta_{ii'} \delta_{jj'} \delta_{kk'} |\mathbf{C}| / d_i, \quad (7.49)$$

The last terms (i.e. $i \in \{2, \dots, N-1\}$) in direct sum on the RHS of eq. (7.48) essentially describe the extent to which the map Λ is still "Schur orthogonal" to $\phi_2, \dots, \phi_{N-1}$. Analogously, the first two terms in the direct sum on the RHS of eq. (7.48) capture how the map Λ obeys the orthogonality relations (eq. (7.49)) with ϕ_0 and ϕ_1 . In other words, the twirl matrix

$\mathcal{T}_{\Pi,\Lambda}$ encapsulate all the “coefficient inner products” (which become the coefficient orthogonality relations stipulated by eq. (7.49) in the ideal case) between the map Λ and the irreps ϕ_i .

7.3.3 The convergence of sequence-asymptotic 2-designs

The Clifford twirl $\mathcal{T}_{\Pi,L}$ forms a rank-2 projector on depolarizing channels. Now, consider the weaker twirl

$$\mathcal{T}_{\Omega,L} = \sum_g \Omega(g) \phi_0^*(g) \otimes L(g) \oplus \sum_g \Omega(g) \phi_1^*(g) \otimes L(g) \oplus \sum_{i=2}^{i=N-1} \sum_g \Omega(g) \phi_i^*(g) \otimes L(g). \quad (7.50)$$

It is not too much effort to show that iterating this twirl converges to $\mathcal{T}_{\Pi,L}$. The idea is to look at individual blocks in the direct sum in eq. (7.50), and use Schur’s lemma.

Lemma 8: Bounding the spectral norm of the twirling blocks

Let ϕ_i and ϕ_j be two irreps of the n -qubit Clifford group \mathbf{C} . Consider a generating gate-set \mathbf{H} of \mathbf{C} , and a probability distribution Ω over elements of \mathbf{H} (with full support over \mathbf{H}). Consider the following matrix

$$M = \sum_g \Omega(g) \phi_i^*(g) \otimes \phi_j(g). \quad (7.51)$$

If ϕ_i and ϕ_j are inequivalent, the spectral norm of M is strictly smaller than 1. If $\phi_i = \phi_j$, M has exactly one singular value (which is also an eigenvalue) of 1, and the other singular values are strictly smaller than 1.

Proof. The proof follows almost immediately from Schur’s lemma. Consider the positive semidefinite matrix

$$M^\dagger M = \sum_{g,h} \Omega(h) \Omega(g) \phi_i^*(h^{-1}g) \otimes \phi_j(h^{-1}g), \quad (7.52)$$

for which the singular spectrum obeys $\sigma_i(M^\dagger M) = \sigma_i^2(M)$. First, notice that by the triangle inequality, the spectral norm of $M^\dagger M$ is upper bounded by 1 (hence it applies to M as well):

$$\|M^\dagger M\|_\infty \leq \sum_{g,h} \Omega(h) \Omega(g) \|\phi_i^*(h^{-1}g) \otimes \phi_j(h^{-1}g)\|_\infty = 1. \quad (7.53)$$

Now suppose that there exists a unit vector v for which $M^\dagger M v = v$. Obviously, since $M^\dagger M$ is a convex sum of unitaries, this can only happen if

$$\phi_i^*(h^{-1}g) \otimes \phi_j(h^{-1}g) v = v \quad \forall h, g \in \mathbf{H}. \quad (7.54)$$

WOLOG, let $v = \text{col}(\mathcal{A})$, where $\mathcal{A} \in M_{d_j \times d_i}(\mathbb{C})$. The above expression can be re-expressed as

$$\phi_j(h^{-1}g) \mathcal{A} = \mathcal{A} \phi_i(h^{-1}g) \quad \forall h, g \in \mathbf{H}. \quad (7.55)$$

Since, the elements of \mathbf{H} generate the group \mathbf{C} , it follows that

$$\phi_j(g)\mathcal{A} = \mathcal{A}\phi_i(g) \quad \forall g \in \mathbf{C}. \quad (7.56)$$

Consider the case where ϕ_i and ϕ_j are inequivalent irreps. It follows from Schur's lemma that $\mathcal{A} = 0$ or, to put it simply, that $M^\dagger M$ has no singular value of 1.

Consider the case where $\phi_i = \phi_j$. It follows from Schur's lemma that $\mathcal{A} = \mathbb{I}_{d_i} / \sqrt{d_i}$ is the unique non-trivial solution of eq. (7.56) (up to scaling factor). \square

The statement about the singular values of the twirl $\mathcal{T}_{\Omega,L}$ can be translated into a statement regarding its eigenvalues. Let the eigenvalues corresponding to the i^{th} block $\sum_g \Omega(g)\phi_i(g) \otimes L(g)$ of the twirl matrix $\mathcal{T}_{\Omega,L}$ be labeled as $\lambda_k^{(i)}$, and let $|\lambda_0^{(i)}| \geq \dots \geq |\lambda_{d_i^2-1}^{(i)}|$. Similarly, let the singular values corresponding to the i^{th} block $\sum_g \Omega(g)\phi_i(g) \otimes L(g)$ of the twirl matrix $\mathcal{T}_{\Omega,L}$ be labeled as $\sigma_k^{(i)}$, and let $\sigma_0^{(i)} \geq \dots \geq \sigma_{d_i^2-1}^{(i)}$.

For $i > 1$, it follows from lemma 8 that the spectral norm of the blocks is strictly smaller than 1. Since the spectral radius is upper bounded by the spectral norm, $|\lambda_k^{(i)}| < 1$ for $i > 1$.

For the first 2 blocks ($i \in \{0, 1\}$), there exists two eigenvalues of 1, one for each block:

$$\sum_g \Omega(g)[\phi_0(g) \oplus \phi_1(g)] \otimes L(g) \text{col}(1 \oplus 0\phi_1(e)) = \text{col}(1 \oplus 0\phi_1(e)), \quad (7.57a)$$

$$\sum_g \Omega(g)[\phi_0(g) \oplus \phi_1(g)] \otimes L(g) \text{col}(0 \oplus \phi_1(e)) = \text{col}(0 \oplus \phi_1(e)). \quad (7.57b)$$

It follows from Weyl's inequality between eigenvalues and singular values that

$$|\lambda_0^{(0)} \lambda_1^{(0)}| \leq \sigma_0^{(0)} \sigma_1^{(0)}. \quad (7.58)$$

Since $\lambda_0^{(0)} = \sigma_0^{(0)} = 1$, $|\lambda_1^{(0)}| \leq \sigma_1^{(0)}$. From lemma 8, $\sigma_1^{(0)} < 1$, hence $|\lambda_1^{(0)}| < 1$. A similar argument applies to the second block.

To put it simply,

$$\mathcal{T}_{\Omega,L} = \underbrace{\sum_g \Omega(g)\phi_0^*(g) \otimes L(g)}_{\substack{\{\lambda_k^{(1)}\} = 1 \cup \{\lambda_k^{(1)}\}_{k>0} \\ |\lambda_k^{(1)}| < 1, \forall k > 0}} \oplus \underbrace{\sum_g \Omega(g)\phi_1^*(g) \otimes L(g)}_{\substack{\{\lambda_k^{(2)}\} = 1 \cup \{\lambda_k^{(2)}\}_{k>0} \\ |\lambda_k^{(2)}| < 1, \forall k > 1}} \oplus \sum_{i=2}^{i=N-1} \underbrace{\sum_g \Omega(g)\phi_i^*(g) \otimes L(g)}_{|\lambda_k^{(i)}| < 1}. \quad (7.59)$$

The eigenspace spanned by the eigenvalues of 1 is the same as the row space of the Clifford twirl $\mathcal{T}_{\Pi,L}$. In particular, this implies that the twirl $\mathcal{T}_{\Omega,L}^m$ converges to the Clifford twirl $\mathcal{T}_{\Pi,L}$ as m goes to infinity. To make a more precise statement, consider a $d^2 \times d_R$ matrix A

$$A = [\mathcal{A}^{(0)} \mid \mathcal{A}^{(1)} \mid \dots \mid \mathcal{A}^{(N-1)}], \quad (7.60)$$

where $\mathcal{A}^{(i)}$ are $d^2 \times d_i$ sectors. Then, it follows immediately from lemma 8 and the compatibility of the spectral norm with $\|\cdot\|_2$ that

$$\|(\mathcal{T}_{\Omega,L}^m - \mathcal{T}_{\Pi,L})[\mathcal{A}]\|_2 \leq \left(\sigma_1^{(0)}\right)^m \|\mathcal{A}^{(0)}\|_2 + \left(\sigma_1^{(1)}\right)^m \|\mathcal{A}^{(1)}\|_2 + \sum_{i>1} \left(\sigma_0^{(i)}\right)^m \|\mathcal{A}^{(i)}\|_2. \quad (7.61)$$

Because of this convergent property, the pair (Ω, \mathbf{H}) – where Ω is a probability distribution with full support of elements of \mathbf{H} , and where \mathbf{H} generates a group 2-design – is labeled as a sequence-asymptotic 2-design.

The distribution

$$\Omega_m(g) := \sum_{\substack{h_i \\ h_{m:1}=g}} \Omega(h_m) \cdots \Omega(h_1), \quad (7.62)$$

which is obtained when considering random sequences of length m where each element $h_i \in \mathbf{H}$ is sampled independently at random from Ω , doesn't generally converge to a uniform distribution over the group \mathbf{C} . Recall that in the simple NIST RB protocol (protocol 2), the distribution Ω_m oscillates between a support over \mathbf{C}_{12} and a support over $\sqrt{Z}\mathbf{C}_{12}$. The convergence of Ω_m is not needed for obtaining a convergent twirling. Instead of requiring the usage of randomizing gate-sets that ensure the unnecessary convergence of Ω_m , the direct RB protocol appeals to sequence-asymptotic 2-designs, which are sufficient (and necessary) to ensure that $\forall \epsilon > 0, \exists N \in \mathbb{N}$ s.t. the twirl $\mathcal{T}_{\Omega,L}^m$ forms an ϵ -approximate 2-design for all $m > N$.

7.3.4 Noisy twirls over sequence-asymptotic 2-designs

Acceptable regime for direct RB

The noisy twirl $\mathcal{T}_{\Omega,\Lambda}$ is a perturbation away from the error-free twirl $\mathcal{T}_{\Omega,L}$. To quantify the perturbation, consider a realistic error model where (see theorem 16)

$$\sum_g \Omega(g) \|\Lambda(g) - L(g)\|_\infty \leq O\left(\sqrt{\sum_g \Omega(g) r_{\text{coh}}(g)}\right). \quad (7.63)$$

A simple usage of the Weyl's inequality suffices to show that the singular values of $\mathcal{T}_{\Omega,\Lambda}$ only deviate from the singular values of $\mathcal{T}_{\Omega,L}$ by order $O(\sqrt{\sum_g \Omega(g) r_{\text{coh}}(g)})$.

The direct RB protocol is expected to perform well in the regime where the spectral gap for singular values, $\Delta := 1 - \|\mathcal{T}_{\Omega,L} - \mathcal{T}_{\Pi,L}\|_\infty$, is significantly larger than $O(\sqrt{\sum_g \Omega(g) r_{\text{coh}}(g)})$. Notice that for Clifford RB (protocol 1), the value of the gap is 1.

The two eigenvalues perturbed from 1 are of special interest. The eigenvalue of 1 corresponding to the first block of the twirl stays unchanged under trace-preserving error maps:

$$\text{col}^\dagger(\phi_0^*(e) \oplus 0\phi_1^*(e)) \sum_g \Omega(g) [\phi_0(g) \oplus 0\phi_1(g)] \otimes \Lambda(g) = \text{col}^\dagger(\phi_0(e) \oplus 0\phi_1(e)). \quad (7.64)$$

The generalization to include leakage should be straightforward [WF14; WBE14; WBE15].

The eigenvalue of 1 corresponding to the second block of the twirl is perturbed to p , for which the value can be approximated through first order perturbation theory:

$$\begin{aligned}
p &\approx \frac{\text{col}^\dagger(0 \oplus \phi_1(e)) \sum_g \Omega(g) [0 \oplus \phi_1^*(g)] \otimes \Lambda(g) \text{col}(0 \oplus \phi_1(e))}{d^2 - 1} && \text{(First order pert.)} \\
&= \sum_g \Omega(g) \frac{\text{Tr} [\Lambda(g) \phi_1(g) L(g^{-1})]}{d^2 - 1} \\
&= \sum_g \Omega(g) f_{\text{tr}}(\Lambda(g), L(g)) . && (7.65)
\end{aligned}$$

In the regime where the (singular) spectral gap Δ is significantly larger than $O(\sqrt{\sum_g \Omega(g) r_{\text{coh}}(g)})$, $|p|$ is ensured to be larger than all the other eigenvalues, since their magnitude must deviate from 1 by at least $O(\sqrt{\sum_g \Omega(g) r_{\text{coh}}(g)})$. Since the second block of the twirl has a real representation (recall that $\phi_{\text{Bloch}}(g)$ has a real representation), the eigenvalues of the second block must either be real or come in conjugate pairs. Since the magnitude of p is greater than the other eigenvalues, p must be real.

Interpretation of the decay constant p

The expression eq. (7.65) is an approximation of p . Here, I show that p exactly connects with a fidelity through a gauge transformation:

$$p = \sum_g \Omega(g) f_{\text{tr}}(\Lambda(g), \mathcal{B}L(g)\mathcal{B}^{-1}) . \quad (7.66)$$

Consider two right eigenvectors of the first two block of the twirl matrix:

$$\sum_g \Omega(g) [\phi_0^*(g) \oplus 0\phi_1^*(g)] \otimes \Lambda(g) \text{col}(\mathcal{A}_0) = \text{col}(\mathcal{A}_0) \quad (7.67a)$$

$$\sum_g \Omega(g) [0\phi_0^*(g) \oplus \phi_1^*(g)] \otimes \Lambda(g) \text{col}(\mathcal{A}_1) = p \text{col}(\mathcal{A}_1) , \quad (7.67b)$$

with $\text{Tr } \mathcal{A}_0 = 1$, $\text{Tr } \mathcal{A}_1 = d^2 - 1$. Recall that

$$\mathcal{A}_0 \approx \phi_0(e) \oplus 0\phi_1(e) \quad (7.68a)$$

$$\mathcal{A}_1 \approx 0\phi_0(e) \oplus \phi_1(e) . \quad (7.68b)$$

There is more constraints on the form of \mathcal{A}_0 and \mathcal{A}_1 . First realize that replacing \mathcal{A}_i by $\mathcal{A}_i(\delta_{1i}\phi_0(e) \oplus \delta_{2i}\phi_1(e))$, where δ_{ij} denotes the Kronecker delta function, leaves eqs. (7.67a) and (7.67b) invariant. Second, notice that due to the trace preserving condition

$$(L(e) \otimes [\phi_0(e) \oplus 0\phi_1(e)]) \left(\sum_g \Omega(g) [0\phi_0^*(g) \oplus \phi_1^*(g)] \otimes \Lambda(g) \right) (L(e) \otimes [\phi_0(e) \oplus 0\phi_1(e)]) = 0 , \quad (7.69)$$

which grants the second block of the twirl matrix a lower triangular block structure. The lower triangular structure in turns implies that

$$\mathcal{A}_1 = [0\phi_0(e) \oplus \phi_1(e)] \mathcal{A}_1 [0\phi_0(e) \oplus \phi_1(e)] . \quad (7.70)$$

Hence, \mathcal{A}_0 and \mathcal{A}_1 are of the form

$$\mathcal{A}_0 = \left[\begin{array}{c|c} 1 & 0_{1 \times (d^2-1)} \\ \hline b & 0_{(d^2-1) \times (d^2-1)} \end{array} \right] , \quad (7.71a)$$

$$\mathcal{A}_1 = \left[\begin{array}{c|c} 0 & 0_{1 \times (d^2-1)} \\ \hline 0_{(d^2-1) \times 1} & \mathcal{B}_{\text{trls}} \end{array} \right] , \quad (7.71b)$$

$$(7.71c)$$

where b is a $(d^2 - 1) \times 1$ vector and $\mathcal{B}_{\text{trls}}$ is a $(d^2 - 1) \times (d^2 - 1)$ matrix. For small perturbations, $\mathcal{B}_{\text{trls}}$ is expected to be invertible.

For small perturbations, $\mathcal{B} := \mathcal{A}_0 + \mathcal{A}_1$ is invertible, and defines a valid gauge transformation:

$$\mathcal{B}L(g)\mathcal{B}^{-1} . \quad (7.72)$$

Notice that \mathcal{B}^{-1} is of the form

$$\mathcal{B}^{-1} = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & \mathcal{B}_{\text{trls}}^{-1} \end{array} \right] \left[\begin{array}{c|c} 1 & 0 \\ \hline -b & \mathbb{I}_{d^2-1} \end{array} \right] \quad (7.73)$$

From there,

$$\begin{aligned} \sum_g \Omega(g) \Phi(\Lambda(g), \mathcal{B}L(g)\mathcal{B}^{-1}) &= \sum_g \Omega(g) \frac{\text{Tr} [\Lambda(g)\mathcal{B}L(g)\mathcal{B}^{-1}]}{d^2} \\ &= \frac{\text{Tr} [(\mathcal{A}_0 + p\mathcal{A}_1)\mathcal{B}^{-1}]}{d^2} \\ &= \frac{\text{Tr} [1 \oplus p\phi_1(e)]}{d^2} \\ &= \frac{1}{d^2} + \frac{d^2 - 1}{d^2} p . \end{aligned} \quad (7.74)$$

7.3.5 A signal analysis for direct RB under gate-dependent effects

The main motivation behind the appearance of $R(g)$ (defined in eq. (7.37)) in the definition of twirls (definition 10) stems from the following lemma and its corollary:

Lemma 9: $\{R(g)\}$ are linearly independent

$\{R(g)\}$, where the representation R is defined through eq. (7.37), forms a linearly independent set of $d_R \times d_R$ matrices ($d_R = \sum_i d_i$). That is

$$\sum_g c_g R(g) = 0, \quad (7.75)$$

if and only if the coefficients $c_g = 0 \forall g \in \mathbf{C}$.

Proof. First, let's upper bound the amount of parameters of $\{R(g)\}$. The matrices $R(g)$ are a direct sum of $d_i \times d_i$ irreducible matrices. This means that there are $\sum_i d_i^2$ entries that are not always 0. Representation theory tells us that $\sum_i d_i^2 = |\mathbf{C}|$ [Art, p. 300, Theorem 10.4.6 c)].

This means that there are at most $|\mathbf{C}|$ independent parameters. In the next part of the proof, we show that there are at least $|\mathbf{C}|$ independent parameters. It suffices to show the existence of $|\mathbf{C}|$ orthogonal matrix elements obtained from linear combinations of $\{R(g)\}$. This is quickly obtained from the Schur orthogonality relations of irrep entries, eq. (7.49). By picking the coefficients $c_g = [\phi_i(g)]_{jk}^*$, the matrix $\sum_g c_g R(g)$ contains zero entries, except for the element on the row $d_0 + \dots + d_{i-1} + j$ and column $d_0 + \dots + d_{i-1} + k$, which is $|\mathbf{C}|/d_i$. This procedure allows to isolate $\sum_i d_i^2 = |\mathbf{C}|$ orthogonal matrices. \square

The appeal behind the representation $R(g)$ is that since $\{R(g)\}$ forms a linearly independent set, there exists a linear function that maps $R(g)$ to $\Lambda(g)$:

Corollary 3: Linear mapping between $R(g)$ and $\Lambda(g)$

Let the representation R be defined through eq. (7.37), and consider an arbitrary error map (possibly gate-dependent) acting on Clifford gates $\Lambda : U(d) \rightarrow M_{d^2}(\mathbf{C})$. There exists a linear function such that

$$\sum_{ij} \chi_{ij} \mathcal{A}_i R(g) \mathcal{B}_j = \Lambda(g), \quad (7.76)$$

where $\{\mathcal{A}_i\}$ is any basis for $d^2 \times d_R$ matrices, and $\{\mathcal{B}_j\}$ is any basis for $d_R \times d^2$ matrices.

The proof trivially follows from lemma 9. With this result at hand, an exact decay model for the recovery probability (see eq. (7.36)) can be obtained:

Theorem 22: Decay model for the recovery probability

Consider the direct RB protocol described by protocol 3. The average recovery probability takes the following form

$$\mathbb{E} \Pr(\text{"recovery"} | \{g_i\}, m) = A_0(\rho, \mu) + A_1(\rho, \mu) p^m + \sum_i C_i(\rho, \mu) \lambda_i^m, \quad (7.77)$$

where $\{1, p\} \cup \{\lambda_i\}$ are the eigenvalues of the noisy twirl matrix $\mathcal{T}_{\Omega, \Lambda}$.

Proof. Let

$$\Lambda(g) = \sum_{ij} \chi_{ij} \mathcal{A}_i R(g) \mathcal{B}_j \text{ for } g \in \mathbf{C}, \quad (7.78)$$

where $\text{col}(\mathcal{A}_i)$ are right eigenvectors of the noisy twirl matrix $\mathcal{T}_{\Omega, \Lambda}$, and $\text{col}(\mathcal{B}_i^\dagger)$ are left eigenvectors (both associated with the eigenvalue λ_i). Let $\lambda_0 = 1$ and $\lambda_1 = p$.

For simplicity, let the recovery operator be the sequence inverse. Compiling bit flips can be interpreted as changing the POVM. The average direct RB operational sequence can be expressed as

$$\begin{aligned} & \sum \Omega(g_1) \cdots \Omega(g_m) \Pi(h) \Lambda(h) \Lambda(g_m) \cdots \Lambda(g_1) \Lambda(g_1^{-1} \cdots g_m^{-1} h^{-1}) \\ &= \sum \chi_{ij} \Omega(g_1) \cdots \Omega(g_m) \Pi(h) \Lambda(h) \Lambda(g_m) \cdots \Lambda(g_1) \mathcal{A}_i R(g_1^{-1} \cdots g_m^{-1} h^{-1}) \mathcal{B}_j, \end{aligned} \quad (7.79)$$

where Π is the uniform distribution over the Clifford group. By column-vectorizing the matrix form of eq. (7.79), the twirl matrices naturally appear:

$$\begin{aligned} & \text{col} \left[\sum \Omega(g_1) \cdots \Omega(g_m) \Pi(h) \Lambda(h) \Lambda(g_m) \cdots \Lambda(g_1) \Lambda(g_1^{-1} \cdots g_m^{-1} h^{-1}) \right] \\ &= \sum \chi_{ij} (\mathcal{B}_j^T \otimes L(e)) \mathcal{T}_{\Pi, \Lambda} \mathcal{T}_{\Omega, \Lambda}^m \text{col}(\mathcal{A}_i), \\ &= \sum \chi_{ij} (\mathcal{B}_j^T \otimes L(e)) \mathcal{T}_{\Pi, \Lambda} \lambda_i^m \text{col}(\mathcal{A}_i), \\ &= \sum_i \lambda_i^m \text{col} \left[\left(\sum_g \Pi(g) \Lambda(g) \mathcal{A}_i R(g^{-1}) \right) \left(\sum_j \chi_{ij} \mathcal{B}_j \right) \right]. \end{aligned} \quad (7.80)$$

Recall that $\lambda_0 = 1$ and $\lambda_1 = p$. By introducing SPAM procedures, the average recovery probability can be expressed as

$$\mathbb{E} \text{Tr} \mu \Lambda(g)_{m+1:0} [\rho] = A_0(\rho, \mu) + A_1(\rho, \mu) p^m + \sum_{i>1} C_i(\rho, \mu) \lambda_i^m, \quad (7.81)$$

where

$$A_0(\rho, \mu) = \langle \mu | \left(\sum_g \Pi(g) \Lambda(g) \mathcal{A}_0 R(g^{-1}) \right) \left(\sum_j \chi_{0j} \mathcal{B}_j \right) | \rho \rangle \quad (7.82a)$$

$$A_1(\rho, \mu) = \langle \mu | \left(\sum_g \Pi(g) \Lambda(g) \mathcal{A}_1 R(g^{-1}) \right) \left(\sum_j \chi_{1j} \mathcal{B}_j \right) | \rho \rangle \quad (7.82b)$$

$$C_i(\rho, \mu) = \langle \mu | \left(\sum_g \Pi(g) \Lambda(g) \mathcal{A}_i R(g^{-1}) \right) \left(\sum_j \chi_{ij} \mathcal{B}_j \right) | \rho \rangle. \quad (7.82c)$$

□

The impact of the term $\sum_{i>1} C_i(\rho, \mu) \lambda_i^m$ on the fitting model is expected to become quickly negligible for two reasons. First, by design, direct RB is expected to perform when λ_i^m 's decay much faster than p^m . Second, in the noiseless case, $A_1(\rho, \mu) = \langle \mu | 0 \oplus \phi_{\text{Bloch}}(e) | \rho \rangle \approx (d-1)/d$ and $A_0(\rho, \mu) = \langle \mu | 1 \oplus 0 \phi_{\text{Bloch}}(e) | \rho \rangle \approx 1/d$ and $C_i(\rho, \mu) = 0$. In the noisy case, the coefficients $C_i(\rho, \mu)$ are expected to be perturbed from 0, and so is their sum. To put it differently, when

considering the Laplace transform of the recovery signal given by eq. (7.77),

$$\mathcal{L}[\Pr(\text{"recovery"}|\{g_i\}, m)](x) = \frac{A_0(\rho, \mu)}{x} + \frac{A_1(\rho, \mu)}{x - p} + \sum_{i>1} \frac{C_i(\rho, \mu)}{x - \lambda_i}, \quad (7.83)$$

the peak corresponding to p is easily distinguishable from the peaks corresponding to λ_i for $i > 1$ [YC87]. Recall that by compiling random bit flips in the recovery step, the singularity at $x = 0$ can essentially be eliminated [Har+19]. While the distinguished singularity at $x = p$ indicates that p should be efficiently estimated, I leave the rigorous demonstration of a robust estimator \hat{p} for further work.

Chapter 8

Summary and conclusion

The development of randomized benchmarking methods involve three related undertakings:

- i. The design of new circuit families and sampling distributions, for which the resulting raw data (i.e. the collection of “success” and “failure” events) could carry novel information.
- ii. The development of more advanced analytical and statistical tools to translate raw data into valuable error parameters.
- iii. The study of inference techniques allowing to leverage the knowledge of a few figures of merit to quantify the outcome reliability of simple or complex computations.

The work presented in this thesis offers important developments in the three avenues mentioned above. In part I, the emphasis was on the relation between the quality of elementary operations and circuits thereof (item iii.). In chapter 3, which is based on [CWE16], it was first shown that the process infidelity of a composition of errors is upper bounded by a quadratic relation in the circuit’s length m , modulated by the process infidelity of its elements:

$$\delta\Phi(\mathcal{A}_{m:1}) \leq \underbrace{C_1 \sum_{i=1}^m \delta\Phi(\mathcal{A}_i)}_{\text{Linear}} + \underbrace{C_2 \left(\sum_{i=1}^m \sqrt{\delta\Phi(\mathcal{A}_i)} \right)^2}_{\text{Quadratic}} + O \left(\left(\sum_{i=1}^m \delta\Phi(\mathcal{A}_i) \right)^2 \right). \quad (8.1)$$

The relation was shown to be saturated in even dimensions. In the saturation examples, the quadratic signature was obtained through unitary processes. It took the tools derived in chapter 4 to pinpoint physical unitaries as the sole cause of such behavior. Still in chapter 3, the constant C_2 appearing eq. (8.1) was shown to be modulated by the coherence level of errors. In particular, it was shown that

$$|\delta\Phi(\mathcal{A} \circ \mathcal{B}) - \delta\Phi(\mathcal{A}) - \delta\Phi(\mathcal{B})| \leq 2\sqrt{(\delta\Phi(\mathcal{A}) - \delta\Upsilon(\mathcal{A})) (\delta\Phi(\mathcal{B}) - \delta\Upsilon(\mathcal{B}))}. \quad (8.2)$$

This simple result is a good example where the study of the evolution of figures of merit in circuits (item iii.) can provide better analytical tools to translate RB data into more precise information (item ii.). Indeed, eq. (8.2) can be re-expressed to improve the confidence interval on the infidelity of a specific gate of interest by combining the results of three distinct RB experiments (see corollary 2, section 3.7).

In chapter 4, which is based on [CAE19], the relation between the circuit infidelity and the infidelity of circuit components is made much more limpid. It is first shown that there

exists an unambiguous way to factorize any non-catastrophic channel \mathcal{A} into a decoherent and a physical unitary component: $\mathcal{A} = \mathcal{V} \circ \mathcal{D}$. It is then shown that under realistic assumptions, which is rigorously defined through the equability condition (see definition 7), the process infidelity of $\mathcal{A} = \mathcal{V} \circ \mathcal{D}$ to its target \mathcal{U} is expressed as a sum:

$$\delta\Phi(\mathcal{A}, \mathcal{U}) = \underbrace{\delta\Phi(\mathcal{D})}_{\delta\Phi_{\text{decoh}}} + \underbrace{\delta\Phi(\mathcal{V}, \mathcal{U})}_{\delta\Phi_{\text{coh}}(\mathcal{A}, \mathcal{U})} + O(\delta^2\Phi(\mathcal{A}, \mathcal{U})) . \quad (8.3)$$

More remarkably, still under the equability condition, it is shown that the process fidelity of a composition $\mathcal{A}_{m:1}$ of elements $\mathcal{A}_i = \mathcal{V}_i \circ \mathcal{D}_i$ obeys

$$\Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1}) = \Phi(\mathcal{V}_{m:1}, \mathcal{U}_{m:1}) \prod_{i=1}^m \Phi(\mathcal{D}_i) + O(\delta^2\Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1})), \quad (8.4)$$

which fully generalizes eq. (8.1). In particular, in the regime where $\delta^2\Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1})$ is negligible, eq. (8.4) rigorously confirms the intuition provided by the saturation examples for eq. (8.1): any non-linear signature in the infidelity must be attributed to unitary processes alone.

The equability condition introduced to obtain eq. (8.4) was also used to derive strong connections between process fidelities – which are akin to inner products – to norms such as the diamond distance and the superoperator spectral norm (see theorems 16 and 17). Those norms frequently appear in quantum circuit error analyses but, as opposed to process fidelities, they aren't efficiently extractable through experimental means, which motivates the appeal of relating them to more pragmatic quantities (item iii.). As shown in [WF14; Wal15; SWS15], the diamond distance ϵ_{\diamond} is expected to generally scale as the square root of the infidelity. What is shown more precisely here is that in the regime where the coherent infidelity $\delta\Phi_{\text{coh}}(\mathcal{A}, \mathcal{U})$ is comparable to the total infidelity $\delta\Phi(\mathcal{A}, \mathcal{U})$,

$$\sqrt{\delta\Phi_{\text{coh}}(\mathcal{A}, \mathcal{U})} \stackrel{\delta\Phi}{\lesssim} \epsilon_{\diamond}(\mathcal{A}, \mathcal{U}) \stackrel{\delta\Phi}{\gtrsim} \sqrt{\Gamma_{\text{coh}}} \sqrt{\delta\Phi_{\text{coh}}(\mathcal{A}, \mathcal{U})}, \quad (8.5)$$

where Γ_{coh} is a strict-sense equability constant appearing in definition 7, and it is not expected to scale with the dimension. Similarly, the spectral norm is upper-bounded by $O(\sqrt{\delta\Phi_{\text{coh}}(\mathcal{A}, \mathcal{U})})$:

$$\max_{M \in M_d(\mathbb{C})} \frac{\|(\mathcal{A} - \mathcal{U}_{\text{target}})[M]\|_2}{\|M\|_2} \stackrel{\delta\Phi}{\lesssim} 2\sqrt{\Gamma_{\text{coh}}} \sqrt{\delta\Phi_{\text{coh}}}. \quad (8.6)$$

Such scaling in the spectral norm will allow to confidently generalize the results obtained in [Car+18], as discussed in the second part of this thesis.

The second part of this thesis treats of the interpretation of RB data (item ii.), and of the design of larger families of benchmarking experiments (item i.). Standard RB is expected to yield, by estimating the ratio of “success” events for multiple circuit lengths, a single exponential decay of the form $Ap^m + B$ [MGE11; MGE12]. This behavior has been proven to occur even in the case of gate-dependent Markovian error models, although the connection between the estimated decay parameter p and a physically meaningful figure of merit was generally left unresolved [Pro+17a; Wal17]. More precisely, the decay parameter p was shown to correspond to an average gate-set fidelity to a targeted gate-set realization

expressed in some potentially nonphysical gauge (defined through the linear transformation $\mathcal{B} \in M_{d^2}(\mathbb{C})$):

$$\frac{1}{d} + \frac{d-1}{d}p = \mathbb{E}_{g \in \mathbf{G}} F(\Lambda(g), \mathcal{B}L(g)\mathcal{B}^{-1}). \quad (8.7)$$

In chapter 5, which is based on [Car+18], the decay constant is reconciled to high precision with a physically meaningful gate-set fidelity in the single qubit case (item ii.). That is, it is shown that there exists a physical gauge transformation, defined by a unitary $U \in U(d)$ such that

$$\frac{1}{d} + \frac{d-1}{d}p = \mathbb{E}_{g \in \mathbf{G}} F(\Lambda(g), L(UgU^{-1})) + O((1-p)^2). \quad (8.8)$$

In [Car+18], eq. (8.8) is conjectured to hold for all dimensions. The conjecture is proved later in chapter 5 by using the polar decomposition of quantum channels, as well as the bound on the spectral norm of superoperators, eq. (8.6) (both shown in chapter 4). This is not merely an existence proof; given an error map Λ and a randomizing 2-design \mathbf{G} , there is a (gauge-fixing) procedure to find the physical target gate-set realization that connects the fidelity with an experimentally observable decay.

In chapter 6, which is based on [CWE15], a group-based generalization of RB, referred to as dihedral benchmarking, is proposed for single qubit operations (item i.). Instead of invoking a 2-design randomizing gate-set, the dihedral group is considered (see eq. (6.2) for the definition). An advantage of considering the dihedral group as a gate-set is that it naturally yields a characterization procedure for the $\pi/8$ gate, an operation which is often proposed to be implemented via drastically different physical procedures than Clifford operations[BK05]. The mathematical toolkit used to ensure the functioning of dihedral benchmarking (i.e. representation theory and character theory) can be also used to extend benchmarking techniques to other group-based protocols [Cro+16; Hel+18; SH18; OWE18; BE18; Erh+19].

Chapter 7 is divided in two main sections. The first one is based on [Boo+19], and consists of a reinvestigation of “NIST RB”, a protocol described in [Kni+08] (item ii.). The experiment had the particularity of using a randomizing gate-set that neither formed a group, nor a 2-design, rendering typical group-based analysis tools unusable. Under the assumption of gate-independent errors, it was formally shown in [Boo+19] that the NIST RB experiment yields, for large enough sequence lengths values, a decay model of the form $Ap^m + B$ where p can be related to a gate-set fidelity. The second section of chapter 7 revolves around the analysis of a more general generator-based protocol, known as direct RB [Pro+18] (items i. and ii.). In particular, it is shown that even in the advent of gate-dependent errors, choosing a randomizing gate-set \mathbf{H} for which the elements generate a 2-design is typically sufficient¹ to yield a decay model of the form

$$Ap^m + B + \sum_i C_i \lambda_i^m, \quad (8.9)$$

where C_i are expected to be small, and where the various $|\lambda_i|^m$ decay noticeably faster than p^m . Moreover, it is shown that the decay constant p connects with a gate-set fidelity to a (potentially nonphysical) target gate-set realization.

¹Provided that the gate-set infidelity is much smaller than the spectral gap featuring in the twirl matrix.

This summarizes the work presented in this thesis. Many further developments based on those ideas can be proposed. To enumerate two promising avenues:

- i. As shown in eq. (8.4), any deviation from a simple fidelity decay is due to a unitary evolution. This realization is expected to enable the design (item i.) of scalable RB-type experiments conceived to estimate the level of coherence in quantum errors.
- ii. The gauge-fixing procedure involved in finding the target gate-set realization in eq. (8.8) doesn't fundamentally rely on 2-designs. In particular, it seems possible that weaker algebraic structures, such as the generators of 1-designs are enough to fix reference gauges. This would be a valuable result for the development of direct RB routines, for which reference gauge-fixing could be ensured at a low operational cost.

Bibliography

- [Art] Michael Artin. *Algebra*. Pearson, 2nd Edition.
- [Bar+14] R. Barends et al. “Rolling quantum dice with a superconducting qubit”. In: *Physical Review A* 90, 030303 (2014), p. 030303. DOI: [10.1103/PhysRevA.90.030303](https://doi.org/10.1103/PhysRevA.90.030303). arXiv: [1406.3364](https://arxiv.org/abs/1406.3364) [quant-ph].
- [Bar+15] R. Barends et al. “Digital quantum simulation of fermionic models with a superconducting circuit”. In: *Nature Communications* 6, 7654 (July 2015), p. 7654. DOI: [10.1038/ncomms8654](https://doi.org/10.1038/ncomms8654). arXiv: [1501.07703](https://arxiv.org/abs/1501.07703) [quant-ph].
- [BE18] Winton G. Brown and Bryan Eastin. “Randomized benchmarking with restricted gate sets”. In: *Physical Review A* 97, 062323 (June 2018), p. 062323. DOI: [10.1103/PhysRevA.97.062323](https://doi.org/10.1103/PhysRevA.97.062323). arXiv: [1801.04042](https://arxiv.org/abs/1801.04042) [quant-ph].
- [BF60] F. L. Bauer and C. T. Fike. “Norms and exclusion theorems”. In: *Numerische Mathematik* 2.1 (1960), p. 137. ISSN: 0029-599X. DOI: [10.1007/BF01386217](https://doi.org/10.1007/BF01386217). URL: <http://link.springer.com/10.1007/BF01386217>.
- [Bha97] Rajendra Bhatia. *Matrix Analysis*. Springer, 1997.
- [BK05] Sergey Bravyi and Alexei Kitaev. “Universal quantum computation with ideal Clifford gates and noisy ancillas”. In: *Physical Review A* 71.2 (Feb. 2005), p. 022316. ISSN: 1050-2947. DOI: [10.1103/PhysRevA.71.022316](https://doi.org/10.1103/PhysRevA.71.022316). arXiv: [0403025v2](https://arxiv.org/abs/0403025v2) [arXiv:quant-ph]. URL: <http://link.aps.org/doi/10.1103/PhysRevA.71.022316>.
- [Blo46] F. Bloch. “Nuclear Induction”. In: *Phys. Rev.* 70 (7-8 1946), pp. 460–474. DOI: [10.1103/PhysRev.70.460](https://doi.org/10.1103/PhysRev.70.460). URL: <https://link.aps.org/doi/10.1103/PhysRev.70.460>.
- [Blu+10] Robin Blume-Kohout et al. “Information-preserving structures: A general framework for quantum zero-error information”. In: 82.6, 062306 (2010), p. 062306. DOI: [10.1103/PhysRevA.82.062306](https://doi.org/10.1103/PhysRevA.82.062306). arXiv: [1006.1358](https://arxiv.org/abs/1006.1358) [quant-ph].
- [Boo+19] Kristine Boone et al. “Randomized benchmarking under different gate sets”. In: *Phys. Rev. A* 99 (3 2019), p. 032329. DOI: [10.1103/PhysRevA.99.032329](https://doi.org/10.1103/PhysRevA.99.032329). URL: <https://link.aps.org/doi/10.1103/PhysRevA.99.032329>.
- [Bro+11] K. R. Brown et al. “Single-qubit-gate error below 10^{-4} in a trapped ion”. In: *Physical Review A* 84.3 (Sept. 2011), p. 030303. ISSN: 1050-2947. DOI: [10.1103/PhysRevA.84.030303](https://doi.org/10.1103/PhysRevA.84.030303). URL: <http://link.aps.org/doi/10.1103/PhysRevA.84.030303>.
- [BW04] P. S. Bourdon and H. T. Williams. “Unital quantum operations on the Bloch ball and Bloch region”. In: *Phys. Rev. A* 69 (2 2004), p. 022314. DOI: [10.1103/PhysRevA.69.022314](https://doi.org/10.1103/PhysRevA.69.022314). URL: <https://link.aps.org/doi/10.1103/PhysRevA.69.022314>.

- [BZ06] Ingemar Bengtsson and Karol Zyczkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, 2006. ISBN: 0521814510.
- [CAE19] Arnaud Carignan-Dugas, Matthew Alexander, and Joseph Emerson. “A polar decomposition for quantum channels (with applications to bounding error propagation in quantum circuits)”. In: *Quantum* 3 (Aug. 2019), p. 173. ISSN: 2521-327X. DOI: [10.22331/q-2019-08-12-173](https://doi.org/10.22331/q-2019-08-12-173). URL: <https://doi.org/10.22331/q-2019-08-12-173>.
- [Cal+18] S. A. Caldwell et al. “Parametrically Activated Entangling Gates Using Transmon Qubits”. In: *Physical Review Applied* 10, 034050 (Sept. 2018), p. 034050. DOI: [10.1103/PhysRevApplied.10.034050](https://doi.org/10.1103/PhysRevApplied.10.034050). arXiv: [1706.06562](https://arxiv.org/abs/1706.06562) [quant-ph].
- [Car+18] Arnaud Carignan-Dugas et al. “From randomized benchmarking experiments to gate-set circuit fidelity: how to interpret randomized benchmarking decay parameters”. In: *New Journal of Physics* 20, 092001 (Sept. 2018), p. 092001. DOI: [10.1088/1367-2630/aadcc7](https://doi.org/10.1088/1367-2630/aadcc7).
- [Cas+16] L. Casparis et al. “Gatemon Benchmarking and Two-Qubit Operations”. In: *Physical Review Letter* 116, 150505 (Apr. 2016), p. 150505. DOI: [10.1103/PhysRevLett.116.150505](https://doi.org/10.1103/PhysRevLett.116.150505).
- [Cha+18] K. W. Chan et al. “Assessment of a Silicon Quantum Dot Spin Qubit Environment via Noise Spectroscopy”. In: *Physical Review Applied* 10, 044017 (Oct. 2018), p. 044017. DOI: [10.1103/PhysRevApplied.10.044017](https://doi.org/10.1103/PhysRevApplied.10.044017). arXiv: [1803.01609](https://arxiv.org/abs/1803.01609) [quant-ph].
- [Cho75] Man-Duen Choi. “Completely positive linear maps on complex matrices”. In: *Linear Algebra and its Applications* 10.3 (1975), pp. 285–290. ISSN: 0024-3795. DOI: [https://doi.org/10.1016/0024-3795\(75\)90075-0](https://doi.org/10.1016/0024-3795(75)90075-0). URL: <http://www.sciencedirect.com/science/article/pii/0024379575900750>.
- [CN97] Isaac L. Chuang and Michael A. Nielsen. “Prescription for experimental determination of the dynamics of a quantum black box”. In: *Journal of Modern Optics* 44.11-12 (Nov. 1997), p. 2455. ISSN: 0950-0340. DOI: [10.1080/09500349708231894](https://doi.org/10.1080/09500349708231894). URL: <http://www.tandfonline.com/doi/abs/10.1080/09500349708231894>.
- [Com+17] Joshua Combes et al. “Logical Randomized Benchmarking”. In: *arXiv e-prints*, arXiv:1702.03688 (Feb. 2017), arXiv:1702.03688. arXiv: [1702.03688](https://arxiv.org/abs/1702.03688) [quant-ph].
- [Cór+13] Antonio D. Córcoles et al. “Process verification of two-qubit quantum gates by randomized benchmarking”. In: *Physical Review A* 87.3 (2013), p. 030301. ISSN: 1050-2947. DOI: [10.1103/PhysRevA.87.030301](https://doi.org/10.1103/PhysRevA.87.030301). URL: <http://link.aps.org/doi/10.1103/PhysRevA.87.030301>.
- [Cro+16] Andrew W. Cross et al. “Scalable randomised benchmarking of non-Clifford gates”. In: *npj Quantum Information* 2, 16012 (Apr. 2016), p. 16012. DOI: [10.1038/npjqi.2016.12](https://doi.org/10.1038/npjqi.2016.12). arXiv: [1510.02720](https://arxiv.org/abs/1510.02720) [quant-ph].
- [CW15] T. Chasseur and F. K. Wilhelm. “Complete randomized benchmarking protocol accounting for leakage errors”. In: *Physical Review A - Atomic, Molecular, and Optical Physics* 92.4 (2015), p. 042333. ISSN: 10941622. DOI: [10.1103/PhysRevA.92.042333](https://doi.org/10.1103/PhysRevA.92.042333). arXiv: [1505.00580v2](https://arxiv.org/abs/1505.00580v2).

- [CWE15] Arnaud Carignan-Dugas, Joel J. Wallman, and Joseph Emerson. “Characterizing universal gate sets via dihedral benchmarking”. In: *Physical Review A* 92, 060302 (Dec. 2015), p. 060302. DOI: [10 . 1103 / PhysRevA . 92 . 060302](https://doi.org/10.1103/PhysRevA.92.060302). arXiv: [1508.06312](https://arxiv.org/abs/1508.06312) [quant-ph].
- [CWE16] Arnaud Carignan-Dugas, Joel J. Wallman, and Joseph Emerson. “Bounding the average gate fidelity of composite channels using the unitarity”. In: *arXiv e-prints*, arXiv:1610.05296 (Oct. 2016), arXiv:1610.05296. arXiv: [1610 . 05296](https://arxiv.org/abs/1610.05296) [quant-ph].
- [Dan+06] Christoph Dankert et al. “Exact and approximate unitary 2-designs: constructions and applications”. In: (2006). arXiv: [0606161v1](https://arxiv.org/abs/0606161v1). URL: <http://arxiv.org/abs/quant-ph/0606161v1>.
- [Dan+09] Christoph Dankert et al. “Exact and approximate unitary 2-designs and their application to fidelity estimation”. In: *Physical Review A* 80.1 (July 2009), p. 012304. ISSN: 1050-2947. DOI: [10 . 1103 / PhysRevA . 80 . 012304](https://doi.org/10.1103/PhysRevA.80.012304). URL: <http://link.aps.org/doi/10.1103/PhysRevA.80.012304>.
- [DCP15] Guillaume Duclos-Cianci and David Poulin. “Reducing the quantum-computing overhead with complex gate distillation”. In: *Physical Review A* 91.4 (Apr. 2015), p. 042315. ISSN: 1050-2947. DOI: [10 . 1103 / PhysRevA . 91 . 042315](https://doi.org/10.1103/PhysRevA.91.042315). arXiv: [1403.5280](https://arxiv.org/abs/1403.5280). URL: <http://arxiv.org/abs/1403.5280>.
- [DHW19] Bas Dirkse, Jonas Helsen, and Stephanie Wehner. “Efficient unitarity randomized benchmarking of few-qubit Clifford gates”. In: *Physical Review A* 99, 012315 (Jan. 2019), p. 012315. DOI: [10 . 1103 / PhysRevA . 99 . 012315](https://doi.org/10.1103/PhysRevA.99.012315). arXiv: [1808.00850](https://arxiv.org/abs/1808.00850) [quant-ph].
- [EAZ05] Joseph Emerson, Robert Alicki, and Karol Życzkowski. “Scalable noise estimation with random unitary operators”. In: *Journal of Optics B: Quantum and Semi-classical Optics* 7.10 (Oct. 2005), S347–S352. ISSN: 1464-4266. DOI: [10 . 1088 / 1464 - 4266 / 7 / 10 / 021](https://doi.org/10.1088/1464-4266/7/10/021). arXiv: [0503243](https://arxiv.org/abs/0503243) [quant-ph]. URL: <http://arxiv.org/abs/quant-ph/0503243>.
- [EBW87] R.R. Ernst, G. Bodenhausen, and A. Wokaun. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. International series of monographs on chemistry. Clarendon Press, 1987. ISBN: 9780198556299. URL: <https://books.google.ca/books?id=XndTnwEACAAJ>.
- [Eme+07] Joseph Emerson et al. “Symmetrized characterization of noisy quantum processes.” In: *Science* 317.5846 (Sept. 2007), pp. 1893–6. ISSN: 1095-9203. DOI: [10 . 1126 / science . 1145699](https://doi.org/10.1126/science.1145699). URL: <http://www.ncbi.nlm.nih.gov/pubmed/17901327>.
- [Erh+19] Alexander Erhard et al. “Characterizing large-scale quantum computers via cycle benchmarking”. In: *arXiv e-prints*, arXiv:1902.08543 (2019), arXiv:1902.08543. arXiv: [1902.08543](https://arxiv.org/abs/1902.08543) [quant-ph].
- [FA99] Akio Fujiwara and Paul Algoet. “One-to-one parametrization of quantum channels”. In: *Phys. Rev. A* 59 (5 1999), pp. 3290–3294. DOI: [10 . 1103 / PhysRevA . 59 . 3290](https://doi.org/10.1103/PhysRevA.59.3290). URL: <https://link.aps.org/doi/10.1103/PhysRevA.59.3290>.

- [Fen+16] Guanru Feng et al. “Estimating the Coherence of Noise in Quantum Control of a Solid-State Qubit”. In: *Physical Review Letter* 117, 260501 (Dec. 2016), p. 260501. DOI: [10.1103/PhysRevLett.117.260501](https://doi.org/10.1103/PhysRevLett.117.260501). arXiv: [1603.03761](https://arxiv.org/abs/1603.03761) [quant-ph].
- [For+15] Simon Forest et al. “Exact synthesis of single-qubit unitaries over Clifford-cyclotomic gate sets”. In: (2015). arXiv: [1501.04944](https://arxiv.org/abs/1501.04944). URL: <http://arxiv.org/abs/1501.04944>.
- [Fow+12] Austin G. Fowler et al. “Surface codes: Towards practical large-scale quantum computation”. In: *Physical Review A* 86.3 (Sept. 2012), p. 032324. ISSN: 1050-2947. DOI: [10.1103/PhysRevA.86.032324](https://doi.org/10.1103/PhysRevA.86.032324). arXiv: [1208.0928](https://arxiv.org/abs/1208.0928). URL: <http://arxiv.org/abs/1208.0928>.
- [GAE07] David Gross, K. Audenaert, and Jens Eisert. “Evenly distributed unitaries: On the structure of unitary designs”. In: *Journal of Mathematical Physics* 48.5 (2007), p. 052104. ISSN: 00222488. DOI: [10.1063/1.2716992](https://doi.org/10.1063/1.2716992). URL: <http://link.aip.org/link/JMAPAQ/v48/i5/p052104/s1&Agg=doi>.
- [Gae+12] J. P. Gaebler et al. “Randomized Benchmarking of Multiqubit Gates”. In: *Phys. Rev. Lett.* 108 (26 2012), p. 260503. DOI: [10.1103/PhysRevLett.108.260503](https://doi.org/10.1103/PhysRevLett.108.260503). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.108.260503>.
- [Gam+12] Jay M. Gambetta et al. “Characterization of Addressability by Simultaneous Randomized Benchmarking”. In: *Physical Review Letter* 109, 240504 (Dec. 2012), p. 240504. DOI: [10.1103/PhysRevLett.109.240504](https://doi.org/10.1103/PhysRevLett.109.240504). arXiv: [1204.6308](https://arxiv.org/abs/1204.6308) [quant-ph].
- [Gam+12] Jay M. Gambetta et al. “Characterization of addressability by simultaneous randomized benchmarking”. In: *Physical Review Letters* 109.24 (Dec. 2012), p. 240504. ISSN: 00319007. DOI: [10.1103/PhysRevLett.109.240504](https://doi.org/10.1103/PhysRevLett.109.240504). arXiv: [arXiv:1204.6308v1](https://arxiv.org/abs/1204.6308v1). URL: <http://link.aps.org/doi/10.1103/PhysRevLett.109.240504>.
- [GB15] Mauricio Gutiérrez and Kenneth R. Brown. “Comparison of a quantum error-correction threshold for exact and approximate errors”. In: *Phys. Rev. A* 91 (2 2015), p. 022335. DOI: [10.1103/PhysRevA.91.022335](https://doi.org/10.1103/PhysRevA.91.022335). URL: <http://link.aps.org/doi/10.1103/PhysRevA.91.022335>.
- [GFC14] Christopher Granade, Christopher Ferrie, and D G Cory. “Accelerated Randomized Benchmarking”. In: *New Journal of Physics* 17.1 (Jan. 2014), pp. 1–6. ISSN: 1367-2630. DOI: [10.1088/1367-2630/17/1/013042](https://doi.org/10.1088/1367-2630/17/1/013042). arXiv: [arXiv:1404.5275v1](https://arxiv.org/abs/1404.5275v1). URL: <http://arxiv.org/abs/1404.5275>.
- [GKS76] Vittorio Gorini, Andrzej Kossakowski, and E. C. G. Sudarshan. “Completely positive dynamical semigroups of N-level systems”. In: *Journal of Mathematical Physics* 17.5 (1976), pp. 821–825. DOI: [10.1063/1.522979](https://doi.org/10.1063/1.522979). eprint: <https://aip.scitation.org/doi/pdf/10.1063/1.522979>. URL: <https://aip.scitation.org/doi/abs/10.1063/1.522979>.
- [Got10] Daniel Gottesman. “An Introduction to Quantum Error Correction and Fault-Tolerant Quantum Computation”. In: *Proceedings of Symposia in Applied Mathematics* 68 (2010), p. 13. arXiv: [0904.2557v1](https://arxiv.org/abs/0904.2557v1).

- [Har+14] T. P. Harty et al. “High-Fidelity Preparation, Gates, Memory, and Readout of a Trapped-Ion Quantum Bit”. In: *Phys. Rev. Lett.* 113 (22 2014), p. 220501. DOI: [10.1103/PhysRevLett.113.220501](https://doi.org/10.1103/PhysRevLett.113.220501). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.113.220501>.
- [Har+19] Robin Harper et al. “Statistical analysis of randomized benchmarking”. In: *arXiv e-prints*, arXiv:1901.00535 (Jan. 2019), arXiv:1901.00535. arXiv: [1901.00535](https://arxiv.org/abs/1901.00535) [quant-ph].
- [Has+18] A. K. Hashagen et al. “Real Randomized Benchmarking”. In: *arXiv e-prints*, arXiv:1801.06121 (Jan. 2018), arXiv:1801.06121. arXiv: [1801.06121](https://arxiv.org/abs/1801.06121) [quant-ph].
- [Hav03] Timothy F. Havel. “Robust procedures for converting among Lindblad, Kraus and matrix representations of quantum dynamical semigroups”. In: *Journal of Mathematical Physics* 44 (Feb. 2003), pp. 534–557. DOI: [10.1063/1.1518555](https://doi.org/10.1063/1.1518555). arXiv: [quant-ph/0201127](https://arxiv.org/abs/quant-ph/0201127) [quant-ph].
- [Hel+18] Jonas Helsen et al. “A new class of efficient randomized benchmarking protocols”. In: *arXiv e-prints*, arXiv:1806.02048 (June 2018), arXiv:1806.02048. arXiv: [1806.02048](https://arxiv.org/abs/1806.02048) [quant-ph].
- [HF17] Robin Harper and Steven T. Flammia. “Estimating the fidelity of T gates using standard interleaved randomized benchmarking”. In: *Quantum Science and Technology* 2 (Mar. 2017), p. 015008. DOI: [10.1088/2058-9565/aa5f8d](https://doi.org/10.1088/2058-9565/aa5f8d). arXiv: [1608.02943](https://arxiv.org/abs/1608.02943) [quant-ph].
- [Kel+14] J. Kelly et al. “Optimal Quantum Control Using Randomized Benchmarking”. In: *Physical Review Letters* 112.24, 240504 (June 2014), p. 240504. DOI: [10.1103/PhysRevLett.112.240504](https://doi.org/10.1103/PhysRevLett.112.240504). arXiv: [1403.0035](https://arxiv.org/abs/1403.0035) [quant-ph].
- [Kim+14] Shelby Kimmel et al. “Robust Extraction of Tomographic Information via Randomized Benchmarking”. In: *Physical Review X* 4.1 (Mar. 2014), p. 011050. ISSN: 2160-3308. DOI: [10.1103/PhysRevX.4.011050](https://doi.org/10.1103/PhysRevX.4.011050). arXiv: [1306.2348](https://arxiv.org/abs/1306.2348). URL: <http://arxiv.org/abs/1306.2348>.
- [Kni05] Emanuel Knill. “Quantum computing with realistically noisy devices.” In: *Nature* 434.7029 (2005), pp. 39–44. ISSN: 0028-0836. DOI: [10.1038/nature03350](https://doi.org/10.1038/nature03350). arXiv: [0410199](https://arxiv.org/abs/0410199) [quant-ph].
- [Kni+08] E. Knill et al. “Randomized benchmarking of quantum gates”. In: *Physical Review A* 77.1 (Jan. 2008), p. 012307. ISSN: 1050-2947. DOI: [10.1103/PhysRevA.77.012307](https://doi.org/10.1103/PhysRevA.77.012307). URL: <http://link.aps.org/doi/10.1103/PhysRevA.77.012307>.
- [Kra+83] K. Kraus et al. *States, effects, and operations: fundamental notions of quantum theory : lectures in mathematical physics at the University of Texas at Austin*. Lecture notes in physics. Springer-Verlag, 1983. ISBN: 9780387127323. URL: <https://books.google.ca/books?id=fRBBAQAIAAJ>.
- [LC13] Andrew J. Landahl and Chris Cesare. In: (2013). arXiv: [1302.3240](https://arxiv.org/abs/1302.3240). URL: <http://arxiv.org/abs/1302.3240>.
- [Lev+07] Benjamin Levi et al. “Efficient error characterization in quantum information processing”. In: *Physical Review A* 75 (2007), p. 022314. DOI: [10.1103/PhysRevA.75.022314](https://doi.org/10.1103/PhysRevA.75.022314).

- [Lin76] G. Lindblad. “On the generators of quantum dynamical semigroups”. In: *Communications in Mathematical Physics* 48 (June 1976), pp. 119–130. DOI: [10.1007/BF01608499](https://doi.org/10.1007/BF01608499).
- [Mag+12] Easwar Magesan et al. “Efficient Measurement of Quantum Gate Error by Interleaved Randomized Benchmarking”. In: *Physical Review Letters* 109.8 (Aug. 2012), p. 080505. ISSN: 0031-9007. DOI: [10.1103/PhysRevLett.109.080505](https://doi.org/10.1103/PhysRevLett.109.080505). arXiv: [1203.4550](https://arxiv.org/abs/1203.4550). URL: <http://arxiv.org/abs/1203.4550>.
- [Mag+13] Easwar Magesan et al. “Modeling quantum noise for efficient testing of fault-tolerant circuits”. In: *Phys. Rev. A* 87 (1 2013), p. 012324. DOI: [10.1103/PhysRevA.87.012324](https://doi.org/10.1103/PhysRevA.87.012324). URL: <http://link.aps.org/doi/10.1103/PhysRevA.87.012324>.
- [McK+16] David C. McKay et al. “Universal Gate for Fixed-Frequency Qubits via a Tunable Bus”. In: *Physical Review Applied* 6, 064007 (Dec. 2016), p. 064007. DOI: [10.1103/PhysRevApplied.6.064007](https://doi.org/10.1103/PhysRevApplied.6.064007).
- [McK+17a] David C. McKay et al. “Efficient Z gates for quantum computing”. In: *Physical Review A* 96, 022330 (Aug. 2017), p. 022330. DOI: [10.1103/PhysRevA.96.022330](https://doi.org/10.1103/PhysRevA.96.022330). arXiv: [1612.00858](https://arxiv.org/abs/1612.00858) [quant-ph].
- [McK+17b] David C. McKay et al. “Three Qubit Randomized Benchmarking”. In: *ArXiv e-prints*, arXiv:1712.06550 (Dec. 2017), arXiv:1712.06550. arXiv: [1712.06550](https://arxiv.org/abs/1712.06550) [quant-ph].
- [MEK12] Adam M. Meier, Bryan Eastin, and Emanuel Knill. “Magic-state distillation with the four-qubit code”. In: (Apr. 2012). arXiv: [1204.4221](https://arxiv.org/abs/1204.4221). URL: <http://arxiv.org/abs/1204.4221>.
- [Mer+13] Seth T. Merkel et al. “Self-consistent quantum process tomography”. In: *Physical Review A* 87.6 (June 2013), p. 062119. ISSN: 1050-2947. DOI: [10.1103/PhysRevA.87.062119](https://doi.org/10.1103/PhysRevA.87.062119). URL: <http://link.aps.org/doi/10.1103/PhysRevA.87.062119>.
- [MGE11] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson. “Scalable and Robust Randomized Benchmarking of Quantum Processes”. In: *Physical Review Letters* 106.18 (May 2011), p. 180504. ISSN: 0031-9007. DOI: [10.1103/PhysRevLett.106.180504](https://doi.org/10.1103/PhysRevLett.106.180504). URL: <http://link.aps.org/doi/10.1103/PhysRevLett.106.180504>.
- [MGE12] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson. “Characterizing quantum gates via randomized benchmarking”. In: *Physical Review A* 85.4 (Apr. 2012), p. 042311. ISSN: 1050-2947. DOI: [10.1103/PhysRevA.85.042311](https://doi.org/10.1103/PhysRevA.85.042311). URL: <http://link.aps.org/doi/10.1103/PhysRevA.85.042311>.
- [MPF18] Seth T. Merkel, Emily J. Pritchett, and Bryan H. Fong. “Randomized Benchmarking as Convolution: Fourier Analysis of Gate Dependent Errors”. In: *arXiv e-prints*, arXiv:1804.05951 (2018), arXiv:1804.05951. arXiv: [1804.05951](https://arxiv.org/abs/1804.05951) [quant-ph].
- [Muh+15] J T Muhonen et al. “Quantifying the quantum gate fidelity of single-atom spin qubits in silicon by randomized benchmarking”. In: *Journal of Physics: Condensed Matter* 27.15 (2015), p. 154205. URL: <http://stacks.iop.org/0953-8984/27/i=15/a=154205>.

- [Nic+17] John M. Nichol et al. “High-fidelity entangling gate for double-quantum-dot spin qubits”. In: *npj Quantum Information* 3, 3 (Jan. 2017), p. 3. DOI: [10.1038/s41534-016-0003-1](https://doi.org/10.1038/s41534-016-0003-1). arXiv: 1608.04258 [cond-mat.mes-hall].
- [Nie02] Michael A. Nielsen. “A simple formula for the average gate fidelity of a quantum dynamical operation”. In: *Physics Letters A* 303.4 (Oct. 2002), p. 249. ISSN: 03759601. DOI: [10.1016/S0375-9601\(02\)01272-0](https://doi.org/10.1016/S0375-9601(02)01272-0). URL: <http://linkinghub.elsevier.com/retrieve/pii/S0375960102012720>.
- [OWE18] E. Onorati, A. H. Werner, and J. Eisert. “Randomized benchmarking for individual quantum gates”. In: *arXiv e-prints*, arXiv:1811.11775 (2018), arXiv:1811.11775. arXiv: 1811.11775 [quant-ph].
- [PCZ97] J. F. Poyatos, J. Ignacio Cirac, and P. Zoller. “Complete Characterization of a Quantum Process: The Two-Bit Quantum Gate”. In: *Physical Review Letters* 78.2 (Jan. 1997), p. 390. ISSN: 0031-9007. DOI: [10.1103/PhysRevLett.78.390](https://doi.org/10.1103/PhysRevLett.78.390). URL: <http://link.aps.org/doi/10.1103/PhysRevLett.78.390>.
- [PG+06] D. Pérez-García et al. “Contractivity of positive and trace-preserving maps under L_p norms”. In: *Journal of Mathematical Physics* 47.8 (Aug. 2006), pp. 083506–083506. DOI: [10.1063/1.2218675](https://doi.org/10.1063/1.2218675). eprint: math-ph/0601063.
- [Pro+17a] T. Proctor et al. “What Randomized Benchmarking Actually Measures”. In: *Physical Review Letters* 119.13, 130502 (Sept. 2017), p. 130502. DOI: [10.1103/PhysRevLett.119.130502](https://doi.org/10.1103/PhysRevLett.119.130502). arXiv: 1702.01853 [quant-ph].
- [Pro+17b] Timothy Proctor et al. “What Randomized Benchmarking Actually Measures”. In: *arXiv e-prints*, arXiv:1702.01853v1 (Feb. 2017). arXiv: 1702.01853v1 [quant-ph].
- [Pro+18] Timothy J. Proctor et al. “Direct randomized benchmarking for multi-qubit devices”. In: *arXiv e-prints*, arXiv:1807.07975 (July 2018), arXiv:1807.07975. arXiv: 1807.07975 [quant-ph].
- [Puz+14] Daniel Puzzuoli et al. “Tractable simulation of error correction with honest approximations to realistic fault models”. In: *Phys. Rev. A* 89 (2 2014), p. 022306. DOI: [10.1103/PhysRevA.89.022306](https://doi.org/10.1103/PhysRevA.89.022306). URL: <http://link.aps.org/doi/10.1103/PhysRevA.89.022306>.
- [QK18] Jiaan Qi and Hui Khoon Ng. “Comparing randomized benchmarking figure with average infidelity of quantum gate-set”. In: *arXiv e-prints*, arXiv:1805.10622 (May 2018), arXiv:1805.10622. arXiv: 1805.10622 [quant-ph].
- [RSW02] Mary Beth Ruskai, Stanislaw Szarek, and Elisabeth Werner. “An analysis of completely-positive trace-preserving maps on M_2 ”. In: *Linear Algebra and its Applications* 347.1 (2002), pp. 159–187. ISSN: 0024-3795. DOI: [https://doi.org/10.1016/S0024-3795\(01\)00547-X](https://doi.org/10.1016/S0024-3795(01)00547-X). URL: <http://www.sciencedirect.com/science/article/pii/S002437950100547X>.
- [Ser] Jean-Pierre Serre. *Linear Representations of Finite Groups*. Springer, New-York, NY.
- [SH18] Daniel Stilck França and Anna-Lena Hashagen. “Approximate Randomized Benchmarking for Finite Groups”. In: *arXiv e-prints*, arXiv:1803.03621 (Mar. 2018), arXiv:1803.03621. arXiv: 1803.03621 [quant-ph].

- [She+16] Sarah Sheldon et al. “Characterizing errors on qubit operations via iterative randomized benchmarking”. In: *Phys. Rev. A* 93 (1 2016), p. 012301. DOI: [10.1103/PhysRevA.93.012301](https://doi.org/10.1103/PhysRevA.93.012301). URL: <http://link.aps.org/doi/10.1103/PhysRevA.93.012301>.
- [SWS15] Yuval R. Sanders, Joel J. Wallman, and Barry C. Sanders. “Bounding Quantum Gate Error Rate Based on Reported Gate Fidelity”. In: (2015). arXiv: [1501.04932](https://arxiv.org/abs/1501.04932). URL: <http://arxiv.org/abs/1501.04932>.
- [Tak+16a] K. Takeda et al. “A fault-tolerant addressable spin qubit in a natural silicon quantum dot”. In: *Science Advances* 2 (Aug. 2016), e1600694–e1600694. DOI: [10.1126/sciadv.1600694](https://doi.org/10.1126/sciadv.1600694). arXiv: [1602.07833](https://arxiv.org/abs/1602.07833) [[cond-mat.mes-hall](#)].
- [Tak+16b] Maika Takita et al. “Demonstration of Weight-Four Parity Measurements in the Surface Code Architecture”. In: *Physical Review Letter* 117, 210505 (Nov. 2016), p. 210505. DOI: [10.1103/PhysRevLett.117.210505](https://doi.org/10.1103/PhysRevLett.117.210505).
- [Tar] James Edward Tarlton. “Probing qubit memory errors at the 10e-5 level”. PhD thesis, pp. 10–4.
- [Vel+14] M. Veldhorst et al. “An addressable quantum dot qubit with fault-tolerant control-fidelity”. In: *Nature Nanotechnology* 9 (Dec. 2014), pp. 981–985. DOI: [10.1038/nnano.2014.216](https://doi.org/10.1038/nnano.2014.216). arXiv: [1407.1950](https://arxiv.org/abs/1407.1950) [[cond-mat.mes-hall](#)].
- [Vel+15] M. Veldhorst et al. “Spin-orbit coupling and operation of multivalley spin qubits”. In: *Physical Review B* 92, 201401 (Nov. 2015), p. 201401. DOI: [10.1103/PhysRevB.92.201401](https://doi.org/10.1103/PhysRevB.92.201401). arXiv: [1505.01213](https://arxiv.org/abs/1505.01213) [[cond-mat.mes-hall](#)].
- [Wal15] J. J. Wallman. “Bounding experimental quantum error rates relative to fault-tolerant thresholds”. In: *ArXiv e-prints* (Nov. 2015). arXiv: [1511.00727](https://arxiv.org/abs/1511.00727) [[quant-ph](#)].
- [Wal+15] Joel J. Wallman et al. “Estimating the Coherence of Noise”. In: *New J. Phys.* 17 (2015), p. 113020. ISSN: 1367-2630. DOI: [10.1088/1367-2630/17/11/113020](https://doi.org/10.1088/1367-2630/17/11/113020). arXiv: [arXiv:1503.0786](https://arxiv.org/abs/1503.0786). URL: <http://dx.doi.org/10.1088/1367-2630/17/11/113020>.
- [Wal17] J. J. Wallman. “Randomized benchmarking with gate-dependent noise”. In: *ArXiv e-prints* (Mar. 2017). arXiv: [1703.09835](https://arxiv.org/abs/1703.09835) [[quant-ph](#)].
- [Wan+18a] Tenghui Wang et al. “Experimental realization of a fast controlled-Z gate via a shortcut-to-adiabaticity”. In: *arXiv e-prints*, arXiv:1811.08096 (Nov. 2018), arXiv:1811.08096. arXiv: [1811.08096](https://arxiv.org/abs/1811.08096) [[quant-ph](#)].
- [Wan+18b] Tenghui Wang et al. “The experimental realization of high-fidelity ‘shortcut-to-adiabaticity’ quantum gates in a superconducting Xmon qubit”. In: *New Journal of Physics* 20, 065003 (June 2018), p. 065003. DOI: [10.1088/1367-2630/aac9e7](https://doi.org/10.1088/1367-2630/aac9e7). arXiv: [1804.08247](https://arxiv.org/abs/1804.08247) [[quant-ph](#)].
- [Wat12] John Watrous. “Simpler semidefinite programs for completely bounded norms”. In: *arXiv e-prints*, arXiv:1207.5726 (2012), arXiv:1207.5726. arXiv: [1207.5726](https://arxiv.org/abs/1207.5726) [[quant-ph](#)].

- [WBE14] Joel J. Wallman, Marie Barnhill, and Joseph Emerson. “Characterization of Leakage Errors via Randomized Benchmarking”. In: (Dec. 2014), p. 4. arXiv: 1412.4126. URL: <http://arxiv.org/pdf/1412.4126><http://arxiv.org/abs/1412.4126>.
- [WBE15] Joel J. Wallman, Marie Barnhill, and Joseph Emerson. “Robust Characterization of Loss Rates”. In: *Phys. Rev. Lett.* 115.6 (2015), p. 060501. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.115.060501. URL: <http://link.aps.org/doi/10.1103/PhysRevLett.115.060501>.
- [WBE16] Joel J Wallman, Marie Barnhill, and Joseph Emerson. “Robust characterization of leakage errors”. In: *New J. Phys.* 18.4 (2016), p. 043021. ISSN: 1367-2630. DOI: 10.1088/1367-2630/18/4/043021. URL: <http://stacks.iop.org/1367-2630/18/i=4/a=043021?key=crossref.7ef5a2d1daa3b6b75f4a9c8cb1f0f7c9>.
- [WE16] Joel J. Wallman and Joseph Emerson. “Noise tailoring for scalable quantum computation via randomized compiling”. In: *Physical Review A* 94.5 (2016), p. 052325. ISSN: 2469-9926. DOI: 10.1103/PhysRevA.94.052325. URL: <http://link.aps.org/doi/10.1103/PhysRevA.94.052325>.
- [WF14] Joel J Wallman and Steven T Flammia. “Randomized benchmarking with confidence”. In: *New Journal of Physics* 16.10 (Oct. 2014), p. 103032. ISSN: 1367-2630. DOI: 10.1088/1367-2630/16/10/103032. arXiv: 1404.6025. URL: <http://arxiv.org/abs/1404.6025>.
- [WG18] Christopher J Wood and Jay M Gambetta. “Quantification and characterization of leakage errors”. In: *Physical Review A* 97.3 (2018), p. 032306.
- [WG93] Bo-Ying Wang and Ming-Peng Gong. “Some eigenvalue inequalities for positive semidefinite matrix power products”. In: *Linear Algebra and its Applications* 184 (1993), pp. 249–260. ISSN: 0024-3795. DOI: [https://doi.org/10.1016/0024-3795\(93\)90382-X](https://doi.org/10.1016/0024-3795(93)90382-X). URL: <http://www.sciencedirect.com/science/article/pii/002437959390382X>.
- [Xia04] Wen Xiao-Gang. *Quantum Field Theory of Many-Body Systems: From the Origin of Sound to an Origin of Light and Electrons*. Oxford University Press, 2004.
- [Xia+15] T. Xia et al. “Randomized Benchmarking of Single-Qubit Gates in a 2D Array of Neutral-Atom Qubits”. In: *Physical Review Letter* 114, 100503 (Mar. 2015), p. 100503. DOI: 10.1103/PhysRevLett.114.100503. arXiv: 1501.02041 [quant-ph].
- [Xue+18] X. Xue et al. “Benchmarking Gate Fidelities in a Si/SiGe Two-Qubit Device”. In: *arXiv e-prints*, arXiv:1811.04002 (2018), arXiv:1811.04002. arXiv: 1811.04002 [quant-ph].
- [Yan+18] C. H. Yang et al. “Silicon qubit fidelities approaching incoherent noise limits via pulse optimisation”. In: *arXiv e-prints*, arXiv:1807.09500 (July 2018), arXiv:1807.09500. arXiv: 1807.09500 [cond-mat.mes-hall].
- [YC87] E. Yeramian and P. Claverie. “Analysis of multiexponential functions without a hypothesis as to the number of components”. In: *Nature* 326.6109 (1987), pp. 169–174. ISSN: 1476-4687. DOI: 10.1038/326169a0. URL: <https://doi.org/10.1038/326169a0>.

- [Yon+18] Jun Yoneda et al. "A quantum-dot spin qubit with coherence limited by charge noise and fidelity higher than 99.9%". In: *Nature Nanotechnology* 13 (Feb. 2018), pp. 102–106. DOI: [10 . 1038 / s41565 - 017 - 0014 - x](https://doi.org/10.1038/s41565-017-0014-x). arXiv: [1708 . 01454](https://arxiv.org/abs/1708.01454) [[cond-mat.mes-hall](#)].
- [Zha+18] Zhenxing Zhang et al. "Single-shot realization of nonadiabatic holonomic gates with a superconducting Xmon qutrit". In: *arXiv e-prints*, arXiv:1811.06252 (Nov. 2018), arXiv:1811.06252. arXiv: [1811.06252](https://arxiv.org/abs/1811.06252) [[quant-ph](#)].