Doktorarbeit

Random Processes over the Unitary Group

Mixing Properties and Applications in Quantum Information

Emilio Federico Onorati

Jahr 2018

Im Fachbereich Physik der Freien Universität Berlin eingereichte Dissertation
zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften
Dr. rer. nat.

Dahlem Center for Complex
Quantum Systems

Fachbereich Physik

Freie Universität Berlin

Erstgutachter: Prof. Dr. Jens Eisert

Zweitgutachter: Prof. Dr. Wolfgang König

Drittgutachter: Prof. Dr. Michał Horodecki

Datum der Disputation: 24. July 2019

Publications of this author contained in the thesis

- E. Onorati, O. Buerschaper, M. Kliesch, W. Brown, A. H. Werner, J. Eisert, *Mixing properties of stochastic quantum Hamiltonians*, Communications in Mathematical Physics, November 2017, Volume 355, Issue 3.
  doi: 10.1007/s00220-017-2950-6
- E. Onorati, A. H. Werner, J. Eisert, *Randomized benchmarking for individual quantum gates*, Physical Review Letters, August 2019, Volume 123, Issue 6.
  doi: 10.1103/PhysRevLett.123.060501
- J. Bouda, E. Onorati, J. Eisert, *Quantum encryption with weakly random sources*, in preparation.

Talks related to the thesis given by this author

- *Mixing properties of stochastic quantum Hamiltonians*, Theory of Quantum Computation, Communication and Cryptography, Paris, France, June 2017.
- *Randomized Benchmarking of small gates sets*, Deutsche Physikalische Gesellschaft, Mainz, Germany, March 2017.
- *Mixing properties of local diffusion processes on the unitary group*, Deutsche Physikalische Gesellschaft, Hannover, Germany, March 2016.
- *Random Hamiltonians, random circuits and unitary designs*, Deutsche Physikalische Gesellschaft, Heidelberg, Germany, March 2015.
- *Decoupling theorem and hypothesis-testing entropy*, QUTE Europe Summer School, Smolenice, Slovakia, August 2014.

Contents

1 Introduction **1**
1.1 Structure of this thesis . . . 5

2 Preliminaries **7**
2.1 Notation and fundamental structures in quantum information . . . 7
2.2 Probability theory . . . 13
2.3 Group and representation theory . . . 17

3 Measures over the unitary group and unitary designs **23**
3.1 The Haar measure and unitary designs . . . 24
3.2 Universality of Hadamard, CNOT and T-gate . . . 27
3.3 The Clifford group as a unitary 3-design . . . 29
3.4 Encryption from an imperfect source of randomness . . . 30

4 Decoupling **33**
4.1 Decoupling theorem and quantum entropy measures . . . 34
4.2 Decoupling linked to hypothesis-testing . . . 36

5 Random quantum circuits **41**
5.1 Approximate unitary 2-designs with random walk on Pauli basis . . . 42
5.2 Approximate unitary designs from Hamiltonian gaps . . . 50
5.3 Decoupling with random quantum circuits . . . 52

6 Randomized benchmarking **59**
6.1 The randomized benchmarking protocol with Clifford group . . . 60
6.2 Interleaved randomized benchmarking and other protocols . . . 65
6.3 Randomized benchmarking for individual quantum gates . . . 68
6.3.1 Example: Tensor copies of T-gate . . . 75

7 Brownian motion over the unitary group **80**
7.1 Brownian motion on the unitary group is a k-design . . . 85
7.1.1 Proof of Theorem 7.7 . . . 86
7.1.2 Local gap . . . 90
7.1.3 Hamiltonian driving . . . 94
7.1.4 More general interaction graphs . . . 95
7.1.5 Example: White noise in the Pauli basis . . . 96
7.2 Fast decoupling induced by Brownian motion on the unitary group . . . 96
7.2.1 Proof of Theorem 7.19 . . . 98
7.3 Fast scrambling and other applications . . . 108

8 Conclusions and outlook **112**

Acknowledgements and final words **114**

Bibliography **115**

Appendix: Mathematica notebook **123**

Chapter 1

Introduction

Quantum information is an interdisciplinary field laying at the intersection of quantum mechanics, mathematics, computer science and information theory: it sounds hence natural to review the contributions of these disciplines as an introduction to quantum information and to this thesis. Quantum mechanics represents one of the pinnacles of physics; at the doors of the twentieth century, Max Planck hypothesized that light is emitted in the form of discrete particles of energy. Albert Einstein in 1905, the *annus mirabilis*, developed further this intuition in order to explain the *photoelectric effect paradox* [1], unsolved according to the laws of classical physics, by interpreting light itself as made of fundamental particles, later referred to as *photons*. The discretized nature of light was formalized by the *Planck-Einstein-relation* describing the energy of these fundamental elements as the product of the *Planck constant* and the frequency of the electromagnetic wave, $E = h\nu$. In 1924 Louis de Broglie did the converse: he associated to a massive particle like the electron a wavelength λ according to a relation involving again the Plank constant together with the momentum p of the particle itself, $\lambda = h/p$. These formulae embody the wave-particle duality that characterizes this physical theory and allowed to explain Nature in contexts where the classical interpretation failed. Quantum mechanics acknowledges other illustrious fathers: Max Born, Wolfgang Pauli and Werner Heisenberg, known respectively for instance for the *Born rule* [2], the *Pauli exclusion principle* and the *Heisenberg uncertainty principle* [3], the latter stating that position and momentum of a quantum object cannot be measured simultaneously with arbitrary precision: this is a constraint set by Nature itself, not by experimental instruments! The most famous equation was provided by Erwin Schrödinger in 1926 [4], whose solutions are given by *wavefunctions* describing the allowed states of a quantum system under the influence of a certain *Hamiltonian* operator. The wavefunction is indeed one of the central objects of the theory: according to the so-called *Copenhagen interpretation*, formulated by Niels Bohr and Werner Heisenberg, it embeds every information regarding the physical system that one can retrieve through a measure. This action provides probabilistic results in the classical world, and makes the wavefunction *collapse* into the eigenfunction related to the observed quantity. However, before the measurement has been performed, the wavefunction usually describes a *superposition* of states, as explained by the the *Schrödinger cat* epitome [5]. Other interpretations of quantum mechanics have been provided, for instance the fascinating Hugh Everett's *many-world interpretation* [6], but none is as widely accepted as the one given by Bohr and Heisenberg. Quantum mechanics has been object of debate among its protagonists, Einstein himself has been one of the most active critics of the theory. Convinced that it did

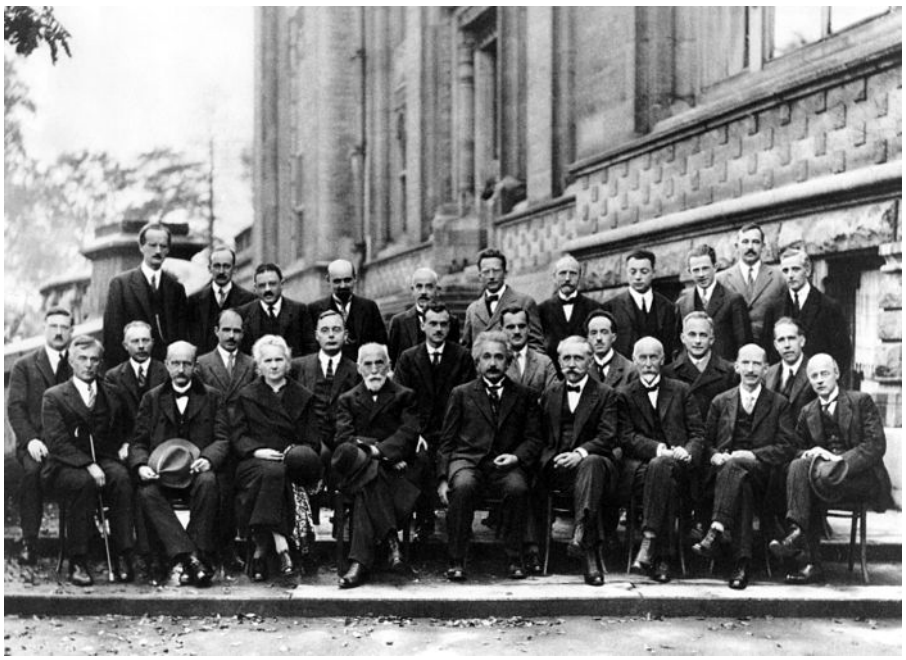


Figure 1.1: Participants of the fifth Solvay Conference, Brussels, 1927.
rear: A. Piccard, E. Henriot, P. Ehrenfest, E. Herzen, Th. de Donder, E. Schrödinger, J. E. Verschaffelt, W. Pauli, W. Heisenberg, R. H. Fowler, L. Brillouin \ middle: P. Debye, M. Knudsen, W.L. Bragg, H. A. Kramers, P. A. M. Dirac, A. H. Compton, L. de Broglie, M. Born, N. Bohr \ front: I. Langmuir, M. Planck, M. Curie, H. A. Lorentz, A. Einstein, P. Langevin, Ch. E. Guye, C. T. R. Wilson, O. W. Richardson. (picture taken from Solvay Conference, Wikipedia, by Benjamin Couprie, https://en.wikipedia.org/wiki/File:Solvay_conference_1927.jpg)

not provide a complete description of Nature, together with Boris Podolsky and Nathan Rosen he constructed a thought experiment in 1935, the *EPR paradox* [7], aimed at proving that the wavefunction does not contain all information of a physical system, hence contradicting Copenhagen interpretation. The three authors argued that by separating to an arbitrary distance two initially correlated electrons and accepting conservation law for total momentum, measuring the spin of one of these particles must *instantaneously* affect the spin of the other one. This phenomenon, now defined as *quantum entanglement*, would however imply faster-than-light communication (this implication is actually not true as discussed in later contributions) that would violate special relativity, and thus the paradox arises. To resolved it, Einstein, Podolsky and Rosen proposed the existence of *hidden variables*, values existing before being the outcome of a measurement, introducing the notion of *local realism*. John Stewart Bell used this approach to derive constrains – experimentally detectable – called *Bell's inequalities* [8] and argued that such a theory could not reproduce all of the predictions of quantum mechanics. A number of tests, such as the ones designed by Alain Aspect, showed that those inequalities are experimentally violated, hence demonstrating hidden variables formulation to be incorrect.
Modern computer science may be considered to germinate from the David Hilbert's attempt to provide a more consistent and complete formulation of mathematics, tackling foundation problems of this subject. In year 1900 he presented in Paris, at the Second International Congress, twenty-three problems and continued during the first part of the twentieth century with the so-called Hilbert's Program. In 1928 he posed three questions among others [9, 10]. Is mathematics *complete*, in the sense that every mathematical statement can be either proved or disproved? Is mathematics *consistent*, i.e., no contradiction can be proved in its formalism? Is mathematics

decidable, that is, does an algorithm exist so that it can be applied to any mathematical assertion and decide whether this is true or false (this dilemma was called *Entscheidungsproblem*)? Three years later Kurt Gödel answered the first two questions in a negative way: in his *incompleteness Theorems* [11], he showed that a system with elementary arithmetics rules is either incomplete or inconsistent; additionally, a systems of axioms cannot prove its own consistency. Gödel's work inspired the answer to the third Hilbert's question, given independently in the same year by Alonzo Church [12] and Alan Turing [13]. The latter formalized at a mathematical level the universal Turing machine and showed the existence of problems that such a machine could not solve. This abstract computer is conversely supposed to be able to perform any real-world calculation, up to resources limitation, according to the *Church-Turing Thesis*. In the 1940s started the race for building an electronic digital computer, whose descendants are today's computer machines; while interesting, we are not going to deepen this topic.
If Turing can be considered the father of modern computer science, Claude Shannon is the equivalent for information theory. With his famous work, *A Mathematical Theory of Communication* published in 1948 [14], Shannon provided a quantitative definition of Information, and explored communication through channels with two important coding theorems; the first one on noiseless channels measures the amount of physical resources required to store the output from an information source, the second one gives a bound on how much information is possible to transmit reliably through a noisy channel. Information and physics are indeed strongly tied together since the former can only exist and being transmitted in physical form. In viewpoint of Landauer's Principle, erasure of information will produce heat, and the amount thereof cannot be reduced under a precise lower bound.

Computer science approached the quantum world in the 1970s and the link has consolidated during the 1980s. In the one hand, according to the *Moore's law* formulated in 1965 [15] (suprisingly accurate over almost half a century), the number of transistors per integrated circuit doubles every year, and so the increasingly small size of circuit components will unavoidably run up against quantum effects. On the other hand, going back to the Church-Turing thesis, in 1985 David Deutsch asked whether there exists a universal computing device able to simulate every physical system. For this end, being quantum mechanics the fundamental theory of physics, it was necessary to look at designing a quantum computer. In addition to Deutsch, among the first contributors in this new field we cite Alexander Holevo, R. P. Poplavskii, Yuri Manin, Roman Ingarden, Steven Wiesner, Charles Bennet, Paul Benioff and Richard Feynman.
The most prominent question in quantum information that has puzzled the community from the beginning of this emerging domain is the following: can quantum computers beat their classical counterparts, i.e., can they solve problems that classical computers cannot and be faster in finding a solution to problems that also classical machines can handle? Answering this question – we refer to this goal as *quantum supremacy* [16] – has been unfortunately hard. One of the most encouraging results has been provided in 1994 by Peter Shor [17], who designed an algorithm for quantum computers able to factor rapidly large integer numbers, allowing for an efficient solution of both the factoring and the discrete logarithm problems, which are supposed to have no efficient solution on classical machines. A powerful implication of Shor's algorithm is the possibility to break easily a widespread security cryptosystem known as RSA [18]. The quest for quantum supremacy is still ongoing [19, 20] and in particular, at the time of writing this work, Google announced the release of a groundbreaking result, revealing a 49-qubit processor expected to perform calculations beyond the reach of the most powerful classical supercomputer of these days [21]!

Now leaving quantum information roots and perspectives behind, we shall get closer to the topic of this work and introduce the subject of its investigation. The unitary group $\mathbb{U}(N)$ is defined as the group of operators on $\mathbb{C}(N)$ such that $UU^\dagger = U^\dagger U = \mathbb{1},\ \forall U \in \mathbb{U}(N)$, and it is usually

identified with the group of $N \times N$ unitary matrices satisfying the same property. This group in of fundamental importance for quantum mechanics – and so of quantum information – since one of its four postulates says [22, 23]:

> *The evolution of a* closed *quantum system is described by a* unitary transformation. *That is, the state ψ of the system at time t_1 is related to the state ψ' of the system at time t_2 by a unitary operator U which depends only on times t_1 and t_2, i.e.,*
>
> $$|\psi'\rangle = U\,|\psi\rangle\,. \tag{1.1}$$

With *closed* system we refer to a system that does not interact in any way with another one and so can be treated as isolated; we will investigate this topic in Chapter 4.

Actually, any quantum map can be understood in terms of a unitary evolution on a larger system according to the following construction. For this, we will denote the quantum physical systems as *Hilbert spaces* and compose them with the tensor product. Please refer to Chapter 2 for details.

Theorem 1.1 (Stinespring dilation, cfr. [23])**.** *Let $\mathcal{E}$ be a CPTP map from a system A to a system B. Then there exists a complementary system R and an isometry, i.e., a metric-preserving operator $\widetilde{U}$, such that*

$$\mathcal{E} : \mathcal{H}_A \to \mathcal{H}_B \tag{1.2}$$

$$\rho_A \mapsto \mathcal{E}(\rho_A) = \mathrm{Tr}_R[\widetilde{U}^\dagger\,\rho_A\,\widetilde{U}]. \tag{1.3}$$

The isometry $\widetilde{U}$ is called Stinespring dilation.

With this, we can further embed $\widetilde{U}$ into a unitary operator U in $\mathcal{H}_A \otimes \mathcal{H}_R$, i.e., U is unitary and for some fixed $w_0 \in \mathcal{H}_R$ satisfies $U(v \otimes w_0) = \widetilde{U}v,\ \forall v \in \mathcal{H}_A$. It follows that we can write

$$\mathcal{E}(\rho_A) = \mathrm{Tr}_R[U^\dagger\,(\rho_A \otimes |w_0\rangle\,\langle w_0|)\,U] \tag{1.4}$$

and so interpret $\mathcal{E}$ as an evolution compatible to the postulate for an extended system $\mathcal{H}_A \otimes \mathcal{H}_R$.

It is therefore of great interest in the context of quantum information to investigate properties of the group ruling evolutions of quantum systems; besides its mathematical importance, the goal of this work is to demonstrate that continuous-time stochastic processes on this group lead to phenomena such as decoupling and scrambling, particularly relevant for applications. In this direction a different framework, namely, random quantum circuits, has already been studied [24–28]; those are settings where at each step of the circuit a random unitary gate is applied on a pair of qubits and indeed they are able to reproduce an approximate uniform distribution of unitary operators on the entire system, decouple a subsystem or scramble information. An extensive literature review will be provided in Chapter 5. On the other hand, a geometric formulation of quantum computing in terms of a Riemann manifold was developed in 2006 by Michael Nielsen and his colleagues [29–31], but does not seem to become established in the field of quantum information yet. Again, stochastic processes and Brownian motions over Lie groups have been rigorously defined [32–36] but without a quantum information perspective.

The principal goal of this work is to unify these different and apparently disconnected frameworks under a common and comprehensive mathematical formulation. More precisely, we are going to show that a Brownian motion over the unitary group defined through differential increments of its Lie algebra induces a distribution of unitaries with the same properties displayed by the one induced by random quantum circuits, that is, it produces unitary designs approximations and

decouples one system form another, with compatible scaling in the system size and in the degree of the moments. This implies a strong and fascinating link between continuous-time random processes driven by stochastic Hamiltonian evolutions and setting with random quantum gates, which beside its mathematical beauty can be of relevant interest for experimental implementations in presence of noisy Hamiltonians or in order to understand exotic phenomena such as black holes scrambling. Indeed, by expressing the scrambling condition in form of a decoupling theorem, in Section 7.3 we provide an analytic argument to support the conjecture proposed by Yasuhiro Sekino and Leonard Susskind [37], that is, black holes are the fastest scramblers in Nature, destroying information in logarithmic time with respect to the degrees of freedom.
To prove our results regarding approximate unitary designs, we will make use of mathematical tools from representation theory and probability. In order to estimate the speed of the local diffusion over the unitary group, we will work with a mixed tensor representation of the Lie algebra of the special unitary group in order to express the generator of the moments induced by Brownian motion in terms of a peculiar object of the Lie algebra, the Casimir element. Using an argument based on Young diagrams we will then identify the irreducible components of the latter, whose eigenvalues characterize the convergence of the diffusion towards the uniform distribution, or more precisely, towards its moments. To show fast decoupling in presence of Brownian motion, we will conversely make use of a random walk over Pauli basis weights already employed in refs. [24, 28], proving an almost linear time scaling in system size for the convergence to the uniform distribution of this Markov chain. This will involve intriguing tricks of probability theory, such as the Gambler's ruin and the shuffling of cards investigated by Persi Diaconis [38–40], now Professor of statistics and mathematics at Stanford and previously professional magician!

With his thesis this author hopes to provide a contribution at different planes. It aims to enhance the understanding of fundamental aspects of quantum information theory regarding one of its axioms and to unify different frameworks under a common formalism. It illustrates applications for randomness involving the unitary group in the task of benchmarking quantum gates, to allow quantum cryptography in the presence of imperfect sources of randomness and to strengthen the black holes scrambling conjecture. Moreover, it supplies an example of applications in quantum information of mathematical tools from probability and representation theory in addition to mathematical results such as demonstrating a spectral gap for local diffusion over the unitary group.

1.1 Structure of this thesis

The present work is based on the following projects developed during the PhD studies of this author: *Mixing properties of stochastic quantum Hamiltonians* [41], here included in **Sections** 7.1, 7.2 and 7.3, *Randomized benchmarking for individual quantum gates* [42], illustrated in **Section** 6.3, *Quantum encryption with weakly random sources*, here in **Section** 3.4. This author is the principal investigator of the first two projects and contributed extensively in all parts of both, conversely he contributed only partially to the third project, in particular for the reduction of the involved concentration bounds from the unitary group to the special unitary subgroup (not discussed in this work). Everything outside the mentioned sections is not novel material, but part of the literature of quantum information and related fields.

Now describing the structure of the thesis, **Chapter** 2 contains the introductory material needed in order to understand the successive pages and is divided in three parts. In the first section we define the basic structures of quantum mechanics in terms of the quantum information formalism, that is, Hilbert spaces, density operators, quantum channels, norms, and entropy quantities; note

that this does not constitute a complete summary of all structures of quantum mechanics but contains only those necessary for this work. In the second part we include a description of Markov chains, also in the continuous-time version, and list the concentration bounds used (sometimes implicitly) throughout the thesis. Finally we include a paragraph on Diaconis shuffling cards, a peculiar topic connected to decoupling results in relation to random quantum circuits and Brownian motion. In the third section we present in a self-contained way group and representation theory material, with a particular focus on irreducible representations and Schur's Lemma. The last paragraph is a brief recapitulation of the principal definitions of Lie groups and Lie algebras. In **Chapter** 3 we introduce the concept of distributions over the unitary group, with an emphasis on the Haar measure and other distributions approximating it, the so-called approximate unitary designs. We also define the notion of universality for a unitary distribution, with the particular example in Section 3.2 constituted by the set of three important unitary operators, namely, Hadamard, CNOT and T gates. In Section 3.3 we discuss the Clifford group as a unitary 2- and 3-design, and then conclude the chapter with a novel application of unitary designs in the context of encryption in presence of an imperfect source of randomness (cfr. Section 3.4). **Chapter** 4 is devoted to one of the most important topics of this work, that is, decoupling from the environment for a quantum system affected by a random unitary evolution and a subsequent quantum channel transformation. We present decoupling theorems in different fashions, that is, in terms of different entropy quantities such as the quantum collision entropy and the hypothesis-testing entropy, a measure for the probability of correctly distinguish two quantum states. In **Chapter** 5 we review the literature regarding random quantum circuits, a central construction for the development of this work. We will illustrate how subsequent implementations of local random gates lead to the notions investigated in the previous chapters, that is, approximate unitary designs and decoupling. A particular emphasis will be given to the fascinating mathematical tools used to obtain these results, ranging from convergence of random walks on Pauli strings (cfr. Sections 5.1 and 5.3) to spectral gap bounds for induced Hamiltonians (see Section 5.2). **Chapter** 6 presents one of the most relevant emerging techniques of quantum information aimed at gauging the experimental implementation of quantum gates. In Section 6.1 we illustrate how to derive an estimation of the average fidelity of Clifford gates through a polarization parameter obtained by taking the average over random samples of gate sequences for various lengths and fitting the according survival probability with respect to the sought parameter. Subsequently, we consider an interleaved randomized benchmarking protocol to extract an estimation for a singular target Clifford gate and use it to link randomized benchmarking to quantum process tomography (cfr. Section 6.2). We conclude the chapter with Section 6.3 containing a new protocol to benchmark individually tensor compositions of local gates outside the Clifford group by exploiting their local and permutations symmetries. As a relevant example, we apply this scheme to tensor copies of the T-gate (see Subsection 6.3.1). **Chapter** 7 encompasses the core of this work, namely, the study of mixing properties of Brownian motion over the unitary group. After a definition of this stochastic continuous-time process as an injection of differential increments over the Lie algebra into the group through the exponential map, in Section 7.1 we prove that the diffusion induces an approximate unitary design of arbitrary degree with linear dependence in the number of qudits and polynomial dependence with respect to the design degree. Section 7.2 includes the second main result – and the proof thereof – for Brownian motion over unitaries, that is, fast decoupling of an affected quantum systems with respect to the environment. The last part of the chapter, Section 7.3, illustrates two applications of the results, black holes scrambling and dissipative dynamics, and mentions other possible connected settings. We summarize the content of the thesis in the Conclusion **Chapter** 8 with some outlook on potential improvements and follow-up projects.

CHAPTER 2

Preliminaries

In this chapter we provide the tools required to understand the literature review and the novel results contained in this work. First, we have a short description of the formulation of quantum information theory and its principal objects, e.g., density operators and quantum channels, and state its fundamental principles. These notions are common knowledge in the field, and we will mainly follow refs. [22, 23].
Since all the mathematical techniques used to discuss previous results and prove the new ones rely on probability and group theory, the two subsequent sections will outline the necessary constructions of these two mathematical fields.

2.1 Notation and fundamental structures in quantum information

We identify the physical quantum system A by a *Hilbert space* $\mathcal{H}_A$; we only consider finite-dimensional systems and we denote the dimension of $\mathcal{H}_A$ by $|A|$. The set of homomorphisms (i.e., linear maps) between two systems A and B are denoted by $\mathrm{Hom}(\mathcal{H}_A, \mathcal{H}_B)$ and the set of endomorphisms, that is, homomorphisms from a Hilbert space onto itself, by $\mathrm{End}(\mathcal{H}_A) = \mathrm{Hom}(\mathcal{H}_A, \mathcal{H}_A)$. Density operators, defined in the next paragraph, are endomorphisms fully describing the state of a quantum system (first postulate of quantum information) and hence are one of the central objects of study in quantum information. The Hilbert-Schmidt (inner) product for $X, Y \in \mathrm{End}(\mathcal{H}_A)$ is given by $\langle X, Y\rangle \coloneqq \mathrm{Tr}[X^\dagger Y]$ with the induced metric $\|X\| = \sqrt{\langle X, X\rangle}$.

Density operators A *density operator* on $\mathcal{H}$ is a normalized semi-definite operator ρ, in other words, the set of all density operator is given by

$$\mathcal{S}(\mathcal{H}) \coloneqq \{\, \rho \in \mathrm{End}(\mathcal{H}) : \rho \geq 0 \text{ and } \mathrm{Tr}\,\rho = 1 \,\}\,; \tag{2.1}$$

one can also define *subnormalized* states as

$$\mathcal{S}_{\leq}(\mathcal{H}) \coloneqq \{\, \rho \in \mathrm{End}(\mathcal{H}) : \rho \geq 0 \text{ and } \mathrm{Tr}\,\rho \leq 1 \,\}\,. \tag{2.2}$$

A density operator ρ is *pure* if and olny if it can be writte as $\rho = |\psi\rangle\langle\psi|$ for some $\psi \in \mathcal{H}$. Conversely, it follows from the spectral decomposition theorem that any density operator can be

written in the form

$$\rho = \sum_x p_X(x) \left|e_x\right\rangle \left\langle e_x\right|, \tag{2.3}$$

where p_X is the probability mass function defined by the eigenvalues of ρ and $\{ e_x \}_x$ are the corresponding eigenvectors. This helps us to interpret any density operator as a statistical mixture of pure states with corresponding probability outcome $p_X(x)$. In particular, the state $\rho \in \mathcal{S}(\mathcal{H}_A)$ with the form $\rho = \frac{\mathbb{1}}{|A|}$ is called *fully mixed*.

Composition of subsystems For systems composed by n d-level subsystems (that we refer to as *qudits*) we construct the global d^n-dimensional system through the *tensor product* $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_n$ (fourth postulate of quantum information). In many cases, one considers 2-level quantum subsystems called *qubits*, usually denoted as a *superposition* of *computational basis states*, that is,

$$\left|\psi\right\rangle = \alpha \left|0\right\rangle + \beta \left|1\right\rangle, \tag{2.4}$$

with $\alpha, \beta \in \mathbb{C}$ so that $|\alpha|^2 + |\beta|^2 = 1$. As a basis for endomorphism space on qubit systems, the most common choice is given by *Pauli matrices*, together with the identity $\sigma_0 = \mathbb{1}_2$. These are given by

$$\sigma_1 = X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_2 = Y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \qquad \sigma_3 = Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{2.5}$$

which are Hermitian and traceless. By multipling them by the imaginary unit i, they become anti-Hermitian and constitute a basis for the Lie algebra $\mathfrak{su}(2)$, as we will discuss later. For n-qubit systems, we obtain a basis by composing n Pauli matrices through the tensor product, i.e.,

$$\mathcal{P}_n = \left\{ \bigotimes_{j=1}^{n} \sigma_{k_j} : k_j \in \{ 0, 1, 2, 3 \} \right\}, \tag{2.6}$$

and we will call the basis elements *Pauli strings* from now on. Density operators on a bipartite system $\mathcal{H}_{AB} \coloneqq \mathcal{H}_A \otimes \mathcal{H}_B$, are not only the ones given by the simple tensor product $\rho_A \otimes \rho_B$ between possible states of the two individual subsystems. Such states are said to be *independent* (or alternatively called *product states*): a measurement on the system A will give us no information about the state of system B and vice versa.
Still, there exist correlated states for which this is no longer the case, and hence measuring one of the two subsystems will provide us information about the other one. If a density ρ_{AB} can be decomposed as a convex combination of tensor product states, i.e.

$$\rho_{AB} = \sum_j p_j \, \rho_A^j \otimes \rho_B^j, \tag{2.7}$$

then it is said to be *separable*. Conversely, states that cannot be written in such a form contain an even "stronger" form of quantum correlation and are said to be *entangled*.

For a bipartite density operator ρ_{AB} we can derive the reduced operator ρ_A on $\mathcal{H}_A$ by taking its *partial trace* $\rho_A \coloneqq \mathrm{Tr}_B \, \rho_{AB}$. The reduced state fully characterizes all observable properties of the subsystem A since it is the (unique) operator that provides the correct statistics for the outcomes of measurements made on subsystem A. We note that the reduced state ρ_A of a pure bipartite state ρ_{AB} is not necessarily pure; conversely any mixed density operator can be seen as a reduced state of a pure state on a larger system. Such a construction, not unique in general, is called *purification*.
For a general $M_A \in \mathrm{End}(\mathcal{H}_A)$, we write $M_A \equiv M_A \otimes \mathbb{1}_B$ for the embedding on any $\mathcal{H}_{AB}$.

Quantum channels Another fundamental structure is given by *quantum channels*; these are maps that bring any density operator into another one, that is,

$$\begin{aligned} \mathcal{C} : \mathrm{End}(\mathcal{H}_A) &\to \mathrm{End}(\mathcal{H}_B) \\ \rho_A &\mapsto \rho_B, \end{aligned} \tag{2.8}$$

with the following properties.

[1] The map $\mathcal{C}$ must preserve the convex mixture of the set of density operators, namely, for any $p \in [0,1]$,

$$\mathcal{C}(p\rho_1 + (1-p)\rho_2) = p\,\mathcal{C}(\rho_1) + (1-p)\rho_2. \tag{2.9}$$

This is equivalent to the requirement for the channel $\mathcal{C}$ to be linear.

[2] The map $\mathcal{C}$ must be *completely positive*.
A map $\mathcal{C} : \mathrm{End}(\mathcal{H}_A) \to \mathrm{End}(\mathcal{H}_B)$ is said to be *positive* if $\mathcal{C}(M) \geq 0$ for all $M \geq 0$. Now, in order to ensure that a state ρ_{AR} of a larger system $\mathcal{H}_{AR}$ is again positive semi-definite when $\mathcal{C}$ is applied, we require that the map $\mathcal{C} \otimes \mathcal{I}_R$ is positive for any additional reference system R. In this case, the map $\mathcal{C}$ is said to be *completely positive*.

[3] The new state should be normalized, thus we require $\mathrm{Tr}[\mathcal{C}(M_A)] = \mathrm{Tr}[M_A]$ for all $M_A \in End(\mathcal{H}_A)$. A map fulfilling this condition is said to be *trace preserving*.

The evolution of a quantum system is hence described by linear, completely positive and trace preserving (CPTP) maps.

Representations of quantum channels It is often beneficial to represent quantum channels in other terms. Depending on the situation, one may want to use one of the following representations.
Any CPTP map $\mathcal{C}$ can be written in terms of *operator-sum representation*, that is, for any $M \in \mathrm{End}(\mathcal{H})$ there exists a set of (*Kraus*) operators $\{ K_j \}_j$ under the constraint $\sum_j K_j^\dagger K_j = \mathbb{1}$ such that

$$\mathcal{C}(M) = \sum_j K_j M K_j^\dagger. \tag{2.10}$$

For two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, the *Choi-Jamiolkowski map* $\mathcal{J}$ takes CP map $\mathcal{T}_{A\to B} \in \mathrm{Hom}(\mathrm{End}(\mathcal{H}_A), \mathrm{End}(\mathcal{H}_B))$ into positive semi-definite operators on $\mathcal{H}_{A'B}$ (and vice versa), where $\mathcal{H}_A \cong \mathcal{H}_{A'}$. This reduces the study of CP maps to the study of density operators.
Let $\{j_A\}_{j=1}^{|A|}$ be an orthonormal basis of $\mathcal{H}_A \cong \mathcal{H}_{A'}$, with the maximally entangled vector

$$|\Psi_{A'A}\rangle = \frac{1}{\sqrt{|A|}} \sum_j |j\rangle_{A'} \otimes |j\rangle_A. \tag{2.11}$$

The Choi-Jamiolkowski map $\mathcal{J}$ from $\mathrm{Hom}(\mathrm{End}(\mathcal{H}_A), \mathrm{End}(\mathcal{H}_B))$ to $\mathrm{End}(\mathcal{H}_{A'B})$ is then defined by

$$\mathcal{J}(\mathcal{T}_{A\to B}) \coloneqq (\mathcal{I}_{A'} \otimes \mathcal{T}_{A\to B})(|\Psi\rangle\langle\Psi|_{A'A}). \tag{2.12}$$

This mapping is an isomorphism, and its inverse for $\tau_{A'B} \in \mathrm{End}(\mathcal{H}_{A'B})$ is given by

$$\mathcal{J}^{-1}(\tau_{A'B}) : \; M_A \mapsto |A| \; \mathrm{Tr}_{A'}\left[(\mathcal{E}_{A\to A'}(M_A) \otimes \mathbb{1}_B)\tau_{A'B}\right], \tag{2.13}$$

where $\mathcal{E}_{A\to A'} : \mathrm{End}(\mathcal{H}_A) \to \mathrm{End}(\mathcal{H}_{A'})$ is defined as

$$\mathcal{E}_{A\to A'}(M_A) \coloneqq \sum_{i,j} |i\rangle_{A'} \langle j|_A \, M_A \, |i\rangle_A \langle j|_{A'}. \tag{2.14}$$

In this work, preservation of the eigenvalues structure is fundamental, and so the above representation is not suitable. We will instead make use of the following two constructions. We can identify $\mathrm{End}(\mathcal{H}_A) \simeq \mathbb{C}^{|A|^2}$, with $|j\rangle\langle k| \mapsto |j\rangle \otimes |k\rangle$. Any map $\mathcal{C}$ acting as $\mathcal{C}(M_A) = \sum_j X_j M_A Y_j$ can accordingly be written as $\mathcal{C} \simeq \sum_j X_j \otimes Y_j^T$ and we call such a formulation *vectorisation isomorphism*.
When working in **normalized** Pauli basis $\left\{ \tilde{\sigma}_k \coloneqq \frac{1}{\sqrt{2}^n} \bigotimes_{j=1}^n \sigma_{k_j} : k_j \in \{0,1,2,3\} \right\}$ (where we slightly abused the notation and re-arranged the vectors $(k_1, \ldots, k_n)$ as numbers from 1 to 4^n labeled by k), it can be useful to make use instead of the *Pauli-Liouville representation*, as we will do when discussing randomized benchmarking in Chapter 6. We can express any density operator ρ and quantum channel $\mathcal{C}$ as a linear combination of projections onto the Pauli basis, i.e.

$$\rho = \sum_{k=1}^{4^n} \rho_k \tilde{\sigma}_k \qquad \text{and} \qquad \mathcal{C}(\rho) = \sum_{k=1}^{4^n} \mathcal{C}(\tilde{\sigma}_k)\rho_k, \tag{2.15}$$

where $\rho_k \coloneqq \langle \tilde{\sigma}_k, \rho \rangle$ and so we can represent them as

$$|\rho\rangle = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \cdots \\ \rho_{4^n} \end{pmatrix} \qquad \text{and} \qquad \mathcal{C}_{k\ell} = \langle \tilde{\sigma}_k, \mathcal{C}(\tilde{\sigma}_\ell) \rangle. \tag{2.16}$$

In this way, we may represent $\mathcal{C}(\rho)$ as a matrix-vector multiplication $\mathcal{C}\,|\rho\rangle$ and the concatenation of two channels $\mathcal{D}$ and $\mathcal{C}$ as a matrix multiplication $\mathcal{D}\mathcal{C}$. We can analogously represent a map $M \in \mathrm{End}(\mathcal{H}_A)$ in the form

$$\langle M| = (M_1 \; M_2 \ldots M_{4^n}) \qquad \text{with} \qquad M_k = \langle M, \tilde{\sigma}_k \rangle. \tag{2.17}$$

When M is a measurement, the probability to obtain an outcome described by M when measuring ρ is $\langle M, \rho \rangle$ (cfr. third postulate of quantum information).

Norms and distances between states In order to quantify the distance (or closeness) between two quantum states, one can make use of the metric on $\mathrm{End}(\mathcal{H})$ (often defined with an additional factor $\frac{1}{2}$) induced by the 1-norm, given by

$$\|M\|_1 \coloneqq \mathrm{Tr}\sqrt{M^\dagger M} = \sum_j s_j(M), \tag{2.18}$$

where we denote by $s_j(M)$ the j-th singular value of M. The so-called *trace distance* is then defined as

$$\delta(\rho, \sigma) = \frac{1}{2} \|\rho - \sigma\|_1 . \tag{2.19}$$

This distance characterizes the maximum probability to successfully distinguish two states.
Another commonly used measure for the closeness between two density operators $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ is the *fidelity*

$$F(\rho, \sigma) \coloneqq \left\| \sqrt{\rho}\sqrt{\sigma} \right\|_1 . \tag{2.20}$$

Note that F is always between 0 and 1, and that $F(\rho, \rho) = 1$. For pure states $\psi = |\psi\rangle\langle\psi|$ and $\phi = |\phi\rangle\langle\phi|$, the fidelity takes the simple form

$$F(\psi, \phi) = |\langle \psi, \phi \rangle| . \tag{2.21}$$

A number of properties can be mentioned for the fidelity distance: Uhlmann's Theorem, monotonicity, contractivity and (strong) concavity, but none of these is relevant to this work so we are

not going to illustrate them.
The trace distance and the fidelity are related to each other; in particular, for pure states it stands

$$F(\psi,\phi)^2 + \delta(\psi,\phi)^2 = 1, \tag{2.22}$$

while the relation for general density operators reads

$$1 - F(\rho,\sigma) \leq \delta(\rho,\sigma) \leq \sqrt{1 - F(\rho,\sigma)^2}. \tag{2.23}$$

To extend the argument to quantum channels, we define the *channel fidelity* for two CPTP maps $\mathcal{C}_1, \mathcal{C}_2$ on a given density operator ρ simply by using the previous definition, i.e.

$$\mathcal{F}_{\mathcal{C}_1,\mathcal{C}_2}(\rho) \coloneqq F(\mathcal{C}_1(\rho), \mathcal{C}_2(\rho)) = \left\| \sqrt{\mathcal{C}_1(\rho)} \sqrt{\mathcal{C}_2(\rho)} \right\|_1 . \tag{2.24}$$

This quantity, for a quantum channel $\mathcal{E}$ and a unitary operation $\mathcal{U}$ and taking a pure state $\phi = |\phi\rangle\langle\phi|$ as input, is called *gate fidelity* and can be simplified to

$$\mathcal{F}_{\mathcal{E},\mathcal{U}}(\phi) \coloneqq \mathrm{Tr}\left[\mathcal{U}(\phi)\,\mathcal{E}(\phi)\right] \tag{2.25}$$

and defining $\Lambda = \mathcal{U}^\dagger \circ \mathcal{E}$ one has

$$\mathcal{F}_{\mathcal{E},\mathcal{U}}(\phi) = \mathcal{F}_{\Lambda,\mathcal{I}}(\phi) = \mathrm{Tr}\left[\phi\,\Lambda(\phi)\right], \tag{2.26}$$

that can be interpreted as quantifying the noise channel Λ for an implementation $\mathcal{E}$ of $\mathcal{U}$.
The *average gate fidelity* is then obtained by integrating this quantity over the Haar measure on pure states, that is,

$$\mathbb{E}(\mathcal{F}_{\mathcal{E},\mathcal{U}}) = \mathbb{E}(\mathcal{F}_{\Lambda,\mathcal{I}}) \coloneqq \int_{\mathrm{Haar}} \mathrm{Tr}\left[\phi\,\Lambda(\phi)\right]\,\mathrm{d}\phi. \tag{2.27}$$

Conversely, the *entanglement fidelity* of a quantum channel $\mathcal{E}$, defined as

$$F_{\mathrm{ent}}(\mathcal{E}) \coloneqq \langle\Psi|\,(\mathcal{I}\otimes\ \mathcal{E})(\Psi)\,|\Psi\rangle\,, \tag{2.28}$$

with $|\Psi\rangle$ being a maximally entangled state vector, can be written as [43]

$$F_{\mathrm{ent}}(\mathcal{E}) = d^{-3}\sum_j \mathrm{Tr}[V_j^\dagger \mathcal{E}(V_j)], \tag{2.29}$$

for any orthonormal basis $\{\,V_j\,\}_j$ such that $\mathrm{Tr}[V_j V_k] = d\,\delta_{j,k}$ (in the case of n qubits, $d = 2^n$). Hence, it is simple to obtain when the trace of the quantum channel is known. The average gate fidelity of $\mathcal{E}$ is then linked to this quantity by [43]

$$\mathbb{E}(\mathcal{F}_{\mathcal{E},\mathcal{I}}) = \frac{dF_{\mathrm{ent}}(\mathcal{E}) + 1}{d+1} = \frac{\sum_j \mathrm{Tr}[V_j^\dagger \mathcal{E}(V_j)] + d^2}{d^2(d+1)}. \tag{2.30}$$

Additionally, the *diamond norm* [44] of a CPTP map $\mathcal{T}$ is defined to be

$$\|\mathcal{T}\|_\diamond \coloneqq \sup_d \sup_{X\neq 0} \frac{\|(\mathcal{T}\otimes\mathcal{I}_d)\,X\|_1}{\|X\|_1}, \tag{2.31}$$

and it is the most common choice to quantify the magnitude of a quantum channel from a physical perspective, since it considers the map as affecting only a part of an arbitrarily larger system. The *infinity norm* $\|\mathcal{T}\|_\infty$ is given by the largest singular value of $\mathcal{T}$ and instead used in mathematical and technical contexts; in our case, for instance, it quantifies the gap ruling

convergence and mixing times for moment operators.

The average gate fidelity can also be linked to the diamond norm, for instance with the following lower bound (which is actually true for any gate fidelity on arbitrary density operators) [45]

$$\mathbb{E}(\mathcal{F}_{\mathcal{E}_1,\mathcal{E}_2}) \geq 1 - \|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond, \tag{2.32}$$

and additionally [46, Theorem 9]

$$r(d+1)/d \leq \frac{1}{2}\|\Lambda - \mathcal{I}\|_\diamond \leq \sqrt{d(d+1)r}, \tag{2.33}$$

where $r := 1 - \mathbb{E}(\mathcal{F}_{\Lambda,\mathcal{I}})$ is the *average error rate for* Λ.

Entropy quantities Entropy quantities are essential in both classical and quantum information theory because they measure the uncertainty (or the knowledge) of an observer regarding a certain (quantum) system. This subfield has a deep and extensive literature and is continuously expanding, so we will restrict ourself to the basic definitions. Entropy measures are used to characterize processes or formulations such as decoupling theorems, as we shall see in the related Chapter 4.

The *von Neumann entropy* is the quantum mechanical generalization of the classical Shannon entropy and it is defined as

$$H(\rho) := -\operatorname{Tr}(\rho \log \rho). \tag{2.34}$$

The next step is to introduce the *quantum relative entropy of* $\rho \in \mathcal{S}_\leq(\mathcal{H})$ *with respect to* $\sigma \geq 0$, given by

$$D(\rho\|\sigma) := \operatorname{Tr}(\rho \log \rho) - \operatorname{Tr}(\rho \log \sigma) \tag{2.35}$$

if $supp(\rho) \subseteq supp(\sigma)$ and $+\infty$ otherwise.
We can then use this quantity to define *the conditional von Neumann entropy for* $\rho_{AB} \in \mathcal{S}_\leq(\mathcal{H}_{AB})$ *with respect to* $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$ as

$$H(A|B)_{\rho|\sigma} := -D(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B); \tag{2.36}$$

from this we can also define

$$H(A|B)_\rho := \sup_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} H(A|B)_{\rho|\sigma}. \tag{2.37}$$

We can now introduce min- and max-entropy. More precisely, we first define the *quantum relative max-entropy for* $\rho \in \mathcal{S}_\leq(\mathcal{H})$ *with respect to* $\sigma \geq 0$ by

$$D_{\max}(\rho\|\sigma) := \inf\left\{ \lambda \in \mathbb{R} : 2^\lambda \sigma \geq \rho \right\}. \tag{2.38}$$

The *conditional min-entropy* H_{min} *of* $\rho_{AB} \in \mathcal{S}_\leq(\mathcal{H}_{AB})$ *with respect to* $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$ is given by

$$H_{\min}(A|B)_{\rho|\sigma} := -D_{\max}(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) = \sup\left\{ \lambda : \ 2^{-\lambda}\mathbb{1}_A \otimes \sigma_B \geq \rho_{AB} \right\} \tag{2.39}$$

and

$$H_{\min}(A|B)_\rho := \sup_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} H_{\min}(A|B)_{\rho|\sigma}. \tag{2.40}$$

In particular, if $\mathcal{H}_B$ is the trivial space, it follows that

$$H_{\min}(A)_\rho = -\log \|\rho_A\|_\infty. \tag{2.41}$$

In analogous way we define the max-entropy; the *quantum relative min-entropy for* $\rho \in \mathcal{S}_{\leq}(\mathcal{H})$ *with respect to* $\sigma \geq 0$ is defined by

$$D_{\min}(\rho\|\sigma) := -\log(F(\rho,\sigma)^2). \tag{2.42}$$

The *conditional max-entropy* $H_{\max}$ *of* $\rho_{AB} \in \mathcal{S}_{\leq}(\mathcal{H}_{AB})$ *with respect to* $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$ is then

$$H_{\max}(A|B)_{\rho|\sigma} := -D_{\min}(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) = \log(F(\rho_{AB}, \mathbb{1}_A \otimes \sigma_B)^2) \tag{2.43}$$

and also

$$H_{\max}(A|B)_{\rho} := \sup_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} H_{\max}(A|B)_{\rho|\sigma}; \tag{2.44}$$

if $\mathcal{H}_B$ is the trivial space, then

$$H_{\max}(A)_{\rho} = 2\log(\mathrm{Tr}\sqrt{\rho_A}). \tag{2.45}$$

The non-smooth version of decoupling theorem relies on the *quantum conditional collision entropy*. Let $\rho_{AB} \in \mathcal{S}_{\leq}(\mathcal{H}_{AB})$ and $\sigma_B \in \mathcal{H}_B, \sigma_B \geq 0$, then the quantum conditional collision entropy of A conditioned on B of the state ρ_{AB} relative to σ_B is defined as

$$H_2(A|B)_{\rho|\sigma} := -\log \mathrm{Tr}\left[\left\{(\mathbb{1}_A \otimes \sigma_B)^{-1/4}\rho_{AB}(\mathbb{1}_A \otimes \sigma_B)^{-1/4}\right\}^2\right]. \tag{2.46}$$

Taking the supremum over all normalized σ_B, we write

$$H_2(A|B)_{\rho} := \sup_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} H_2(A|B)_{\rho|\sigma}. \tag{2.47}$$

The *smooth* entropy versions are defined by taking the extremum over a set of nearby states, where the notion of "nearby" is expressed in terms of the *purified distance* (cfr. ref. [23]). For $\varepsilon \geq 0$, the *smooth conditional max- and min-entropy of* A *given* B for $\rho_{AB} \in \mathcal{S}_{\leq}(\mathcal{H}_{AB})$ are defined as

$$H^{\varepsilon}_{\max}(A|B)_{\rho} = \inf_{\bar{\rho}_{AB} \in \mathcal{B}^{\varepsilon}(\rho_{AB})} H_{\max}(A|B)_{\bar{\rho}} \tag{2.48}$$

and

$$H^{\epsilon}_{\min}(A|B)_{\rho} = \sup_{\bar{\rho}_{AB} \in \mathcal{B}^{\varepsilon}(\rho_{AB})} H_{\min}(A|B)_{\bar{\rho}}. \tag{2.49}$$

This latter quantities are used to characterize the upper bound of decoupling Theorem in eq. (4.18) at the end of Section 4.1.
Another entropy measure called *hypothesis-testing entropy* will be introduced later in Section 4.2.

2.2 Probability theory

Brownian motion, one of the core constructions in this work, is arguably the most important and widely studied random process in probability theory. Looking at diffusion over the unitary group requires a set of stochastic definitions and results that we are going to briefly explain in the following.

Markov chains A (discrete) stochastic process with a sequence of random variables $X_1, X_2, \dots$ whose next step depends solely on the current state is called a *Markov chain*. We consider a finite set of values $\Lambda = \{\lambda_1, \lambda_2, \dots\}$ which the variables X_n can assume during the process and denote it as *state space*. For the variable X_n we can then assign a *probability distribution* $\omega_n = (\omega_n^1, \omega_n^2, \dots)$ where $\omega_n^k = \mathbb{P}(X_n = \lambda_k)$. If the state space is finite, the transition from X_n to X_{n+1} can be described by a *transition matrix* P_n with entries

$$p_{k,\ell} = \mathbb{P}(X_{n+1} = \lambda_\ell | X_n = \lambda_k) = \mathbb{P}(X_{n+1} = \ell | X_n = k) \tag{2.50}$$

such that we have

$$\omega_{n+1} = \omega_n \, P_n. \tag{2.51}$$

If the process is *time-homogeneous*, then each transition is governed by the same transition matrix P at each step, and thus

$$\omega_n = \omega_0 \, P^n. \tag{2.52}$$

The *stationary distribution* of the process ω satisfies

$$\omega = \omega \, P \tag{2.53}$$

and can hence be regarded as a fixed point of the chain. If, for all $\lambda_k, \lambda_\ell \in \Lambda$, the following relation (called *detailed balance condition*) stands,

$$\omega^k p_{k,\ell} = \omega^\ell p_{\ell,k}, \tag{2.54}$$

the chain is *reversible*. A reversible Markov chain on a graph is called *random walk*.

A chain is said to be *ergodic* if it is *irreducible* and *aperiodic*. Irreducibility means, in rough words, that any state of the chain can be reached in finite time, regardless of the starting value, so that it exists $m < \infty$ such that for all λ_k and λ_ℓ

$$\mathbb{P}\left(X_{n+m} = \ell | X_n = k\right) > 0. \tag{2.55}$$

More formally, we say that there exists only one *communicating class*. For a state λ_k of the chain, the greatest common divisor of the set of all times m when a return to the state λ_k is possible is called *period* of λ_k. A state is said to be aperiodic if its period is 1, that is, for all n

$$\gcd\left\{ m : \mathbb{P}\left(X_{n+m} = k | X_n = k\right) \right\} = 1, \tag{2.56}$$

and if every state is aperiodic, so is said to be the chain. Note that all states of an irreducible chain have the same period and hence by proving aperiodicity for a single state one can infer this property to the full state space.
A fundamental result in this regard is that every ergodic chain has a unique stationary distribution toward which it will eventually converge. We refer as the *mixing time* of the chain to the number of steps required to reach closeness to the stationary distribution. For two arbitrary distributions ω and η, the *total variation distance* is given by

$$\|\omega - \eta\|_{TV} = \frac{1}{2} \|\omega - \eta\|_1 = \frac{1}{2} \sum_j |(\omega)_j - (\eta)_j| \,. \tag{2.57}$$

Then the *mixing time* is defined as

$$\tau(\varepsilon) \coloneqq \max_{\omega_0} \min_{t \geq 0} \left\{ t \ : \left\| \omega_0 P^t - \omega \right\|_{TV} \leq \varepsilon \right\} \ , \tag{2.58}$$

where ω_0 is the initial probability distribution and ω the stationary distribution.

In this work, we consider Markov chain on Pauli strings (that we also address to as *random walk on Pauli basis*) induced by Brownian motion. Since the latter is a continuous-time stochatic process, the random walk will also be a continuous-time, described by *jumps* from a Pauli string to another one, spaced out by *waiting times* where the process does not move. Specifically, the function $\{ N(t) \ : \ t \geq 0 \}$ counting the number of jumps occurred up to the positive time t defines a *Poisson process*, given that the following properties are satisfied:

[1] $N(0) = 0$,

[2] the increments are independent and stationary,

[3] each increment $N(t+\Delta t) - N(t)$ is distributed as a Poisson random variable with parameter (mean) $\mu \Delta t$.

The last condition implies that $\mathbb{E}[N(t)] = \mu t$ and, in particular, the probability that two or more jumps occur in the time interval is negligible when this is small.

The waiting time W between two consecutive jumps is then described by an *exponential distribution*, having for $\mu > 0$ a *cumulative distribution function*

$$\mathbb{P}(W \leq t) = 1 - \mathrm{e}^{-\mu t} \tag{2.59}$$

and a *probability density function*

$$f(t) = \mu \, \mathrm{e}^{-\mu t}. \tag{2.60}$$

In order to study the mixing time of this chain, we will project the random walk on Pauli basis onto a one-dimensional *simple* random walk on the discrete state space $\Lambda = \{ 1, 2, \ldots, n \}$ with jumps restricted to adjacent lattice points. This induced chain counts the number of single-qubit identity elements (*weights*) of the Pauli strings that the original chain is visiting.
We will also put conditions on reaching certain points of the state space or reverting back to already visited values, establishing that the walk has then ended. In this case we refer to random walks with *absorbing barrier*. We can introduce the argument by mentioning here a famous application for this concept: the *Gambler's ruin*. Let us consider a player starting with an initial amount of money being x Euro, winning at each bet 1 Euro with probability p and losing 1 Euro with probability $q = 1 - p$. The player wins if he reaches a total amount of N Euro, and goes bankrupt if he loses all of his money. Let us denote the probability of winning starting with x Euro by $\mathbb{P}_x$, and so the probability of going bankrupt will be $1 - \mathbb{P}_x$. Then we have

$$\mathbb{P}_x = \begin{cases} \frac{1-(q/p)^x}{1-(q/p)^N} & \text{if } p \neq q \\ \frac{x}{N} & \text{if } p = q = 1/2. \end{cases} \tag{2.61}$$

In particular, if the gambler is allowed to play forever until ruined, if $p > 1/2$ there is a positive probability that he will never get ruined, otherwise, if $p \leq 1/2$, ruin is unavoidable. Mathematically,

$$\lim_{N \to \infty} \mathbb{P}_x = \begin{cases} 1 - (q/p)^x & \text{if } p > 1/2 \\ 0 & \text{if } p \leq 1/2. \end{cases} \tag{2.62}$$

In this work we will apply the following result for a random walk with absorbing barrier.

Lemma 2.1 (Lemma A.6 in ref. [28])**.** *Consider an asymmetric simple random walk with probability of moving forwards p and probability of going backwards $q = 1 - p$ that starts at $a > 0$ and has an absorbing barrier at the origin. The probability that the walk eventually absorbs at the origin is 1 if $p \leq q$ and $(q/p)^a$ otherwise.*

Hoeffding's inequalities Hoeffding's inequalities are an additive form of *Chernoff bounds*, which are concentration inequalities on the tail probability of the sum of independent (but not necessarily identically distributed) random variables. Hoeffding's inequalities are exponentially decreasing and are formulated as follows [47, Theorem 1]. Let us first assume $X_1, \ldots, X_n$ to be i.i.d. Bernoulli random variables for all j and denote $\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$. Then for all $\varepsilon > 0$,

$$\mathbb{P}\left(\overline{X} - \mathbb{E}\left[\overline{X}\right] \geq \varepsilon\right) \leq \left\{\left(\frac{\mathbb{E}\left[\overline{X}\right]}{\mathbb{E}\left[\overline{X}\right] + \varepsilon}\right)^{\mathbb{E}\left[\overline{X}\right]+\varepsilon} \left(\frac{1 - \mathbb{E}\left[\overline{X}\right]}{1 - \mathbb{E}\left[\overline{X}\right] - \varepsilon}\right)^{1-\mathbb{E}\left[\overline{X}\right]-\varepsilon}\right\}^{n} \tag{2.63}$$

$$\leq \exp\{-2\varepsilon^2 n\} \tag{2.64}$$

and

$$\mathbb{P}\left(\overline{X} - \mathbb{E}\left[\overline{X}\right] \leq -\varepsilon\right) \leq \left\{\left(\frac{\mathbb{E}\left[\overline{X}\right]}{\mathbb{E}\left[\overline{X}\right] - \varepsilon}\right)^{\mathbb{E}\left[\overline{X}\right]-\varepsilon} \left(\frac{1 - \mathbb{E}\left[\overline{X}\right]}{1 - \mathbb{E}\left[\overline{X}\right] + \varepsilon}\right)^{1-\mathbb{E}\left[\overline{X}\right]+\varepsilon}\right\}^{n} \tag{2.65}$$

$$\leq \exp\{-2\varepsilon^2 n\}; \tag{2.66}$$

this can be combined to

$$\mathbb{P}\left(|\overline{X} - \mathbb{E}\left[\overline{X}\right]| \geq \varepsilon\right) \leq 2\exp\{-2\varepsilon^2 n\}. \tag{2.67}$$

We can extend the bound for independent random variables $Y_1, \ldots, Y_n$ with $a_j \leq Y_j \leq b_j$, namely [47, Theorem 2],

$$\mathbb{P}\left(\overline{Y} - \mathbb{E}\left[\overline{Y}\right] \geq \varepsilon\right) \leq \exp\left\{-2\varepsilon^2 n^2 / \sum_{j=1}^{n}(b_j - a_j)^2\right\} \tag{2.68}$$

and equivalently

$$\mathbb{P}\left(\overline{Y} - \mathbb{E}\left[\overline{Y}\right] \leq -\varepsilon\right) \leq \exp\left\{-2\varepsilon^2 n^2 / \sum_{j=1}^{n}(b_j - a_j)^2\right\}, \tag{2.69}$$

such that again we can put them together as

$$\mathbb{P}\left(|\overline{Y} - \mathbb{E}\left[\overline{Y}\right]| \geq \varepsilon\right) \leq 2\exp\left\{-2\varepsilon^2 n^2 / \sum_{j=1}^{n}(b_j - a_j)^2\right\}. \tag{2.70}$$

Specifically, through this work we will make use of the following Chernoff-Hoeffding's inequality.

Lemma 2.2 (Lemma A.4 in ref. [28])**.** *Let X_k be the random variable giving the position of a random walk after k steps starting at the origin with probability $p_j \geq p$ of moving right and probability $q_j \leq p$ of moving left at step j and let $\mu = p - (1-p) = 2p - 1$. Then for any $\eta > 0$*

$$\mathbb{P}\left(X_k \geq \mu k + \eta\right) \leq \exp\left\{-\frac{\eta^2}{2k}\right\} \qquad \textit{and} \qquad \mathbb{P}\left(X_k \geq \mu k - \eta\right) \leq \exp\left\{-\frac{\eta^2}{2k}\right\}. \tag{2.71}$$

Diaconis shuffling of cards Convergence of Markov chains is an important object of study, where for instance *log-Sobolev constants* play a major role [48]. A very useful instance is given by the random walk on the symmetric group induced by random transpositions, with very curious applications on shuffling a deck of cards [39]. More formally, we consider the symmetric group S_n as the state space and write transition matrix given by random transpositions as

$$\mathcal{T}(\pi,\rho) = \begin{cases} 1/n & \text{if } \pi\rho^{-1} = \mathcal{I} \\ 2/n^2 & \text{if } \pi\rho^{-1} \text{ is a transposition} \\ 0 & \text{otherwise.} \end{cases} \tag{2.72}$$

We then ask how many steps of the chain are required in order to get arbitrarily close to the uniform stationary distribution $\omega(\pi) = 1/n!$ (in the language of cards, this means how many times we should swap two cards at random until the deck is completely shuffled). Diaconis and Shahshahani provided an answer in ref. [38, Theorem 1]; for $n \geq 10$ the mixing time of $\mathcal{T}$ to a given precision ε is $\tau(\varepsilon) = O(n \log \frac{n}{\varepsilon})$.
As an interesting remark, this convergence presents a cutoff phenomenon [40]. Taking for example the particular case of Gilbert-Shannon-Reeds shuffling, in Theorem 1 of this work of Diaconis is proved that with $k = \frac{3}{2}\log_2 n + \theta$ steps the distance to the uniform distribution is bounded for any initial distribution ω_0 by

$$\|\omega_0 P_{\text{GSR}}^k - \omega\|_{TV} = 1 - 2\Phi\left(-2^{-\theta}/4\sqrt{3}\right) + O(1/\sqrt{n}), \tag{2.73}$$

where $\Phi(z) := \int_{-\infty}^{z} \mathrm{e}^{-t^2/2}/\sqrt{2\pi}\,\mathrm{d}t$. The cutoff behavior around $\theta = 0$ is illustrated in Figure 2.1. This phenomenon can be observed in connection to the (high) degeneracy of the second-highest eigenvalue of these chains.
Another relevant setting is the uniform sample with replacement from an urn with n balls, or equivalently, the collection of n different coupons (indeed, this problem is called *coupon collector's argument*), one of them received uniformly at random after each time one buys the magazine. Let ℓ be the number of draws required until each ball has been extracted (or coupon obtained) at least once. Then [39, Lemma 2] for $c \geq 0$ and $n \geq 1$

$$\mathbb{P}\left(\ell > n\log n + cn\right) \leq \mathrm{e}^{-c}. \tag{2.74}$$

2.3 Group and representation theory

In order to study the diffusion of a continuous-time stochastic process over the unitary set of matrices and to exploit operators symmetries for the novel randomized benchmarking protocol that we are going to present in Chapter 6, it is convenient to switch to the representation theory framework and make use of the powerful results that we are going to recapitulate in the following paragraphs.

Definition 2.3 (Group)**.** *A* group G *is a set of elements equipped with a binary operation satisfying the following properties:*
Closure*: For all* $g, h \in G,\ g \cdot h \in G$.
Associativity*: For all* $g, h, k \in G,\ (g \cdot h) \cdot k = g \cdot (h \cdot k)$.
Identity element*: There exist a unique identity element,* e*, such that for all* $G \in G\ e \cdot g = g \cdot e = g$.
Inverse*: for every element* $g \in G$ *there exist an inverse element* g^{-1} *such that* $g^{-1} \cdot g = g \cdot g^{-1} = e$.

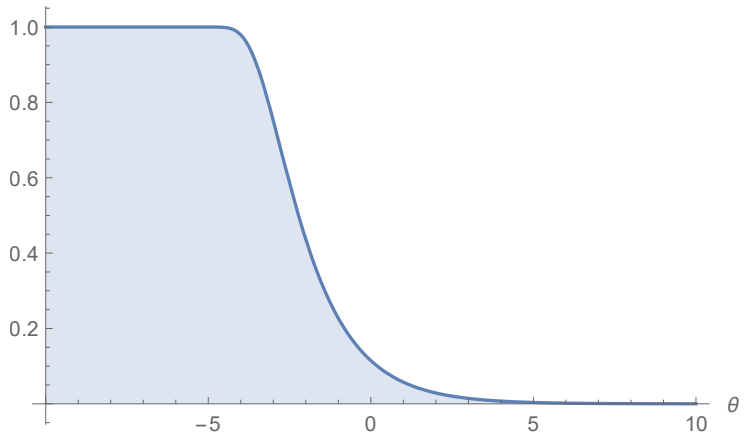


Figure 2.1: Cutoff around θ

.

If two group can be linked by a group isomorphism, they are said to be *isomorphic*. They will then have many properties in common, in particular the same multiplication table or character table (which we will introduce later). One can therefore easily obtain information about a group if it is possible to find a group isomorphism connecting it to another well-known group; this is indeed what we do in our protocol to deal with the local symmetry groups. A subset $H \subset G$ is called *subgroup* if all above conditions are still satisfied, e.g., the subset is closed with respect to the group operation. A subgroup N such that $g^{-1}n\,g \in N$ for all $n \in N,\ g \in G$ is said to be *normal* and this is denoted by $N \lhd G$. The set of all elements of G commuting with any element of the group,

$$Z(G) \coloneqq \{\, z \in G : g^{-1}zg = z \text{ for all } g \in G \,\}\,, \tag{2.75}$$

is a normal subgroup and is called *center of G*.

One can define a *(left) group action of G on a set M* by a function

$$\begin{aligned} \phi : \quad & G \times M \to M, \\ & (g, m) \mapsto \phi(g, m) \end{aligned} \tag{2.76}$$

that fulfills the following two axioms:
Identity: for all $m \in M,\ \phi(e, m) = m$,
Compatibility: for all $m \in M, g, h \in G\ \phi(g, \phi(h, m)) = \phi(g \cdot h, m)$.
With this definition, we can furthermore establish the following.

Definition 2.4 (Orbit)**.** *An* orbit *$G.m$ of an element $m \in M$ is given by all elements in M obtained by the action of G, i.e.,*

$$G.m \coloneqq \{\, \phi(g, m) : g \in G \,\}\,. \tag{2.77}$$

The action of G on M induces a *partition* of the set M itself, i.e., it regroups the elements into subsets such that every element $m \in M$ is contained in one and only one of those.

Definition 2.5 (Stabilizer subgroup)**.** *The* stabilizer subgroup *of G with respect to m is the set of all elements on G such that*

$$G_m \coloneqq \{\, g \in G : \phi(g, m) = m \,\}\,. \tag{2.78}$$

It is always possible to couple two groups to generate a new one. This is indeed what we looking for, having to combine symmetry of the local gates together with the invariance with respect to permutations thereof.

Definition 2.6 (Direct product)**.** *Given two groups G and H, the* direct product $G \times H$ *is a* Cartesian product *of ordered pairs (g, h), with $g \in G, h \in G$ equipped with a binary operation acting component-wise, that is,*

$$(g_1, h_1) \cdot (g_2, h_2) = (g_1 \cdot g_2, h_1 \cdot h_2). \tag{2.79}$$

This new structure satisfies all axioms of closure, associativity, existence of identity – given by (e_G, e_H) – and inverse element – (g^{-1}, h^{-1}) being the inverse of (g, h) – and so it is a group. An alternative way to construct a new group from is given by the *semi-direct product*.

Definition 2.7 (Outer semi-direct product)**.** *Let N, H be groups, $\varphi : H \to \mathrm{Aut}(N)$ be a homomorphism from H to the set of automorphisms of N. Then the* (outer) semi-direct product with respect to ϕ*, denoted by $G = N \rtimes_\varphi H$, is the group whose underlying set are the pairs $(n, h) \in N \times H$ equipped with an operation defined as*

$$\bullet : G \times G \to G \tag{2.80}$$

$$\begin{aligned} ((n_1, h_1), (n_2, h_2)) &\mapsto (n_1, h_1) \bullet (n_2, h_2) \\ &= (n_1 \cdot \varphi_{h_1}(n_2), h_1 \cdot h_2), \end{aligned} \tag{2.81}$$

where $n_1, n_2 \in N, h_1, h_2 \in H$.

This structure is again a group according to the defining axioms, with identity element (e_N, e_H) and inverse $(n, h)^{-1} = (\varphi_{h^{-1}}(n^{-1}), h^{-1})$. Note that the the set $\{(n, e_H) : n \in N\}$ is a normal subgroup of G isomorphic to N.

It is also possible to go the other way around and obtain from a group G and a normal subgroup N a new group called *quotient group*, denoted by G/N. This is the set of all cosets of N in G, i.e.,

$$G/N \coloneqq \{\, gN : g \in G \,\}, \tag{2.82}$$

where gN is the *left coset of N in G* , namely,

$$gN \coloneqq \{\, gn : n \in N \,\}. \tag{2.83}$$

The latter definition stands for all subgroup N, not necessarily normal, however when N is normal the left coset and the right coset (defined analogously) coincide. The set G/N is then a group under the operation $(gN) \cdot (hN) = (gh)N$.
We conclude this paragraph with the following definition of the *canonical projection* which is involved in the construction of irreducible representations of a semi-direct product group.

Definition 2.8 (Canonical projection)**.** *Let $N \lhd G$. The group homomorphism*

$$\tau : G \to G/N, \tag{2.84}$$

$$g \mapsto gN \tag{2.85}$$

is called canonical projection.

We are now going to introduce representations, the core mathematical objects which the respective theory is named after.

Definition 2.9 (Representation)**.** *A* representation *of a group G on a vector space V is a group homomorphism onto the general linear group on V,i.e., a map*

$$\pi : H \to \mathrm{GL}(V), \tag{2.86}$$

$$g \mapsto \pi(g) \tag{2.87}$$

such that

$$\pi(g) \cdot \pi(h) = \pi(g \cdot h). \tag{2.88}$$

A representation is said to be *faithful* if it is injective, and its dimension corresponds to the dimension of the vector field V. A subspace $W \subset V$ is said to be *invariant* if, for all $g \in G$ and $w \in W$,

$$\pi(g)w \in W. \tag{2.89}$$

Furthermore, a representation is said to be *irreducible* if the only invariant subspaces are $\{ 0 \}$ and V itself; often, this is abbreviated as *irrep*. Every complex representation of a finite group is *completely reducible*, i.e., it can be decomposed as a direct sum of irreducible representations. This property, together with Schur's Lemma, makes irreducible representations and their *characters* a central object in the theory and will also be particularly relevant in our work.

Definition 2.10 (Character of a representation)**.** *The* character χ_π of a representation π of a group G on V *is given by*

$$\chi_\pi(g) = \mathrm{Tr}[\pi(g)]. \tag{2.90}$$

The dimension of a representation corresponds then to its character at the identity element, $\chi_\pi(e)$. For finite group, the number of irreducible representations is again finite, and the following result is useful to check if all irreducible representations of a given group have been found.

Proposition 2.11 (Group order and irreducible representations dimension)**.** *The order of a group G and the dimension of its irreducible representations are linked by*

$$|G| = \sum_{\alpha \,:\, \pi_\alpha \mathrm{irrep}} \chi_{\pi_\alpha}(e)^2. \tag{2.91}$$

One of the most important properties for character of irreducible representations is the following orthogonality relation.

Proposition 2.12 (Orthogonality formula)**.** *Let $\{ \chi_j \}_j$ be the set of characters of all irreducible representations of a group G. Then*

$$\frac{1}{|G|} \sum_{g \in G} \chi_j^*(g) \chi_k(g) = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k. \end{cases} \tag{2.92}$$

From this, follows one of the key results in representation theory is the formula for multiplicities, used to decompose a representation into its irreducible components.

Proposition 2.13 (Multiplicity formula)**.** *Let χ_j be the character of the irreducible representation π_j and ϕ the character of the representation π of a group G. Then*

$$\frac{1}{|G|} \sum_{g \in G} \chi_j^*(g) \phi(g) = m_j, \tag{2.93}$$

where m_j is the multiplicity of the irreducible representation π_j in the decomposition of π, so that π is similar to a block diagonal matrix in the form

$$\pi(g) \simeq \bigoplus \pi_j(g) \otimes \mathbb{1}_{m_j} \qquad \forall g \in G, \tag{2.94}$$

with $\mathbb{1}_{m_j}$ being the identity matrix on $\mathbb{C}^{m_j}$.

Schur's Lemma We write here one of the most important results in representation theory, namely, *Schur's Lemma*. We will restrict it on finite-dimensional representations case.

Lemma 2.14 (Schur's Lemma)**.** *Let π_j and π_k be two irreducible representations of a finite group G of dimension m and n respectively, and M an $m \times n$ matrix. If*

$$\pi_j(g)\, M\, \pi_k^{-1}(g) = M \qquad \forall g \in G \tag{2.95}$$

then π_j and π_k are equivalent irreducible representations or $M = 0$.

Furthermore, if

$$\pi_j(g)\, M\, \pi_j^{-1}(g) = M \qquad \forall g \in G \tag{2.96}$$

then $M = \mu \mathbb{1}$, i.e., it is a scalar matrix.

Lie groups and Lie algebras Lie algebras and Lie groups are particularly relevant constructions because of their connection to physical models, such as continuous-time stochastic Hamiltonians. In the following, we will mainly follow ref. [49] to illustrate the most important definitions. We will restrict ourself to the study of the finite-dimensional case and focus on matrix groups.

Definition 2.15. *A* Lie group *is a smooth manifold G which is also a group and such that the group product $G \times G \to G$ and the inverse map $G \to G$ are smooth.*

A *matrix Lie group* is a group of invertible square matrices with complex entries and, as the name itself suggests, it is a Lie group according to the above definition. An alternative formulation can be given through converging matrix sequences, which however we do not need to introduce. Next, we define Lie algebras as follows.

Definition 2.16. *A finite-dimensional* Lie algebra $\mathfrak{g}$ *is a finite-dimensional vector space equipped with a bilinear, skew-symmetric operation $[\cdot, \cdot]$ called* Lie braket *satisfying the Jacobi identity*

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0 \qquad \textit{for all } X, Y, Z \in \mathfrak{g}. \tag{2.97}$$

A subalgebra $\mathfrak{h}$ *is a subspace of $\mathfrak{g}$ such that $[H, L] \in \mathfrak{h}$ for all $H, L \in \mathfrak{h}$, i.e., $\mathfrak{h}$ is closed with respect to the Lie braket.*

To connect the two constructions, one makes use of the *matrix exponential*, given by the usual power series; for a square matrix X, this is given by $\mathrm{e}^X = \sum_{k=0}^{\infty} \frac{X^k}{k!}$. We then have the following relation.

Definition 2.17. *Let G be a matrix Lie group. The Lie algebra $\mathfrak{g}$ of G is the set of all matrices X such that $\mathrm{e}^{tX} \in G$ for all real values t.*

The Lie braket of a Lie algebra of a matrix Lie group is then explicitly given by the commutator $[X, Y] = XY - YX$.

When restricted to the lie algebra $\mathfrak{g}$ of a matrix Lie group G, the matrix exponential $\mathrm{e} : \mathfrak{g} \to G$ is called *exponential map*. One can ask whether this map is a universal cover of the Lie group. This is in general not the case: one can find counter-examples for Lie group elements which cannot be written by exponentiating an element of the corresponding Lie algebra. However, the answer changes when considering a neighborhood around the identity $\mathbb{1}_G$; indeed, for a matrix Lie group G and its Lie algebra $\mathfrak{g}$, there exist a neighborhood U of 0 in $\mathfrak{g}$ and a neighborhood V of $\mathbb{1}_G$ in G such that the exponential map $\mathrm{e} : U \to V$ is a *homeomorphism*. One can hence

identify the two structures when working closely to the identity.

The Lie algebra can also be understood as the *tangent space* of its Lie group at some point $g \in G$, given by the differential $\mathrm{d}\gamma/\mathrm{d}t|_{t=0}$ of smooth curves $\gamma : \mathbb{R} \to G$ such that $\gamma(0) = g$. The tangent space at the identity is also equivalent to the tangent space at any other point of the manifold and one can prove this by using an argument based on left-invariant vector fields; hence, we can restrict our considerations to the Lie algebra at the identity.

Now more concretely, in this work we study properties of $N \times N$ invertible matrices U with $U^\dagger U = UU^\dagger = \mathbb{1}$, called the *unitary group* $\mathbb{U}(N)$. This is a matrix Lie group whose Lie algebra $\mathfrak{u}(N)$ is given by the space of $N \times N$ anti-Hermitian matrices, i.e., all matrices X such that $X^\dagger = -X$. The *special unitary group* $\mathbb{SU}(N)$ is the subgroup of $\mathbb{U}(n)$ given by all unitary matrices with determinant equal to 1. Its Lie algebra $\mathfrak{su}(N)$ is characterized by the space of $N \times N$ anti-Hermitian matrices with vanishing trace.

CHAPTER 3

Measures over the unitary group and unitary designs

In this section we introduce the notions of *measures on the unitary group* and of *unitary designs* that we will extensively use throughout this work.

When randomness is involved, the unitary evolution is no more completely determined so that we assign a probability for a given evolution U to take place and affect the system; this leeds us to the concept of *probability distribution over the unitary group*. In order to discuss these topics in the following sections, we first need some formal definitions on topology and probability measures in order to deal with uncountable sets such as the unitary group. We will use the terms *probability measure on* $\mathbb{U}$ and *probability distribution on* $\mathbb{U}$ interchangeably; the former expression is better suited in mathematical frameworks, the latter is predominant in the quantum information literature.

Definition 3.1 (topology and topological space)**.** *Let X be a a non-empty set. A* topology $\mathcal{T}$ *on X is a collection of sets such that*

(i) X *and* $\emptyset \in \mathcal{T}$

(ii) the collection is closed under arbitrarily large unions

(iii) the collection is closed under pairwise intersection, that is, $A, B \in \mathcal{T} \Rightarrow A \cap B \in \mathcal{T}$

The pair $(X, \mathcal{T})$ *is called* topological space*.*

The definition for σ-algebras is different, since the complement of a set is again in the collection and is closed under *countable* many unions.

Definition 3.2 (algebra, σ-algebra and measurable space)**.** *Let X be a non-empty set. A collection $\mathcal{A}$ of subsets of X is an* algebra *if the following axioms are fulfilled*

(i) $X \in \mathcal{A}$

(ii) $A \in \mathcal{A} \Rightarrow X \setminus A \in \mathcal{A}$

(iii) $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$

Furthermore, an algebra that is closed under countable many unions, i.e., satisfies

(iv) $\bigcup_n A_n \in \mathcal{A}$,

is called σ-algebra *and together with* X *the pair* $(X, \mathcal{A})$ *is a* measurable space.

With these two definitions, we can introduce a new algebra that we will use when defining the Haar measure on the unitary group.

Definition 3.3 (Borel algebra and Borel set)**.** *Let* $(X, \mathcal{T})$ *be a topological space. The smallest* σ*-algebra containing all open sets in* $\mathcal{T}$ *is called* Borel algebra*. Elements of this algebra are* Borel sets*.*

It is now time for defining measures on such topological constructions. We start by generic definitions with the final aim being the definition of measures over the unitary group.

Definition 3.4 (measure, probability measure and probability space)**.** *A* measure *on a measurable space* $(X, \mathcal{A})$ *is a map* $\mu : \mathcal{A} \to [0, +\infty]$ *that is countably additive, i.e.,*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n),$$

for any countable collection $\{ A_n \}_n$ *of pairwise disjoint elements in the* σ*-algebra.*
A measure $\mu : \mathcal{A} \to [0, 1]$ *that additionally satisfies* $\mu(X) = 1$ *is said to be a* probability measure*, and the triple* $(X, \mathcal{A}, \mu)$ *is called* probability space*.*

A measure can also be *discrete*. More precisely, a measure μ is discrete with respect to a measure ν if there exists an at most countable subset $Y \subset X$ such that any one-point subset of Y is μ-measurable and $\nu(Y) = 0$. Informally, we can say that a discrete measure on the unitary group is a set of unitary operators (such as unitary gates of a random circuit) equipped with a probability distribution.

3.1 The Haar measure and unitary designs

We first introduce a special class of measures over Borel algebras, which the Haar measure is part of.

Definition 3.5 (Radon measure)**.** *Let* $(X, \mathcal{T})$ *be a topological space and* $\mathcal{B}$ *its Borel algebra. A* Radon measure *on* X *is a measure* $\mu : \mathcal{B} \to [0, +\infty]$ *such that*

(i) for any compact set $K \subset X$, $\mu(K) < \infty$

(ii) for any $B \in \mathcal{B}$, $\mu(B) = \inf\{\mu(V) : B \subset V \text{ and } V \text{ open}\}$

(iii) for any open set $V \subset X$, $\mu(V) = \sup\{\mu(K) : K \subset V \text{ and } K \text{ compact}\}$.

We are now finally ready to introduce the most important and extensively used measure on the unitary group, namely, the *Haar measure*. This will be our main reference when discussing probability distributions induced by stochastic processes over $\mathbb{U}$.

Definition 3.6 (Haar measure, see ref. [50])**.** *The Haar measure is the unique (up to a strictly positive scalar factor) Radon measure which is non-zero on non-empty open sets and is left- and right-invariant, i.e.,*

$$\mu_{\text{Haar}}(U) > 0 \quad \text{for any non-empty open set } U \subset \mathbb{U} \tag{3.1}$$

and

$$\mu_{\text{Haar}}(B) = \mu_{\text{Haar}}(uB) = \mu_{\text{Haar}}(Bu), \tag{3.2}$$

for any $u \in \mathbb{U}$ *and Borel set* B *of* $\mathbb{U}$*, where the left- and right-translate of* B *with respect to* u *is given by*

$$uB = \{\, u\, b : b \in B \,\} \qquad \text{and} \qquad Bu = \{\, b\, u : b \in B \,\}. \tag{3.3}$$

In simpler words, this means that the measure is unaffected under translations of a given Borel set and for this reason it is also called *uniform measure*. Proofs involving existence and uniqueness can be found in ref. [51].
The Haar measure is widely used in quantum information, from decoupling theorems [52, 53] to the study of quantum channels [54]. One of the fundamental properties of this particular measure is characterized as the *Schur-Weyl duality* [55, 56] that we are going to mention after introducing moment operators with respect to probability measures.

Definition 3.7 (k-th moment operator)**.** *The* k-th moment operator M_μ^k *on* $\mathcal{L}(\mathcal{H}^{\otimes k})$ *with respect to a distribution* μ *on* $\mathbb{U}(N)$ *is given by*

$$X \mapsto M_\mu^k(X) \coloneqq \mathbb{E}_\mu \left[U^{\otimes k} X \, (U^\dagger)^{\otimes k} \right]. \tag{3.4}$$

The Schur-Weyl duality says that M_{Haar}^k is an orthogonal projection – with $\langle A, B\rangle \coloneqq \text{Tr}[A^* B]$ as inner product – onto the span of operators representing a permutation of the k tensor copies of $\mathcal{H}$ (see ref. [56, Proposition 2.2] for a complete description). This means that all elements of this space are eigenvectors with unit eigenvalue, while the complement space belongs to the kernel. In the first two cases, for instance, we have

$$M_{\text{Haar}}^1(X) = \frac{\mathbb{1}}{N} \text{Tr}[X] \tag{3.5}$$

and

$$M_{\text{Haar}}^2(X) = \alpha \mathbb{1} + \beta \mathbb{F}, \tag{3.6}$$

where N denotes the dimension of $\mathcal{H}$, $\mathbb{F}$ is the flip operator permuting the two copies of the Hilbert space, i.e., $\mathbb{F}(|a\rangle\langle b| \otimes |c\rangle\langle d|) = |c\rangle\langle b| \otimes |a\rangle\langle d|$, and α and β satisfy $\text{Tr}[X] = \alpha N^2 + \beta N$ and $\text{Tr}[\mathbb{F}X] = \alpha N + \beta N^2$ (see ref. [52]).
In the particular case of $\mathbb{U}(4)$, i.e., the two-qubit case, when applied on the tensor product of two Pauli matrices $\sigma_\alpha \otimes \sigma_\beta$ with $\alpha, \beta \in \{\, 0, 1, 2, 3 \,\}^2$, the second moment operator of the Haar measure gives [24, 28]

$$M_{\text{Haar}}^2(\sigma_\alpha \otimes \sigma_\beta) = \begin{cases} \mathbb{1} \otimes \mathbb{1} & \text{if } \alpha = \beta = \{0,0\} \\ \frac{1}{15} \sum_{\alpha \neq \{0,0\}} \sigma_\alpha \otimes \sigma_\alpha & \text{if } \alpha = \beta \neq \{0,0\} \\ 0 & \text{if } \alpha \neq \beta \end{cases} \tag{3.7}$$

We are now ready to introduce one of the core concepts of this work: exact and approximate unitary designs in terms of M_μ^k. Unitary designs have a wide range of applications in quantum algorithm design [57–59], in quantum state and process tomography [60, 61], and in notions of benchmarking [62, 63] – basically as a powerful tool for *partial de-randomisation*. Conceptually, they feature strongly in descriptions of equilibration, thermalisation and scrambling [37, 64, 65].

Definition 3.8 (exact and approximate unitary designs)**.** *Let* μ *be a distribution over the unitary group* $\mathbb{U}(N)$*. Then* μ *is an* ε-approximate unitary k-design *if*

$$\left\| M_\mu^k - M_{\text{Haar}}^k \right\|_\diamond \leq \varepsilon. \tag{3.8}$$

For $\varepsilon = 0$*, the distribution* μ *is also called an* exact unitary k-design*.*

Physically implementing an exact unitary design is in general neither an obvious nor an efficient task. Fortunately, for a plethora of applications, exactness of a design is not required. Instead, we are usually interested in obtaining *approximate unitary designs*, i.e., (discrete) distributions which behave similarly to the Haar measure and which can be implemented efficiently. Also, we would like to note that there are different formal definitions of unitary designs, each of which being equipped with a different interpretation and being relevant in a number of context. For instance, in the literature regarding randomized benchmarking, an exact unitary 2-design (which is the most relevant design for protocols and applications in this domain) is said to satisfy the following condition.

Condition 3.9 (twirling)**.** *Any exact unitary 2-design μ satisfies the so-called* twirling condition*,*

$$\mathbb{E}_{\mu}[U\,\Lambda(U^{\dagger}\,X\,U)\,U^{\dagger}] = \mathbb{E}_{\mathrm{Haar}}[U\,\Lambda(U^{\dagger}\,X\,U)\,U^{\dagger}], \tag{3.9}$$

for all quantum channels Λ and operators X.

Further properties of twirled channels, such as depolarization of a quantum state, will be discussed later on in Section 6.1 where they will be directly employed in the task of characterizing experimental implementations of unitary gates.

Observation 3.10. *Being a unitary k-design implies being a unitary $(k-1)$-design, since as definition it stands that*

$$\mathbb{E}_{\mu}\left[U^{\otimes k}\,X\,U^{\otimes k}\right] = \mathbb{E}_{\mathrm{Haar}}\left[U^{\otimes k}\,X\,U^{\otimes k}\right] \tag{3.10}$$

for all operators $X \in \mathcal{L}(\mathcal{H}^{\otimes k})$. We can choose $X = Y \otimes \mathbb{1}$ with Y being an arbitrary operator in $\mathcal{L}(\mathcal{H}^{\otimes k-1})$. Then follows

$$\mathbb{E}_{\mu}\left[U^{\otimes k-1}\,Y\,U^{\otimes k-1}\right] = \mathbb{E}_{\mathrm{Haar}}\left[U^{\otimes k-1}\,Y\,U^{\otimes k-1}\right], \tag{3.11}$$

i.e., μ is unitary $(k-1)$-design.

An effective way to obtain a bound on ε in eq.(3.8) is to analyze the gap of the moment operator M_{μ}^{k}, leading to the following definition.

Definition 3.11 (tensor product expanders)**.** *A distribution μ on the unitary group $\mathbb{U}(N)$ is a quantum (λ, k)-tensor product expander if*

$$\left\| M_{\mu}^{k} - M_{\mathrm{Haar}}^{k} \right\|_{\infty} \leq \lambda. \tag{3.12}$$

The following lemma links this definition to the one of designs.

Lemma 3.12 (criterion for being an approximate unitary design, cfr. [66, Lemma 2.2.14])**.** *Let μ be a distribution on $\mathbb{U}(N)$. If μ is a quantum (λ, k)-tensor product expander, then μ is also an ε-approximate k-design with $\varepsilon = N^{k}\lambda$.*

An important connection between eigenvalues and eigenspaces can be made for *universal distributions* as we will see in Lemma 3.14 below. In order to amplify closeness of a distribution μ on $\mathbb{U}(N)$ to the Haar measure, one can convolute it ℓ times with itself and obtain a new measure $\mu^{\star\ell}$ on $\mathbb{U}(N)$. Importantly, it holds that

$$M_{\mu^{\star\ell}}^{k} = (M_{\mu}^{k})^{\ell}\,. \tag{3.13}$$

If the support of $\mu^{\star\ell}$ becomes dense in $\mathbb{U}(N)$ for large ℓ we call μ *universal*. More precisely, a universal distribution can be defined as follows.

Definition 3.13 (universal distribution)**.** *Let μ be a distribution on $\mathbb{U}(N)$. Then μ is said to be* universal *if for all $V \in U(N)$ and any $\delta > 0$ there exists a positive integer ℓ such that*

$$\mu^{\star\ell}\left(B_\delta(V)\right) > 0, \tag{3.14}$$

where $B_\delta(V)$ is the neighbourhood of V with radius $\delta > 0$.

Here, the canonical way to capture the radius is in terms of the geodesic distance on $\mathbb{U}(N)$. It should be clear, however, that any other equivalent metric gives rise to the same definition of universality. This definition can be seen as a generalization of a universal gate set: if μ is the uniform distribution over finitely many unitaries then this set of unitaries is universal if and only if μ is universal.

Before concluding this part of the section with a small discussion about universality of discrete sets, we mention an important result regarding the convergence of all moment operators for any universal distribution to the ones of the Haar measure.

Lemma 3.14 (Lemma 3.7 in ref. [28])**.** *Let μ be a distribution on $\mathbb{U}(N)$. Then all eigenvectors of M^k_{Haar} with unit eigenvalue are eigenvectors of M^k_μ with unit eigenvalue. Additionally, if μ is universal then μ is k-copy gapped for any positive integer k. This means that*

$$\left\| M^k_\mu - M^k_{\text{Haar}} \right\|_\infty < 1. \tag{3.15}$$

As a consequence, $M^k_{\mu^{\star\ell}}$ converges to M^k_{Haar} for $\ell \to \infty$. For many practical applications, however, a bound on the convergence rate is needed. One of the central results of this work is to extend such a bound from quantum circuits [27] to locally generated Brownian motion on $\mathbb{U}(N)$. As we will discuss at the beginning of Chapter 7, *Brownian motion* over the unitary group induces a probability distribution on $\mathbb{U}$ at any time T; precise details will be given later on.

We now focus on a particular set that is extremely useful in quantum information in order to approximate any unitary within any arbitrary accuracy ε.

3.2 Universality of Hadamard, CNOT and T-gate

A powerful result which allows implementation of arbitrary gates, up to arbitrary approximations, is the universality of the following gate set, $\{\,\mathrm{H,T,CNOT}\,\}$, represented by

$$\left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{e}^{\mathrm{i}\pi/4} \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \right\}. \tag{3.16}$$

This particular set is hence of great interest in the quantum computing domain and used in many quantum circuit settings [67].

We will present the argument in ref. [22], but we will illustrate only the fact that the Hadamard gate and T-gate can be combined to approximate any single-qubit unitary operation to arbitrary precision and invite the interested reader to refer to the book of Nielsen and Chuang for the rest of the proof to extend universality on $\mathbb{U}(2^n)$ for an arbitrary number n of qubits with the addition of the CNOT gate.

By construction, the T-gate is – up to a negligible global phase – a rotation around the $\hat{z}$-axis of the Bloch sphere by an angle $\pi/4$,

$$\mathrm{T} = \mathrm{e}^{\,i\pi/8} R_z(\pi/4) \equiv \mathrm{e}^{\,i\pi/8}\ \exp\{-\mathrm{i}\pi/8\ Z\}. \tag{3.17}$$

Analogously, the composition HTH is a rotation around the $\hat{x}$-axis of the Bloch sphere by $\pi/4$ radiants. Up to a global phase, the multiplication of the two rotations gives

$$\exp\{-\mathrm{i}\pi/8\ Z\}\ \exp\{-\mathrm{i}\pi/8\ X\} = \cos^2\frac{\pi}{8}\mathbb{1} - \mathrm{i}\sin\frac{\pi}{8}\left[\cos\frac{\pi}{8}(X+Z) + \sin\frac{\pi}{8}Y\right]. \tag{3.18}$$

This corresponds to a rotation by an angle $\theta = 2\arccos\left(\cos^2\frac{\pi}{8}\right)$ along a unit vector $\hat{n}$ collinear to $\vec{n} = (\cos\frac{\pi}{8}, \sin\frac{\pi}{8}, \cos\frac{\pi}{8})$; we denote it by $R_{\hat{n}}(\theta)$. The crucial point is that θ can be shown to be an *irrational* multiple of 2π. Now, by defining an angle $\theta_k = k\theta \mod 2\pi \in [0, 2\pi)$ and an accuracy δ, we show that a repetition of its rotations fills the interval $[0, 2\pi)$ with points such that none of them is distant by more than δ from its nearest neighbor. Let us take $N > 2\pi/\delta$, then by pigeonhole argument there exist two rotations θ_k and θ_j with $j, k \in \{1, 2, \ldots, N\}$ such that $\theta_k - \theta_j = \theta_{(k-j)} \leq 2\pi/n < \delta$, and, since θ is an irrational multiple of 2π, $\theta_k - \theta_j \neq 0$. Hence, rotations $\theta_{\ell(k-j)},\ \ell = 1, 2, \ldots$ fill up the interval $[0, 2\pi]$ with points that are no more than δ far apart.
Hence, let us fix $\delta = \varepsilon/3$, then for any α there exists L such that

$$\max_{\psi} \|(R_{\hat{n}}(\alpha) - R^L_{\hat{n}}(\theta))\,|\psi\rangle\| < \frac{\varepsilon}{3}. \tag{3.19}$$

Furthermore, one can show that conjugation by Hadamard gate can translate the rotational axis on the Bloch sphere from $\hat{n}$ to $\hat{m}$ collinear to $\vec{m} = (\cos\frac{\pi}{8}, -\sin\frac{\pi}{8}, \cos\frac{\pi}{8})$, for any angle α. Formally,

$$HR_{\hat{n}}(\alpha)H = R_{\hat{m}}(\alpha). \tag{3.20}$$

By the same argument as above, we have

$$\max_{\psi} \|(R_{\hat{m}}(\alpha) - R^L_{\hat{m}}(\theta))\,|\psi\rangle\| < \frac{\varepsilon}{3}. \tag{3.21}$$

Now, making use of [22, Theorem 4.1], follows that any single qubit unitary U can be written as

$$U = \mathrm{e}^{\mathrm{i}\alpha} R_{\hat{n}}(\beta) R_{\hat{m}}(\gamma) R_{\hat{n}}(\omega). \tag{3.22}$$

Using the fact that errors add at most linearly, we conclude

$$\max_{\psi} \|(U - R^{L_1}_{\hat{n}}(\theta) H R^{L_2}_{\hat{n}}(\theta) H R^{L_3}_{\hat{n}}(\theta))\,|\psi\rangle\| < \varepsilon \tag{3.23}$$

for some integers L_1, L_2, L_3. Hence, any single qubit gate U can be approximated, up to a negligible phase factor, by combinations of Hadamard and T-gates.

In ref. [22, Section 4.5.2] it is proven that a combination of single qubit gates together with the CNOT gate can implement arbitrary two-qubit unitary operations, while in Section 4.5.1 it is pointed out that arbitrary local two-qubit gates can be composed to obtain any unitary n-qubit gate. Hence, the set {H,T,CNOT} is universal in the sense of Definition 3.13.

While this is a very powerful result explaining the success of local quantum circuits, it is unfortunately not efficient: there exist unitary transformations requiring $O(n^2 4^n \log^c(n^2 4^n/\varepsilon))$ gates from this specific universal set to be approximate within a precision ε.

3.3 The Clifford group as a unitary 3-design

The *Clifford group* is one of the most relevant groups involved in this work and many others in quantum information theory. It is defined as the *normalizer* subgroup in $\mathbb{U}(2^n)$ of the Pauli group $\mathcal{P}_n^\star$ on n qubits generated by the set of Pauli matrices, i.e.,

$$\mathcal{C}_n \coloneqq \{\, C \in \mathbb{U}(n) : C^\dagger P C \in \mathcal{P}_n^\star \quad \forall P \in \mathcal{P}_n^\star \,\}\,. \tag{3.24}$$

The dimension of this group is [68, Corollary 5.6]

$$|\mathcal{C}_n| = 8 \prod_{j=1}^{n} 2(4^j - 1)4^j = 2^{n^2+2n+3} \prod_{j=1}^{n} (4^j - 1) \tag{3.25}$$

or, when neglecting global phase factors which are irrelevant when acting over conjugation, it is reduced to [69]

$$|\mathcal{C}_n| = \prod_{j=1}^{n} 2(4^j - 1)4^j = 2^{n^2+2n} \prod_{j=1}^{n} (4^j - 1). \tag{3.26}$$

It has been proven – by induction – that the whole group is generated by the set $\{$ H,P,CNOT $\}$ [70, 71], where

$$\mathrm{P} = \mathrm{T}^2 = \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{i} \end{pmatrix}. \tag{3.27}$$

It is very interesting to remark that substituting the P-gate with the T-gate we achieve a universal gate set, as mentioned in the previous subsection 3.2, and this is one of the main reasons why the T-gate is of great interest for fundamental aspects of quantum information and applications thereof.

In a recent result it has been proven – in two different approaches – that the Clifford group is an exact unitary 3-design and that this is the highest level that this set can reach, i.e., it is not a unitary 4-design [72, 73]. While this result is certainly important from a theoretical perspective, unitary 3-designs do not play a particularly relevant role in applications, although used to show that quantum speed-ups occur for most unitary operators [57]. Arguably, the real breakthrough has been the result of Dankert et al. [74, Theorem 1], stating that the Clifford group constitutes an exact unitary 2-design on arbitrarily large n-qubit systems. To this end, one considers quantum channels of the form $\Lambda(X) = AXB$ for $A, B \in \mathrm{End}(\mathbb{C}^{2n})$ and show to satisfy the twirl condition

$$\mathbb{E}_{\mathcal{C}_n}\left[U^\dagger A U X U^\dagger B U\right] = \mathbb{E}_{\mathrm{Haar}}\left[U^\dagger A U X U^\dagger B U\right], \tag{3.28}$$

where in ref. [75] the RHS has been reformulated as

$$\mathbb{E}_{\mathcal{C}_n}\left[U^\dagger A U X U^\dagger B U\right] = \frac{\mathrm{Tr}[AB]\,\mathrm{Tr}[X]}{4^n}\mathbb{1} + \frac{2^n\,\mathrm{Tr}[A]\,\mathrm{Tr}[B] - \mathrm{Tr}[AB]}{2^n(4^n-1)}\left(X - \frac{\mathrm{Tr}[X]}{2^n}\mathbb{1}\right). \tag{3.29}$$

To evalute the LHS, one executes two twirls, the first one on the set of n-qubit Pauli matrices and subsequently another one on representative elements of the quotient group $\mathcal{C}_n/\mathcal{P}_n$. The first twirl gives (we label elements of the n-qubit Pauli set with an index j such that $\sigma_1 = \mathbb{1}_n$ and use Pauli-Liouville representation for A and B)

$$\frac{1}{|\mathcal{P}_n|}\sum_{\sigma\in\mathcal{P}_n} \sigma A \sigma X \sigma B \sigma = \frac{1}{4^n}\sum_{j=1}^{4^n} r_j\,\sigma_j X \sigma_j \tag{3.30}$$

for a certain set of constants $\{ r_j \}_j$ fulfilling the conditions $r_1 = \mathrm{Tr}[A]\,\mathrm{Tr}[B]/4^n$ and $\sum_{j=1}^{4^n} r_j = \mathrm{Tr}[AB]/2^n$. Now, let us pick representative elements $\{ Q_1, \ldots, Q_{|\mathcal{C}_n/\mathcal{P}_n|} \}$ of the quotient group $\mathcal{C}_n/\mathcal{P}_n$ so that, independently of the choice made, from the defining property of the Clifford group as the normalizer of the Pauli group mapping uniformly on elements of the group (except the identity) one has

$$\mathbb{E}_{\mathcal{C}_n}\left[U^\dagger A U X U^\dagger B U\right] = \frac{|\mathcal{P}_n|}{|\mathcal{C}_n|} \sum_{k=1}^{|\mathcal{C}_n/\mathcal{P}_n|} \sum_{j=1}^{4^n} r_j\, Q_k^\dagger \sigma_j Q_k X Q_k^\dagger \sigma_j Q_k \tag{3.31}$$

$$= r_1\, X + \frac{1}{4^n - 1} \sum_{j=2}^{4^n} r_j \sum_{\ell=2}^{|\mathcal{C}_n/\mathcal{P}_n|} \sigma_\ell X \sigma_\ell. \tag{3.32}$$

Using the 1-design identity $\frac{1}{4^n}\sum_{\ell=1}^{4^n} \sigma_\ell X \sigma_\ell = \frac{1}{2^n}\,\mathrm{Tr}[X]\mathbb{1}$ we finally obtain the equivalence in eq. (3.28).

3.4 Encryption from an imperfect source of randomness

In this section we are going to discuss an example on how the quantum realm can offer more than the classical case.

In classical encryption, two parties A and B are communicating over an insecure channel, which a listener L has access over, with the goal of transmitting a secret message M (usually composed by a string of bits 0 and 1) from A to B. To do so, they share a *private key* K unknown to L, to encrypt into a *ciphertext* C the message to be transmitted. The listener is passive, cannot inject anything in the channel or manipulate the encrypted message, but can read it and also knows the function used by A for encryption as well as the decryption function used by B. He also knows the probability distribution of the random source the other two parties have access to. If A and B share a perfect source of unbiased independent random bits, they can use it to generate the private key K. If this secret key is indeed perfectly random, and additionally is at least as long as the *plaintext* message M to be encrypted and has never been used before in whole or in part, then codifying the message into a ciphertext C through modular addition of each digit of the message with a digit of the key, i.e., using a so-called *one-time pad* protocol, ensures perfect security: the transmitted message cannot be decrypted as proven by Shannon in ref. [76]. One of the main issues is that a perfectly random source is very difficult to access, since most pyhsical and computational sources are imperfect and do not output an unbiased distribution. Hence it is relevant to investigate protocols based on these quasi-random sources and quantify how secure they can be. In the work of McInnes and Pinkas [77], two possible weakly random sources are considered. The first one, illustrated by Santha and Varizani [78], is a (SV) source outputting 0 with probability at least $0 \leq \delta < 1/2$ and not more than $1 - \delta$, and outputs otherwise 1. This imperfect source is in a number of situation indistinguishable from the perfectly random one, and thus the former can be used in place of the latter in different tasks such as pseudo-random number generators, randomizing algorithms and stochastic simulation experiments, although as we shall see in the following it does not guarantee a secure communication. The same can be said about another quasi-random source investigated in ref. [79], called *Probability Bounded source* or simply *PRB source*, which is defined with respect to two parameters, ℓ and b. A (ℓ, b)-source, given an arbitrary prefix α of the output sequence and any ℓ-bit string β, the conditional probability that the next ℓ bits emitted by the source is equal to β is at most 2^{-b}, that is, $\mathbb{P}(\beta|\alpha) \leq 2^{-b}$.

Now, if A and B have both access to a SV source, it can be proven that any correct cryptosystem, that is, a system where the probability that B decrypts the original message is close to 1 and independent to the key and message length, is not secure; more precisely for any encryption of

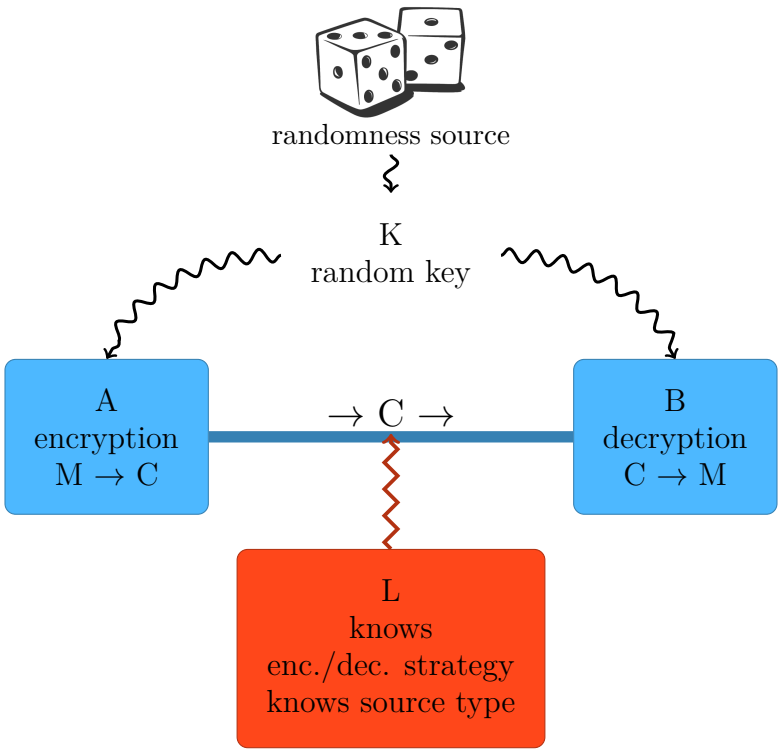


Figure 3.1

a single bit message and decryption functions there exist a SV source with parameter δ and a strategy for the listener such that the probability that he recovers the same decrypted message from C as B is greater or equal to $1/2 + p(\delta)$, for some positive value p depending on δ but not on the length of the key or the one of the ciphertext. For a parameter $\delta \geq 0.45$, [77, Theorem 3] tells us a specific value for the upper bound, that is,

$$\mathbb{P}\left(L \text{ outputs the same as } B\right) \geq 1/2 + \frac{1/2 - 3/2\delta + \delta^2}{2.76}. \tag{3.33}$$

For a PRB source, this bound gets even more pronounced with a constant advantage for $(\ell, \ell-1)$-sources and complete breaking for $(\ell, \ell-2)$-sources. Let us assume that A encrypts a single bit message to B using a n bit private key obtained by a shared PRB source. Then, for every $0 \leq c \leq n$ there exist an $(n, n-c)$-source and a strategy for the listener such that (cfr. [77, Theorem 1])

$$\mathbb{P}\left(L \text{ outputs the same as } B\right) \geq \begin{cases} 1 & \text{for } 2 \leq c \leq n \\ \frac{1}{2} + \frac{2^c}{8} & \text{for } 2 - \log_2 3 \leq c \leq 2 \\ 2^{c-1} & \text{for } 0 \leq c \leq 2 - \log_2 3. \end{cases} \tag{3.34}$$

While secure communication based on a shared quasi-random source is hence not possible, in the same work McInnes and Pinkas show that by providing the parties with an additional public and perfectly random source it is possible to obtain a secure one-time pad.

In quantum mechanics, both plaintext and ciphertext are quantum objects, while the key is still classical. The standard restriction imposed is that the dimensions of plaintext and ciphertext

have to be the same, implying that the encryption operation is a unitary transformation. Taking the formalism of *Private Quantum Channel* (PQC in short) introduced in ref. [80], we assume to have N possible keys $\{ k_j \}_{j=1}^N$, where each k_j is distributed according to a probability p_j, with $\sum_j p_j = 1$ and is linked to a unitary operator U_j. Suppose that A wants to send a pure state ψ to B; first of all, A couples his system with a reference system R in a state ρ_R and then applies the unitary transformation U_j according to the private key k_j and finally sends the resulting state to B. This second party knows the private key and so can apply the inverse transformation $U_j^\dagger$ and trace out the reference system, recovering ψ. The listener L knows the strategy, the probability distribution of the keys and also the correspondence between keys and unitary encryptions, but does not know the private key, so from his point of view the state is given by the operation considering all possibilities weighted according to the probability distribution of the keys, i.e.,

$$\rho_L = \sum_j p_j U_j(|\psi\rangle \langle\psi| \otimes \rho_R)U_j^\dagger. \tag{3.35}$$

The condition for a PQC is that ρ_L is the same for all possible quantum pure states representing the set of plaintexts that Alice can encrypt (by linearity this is then true for all mixed states over this set), hence it provides no information to L. Note that, for encryption without ancilla, $\rho_L = \frac{\mathbb{1}}{2^n}$ whenever $\frac{\mathbb{1}}{2^n}$ is in the span of the input set [80, Theorem 4.3].
To reconnect this argument with the present work, we remark that by using an exact unitary 1-design, from Schur-Weyl duality it follows that any state will be encrypted into the fully mixed state and so provides secure transmission for a PQC. Again, we ask whether secure communication can be achieved with an imperfect random source in the quantum case. In our work *Quantum encryption with weakly random sources*, we show that there always exists an approximate unitary k-design supplied by a weakly random source outputting a random variable X, namely,

Theorem 3.15. *For arbitrary $c \in \mathbb{N}_0$ (min-entropy loss parameter), arbitrary k, an arbitrarily small $\varepsilon > 0$ and an arbitrary random variable $X = \{ x_1, \ldots, x_N \}$ with (sufficiently large) $n = \log N$ and $H_{\min}(X) \geq n - c$, there exists a set of unitaries $\{U_j\}_{j=1}^N$ such that the distribution $\{ x_j, U_j \}_j$ is an ε-approximate unitary k-design.*

This theorem shows that, in a sharp contrast to the classical world, in the quantum world we can tolerate an arbitrary min-entropy loss, with an arbitrary precision.
On the other hand, it does not imply that we can implement practical encryption with weak random source. The price we pay for small ε in Theorem 3.15 is enlarging the number of unitaries we use while keeping the min-entropy loss fixed.

Chapter 4

Decoupling

The concept of decoupling is related to the one of correlation. The latter has been studied and researched in both theoretical and experimental fields. This has led us to a relatively good understanding of the topic and to methods to create correlated states. Decoupling is the opposite task: we will analyze conditions under which an initial (correlated) state describing the joint system between two subsytems, A and E, after a physical evolution, is decoupled, meaning that after the process the two subsystems are uncorrelated.

As definition, we say that a subsystem A is *decoupled* from another subsystem E (in most of the cases, we identify the latter as the *environment*), if the state ρ_{AE} of the joint system has the form $\rho_{AE} = \rho_A \otimes \rho_E$. In terms of quantum information, such a state implies that system A does not contain any information about the other system E, and vice versa.

Specifically, a decoupling theorem considers an initial state ρ_{AE} of a system A that may be correlated to the system E. Then, the system A undergoes an evolution under a random unitary U_A followed by a CPM $\mathcal{T}_{A\to B}$ without any interaction with E, such that from the input system A we obtain an output system B (see Fig 4.1). The decoupling theorem then quantifies – in terms of entropic quantities – how uncorrelated the output system B is from the environment E (on the average over the choices of the unitary) and consequently gives us conditions under which a final decoupled state is obtained.

Decoupling is employed for numerous applications; many of them have in common that decoupling of a system B from a system E is used to show that B is then maximally entangled with a complementary system R. In fact, under the assumpion that R is chosen such that the joint state ρ_{BER} is pure, if $\rho_{BE} = \rho_B \otimes \rho_E$ then there exists a subsystem R' of R such that $\rho_{BR'}$ is pure. Additionally, if ρ_B is fully mixed, then $\rho_{BR'}$ is maximally entangled.

In quantum information theory, this argument occurs in *quantum state merging* [81, 82]. As an example, let us consider a two-player game: Alice and Bob share a quantum system described by a density operator ρ_{AB} and each player controls a subsystem whose state corresponds to the partial trace ρ_A and ρ_B respectively. Alice then communicates to Bob partial quantum information such that he obtains the full state ρ_{AB}: she effectively merges her state with the one of Bob.

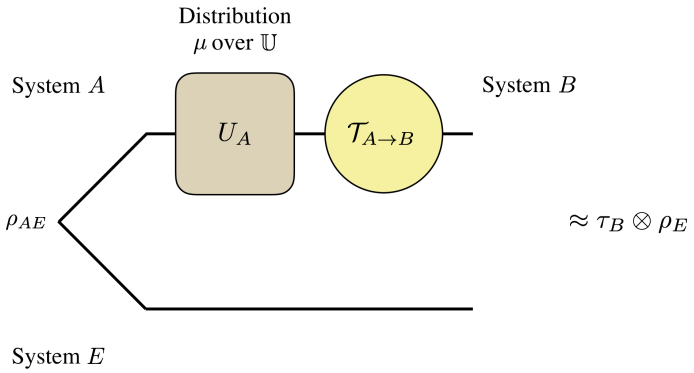


Figure 4.1: In the decoupling theorem, an initial bipartite state ρ_{AE} is affected by a unitary evolution U_A chosen at according to a certain distribution μ. Then, subsystem A is mapped to another subsystem B through a completely positive map $\mathcal{T}_{A\to B}$. Finally, the distance between the final state and the product state $\tau_B \otimes \rho_E$ is characterized by entropy measures.

In quantum information, a state transmission is considered faithful if, although the state merging protocol may depend on the density operator of the source, it succeeds with high probability for any pure state that has been sent. As stated in ref. [82], an equivalent (and elegant) way to interpret this criterion for our two-player game is to imagine that ρ_{AB} is part of a pure state $|\psi_{ABR}\rangle$, where we consider also an additional reference system R. Alice's goal is to transfer her state ρ_A to Bob, but we also demand that, after the protocol, the total state still has high fidelity with $|\psi_{ABR}\rangle$ (meaning that they are nearly identical).
Now, the essential element of the state merging is that ρ_R must be unchanged, and Alice must decouple her state from R: this is a scenario where we can apply a decoupling theorem.
There are also other protocols where decoupling theorems can be brought into play, for instance erasure processes and other fields where randomness is involved [83] as well as for channel capacities [84].

Applications for decoupling theorems can also be found in physics. In thermodynamics, evolution of a system towards thermal equilibrium can be understood as a decoupling process from the observer; in ref. [85], considerations on thermalisation are based on quantum entanglement and interacting quantum systems.
In ref. [86] one studies processes to prepare states from density operators which are initially in an arbitrary state: decoupling theorem are then used to identify pure subsystems which can be then manipulated.
In viewpoint of Landauer's principle, correlation and decoupling are related to the amount of work needed to perform irreversible operations (like the erasure of information): the more we know about the system, the less it costs to erase it [87].
Furthermore, in Section 7.3 we will discuss how *scrambling conditions* for black holes dynamics can be formulated via a decoupling theorem.

4.1 Decoupling theorem and quantum entropy measures

Now formally, we provide a bound on decoupling in terms of *quantum collision entropy measures*, defined in eq. (2.46). This is given as

Theorem 4.1 (non-smooth Decoupling Theorem in ref. [52]). *Let $\rho_{AE} \in \mathcal{S}_{\leq}(\mathcal{H}_{AE})$ and $\mathcal{T}_{A\to B}$ a CPM with Choi-Jamiolkowski representation $\tau_{A'B} = \mathcal{J}(\mathcal{T}_{A\to B})$. Then*

$$\mathbb{E}_{\text{Haar}} \left\| \mathcal{T}(U_A \rho_{AE} U_A^\dagger) - \tau_B \otimes \rho_E \right\|_1 \leq 2^{-\frac{1}{2}H_2(A|E)_\rho - \frac{1}{2}H_2(A'|B)_\tau}. \tag{4.1}$$

Proof of Theorem 4.1. To prove this bound, one first applies the inequality

$$\|M\|_1 \leq \sqrt{\text{Tr}\,[\sigma]\,\text{Tr}\,[\sigma^{-1/2} M \sigma^{-1/2} M^\dagger]} \tag{4.2}$$

for any $M \in \text{End}(\mathcal{H}_A)$ and $\sigma \geq 0 \in \mathcal{H}_A$, which in the particular case where M is Hermitian becomes

$$\|M\|_1 \leq \sqrt{\text{Tr}\,[\sigma]\,\text{Tr}\,[\{\sigma^{-1/4} M \sigma^{-1/4}\}^2]}. \tag{4.3}$$

Hence, for $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$ and $\omega_E \in \mathcal{S}(\mathcal{H}_E)$, one can write

$$\left\| \mathcal{T}(U_A \rho_{AE} U_A^\dagger) - \tau_B \otimes \rho_E \right\|_1 \leq \tag{4.4}$$

$$\sqrt{\text{Tr}\,[\left\{ (\sigma_B \otimes \omega_E)^{-1/4} (\mathcal{T}(U_A \rho_{AE} U_A^\dagger) - \tau_B \otimes \rho_E)(\sigma_B \otimes \omega_E)^{-1/4} \right\}^2]}. \tag{4.5}$$

Now let us define the CPM $\widetilde{\mathcal{T}}_{A\to B}(\,\cdot\,) = \sigma_B^{-1/4} \mathcal{T}_{A\to B} \sigma_B^{-1/4}$ and the states $\tilde{\tau}_{A'B} = \mathcal{J}(\widetilde{\mathcal{T}})$ and $\tilde{\rho}_{AE} = \omega_E^{-1/4} \rho_{AE} \omega_E^{-1/4}$. One then rewrites the above as

$$\left\| \mathcal{T}(U_A \rho_{AE} U_A^\dagger) - \tau_B \otimes \rho_E \right\|_1 \leq \sqrt{\text{Tr}\left[\left\{ \widetilde{\mathcal{T}}(U_A \tilde{\rho}_{AE} U_A^\dagger) - \tilde{\tau}_B \otimes \tilde{\rho}_E \right\}^2 \right]}. \tag{4.6}$$

Using Jensen's inequality, one can put the expected value under the square root:

$$\mathbb{E}_{\text{Haar}} \left\| \mathcal{T}(U_A \rho_{AE} U_A^\dagger) - \tau_B \otimes \rho_E \right\|_1 \leq \sqrt{\mathbb{E}_{\text{Haar}}\, \text{Tr}\left[\left\{ \widetilde{\mathcal{T}}(U_A \tilde{\rho}_{AE} U_A^\dagger) - \tilde{\tau}_B \otimes \tilde{\rho}_E \right\}^2 \right]} \tag{4.7}$$

and then using $\mathbb{E}_{\text{Haar}} \left[U_A \tilde{\rho}_{AE} U_A^\dagger \right] = \frac{1}{|A|} \mathbb{1}_A \otimes \tilde{\rho}_E$ we simplify the expression as

$$\mathbb{E}_{\text{Haar}}\, \text{Tr}\left[\left\{ \widetilde{\mathcal{T}}(U_A \tilde{\rho}_{AE} U_A^\dagger) - \tilde{\tau}_B \otimes \tilde{\rho}_E \right\}^2 \right] = \tag{4.8}$$

$$\mathbb{E}_{\text{Haar}}\, \text{Tr}\left[\left\{ \widetilde{\mathcal{T}}(U_A \tilde{\rho}_{AE} U_A^\dagger) \right\}^2 \right] - \text{Tr}\left[(\tilde{\tau}_B \otimes \tilde{\rho}_E)^2 \right]. \tag{4.9}$$

Using what is sometimes informally called the *swap trick*, namely,

$$\text{Tr}\,[(M \otimes N)\mathbb{F}] = \text{Tr}\,[MN] \tag{4.10}$$

for arbitrary $M, N \in \text{End}(\mathcal{H}_A)$ and $\mathbb{F}$ being the flip operator interchanging the two copies of the subsystem A, one rewrites the first term to bring it into a precise form such that one can subsequently apply the Schur-Weyl duality discussed in Section 3.1,

$$\mathbb{E}_{\text{Haar}}\, \text{Tr}\left[\left\{ \widetilde{\mathcal{T}}(U_A \tilde{\rho}_{AE} U_A^\dagger) \right\}^2 \right] = \mathbb{E}_{\text{Haar}}\, \text{Tr}\left[\left\{ \widetilde{\mathcal{T}}^{\otimes 2}(U_A^{\otimes 2} \tilde{\rho}_{AE}^{\otimes 2} (U_A^\dagger)^{\otimes 2}) \right\} \mathbb{F}_{BE} \right] \tag{4.11}$$

$$= \text{Tr}\left[\tilde{\rho}_{AE}^{\otimes 2} \left(\mathbb{E}_{\text{Haar}} \left\{ (U_A^\dagger)^{\otimes 2} (\widetilde{\mathcal{T}}^\dagger)^{\otimes 2}(F_B) U_A^{\otimes 2} \right\} \otimes \mathbb{F}_E \right) \right], \tag{4.12}$$

where one applies the definition of the adjoint of a superoperator in the second equality. The coefficients satisfy

$$\begin{aligned} \alpha |A|^2 + \beta |A| &= \text{Tr}\,[(\widetilde{\mathcal{T}}^\dagger)^{\otimes 2}(F_B)] \\ &= \text{Tr}\,[F_B \widetilde{\mathcal{T}}^{\otimes 2}(\mathbb{1}_{A\bar{A}})] = |A|^2\, \text{Tr}\,[F_B \tilde{\tau}_B^{\otimes 2}] \\ &= |A|^2\, \text{Tr}\,[\tilde{\tau}_B^2] \end{aligned} \tag{4.13}$$

and

$$\begin{aligned}\alpha|A| + \beta|A|^2 &= \mathrm{Tr}\left[(\widetilde{\mathcal{T}}^\dagger)^{\otimes 2}(F_B)F_A\right] = \mathrm{Tr}\left[F_B\widetilde{\mathcal{T}}^{\otimes 2}(F_A)\right] \\ &= |A|^2\,\mathrm{Tr}\left[F_B\,\mathrm{Tr}_{A'\bar{A}'}\left[\tilde{\tau}^{\otimes 2}_{A'B}(F_{A'}\otimes \mathbb{1}_{B\bar{B}})\right]\right] \\ &= |A|^2\,\mathrm{Tr}\left[(\mathbb{1}_{A'\bar{A}'}\otimes F_B)\tilde{\tau}^{\otimes 2}_{A'B}(F_{A'}\otimes \mathbb{1}_{B\bar{B}})\right] \\ &= |A|^2\,\mathrm{Tr}\left[F_{A'B}\tilde{\tau}^{\otimes 2}_{A'B}\right] \\ &= |A|^2\,\mathrm{Tr}\left[\tilde{\tau}^2_{A'B}\right].\end{aligned} \tag{4.14}$$

The solution of this system of equations is

$$\alpha = \mathrm{Tr}\left[\tilde{\tau}^2_B\right]\left(\frac{|A|^2 - \frac{|A|\,\mathrm{Tr}\left[\tilde{\tau}^2_{A'B}\right]}{\mathrm{Tr}\left[\tilde{\tau}^2_B\right]}}{|A|^2-1}\right) \qquad \text{and} \qquad \beta = \mathrm{Tr}\left[\tilde{\tau}^2_{AB}\right]\left(\frac{|A|^2 - \frac{|A|\,\mathrm{Tr}\left[\tilde{\tau}^2_B\right]}{\mathrm{Tr}\left[\tilde{\tau}^2_{A'B}\right]}}{|A|^2-1}\right). \tag{4.15}$$

Applying the inequality

$$\frac{1}{|A|} \leq \frac{\mathrm{Tr}\left[\xi^2_{AB}\right]}{\mathrm{Tr}\left[\xi^2_B\right]} \leq |A| \tag{4.16}$$

for any $\xi_{AB} \geq 0$, we can bound them with simpler expressions as $\alpha \leq \mathrm{Tr}\left[\tilde{\tau}^2_B\right]$ and $\beta \leq \mathrm{Tr}\left[\tilde{\tau}^2_{A'B}\right]$.

The bound in equation (4.7) can be then expressed as

$$\mathbb{E}_{\mathrm{Haar}}\left\|\mathcal{T}(U_A\rho_{AE}U_A^\dagger) - \tau_B\otimes\rho_E\right\|_1 \leq \sqrt{\mathrm{Tr}\left[\tilde{\tau}^2_{A'B}\right]\mathrm{Tr}\left[\tilde{\rho}^2_{AE}\right]} \tag{4.17}$$

and the final expression of the theorem is finally obtained by using the definitions of $\tilde{\tau}_{A'B}$ and $\tilde{\rho}_{AE}$ together with the definition of the collision entropy H_2. □

Depending on the situation, one is generally more interested in a characterization of the bound by more meaningful operational quantities such as the min- and max-entropy. In particular, the following theorem makes use of smooth versions of these entropy measures, defined by the density operator lying in a neighborhood of the "target" quantum state ρ and maximizing (or minimizing) the entropy value. An extensive discussion regarding this topic is outside the scope of this work, please refer to ref. [88] for further reading.
Knowing that the collision entropy is always greater or equal to the min-entropy, one can substitute the former with the latter in Theorem 4.1 and then consider the smooth versions. Hence, for any $\varepsilon > 0$, we have [52, Theorem 3.1]

$$\mathbb{E}_{\mathrm{Haar}}\left\|\mathcal{T}(U_A\rho_{AE}U_A^\dagger) - \tau_B\otimes\rho_E\right\|_1 \leq 12\varepsilon + 2^{-\frac{1}{2}H^\varepsilon_{\min}(A|E)_\rho - \frac{1}{2}H^\varepsilon_{\min}(A'|B)_\tau}. \tag{4.18}$$

4.2 Decoupling linked to hypothesis-testing

Another version of the decoupling theorem can be given in terms of the *hypothesis-testing entropy*. Albeit constituting previous work on the topic by this author [89], it is actually not involved in the remaining part of the thesis and this section can thus be overlooked.

Let us first introduce the *ε-relative entropy* $D^\varepsilon_{\mathrm{H}}(\rho||\sigma)$ of a density operator $\rho \in \mathcal{S}_{\leq}(\mathcal{H})$ relative to $\sigma \geq 0$, defined as

$$2^{-D^\varepsilon_{\mathrm{H}}(\rho||\sigma)} := \frac{1}{\varepsilon}\inf\left\{\langle Q,\sigma\rangle :\ 0 \leq Q \leq \mathbb{1} \cap \langle Q,\rho\rangle \geq \varepsilon\right\}. \tag{4.19}$$

In words, this reflects the minimal probability – rescaled by ε – that a strategy Q to distinguish ρ from σ produces a wrong guess on input σ (i.e., mistaking σ for ρ), while maintaining a minimum success probability ε in correctly identifying ρ.
Note that, for $\varepsilon = 1$, $D_{\mathrm{H}}^{\varepsilon}(\rho||\sigma)$ is equal to the Rényi entropy of order 0. Properties for $D_{\mathrm{H}}^{\varepsilon}$ include:

[1] Positivity: for any $\rho, \sigma \in \mathcal{S}(\mathcal{H})$

$$D_{\mathrm{H}}^{\varepsilon}(\rho||\sigma) \geq 0 \tag{4.20}$$

with equality if $\rho = \sigma$.

[2] Data Processing Inequality (DPI): for any completely positive, trace non-increasing map $\mathcal{E}$

$$D_{\mathrm{H}}^{\varepsilon}(\rho||\sigma) \geq D_{\mathrm{H}}^{\varepsilon}(\mathcal{E}(\rho)||\mathcal{E}(\sigma)). \tag{4.21}$$

[3] Asymptotic Equipartition Property (AEP): let $D(\rho||\sigma) = \mathrm{Tr}\left[\rho(\log\rho - \log\sigma)\right]$ be the relative entropy between ρ, σ.
Then, for any $0 < \varepsilon \leq 1$

$$\lim_{n\to\infty} \frac{1}{n} D_{\mathrm{H}}^{\varepsilon}(\rho^{\otimes n}||\sigma^{\otimes n}) = D(\rho||\sigma). \tag{4.22}$$

The definition of the *conditional $H_{\mathrm{H}}^{\varepsilon}$-entropy of $\rho_{AB} \in \mathcal{S}_{\leq}(\mathcal{H}_{AB})$ with respect to $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$* is deduced from the relative entropy by

$$H_{\mathrm{H}}^{\varepsilon}(A|B)_{\rho|\sigma} \coloneqq -D_{\mathrm{H}}^{\varepsilon}(\rho_{AB}||\mathbb{1}_A \otimes \sigma_B). \tag{4.23}$$

We can take the supremum over all $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$ and write

$$H_{\mathrm{H}}^{\varepsilon}(A|B)_{\rho} \coloneqq \sup_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} H_{\mathrm{H}}^{\varepsilon}(A|B)_{\rho|\sigma}. \tag{4.24}$$

$H_{\mathrm{H}}^{\varepsilon}$ is computable in terms of a semi-definite program, with primal and dual formulation

PRIMAL		DUAL	
minimize	$\frac{1}{\varepsilon}\mathrm{Tr}\left[Q_{AB}\,(\mathbb{1}_A \otimes \sigma_B)\right]$	maximize	$\mu - \frac{1}{\varepsilon}\mathrm{Tr}\left[X_{AB}\right]$
subject to	$Q_{AB} \leq \mathbb{1}_{AB}$	subject to	$\mu\rho_{AB} - \mathbb{1}_A \otimes \sigma_B \leq X_{AB}$
	$\mathrm{Tr}\left[Q_{AB}\rho_{AB}\right] \geq \varepsilon$		$X_{AB} \geq 0$
	$Q_{AB} \geq 0$		

The complementary *slackness conditions* linking primal and dual optimal solutions $\{Q_{AB}\}$ and $\{\mu, X_{AB}\}$ are

$$(\mu\rho_{AB} - \mathbb{1}_A \otimes \sigma_B)Q_{AB} = X_{AB}Q_{AB} \tag{4.25}$$

$$\mathrm{Tr}\left[Q_{AB}\rho_{AB}\right] = \varepsilon \tag{4.26}$$

$$X_{AB}Q_{AB} = X_{AB}. \tag{4.27}$$

From eq. (4.25) we can then deduce that $[Q_{AB}, X_{AB}] = 0$ such that we can also infer $[\mu\rho_{AB} - \mathbb{1}_A \otimes \sigma_B, Q_{AB}] = 0$. Furthermore, one can see that the positive part of $\mu\rho_{AB} - \mathbb{1}_A \otimes \sigma_B$ is in the eigenspace of Q_{AB} with eigenvalue 1.

The bound on decoupling depending on this entropy quantity is given as follows.

Theorem 4.2 (Decoupling Theorem in ref. [89]). *Let $\varepsilon > 0$, $\rho_{AE} \in \mathcal{S}_{\leq}(\mathcal{H}_{AE})$, $\mathcal{T}_{A\to B}$ a CPM with Choi-Jamiolkowski representation $\tau_{A'B} = \mathcal{J}(\mathcal{T})$. Let $\theta_E \in \mathcal{S}(\mathcal{H}_E)$ be the optimal marginal state such that $H_{\mathrm{H}}^{\varepsilon}(A|E)_{\rho} = H_{\mathrm{H}}^{\varepsilon}(A|E)_{\rho|\theta}$. Then*

$$\begin{aligned}\mathbb{E}_{\mathrm{Haar}} \left\| \mathcal{T}(U_A \rho_{AE} U_A^{\dagger}) - \tau_B \otimes \rho_E \right\|_1 \leq \sqrt{\varepsilon}\, 2^{-\frac{1}{2}H_{\min}(A|E)_{\rho|\theta} - \frac{1}{2}H_{\min}(A'|B)_{\tau}} \\ + 2^{-\frac{1}{2}H_{\mathrm{H}}^{\varepsilon}(A|E)_{\rho} - \frac{1}{2}H_{\min}(A'|B)_{\tau}}.\end{aligned} \tag{4.28}$$

For small value ε, the tightness of the bound is determined by two entropic values alone: the sought $H_{\mathrm{H}}^{\varepsilon}(A|E)_{\rho}$ and $H_{\min}(A'|B)_{\tau}$. The first tells us that the harder is to correctly distinguish the input state $\mathbb{1}_A \otimes \theta_E$ from ρ_{AE}, the tighter is the bound on decoupling. The second entropy quantifies how well the CPM $\mathcal{T}_{A\to B}$, identified by the Choi-Jamiolkowski isomorphism $\tau_{A'B}$, conserves correlation. Note that the influences of ρ_{AE} and of $\mathcal{T}_{A\to B}$ on the bound are independent with respect to each other: there is no better suited mapping for some types of states than for others, and vice versa. In particular, for the case where the CPM $\mathcal{T}_{A\to B}$ is chosen as a partial trace erasing some qubits in subsystem A, we have

$$H_{\min}(A'|B)_{\tau} = m - 2m', \tag{4.29}$$

where m is the initial number of qubits in A and m' the number of remaining qubits in subsystem B. This means that the higher is the number of erased qubits for a given system A, the tighter is the bound. Moreover, according to the theorem, (for ε small enough such that the second term of the bound is dominant), if the exponent

$$H_{\mathrm{H}}^{\varepsilon}(A|E)_{\rho} + H_{\min}(A'|B)_{\tau} \tag{4.30}$$

is sufficiently larger than 0, then decoupling is achieved.

The proof of this decoupling theorem takes again as starting point the non-smooth version 4.1 and heavily relies on the next lemma.

Lemma 4.3. *Let $\varepsilon > 0$ and $\rho_{AE} \in \mathcal{S}_{\leq}(\mathcal{H}_{AE})$. Let $\theta_E \in \mathcal{S}(\mathrm{H}_E)$ be the optimal marginal state such that $H_{\mathrm{H}}^{\varepsilon}(A|E)_{\rho} = H_{H}^{\varepsilon}(A|E)_{\rho|\theta}$. Then*

$$2^{-H_2(A|E)_{\rho}} \leq \varepsilon\, 2^{-H_{\min}(A|E)_{\rho|\theta}} + 2^{-H_{\mathrm{H}}^{\varepsilon}(A|E)_{\rho}}. \tag{4.31}$$

In addition, when ε approaches 0, we have as a corollary

Corollary 4.4. *Let $\varepsilon \to 0$ and $\rho_{AE} \in \mathcal{S}_{\leq}(\mathcal{H}_{AE})$. Then*

$$H_2(A|E)_{\rho} \geq H_{\mathrm{H}}(A|E)_{\rho}. \tag{4.32}$$

Proof of Lemma 4.3. Let $\{\mu, X_{AE}\}$ be the dual optimal solution for the SDP for $H_{\mathrm{H}}^{\varepsilon}(A|E)_{\rho}$. We start from the definition of $2^{-H_2(A|E)_{\rho}}$ and try to gain in the formula the expression $\mu\rho_{AE} - \mathbb{1}_A \otimes \theta_E$. We know that this is related to the dual formulation of the SDP for $H_{\mathrm{H}}^{\varepsilon}(A|E)_{\rho}$ and that it is also present in the slackness conditions.

$$2^{-H_2(A|E)_{\rho}} = \min_{\sigma_E \in \mathcal{S}(\mathcal{H}_E)} \mathrm{Tr}\left[(\mathbb{1}_A \otimes \sigma_E)^{-1/2} \rho_{AE} (\mathbb{1}_A \otimes \sigma_E)^{-1/2} \rho_{AE}\right] \tag{4.33}$$

$$= \min_{\sigma_E \in \mathcal{S}(\mathcal{H}_E)} \frac{1}{\mu} \mathrm{Tr}\left[(\mathbb{1}_A \otimes \sigma_E)^{-1/2} \mu\rho_{AE} (\mathbb{1}_A \otimes \sigma_E)^{-1/2} \rho_{AE}\right] \tag{4.34}$$

$$= \min_{\sigma_E \in \mathcal{S}(\mathcal{H}_E)} \frac{1}{\mu} \mathrm{Tr}\left[(\mathbb{1}_A \otimes \sigma_E)^{-1/2} (\mu\rho_{AE} - \mathbb{1}_A \otimes \theta_E + \mathbb{1}_A \otimes \theta_E)(\mathbb{1}_A \otimes \sigma_E)^{-1/2} \rho_{AE}\right]. \tag{4.35}$$

We have some freedom about the choice of σ_E. Any state of the set $\mathcal{S}(\mathcal{H}_E)$ will give an expression equal or greater than eq. (4.35). We therefore decide to take the optimal marginal state θ_E and get

$$2^{-H_2(A|E)_\rho} \leq \frac{1}{\mu} \operatorname{Tr}\left[(\mathbb{1}_A \otimes \theta_E)^{-1/2}(\mu\rho_{AE} - \mathbb{1}_A \otimes \theta_E + \mathbb{1}_A \otimes \theta_E)(\mathbb{1}_A \otimes \theta_E)^{-1/2}\rho_{AE}\right] \tag{4.36}$$

$$= \frac{1}{\mu} \operatorname{Tr}\left[(\mathbb{1}_A \otimes \theta_E)^{-1/2}(\mu\rho_{AE} - \mathbb{1}_A \otimes \theta_E)(\mathbb{1}_A \otimes \theta_E)^{-1/2}\rho_{AE}\right] + \tag{4.37}$$

$$+ \frac{1}{\mu} \operatorname{Tr}\left[(\mathbb{1}_A \otimes \theta_E)^{-1/2}(\mathbb{1}_A \otimes \theta_E)(\mathbb{1}_A \otimes \theta_E)^{-1/2}\rho_{AE}\right] \tag{4.38}$$

$$\leq \frac{1}{\mu} \operatorname{Tr}\left[(\mathbb{1}_A \otimes \theta_E)^{-1/2}(\mu\rho_{AE} - \mathbb{1}_A \otimes \theta_E)(\mathbb{1}_A \otimes \theta_E)^{-1/2}\rho_{AE}\right] + \frac{1}{\mu} \tag{4.39}$$

$$\leq \frac{1}{\mu} \operatorname{Tr}\left[(\mathbb{1}_A \otimes \theta_E)^{-1/2}(\mu\rho_{AE} - \mathbb{1}_A \otimes \theta_E)(\mathbb{1}_A \otimes \theta_E)^{-1/2}\rho_{AE}\right] + 2^{-H_{\mathrm{H}}^{\varepsilon}(A|E)_\rho}, \tag{4.40}$$

where the last inequality is motivated by

$$2^{H_{\mathrm{H}}^{\varepsilon}(A|E)_\rho} = \mu - \frac{1}{\varepsilon} \operatorname{Tr}[X_{AE}] \leq \mu. \tag{4.41}$$

We focus now on the first term in eq. (4.40). Let $0 \leq P_{AE} \leq \mathbb{1}_{AE}$ be the primal optimal solution in the SDP for $H_{\mathrm{H}}^{\varepsilon}(A|E)_\rho$ and define $\overline{P}_{AE} \coloneqq \mathbb{1}_{AE} - P_{AE}$. We can then apply these two operators to the expression $\mu\rho_{AE} - \mathbb{1}_A \otimes \theta_E$ and make use of the slackness conditions. We have

$$\frac{1}{\mu} \operatorname{Tr}\left[(\mathbb{1}_A \otimes \theta_E)^{-1/2}(\mu\rho_{AE} - \mathbb{1}_A \otimes \theta_E)(\mathbb{1}_A \otimes \theta_E)^{-1/2}\rho_{AE}\right] \tag{4.42}$$

$$= \frac{1}{\mu} \operatorname{Tr}\left[(\mathbb{1}_A \otimes \theta_E)^{-1/2}(P_{AE} + \overline{P}_{AE})(\mu\rho_{AE} - \mathbb{1}_A \otimes \theta_E)(\mathbb{1}_A \otimes \theta_E)^{-1/2}\rho_{AE}\right] \tag{4.43}$$

$$= \frac{1}{\mu} \operatorname{Tr}\left[(\mathbb{1}_A \otimes \theta_E)^{-1/2}P_{AE}(\mu\rho_{AE} - \mathbb{1}_A \otimes \theta_E)(\mathbb{1}_A \otimes \theta_E)^{-1/2}\rho_{AE}\right] + \tag{4.44}$$

$$+ \frac{1}{\mu} \operatorname{Tr}\left[(\mathbb{1}_A \otimes \theta_E)^{-1/2}\overline{P}_{AE}(\mu\rho_{AE} - \mathbb{1}_A \otimes \theta_E)(\mathbb{1}_A \otimes \theta_E)^{-1/2}\rho_{AE}\right]. \tag{4.45}$$

The complementary slackness conditions for primal and dual optimal solutions $\{P_{AE}\}$ and $\{\mu, X_{AE}\}$ are given by

$$P_{AE}(\mu\rho_{AE} - \mathbb{1}_A \otimes \theta_E) = P_{AE}X_{AE} \tag{4.46}$$

$$\operatorname{Tr}\left[P_{AE}\rho_{AE}\right] = \varepsilon \tag{4.47}$$

$$P_{AE}X_{AE} = X_{AE}P_{AE} = X_{AE}. \tag{4.48}$$

This means that the positive part of $\mu\rho_{AE} - \mathbb{1}_A \otimes \theta_E$ is in the eigenspace of P_{AE} with eigenvalue 1, and meanwhile that $\overline{P}_{AE}$ annihilates the positive part of $\mu\rho_{AE} - \mathbb{1}_A \otimes \theta_E$, i.e.,

$$\overline{P}_{AE}(\mu\rho_{AE} - \mathbb{1}_A \otimes \theta_E) \leq 0. \tag{4.49}$$

From this follows that the first term in expression (4.40) can be bounded as

$$\frac{1}{\mu} \operatorname{Tr}\left[(\mathbb{1}_A \otimes \theta_E)^{-1/2}(\mu\rho_{AE} - \mathbb{1}_A \otimes \theta_E)(\mathbb{1}_A \otimes \theta_E)^{-1/2}\rho_{AE}\right] \tag{4.50}$$

$$\leq \frac{1}{\mu} \operatorname{Tr}\left[(\mathbb{1}_A \otimes \theta_E)^{-1/2}X_{AE}(\mathbb{1}_A \otimes \theta_E)^{-1/2}\rho_{AE}\right] + 0 \tag{4.51}$$

$$\leq \frac{1}{\mu}\, 2^{-H_{\min}(A|E)_{\rho|\theta}} \operatorname{Tr}\left[(\mathbb{1}_A \otimes \theta_E)^{-1/2}X_{AE}(\mathbb{1}_A \otimes \theta_E)^{-1/2}(\mathbb{1}_A \otimes \theta_E)\right] \tag{4.52}$$

$$\leq \frac{1}{\mu}\, 2^{-H_{\min}(A|E)_{\rho|\theta}} \operatorname{Tr}\left[X_{AE}\right] \tag{4.53}$$

$$\leq \varepsilon\, 2^{-H_{\min}(A|E)_{\rho|\theta}}, \tag{4.54}$$

where we have used that, by definition of the quantum min-entropy, it stands that

$$2^{-H_{\min}(A|E)_{\rho|\theta}} \mathbb{1}_A \otimes \theta_E \geq \rho_{AE} \tag{4.55}$$

while the last inequality follows from

$$0 \leq 2^{H_{\mathrm{H}}^{\varepsilon}(A|E)_{\rho}} = \mu - \frac{1}{\varepsilon} \operatorname{Tr}\left[X_{AE}\right] \tag{4.56}$$

and consequently

$$\operatorname{Tr}\left[X_{AE}\right] \leq \mu\, \varepsilon. \tag{4.57}$$

Substituting (4.54) into expression (4.40) we can finally conclude

$$2^{-H_2(A|E)_{\rho}} \leq \varepsilon\, 2^{-H_{min}(A|E)_{\rho|\theta}} + 2^{-H_H^{\varepsilon}(A|E)_{\rho}}. \tag{4.58}$$

□

As we have seen in eq. (4.12), the notions of second moment operator and unitary 2-design as well as the Schur-Weyl duality come into play in the proof idea of Theorem 4.1, from which the other ones are stemming. From a mathematical perspective, this is the reason why unitary 2-design are strictly involved in decoupling [53]. Hence, all distributions over the unitary group which are exactly unitary 2-designs will lead to the same result.

CHAPTER 5

Random quantum circuits

Quantum circuits are a fundamental tool of quantum information for manipulation of quantum states [22, 90–92]. These are sequences of gate operations, such as the already introduced Hadamard, CNOT or other Clifford elements, applied to an initial density operator. Implemented gates are often two-qubit, since they can produce operators on multiple qubits by concatenation, but more generally, considering a n-qubit initial state, we say that a quantum gate is ℓ-*local* if it has support on ℓ different qubits. We refer as *size* of the circuit to the number of quantum gates it contains. Two or more local gates with different support can be applied simultaneously in one single *step*, and we call *depth* of a circuit the total number of those steps.

In this work, a precise class of random circuit is taken into account: 2-local random quantum circuits (RQC). These are circuits where at each step two qubits are chosen uniformly at random and a 2-qubit gate, drawn according to a certain unitary distribution, is applied. RQC are a powerful setting displaying important properties, such as efficiently approximating unitary designs of polynomial degree, being a fast decoupler in a number of steps scaling almost linearly in system size, and again being used to characterize the accuracy of gate implementations with randomized benchmarking protocols. In the next section we present a milestone towards these achievements is represented by the work of Harrow and Low [28], where they proved that RQC are approximate unitary 2-designs, making use of a Markov chain strategy already employed in ref. [93].

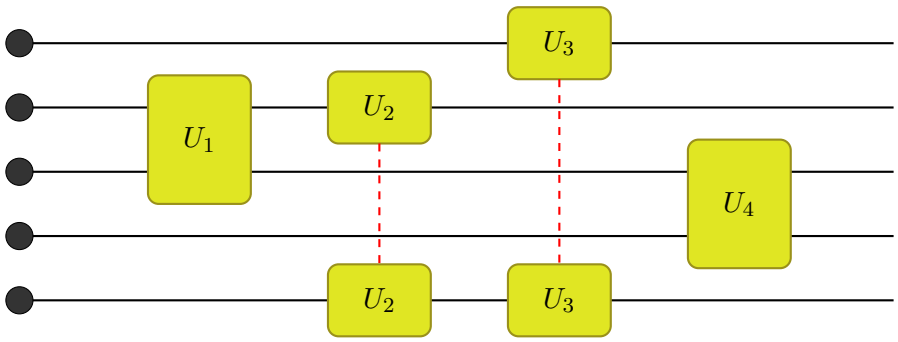


Figure 5.1

5.1 Approximate unitary 2-designs with random walk on Pauli basis

We begin the section by directly stating the main result and then we will proceed with the proof restricted to RQC drawing from the 2-qubit Haar measure, which will be structured around the evolution of Pauli coefficients defined right after the theorem.

Theorem 5.1 (Theorem 2.10 in ref. [28]). *Let μ be a 2-copy gapped distribution on $\mathbb{U}(4)$ and $\mathrm{circ}(\mu)$ be a the distribution on $\mathbb{U}(2^n)$ induced by a random circuit on n qubits drawing gates according to μ. Then there exists $C(\mu)$ depending only on μ such that for any $\varepsilon > 0$ $\mathrm{circ}(\mu)$ is an approximate 2-design after $C(\mu)(n(n+\log 1/\varepsilon))$ steps.*

To prove the theorem the authors have studied the convergence rate of the Pauli basis coefficients towards the uniform distribution. Let us denote the Pauli basis coefficients after T steps of a RQC applied on an initial state ρ as

$$Q^T_{\mathrm{circ}(\mu)}(\alpha,\beta) \coloneqq \frac{1}{4^n}\,\mathrm{Tr}\left[\sigma_\alpha \otimes \sigma_\beta \left(\overbrace{M^{k=2}_{\mathrm{circ}(\mu)} \circ \cdots \circ M^{k=2}_{\mathrm{circ}(\mu)}}^{T \text{ times}}(\rho)\right)\right] \tag{5.1}$$

$$= \frac{1}{4^n}\,\mathrm{Tr}\left[\sigma_\alpha \otimes \sigma_\beta \left(\bigcirc^T M^{k=2}_{\mathrm{circ}(\mu)}(\rho)\right)\right], \tag{5.2}$$

where α and β are n-tuples labeling the n-qubit Pauli strings.
We can now define the following mixing criteria, strictly tied to the above result (cfr. [28, Section 6]).

Lemma 5.2 (Lemma 2.11 in ref. [28]). *Let μ and $\mathrm{circ}(\mu)$ be two distribution as in Theorem 5.1. Let the initial state ρ be such that $Q^0_{\mathrm{circ}(\mu)}(\alpha,\alpha) \geq 0$ fol all α and $\sum_\alpha Q^0_{\mathrm{circ}(\mu)}(\alpha,\alpha) = 1$. Then there exists a constant $C(\mu)$ depending on μ such that for any $\varepsilon > 0$*

[i] for any $T \geq C(\mu) n \log 1/\varepsilon$ (2-norm convergence criterion)

$$\sum_{\alpha,\beta\neq\{0,0\}} \left(Q^T_{\mathrm{circ}(\mu)}(\alpha,\beta) - \frac{\delta_{\alpha,\beta}}{2^n(2^n+1)}\right)^2 \leq \varepsilon \tag{5.3}$$

[ii] for any $T \geq C(\mu) n(n+\log 1/\varepsilon)$ (1-norm convergence criterion)

$$\sum_{\alpha,\beta\neq\{0,0\}} \left|Q^T_{\mathrm{circ}(\mu)}(\alpha,\beta) - \frac{\delta_{\alpha,\beta}}{2^n(2^n+1)}\right| \leq \varepsilon \tag{5.4}$$

and for the special case where μ is the Haar measure over $\mathbb{U}(4)$, the condition is satisfied for $T = O(n \log \frac{n}{\varepsilon})$.

The last statement of this lemma is particularly interesting because it lifts the uncertainty about the constant $C(\mu)$ and at the same time provides a very useful investigation of the Markov chain induced by the random circuit steps. We will summarize in the following the tight analysis for the uniform Haar measure over $\mathbb{U}(4)$. First of all, we prove the suppression of the cross-terms in 1-norm, namely,

Lemma 5.3 (Lemma 5.1 in ref. [28]). *After $T = O(n \log \frac{n}{\varepsilon})$ steps of a random circuit equipped with uniformly distributed 2-qubit unitary gates*

$$\sum_{\alpha\neq\beta} \left|Q^T_{\mathrm{circ(Haar)}}(\alpha,\beta)\right| \leq \varepsilon. \tag{5.5}$$

To show this, one considers the variable H giving the number of different qubits hit by the circuit. Then the probability that this variable is less or equal some value h after T steps is given by

$$\mathbb{P}(H \leq h) \leq \binom{n}{h} \left(\frac{h(h-1)}{n(n-1)}\right)^T \leq \binom{n}{h} (h/n)^T. \tag{5.6}$$

Since the second moment operator with respect to the Haar measure cancels out any tensor product of different single qubit Pauli matrices, i.e., $M^{k=2}_{\text{Haar}}(\sigma_{\alpha_1} \otimes \sigma_{\alpha_2}) = 0$ for $\alpha_1 \neq \alpha_2$ (cfr. eq. (3.7)), every $2n$-qubit Pauli string $\sigma_\alpha \otimes \sigma_\beta$ with $\alpha_j \neq \beta_j$ is immediately suppressed if the qubit j has been hit by the circuit. When h distinct qubits have been it, at most 16^{n-h} different strings survive. Since ρ is a physical state at any time step and therefore $\operatorname{Tr}\rho^2 \leq 1$, it follows $\sum_{\alpha,\beta}(Q^T_{\text{circ(Haar)}}(\alpha,\beta))^2 \leq 1$ and so, if the state has at most N many non zero coefficients,

$$\sum_{\alpha,\beta} \left|Q^T_{\text{circ(Haar)}}(\alpha,\beta)\right| \leq \frac{1}{\sqrt{N}} N = \sqrt{N}. \tag{5.7}$$

Hence it follows

$$\begin{aligned}
\sum_{\alpha\neq\beta} \left|Q^T_{\text{circ(Haar)}}(\alpha,\beta)\right| &\leq \sum_{h=1}^{n-1} \mathbb{P}(H=h) 16^{(n-h)/2} \\
&\leq \sum_{h=1}^{n-1} \mathbb{P}(H\leq h) 4^{(n-h)} \\
&\leq \sum_{h=1}^{n-1} \binom{n}{h} \left(\frac{h}{n}\right)^T 4^{(n-h)} \\
&= \sum_{h=1}^{n-1} \binom{n}{h} \left(1-\frac{h}{n}\right)^T 4^h \qquad h \to n-h \\
&\leq \sum_{h=1}^{n-1} \binom{n}{h} \exp\{-hT/n\} 4^h.
\end{aligned} \tag{5.8}$$

Choosing $T = n \log \frac{n}{\varepsilon}$ finally gives $\sum_{\alpha\neq\beta} \left|Q^T_{\text{circ(Haar)}}(\alpha,\beta)\right| = O(\varepsilon)$.

Now one has to investigate how Pauli coefficients $Q^T_{\text{circ(Haar)}}(\alpha,\alpha)$ evolve under consecutive steps of the random circuit. As we already mentioned, we will make use of a Markov chain analysis to describe the random walk on Pauli matrices induced by the RQC.
Considering eq. (3.7), if we remove the identity state $\mathbb{1}\otimes\mathbb{1}$, it is then possible to reach any state of the chain, meaning that it is an *irreducible* chain. Moreover, the chain contains self loops, being hence *aperiodic*. From these two properties follows that the chain is also *ergodic*, thus converging to a unique stationary distribution (cfr. paragraph on Markov chains in Sec. 2.2). From Lemma 3.14, we know that this is the uniform stationary distribution of the Haar measure, $\omega(x) = \frac{1}{4^n-1}$. Furthermore, the chain is symmetric, and so is *reversible*. All those properties are needed to prove that the chain mixes within ε in $O(n \log \frac{n}{\varepsilon})$ steps.

Now, let us construct the *zero chain*, i.e., the projected Markov chain counting the *weights* – or, in other words, the number of non-zero elements – of the Pauli strings induced by the previous chain. This new chain obeys to the following transition matrix.

Lemma 5.4 (Lemma 5.2 in ref. [28])**.** *The zero chain has transition matrix P on state space*

$\Omega = \{1, 2, \dots\}$ *given by*

$$P(\ell, k) = \begin{cases} 1 - \frac{2\ell(3n-2\ell-1)}{5n(n-1)}\Delta t & k = \ell \\ \frac{2\ell(\ell-1)}{5n(n-1)}\Delta t & k = \ell - 1 \\ \frac{6\ell(n-\ell)}{5n(n-1)}\Delta t & k = \ell + 1 \\ 0 & \textit{otherwise} \end{cases} \tag{5.9}$$

for $1 \leq \ell, k \leq n$.

The proof of this lemma will be completely analogous to the one of the zero chain provided in the main result section for the jumps in the continuous-time stochastic process, so we address the interested reader to Lemma 7.23.

Since the number of possible Pauli strings with weight k is $3^k\binom{n}{k}$, it follows that the stationary distribution of this chain is (cfr. [28, Lemma 5.3])

$$\omega_0(k) = \frac{3^k\binom{n}{k}}{4^n - 1}. \tag{5.10}$$

This can also be proven by direct calculation.

Once again, we construct an additional chain, which we will call *accelerated chain*, i.e., the zero chain conditioned on moving, and consider the previous process as steps of the accelerated chain plus their respective waiting times, that we will prove to be bounded by $O(n \log n)$ for $\varepsilon = \text{poly}(n)$. This new chain has transition matrix (cfr. [28, Definition 5.7])

$$P_{\text{accel}}(\ell, k) = \begin{cases} 0 & k = \ell \\ \frac{\ell-1}{3n-2\ell-1} & k = \ell - 1 \\ \frac{3(n-\ell)}{3n-2\ell-1} & k = \ell + 1 \\ 0 & \text{otherwise.} \end{cases} \tag{5.11}$$

Note that the waiting time at site k is stochastically dominated by a geometric distribution with parameter $\frac{2k}{5n}$, since

$$1 - P(k, k) = \frac{2k(3n - 2k - 1)}{5n(n-1)} \geq \frac{2k}{5n}. \tag{5.12}$$

We divide now the state space Ω in three parts $\Omega_1, \Omega_2, \Omega_3$, partially overlapping, each with an entry space $E_j \subset \Omega_j$ and an exit condition T_j linked to the entry space of the next region. We consider the walk successful if, for each of these regions, it reaches the exit condition within $n \log n$ steps.

Region 1: $\Omega_1 = [1, n^\delta]$ for some $0 < \delta < 1/2$, where the entry condition is the whole space Ω_1 and the exit condition T_1 is satisfied when the walk reaches n^δ. In this phase the bias of the walk is considerable: the probability of increasing the weight at each step of the accelerated chain is very high within this region of Ω, namely, $1 - n^{2\delta-1}$. Conditioning on this event, that we shall denote "Forward", we can bound the waiting time W_1 for a walk starting everywhere in Ω_1 and ending at n^δ using [28, Lemma A.7]. Formally, we can consider a number of steps of order higher than $\log n$, and get

$$\mathbb{P}\left(W_1 > \frac{15}{2}\delta n \log n \mid \text{Forward}\right) \leq 2n^{-\delta}. \tag{5.13}$$

Hence, the probability for the first phase of the walk to complete successfully is close to 1, since

$$\begin{aligned}
&\mathbb{P}\left(\text{the walk in region 1 is successful}\right) \leq \mathbb{P}\left(\text{Forward} \cap W_1 \leq \frac{15}{2}\delta n \log n\right) \\
&\leq \mathbb{P}\left(\text{Forward}\right) \cdot \mathbb{P}\left(W_1 \leq \frac{15}{2}\delta n \log n \mid \text{Forward}\right) && (5.14)\\
&\leq (1-n^{2\delta-1})(1-2n^{-\delta}) && (5.15)\\
&\leq 1-n^{2\delta-1}-2n^{-\delta}. && (5.16)
\end{aligned}$$

Region 2: $\Omega_2 = [n^\delta/2, \theta n]$ for some constant $0 < \theta < 3/4$ with $E_2 = [n^\delta, \theta n]$ and $T_2 = \theta n$. In this region, the probability of moving forward is lower bounded by

$$f(k) = \frac{3(n-k)}{3n-2k-1} \geq \frac{3(1-\theta)}{3-2\theta} =: f. \tag{5.17}$$

Let us define the constants $\mu = 2f - 1$ and $\tilde{\mu} = \mu/\gamma$ for some $\gamma > 2$. Then, using the Chernoff-Hoeffding bound given in Lemma 2.2, one can show that the walk after t accelerated steps will have passed the site $s = \tilde{\mu}t$. Let X_t be the position of the walk at accelerated step t where $X_0 = n^\delta$, so we have [28, Lemma A.10]

$$\mathbb{P}\left(X_t \leq \theta n\right) \leq \exp\{-2/3\ \mu\, \theta n\}. \tag{5.18}$$

If we assume that the walk never goes back beyond $n^\delta/2$, i.e., never leaves Ω_2 from its lower boundary, we bound the probability that after $t = \theta n/\tilde{\mu}$ accelerated steps the waiting time W_2 is "too large" by (cfr. [28, Lemma A.11])

$$\mathbb{P}\left(W_2 \geq \frac{15n\log \theta n}{\mu}\right) \leq \left(\frac{4}{\theta n}\right)^{\frac{3}{2\mu}} + \frac{2\exp\left\{\frac{-\mu n^\delta}{4}\right\}}{1-\exp\{-\mu/2\}}. \tag{5.19}$$

To bound the probability of the event that the walk moves back to weights lower than $n^\delta/2$, one makes use of the *Gambler's ruin* (see the Preliminaries Chapter, Lemma 2.1) and gets

$$\mathbb{P}\left(\text{The walk moves back to } n^\delta/2 - 1\right) \leq \left(\frac{p}{1-p}\right)^{n^\delta/2}. \tag{5.20}$$

Summarizing, the walk in region 2 is successful if, given that the walk never leaves Ω_2, the accelerated chain reaches θn within $t = \theta n/\tilde{\mu}$ steps and the waiting time is not longer than $O(n \log n)$, or more formally

$$\begin{aligned}
&\mathbb{P}\left(\text{the walk in region 2 is successfull}\right) \leq \\
&1 - \left(\frac{4}{\theta n}\right)^{\frac{3}{2\mu}} - \frac{2\exp\left\{\frac{-\mu n^\delta}{4}\right\}}{1-\exp\{-\mu/2\}} - \left(\frac{p}{1-p}\right)^{n^\delta/2}.
\end{aligned} \tag{5.21}$$

Region 3: $\Omega_3 = [\theta n/2, n]$ with $E_3 = [\theta n, n]$. The exit condition is satisfied when the restricted chain has reached its stationary distribution with probability $1-\varepsilon$. As Harrow and Low showed, this happens in $O(n \log \frac{n}{\varepsilon})$ steps.
The transition matrix on Ω_3 is given by

$$P'(\ell,k) = \begin{cases} 0 & \text{if } \ell < \theta n/2 \text{ or } k < \theta n/2 \\ 1 - P(\theta n/2, \theta n/2 + 1) & \text{if } \ell = k = \theta n/2 \\ P(\ell,k) & \text{otherwise.} \end{cases} \tag{5.22}$$

The stationary distribution of P' is then

$$\omega'(k) = \begin{cases} \omega_0(k)/(1-\eta) & \theta n/2 \leq k \leq n \\ 0 & \text{otherwise,} \end{cases} \tag{5.23}$$

where $\eta = \sum_{k=1}^{\theta n/2-1} \omega_0(k)$, and satisfies

$$\sum_{k=\theta n/2}^{n} P'(\ell,k)\omega_0'(\ell) = \omega_0'(k). \tag{5.24}$$

The convergence time of the chain can be retrieved with the following argument based on *log-Sobolev constants*. We are not going to introduce this object in an extensive manner, we simply write its mathematical definition. Given a transition matrix P and its stationary distribution ω (cfr. [28, Definition 4.8], the corresponding log-Sobolev constant is

$$\rho := \min_g \frac{\sum_{k=\neq\ell}(g(\ell)-g(k))^2 P(\ell,k)\omega(k)}{\sum_\ell \omega(\ell)g(\ell)^2 \log \frac{g(\ell^2}{\sum_k \omega(k)g(k)^2}}, \tag{5.25}$$

where the minimum runs over all functions g assigning to elements of Ω a value in $\mathbb{R}$. The mixing time of a finite, reversible and irreducible Markov chain is then given by [94, Lemma 3.7]

$$\tau(\varepsilon) = O\left(\frac{1}{\rho}\log\log\frac{1}{\omega_*} + \frac{1}{\lambda}\log\frac{n}{\varepsilon}\right), \tag{5.26}$$

where ω_* is the smallest value of the stationary distribution and λ the spectral gap. Imagine R_j to be the chain that uniformly mixes the site j, which converges in one step and whose log-Sobolev constant ρ_j is independent with respect to n. Let Q be the chain that chooses randomly a site uniformly mixes it, i.e., Q is the product constructed from the R_j chains. Due to the next lemma, we know then that the chain Q has a gap $1/n$ and $\rho_Q = \rho_j/n$.

Lemma 5.5. *[94, Lemma 3.2] Let $R_j,\ j \in \{1,2,\ldots,n\}$ be Markov chains with gaps λ_j and log-Sobolev constants ρ_j. The product chain Q, whose state space is equal to the product of the spaces for the chains R_j, is constructed by applying randomly at each step one of the chains R_j. The spectral gap of the chain Q is then given by*

$$\lambda = \frac{1}{n}\min_j \lambda_j \tag{5.27}$$

and its log-Sobolev constant by

$$\rho = \frac{1}{n}\min_j \rho_j. \tag{5.28}$$

Now, note that the log-Sobolev constant of the induced zero chain Q_0 counting the number of non-identity elements can only be larger than ρ_Q, since the minimum is taken over a smaller number of functions over the state space. The transition matrix for this chain is given by

$$Q_0(\ell,k) = \begin{cases} \frac{n+2\ell}{4n} & \text{if } k = \ell \\ \frac{\ell}{4n} & \text{if } k = \ell-1 \\ \frac{3(n-\ell)}{4n} & \text{if } k = \ell+1 \\ 0 & \text{otherwise,} \end{cases} \tag{5.29}$$

and its stationary distribution is the same as the one of the P chain. The construction of the restricted chain Q_0' over the region Ω_3 is completely analogous to P', and their stationary distributions are the same. All these observations allow us to say using [28, Theorem 4.6]

$$\rho_{P'} = \rho_{Q_0'} / A, \tag{5.30}$$

with

$$A = \max_{k \geq \theta n/2} \frac{Q_0'(k, k+1)}{P'(a, a+1)} = \max_{k \geq \theta n/2} \frac{5(n-1)}{8k} \leq \frac{5}{4\theta}. \tag{5.31}$$

Hence, we finally obtain a bound for the log-Sobolev constant of P', namely, $\rho_{P'} \geq \frac{4\theta\rho_j}{5n}$, and with the same argument an equivalent bound for the spectral gap, being $\Omega(1/n)$. Using eq. (5.26), we conclude that the mixing time towards the stationary distribution of the restricted chain is $O(n \log \frac{n}{\varepsilon})$. There is still to consider the case where the walk moves back behind weight $\theta n/2$, but the probability of this event can be bounded again by Gambler's ruin to be

$$\mathbb{P}(\text{walk moves back to } \theta n/2) \leq \left(\frac{\theta}{3(2-\theta)}\right)^{\theta n/2}. \tag{5.32}$$

Summarizing, we showed that in each of the three regions the walks on the zero chian P ends successfully with failing probability poly(n) and combining together these results we conclude that the chain mixes within ε towards the stationary distribution in a total amount of steps being $\tau_0 = O\left(n \log \frac{n}{\varepsilon}\right)$.

The next thing to do is to show that, once the zero chain has mixed, the full chain over Pauli strings mixes in a number of steps of equivalent magnitude. Let us consider the collection of sets of Pauli strings, where each set contains all Pauli strings with the same weight. Once the zero chain has converged, we know that all sets are close to their "correct" probability $\frac{3^k \binom{n}{k}}{4^n - 1}$, where k is the weight characterizing the set. However, since the zero chain does not distinguish among strings with the same weight having different positioning of identity elements or different labeling of non-identity elements, we cannot say whether all strings within the same set are uniformly distributed, as required for convergence of the full chain; an additional application of a random permutation together with a cycling of all $\sigma_1, \sigma_2, \sigma_3$ Pauli matrices will ensure the desired result. The cycling is obtained by touching all sites at least once by the full chain, since from (3.7) we see that it uniformly shuffles non-identity elements. By coupon collector's argument (see Preliminaries Chapter and eq. (2.74)), all sites have been hit with probability $1 - \varepsilon$ after $O\left(n \log \frac{n}{\varepsilon}\right)$ steps. This argument provided by Harrow and Low in Theorem 5.6 however is flawed – as pointed out by the subsequent comments of Diniz and Jonathan [95] – when stating that it also ensures that strings with the same elements but different positioning will have the same distribution. These authors presented an alternative proof relying on repeated random transposition chains. As we mentioned at the end of Section 2.2, in a number of articles by Diaconis and his collaborators [38, 39, 96] this chain has been proven to converge toward a random permutation within ε in $O\left(n \log \frac{n}{\varepsilon}\right)$ steps. The main question is then how we can apply those transpositions on the Pauli elements; while this is known to be possible with an additional circuit implementing for instance the Durstenfeld-Knuth shuffle algorithm [97], Diniz and Jonathan showed that this is done by the full chain itself. Since in the following the distinction between Pauli matrices is irrelevant, we construct a new chain C on the state space given by the vertices of an hypercube, $\Omega_C = \{0, 1\}^n$. Consider the set of Pauli strings σ_α represented by $\vec{q} \in \Omega_C$ linked by the bijection $\vec{q} \leftrightarrow \{\alpha \in \{0, 1, 2, 3\}^n : \alpha_j = 0 \Leftrightarrow q_j = 0\} =: S_{\vec{q}}$, then construct the new

chain as a projection

$$C(\vec{p},\vec{q}) := \sum_{\alpha\in S_{\vec{q}}} \frac{1}{4^n} \operatorname{Tr}\left[(\sigma_\alpha\otimes\sigma_\alpha)\ M^{k=2}_{\text{circ(Haar)}}(\sigma_\beta\otimes\sigma_\beta)\right] \qquad \forall\beta\in S_{\vec{p}}. \tag{5.33}$$

with stationary distribution on $\Omega_C/\{0\}^n$ given by

$$\omega_C(\vec{q}) = \frac{3^{|\vec{q}|}}{4^n-1}, \tag{5.34}$$

where $|\vec{q}|$ is the number of 1's in $\vec{q}$. This simpler chain, according to [95, Lemma 1], has the same mixing time of the full chain.
Now, let A_π be the representation of the permutation $\pi\in S_n$ on the space of probability distributions over Ω_C, i.e., given a distribution ϕ,

$$[A_\pi\phi](\vec{q}) = \phi(\pi(\vec{q})). \tag{5.35}$$

Then the uniform distribution of the repeated random transposition chain is given by $S = \frac{1}{n!}\sum_{\pi\in S_n} A_\pi$ and the distance between any uniformly shuffled distribution on Ω_C and the stationary distribution ω_C is equivalent to the one between ϕ_0, i.e., the probability distribution of the zero chain obtained by the projection

$$\phi_0(k) = \sum_{\vec{q}\in\Omega_C:|\vec{q}|=k} \phi(\vec{q}) \qquad \forall k\in\Omega, \tag{5.36}$$

and ω_0. Formally [95, Lemma 2]

$$\|S\phi-\omega_C\|_{TV} = \|\phi_0-\omega_0\|_{TV}. \tag{5.37}$$

As mentioned, the process itself performs random transpositions between identity and non-identity elements. Looking at the projected chain C, we note that this can be written as

$$C = \frac{1}{5}T' + \frac{4}{5}M', \tag{5.38}$$

where $T' = \frac{1}{n(n-1)}T'_{j,k}$ and $M' = \frac{1}{n(n-1)}M'_{j,k}$ are again chains given by 2-local elements whose transition probabilities are collected in the following tables

$T'_{j,k}$	00	01	10	11
00	1	0	0	0
01	0	0	1	0
10	0	1	0	0
11	0	0	0	1

$M'_{j,k}$	00	01	10	11
00	1	0	0	0
01	0	1/4	0	3/4
10	0	0	1/4	3/4
11	0	1/4	1/4	1/2

i.e., C can be seen as a combination of two Markov chains. The chain T' is similar to the random transposition chain $\mathcal{T}$ (cfr. eq. (2.72)) on the state space Ω_C; however, it lacks the identity component: and even (odd) number of steps will always result in an even (odd) permutation of the element $\vec{q}\in\Omega_C$. This means that T' is a non-convergent, periodic chain. To overcome the last issue, we rewrite C as

$$C = \frac{1}{5}T + \frac{4}{5}M, \tag{5.39}$$

where we slightly modified the previous chains in order to incorporate the identity, i.e., $T = \frac{1}{n}\mathcal{I} + \frac{n-1}{n}T'$ and $M = M' + \frac{1}{4n}[T'-\mathcal{I}]$. Although the modified chain T is still reducible, it does converge to the random permutation operator S over Ω_C, namely,

Lemma 5.6 ([95], Lemma 3)**.** *Each initial distribution ν on Ω_C converges under the chain T to its shuffled version $S\nu$. Additionally, the mixing time*

$$\tau_T(\varepsilon) = \max_{\nu} [\min\ z : \|T^z\nu - S\nu\|_{TV}] \leq \varepsilon \tag{5.40}$$

is $O(n \log \frac{n}{\varepsilon})$.

Considering eq. (5.39), we can expect that at each step of the chain C with probability $1/5$ a random transposition is applied, and in the complementary case the chain M. We should hence wait until the required number of transpositions has been applied, shuffling the distribution over Ω_C. In addition, note that the chain M', and so M, is symmetric with respect to permutations of the sites, meaning that it commutes with T. We can therefore imagine that, after a number of applications of the full chain, all steps of the M chain act at first, and then the repeated transpositions will take place.
We have now all the tools that we need to prove the convergence of the full chain. Let us first wait until the zero chain has converged within ε, i.e., the distribution ϕ_0 on Ω satisfies $\|\phi_0 - \omega_0\|_{TV} \leq \varepsilon$. Then according to eq. (5.37) the corresponding state ϕ of the chain C lies in the neighborhood around the stationary distribution ω_C with radious ε. Define a mixing time for C restricted on states in this neighborhood as

$$\tau_C(\varepsilon', \varepsilon) = \max_{\nu \in B_\varepsilon(\omega_C)} [\min\ z : \|C^z\nu - \omega_C\|_{TV}] \leq \varepsilon' \tag{5.41}$$

and use the following result

Lemma 5.7 ([95], Lemma 4)**.** *For all $\varepsilon' > 0$ and all $\varepsilon \geq 0$ and any $\delta > 0$ satisfying $\varepsilon' > \mathrm{e}^{-2\delta^2} + \varepsilon$*

$$\sqrt{\tau_C(\varepsilon', \varepsilon)} < \frac{5}{2}\left[\delta + \sqrt{\delta^2 + \frac{4}{5}\tau_T\left(\varepsilon' - \varepsilon - \mathrm{e}^{-2\delta^2}\right)}\right], \tag{5.42}$$

Choosing $\varepsilon = \varepsilon'/2$ and $\delta = \sqrt{\frac{1}{2}\log(4/\varepsilon')}$ such that

$$\tau_C(\varepsilon', \varepsilon'/2) < \frac{25}{4}\left[\frac{1}{2}\log(4/\varepsilon') + \frac{4}{5}\tau_T(\varepsilon'/4)\right]. \tag{5.43}$$

Since, as we already mentioned, $\tau_T(\varepsilon) = O(n \log \frac{n}{\varepsilon})$, follows $\tau_C(\varepsilon', \varepsilon'/2) = O(n \log \frac{n}{\varepsilon'})$. Summarizing, the mixing of the C chain within ε-approximation, which has the same dynamics as the full chain according to [95, Lemma 1], is upper bounded by the mixing of the zero chain within $\varepsilon/2$ plus the mixing of the C chain within ε starting from a state lying in the neighborhood of ω_C with radius $\varepsilon/2$, i.e.,

$$\tau_C(\varepsilon) \leq \tau_0(\varepsilon/2) + \tau_C(\varepsilon, \varepsilon/2). \tag{5.44}$$

This concludes the proof for the mixing conditions in Lemma 5.2, in both norm criteria, since the 2-norm mixing time is deduced from the one with 1-norm (cfr. [28, Theorem 5.5]). From this we can finally imply that (see [28, Section 6])

$$\|M^{k=2}_{\text{circ(Haar, } O(n \log \frac{n}{\varepsilon}) \text{ steps)}} - M^{k=2}_{\text{Haar}}\|^2_\diamond \leq 2^{4n}\varepsilon^2 \tag{5.45}$$

and hence prove Theorem 5.1.

5.2 Approximate unitary designs from Hamiltonian gaps

Using a different approach, this first result on mixing properties of random circuits has been enhanced by the work of Brandão, Harrow and Horodecki in ref. [27]. These authors showed that, for any k, a single step of random circuit with 2-local gates from the Haar measure on d-level subsystems is a (λ, k)- tensor product expander with λ given by

Theorem 5.8 (cfr. [27], Theorem 5)**.**

$$\lambda = \|M^k_{\text{circ(Haar)}} - M^k_{\text{Haar}}\|_\infty \leq 1 - \left(425 n \lceil \log_d(4k) \rceil^2 d^2 k^5 k^{3.1/\log(d)}\right)^{-1} \tag{5.46}$$

Using the fact that ℓ steps of the circuit corresponds to an ℓ-convolution of the probability distribution of one single step, and hence

$$M^k_{\mu^{\star \ell}} = \left(M^k_\mu\right)^\ell . \tag{5.47}$$

together with Lemma 3.12 follows that

Theorem 5.9 (cfr. [27], Corollary 6, under a different definition for approximate unitary designs)**.** *Random quantum circuits with 2-local gates drawn from the Haar measure are ε-approximate unitary k-designs after*

$$T \geq 425\, n \lceil \log_d(4k) \rceil^2 d^2 k^5 k^{3.1/\log(d)} (nk \log(d) + \log(1/\varepsilon)) \tag{5.48}$$

steps.

It is important to mention that this result applies on a slightly different circuit with respect to the one considered by Harrow and Low: in this case, the 2-qudit gates act always act on adjacent sites, and not on arbitrary pairs of qudits.

Additionally, this result can be extended to a more general class of random circuits with a gate set in $\mathbb{SU}(d^2)$ with *algebraic entries* where, for each gate, its inverse is also contained in the set. Bourgain and Gamburd have shown in ref. [98] that a distribution uniformly picking at random gates from such a set is a tensor-product expander with non-vanishing gap. One can then show that

Corollary 5.10 (cfr. [27], Corollary 7, under a different definition for approximate unitary designs)**.** *Let $d \geq 2$, $G = \{g\}_{j=1}^m \subset S\mathbb{U}(d^2)$ be a universal gate set containing inverses and where each gate is composed of algebraic entries. Then there exists $C(G)$ such that a random circuit drawing gates uniformly at random from G is an ε-approximate unitary k-design after*

$$T \geq C(G)\, n \lceil \log_d(4k) \rceil^2 d^2 k^5 k^{3.1/\log(d)} (nk \log(d) + \log(1/\varepsilon)) \tag{5.49}$$

steps.

The proof is structured in four steps, that we are going to outline in the following.

Relating to spectral gap We consider an Hamiltonian $H_n = \sum_{j=1}^n h_{j,j+1}$, where $h_{j,j+1}$ acts non-trivially on 2 qudits of the system. The *spectral gap*, $\Delta(H_n)$, is defined by the the difference between the second lowest and the lowest eigenvalues of H.
The Hamiltonian terms $h^k_{j,j+1}$ acting on qudits $j, j+1$ are the orthogonal complement of the local k-th moment operators, i.e

$$h^k_{j,j+1} = \mathbb{1} - \mathbb{E}_{\text{Haar}} U^{\otimes k,k}_{j,j+1}, \tag{5.50}$$

where we used the vectorization isomorphism for the moment operator, as introduced in Section 2.1, and subsequently we write $H_n^k = \sum_{j=1}^n h_{j,j+1}^k$.

The gap of the Hamiltonian is related with the tensor product expander of the kth-moment of the random circuit distribution by (see [27, Lemma 16])

$$\|M_{\text{circ(Haar)}}^k - M_{\text{Haar}}^k\|_\infty = 1 - \frac{\Delta(H_n^k)}{n}. \tag{5.51}$$

It is therefore sufficent to lower bound the spectral gap of the Hamiltonian to obtain the mixing time for the unitary design property.

Structure of H_n^k We should now consider the following two traits on eigenspaces of the kernel of H_n^k and their projectors in order to proceed with the third step.

(i) The minimum eigenvalue of H_n^k is zero and the corresponding eigenspace is given by

$$\mathcal{D}^k := \text{span}\left\{ \ |\psi_{\pi,d}\rangle^{\otimes n} : \ |\psi_{\pi,d}\rangle := (\mathbb{1} \otimes V_d(\pi)) \ |\Psi_d\rangle, \pi \in S_k \right\}, \tag{5.52}$$

where Ψ is the maximally entangled state on $(\mathbb{C}^d)^{\otimes 2k}$, S_k is the symmetric group of order k and $V_d(\pi)$ the representation of the permutation $\pi \in S_k$ interchanging copies of $\mathbb{C}^d$ as

$$V_d(\pi)\,|v_1\rangle \otimes \cdots \otimes |v_k\rangle = |v_{\pi^{-1}(v_1)}\rangle \otimes \ldots |v_{\pi^{-1}(v_k)}\rangle$$

(ii) Let D^k be the projector onto $\mathcal{D}^k$. If $k^2 \leq d^n$, then

$$\sum_{\pi \in S_k} |\langle \psi_{\sigma,d} | \psi_{\pi,d} \rangle| \leq 1 + \frac{k^2}{d^n} \qquad \forall \sigma \in S_k \tag{5.53}$$

and

$$\| \sum_{\pi \in S_k} |\psi_{\pi,d}\rangle \langle \psi_{\pi,d}|^{\otimes n} - D^k \|_\infty \leq \frac{k^2}{d^n}. \tag{5.54}$$

Hence, the vectors $|\psi_{\pi,d}\rangle$ spanning the zero space of the Hamiltonian are "almost" pairwise orthogonal.

Lower bounding the spectral gap Using the above properties and Nachtergaele's result [99] for lower bounding the gap of *frustration free* Hamiltonians whose ground space is spanned by matrix product states, Brandão, Harrow and Horodecki showed

Lemma 5.11 (cfr. [27], Lemma 18)**.** *For every integers n and k such that $n \geq \lceil 2.5 \log_d(4k) \rceil$,*

$$\Delta(H_n^k) \geq \frac{\Delta(H_{\lceil 2.5 \log_d(4k) \rceil}^k)}{4 \lceil 2.5 \log_d(4k) \rceil}. \tag{5.55}$$

Bounding convergence using path coupling From the previous steps, we know that lower bounding $H_{\lceil 2.5 \log_d(4k) \rceil}^k$ is everything one needs to characterize the circuit distribution as a tensor product expander. In order to do this, the authors revert back to the random circuit description and bound the tensor product expander in terms of the *Wasserstein distance*; for two probability measures ν_1, ν_2 on $\mathbb{U}(N)$, this is defined as

$$W(\nu_1, \nu_2) := \sup \left\{ \mathbb{E}_{\nu_1} f(U) - \mathbb{E}_{\nu_2} f(U) \quad | \quad f : \mathbb{U}(N) \to \mathbb{R} \quad 1 \text{ - Lipschitz continuous} \right\}. \tag{5.56}$$

The distrance between the Wasserstein distance of a measure ν on $\mathbb{U}(d^n)$ and the Haar measure is an upper bound of the tensor product expander of ν, i.e.,

Lemma 5.12 (cfr. [27], Lemma 20). *For every k and measure ν on $\mathbb{U}(d^n)$ it stands*

$$\|M_\nu^k - M_{\text{Haar}}^k\|_\infty \leq 2k\, W(\nu, \text{Haar}). \tag{5.57}$$

Making use of the *path coupling technique* developed by Oliveira in ref. [100] from the previous method proposed by Bubler and Dyer [101], one can prove the following bound on the Wasserstein distance of a random quantum circuit and the Haar measure.

Lemma 5.13 (cfr. [27], Lemma 19).

$$W\left((\text{circ}(\text{Haar}))^{\star(n-1)\ell}, \text{Haar}\right) \leq \left(1 - \frac{1}{\mathrm{e}^n (d^2+1)^{n-2}}\right)^{\frac{\ell}{n-1}} \sqrt{2} d^{n/2}. \tag{5.58}$$

We have now all necessary tools to prove the main theorem.

Proof of Theorem 5.8. From Lemmas 5.12 and 5.13, eqs. (5.51) and (5.47) we get for every m, k, ℓ

$$1 - \frac{\Delta(H_m^k)}{m} \leq \left(2^{3/2} k d^{m/2}\right)^{\frac{1}{\ell(m-1)}} \left(1 - \frac{1}{\mathrm{e}^m (d^2+1)^{m-2}}\right)^{\frac{1}{(m-1)^2}}. \tag{5.59}$$

In the limit $\ell \to \infty$, we obtain

$$\Delta(H_m^k) \geq m^{-1} \mathrm{e}^{-m} (d^2+1)^{-m}. \tag{5.60}$$

Now, considering Lemma 5.11 and using the last equation setting $m = \lceil 2.5 \log_d(4k) \rceil$, we lower bound the spectral gap for arbirtrary n as

$$\begin{aligned}
\Delta(H_n^k) &\geq \frac{1}{4(\lceil 2.5 \log_d(4k) \rceil)^2} \left(\mathrm{e}(d^2+1)\right)^{-2.5 \log_d(4k) - 1} \\
&\geq \frac{1}{4(\lceil 2.5 \log_d(4k) \rceil)^2 \mathrm{e}(d^2+1)} \mathrm{e}^{-2.5(1+\log(d^2+1)) \frac{\log 4k}{\log d}} \\
&\geq \frac{1}{4(\lceil 2.5 \log_d(4k) \rceil)^2 \mathrm{e}(d^2+1)} (4k)^{-2.5(1+\log(1+d^{-2}) + \log(d^2))/\log d} \\
&\geq \frac{(4k)^{-5-3.1/\log d}}{13.6\, d^2 (\lceil 2.5 \log_d(4k) \rceil)^2},
\end{aligned} \tag{5.61}$$

where in the last inequality we used $\mathrm{e}(1+d^{-2}) \leq 3.4$ and $2.5(1+\log(1+d^{-2})) \leq 3.1$ for $d \geq 2$. Under an further application of Lemma 5.51, we finally obtain Theorem 5.8 (this author cannot confirm the constant 425, which is however irrelevant for scaling behavior). □

5.3 Decoupling with random quantum circuits

As we have seen in Section 4.1, a system affected by a random unitary evolution decouples from the environment whenever the unitary is drawn from a 2-design. A beautiful result from Brown and Fawzi [24] shows that decoupling is also achieved with a RQC in almost linear time with respect to the system size and hence is obtained faster than the unitary 2-design property. Formally, this result is expressed as

Theorem 5.14. *Let $\rho_{AE} \in \mathcal{S}(AE)$ be an initial arbitrary mixed state and $U_t \rho_{AE} U_t^\dagger$ be the corresponding state after the application of t random two-qubit gates on the system A, which is composed of n qubits. Let $\mathcal{T} : \mathcal{S}(A) \to \mathcal{S}(B)$ be a completely positive trace preserving map and let $\tau_{A'B} = I_{A'} \otimes \mathcal{T}_{A \to B}(\Psi_{A'A})$ be the Choi-Jamiolwowski representation of $\mathcal{T}$. Then for any $\delta > 0$ there exists a constant c such that for all n and all $t \geq c\, n \log^2 n$*

$$\mathbb{E}_{\text{circ}(\text{Haar}, t)} \|\mathcal{T}(U_t \rho_{AE} U_t^\dagger) - \tau_B \otimes \rho_E\|_1 \leq \left(\frac{1}{\text{poly}(n)} + 16^{\delta n} \cdot 2^{-H_2(A|B)_\tau - H_2(A|E)_\rho}\right)^{1/2}. \tag{5.62}$$

The following calculations lead to the above theorem but are also fundamental to formalize our novel results on continuous-time evolutions over the unitary group as an analogous decoupling theorem. The proof makes use of a condition on Pauli strings distribution which is weaker than approximate unitary 2-design, from which follows an improvement on the scaling.

Proof of Theorem 5.14. Using an Hölder-type inequality for operators, we write (dropping the subscript $\mathrm{circ}(\mathrm{Haar}, t)$ for the expectation)

$$\begin{aligned}\mathbb{E}\|\mathcal{T}(U_t\rho_{AE}U_t^\dagger) - \tau_B\otimes\rho_E\|_1^2 &\leq \mathbb{E}\,\mathrm{Tr}[\widetilde{\mathcal{T}}(U_t\tilde{\rho}_{AE}U_t^\dagger)^2] - 2\mathbb{E}\,\mathrm{Tr}[\widetilde{\mathcal{T}}(U_t\tilde{\rho}_{AE}U_t^\dagger)\cdot\tilde{\tau}_B\otimes\tilde{\rho}_E] \\ &\quad + \mathrm{Tr}[(\tilde{\tau}_B\otimes\tilde{\rho}_E)^2] \\ &\leq \mathbb{E}\,\mathrm{Tr}[\widetilde{\mathcal{T}}(U_t\tilde{\rho}_{AE}U_t^\dagger)^2] - \mathrm{Tr}[\tilde{\tau}_B^2]\,\mathrm{Tr}[\tilde{\rho}_E^2] + \frac{1}{\mathrm{poly}(n)},\end{aligned} \tag{5.63}$$

where $\tilde{\rho}_{AE} = \rho_E^{-1/4}\rho_{AE}\rho_E^{-1/4}$ and $\widetilde{\mathcal{T}} = \tau_B^{-1/4}\mathcal{T}\tau_B^{-1/4}$. In the following, we omit the $1/\mathrm{poly}(n)$ term to help readability. One can rewrite the first term on the last line of eq. (5.63) using a Pauli basis projection,

$$\widetilde{\mathcal{T}}(U_t\tilde{\rho}_{AE}U_t^\dagger) = \frac{1}{4^n}\sum_{\nu,\xi\in\{0,1,2,3\}^n}\mathrm{Tr}[\sigma_\xi\widetilde{T}(\sigma_\nu)]\sigma_\xi\otimes\mathrm{Tr}_A[\sigma_\nu U_t\tilde{\rho}_{AE}U_t^\dagger], \tag{5.64}$$

as

$$\begin{aligned}\mathrm{Tr}[\widetilde{\mathcal{T}}(U_t\tilde{\rho}_{AE}U_t^\dagger)^2] &= \frac{1}{2^n}\sum_{\xi\in\{0,1,2,3\}^n}\frac{1}{4^n}\mathrm{Tr}[\sigma_\xi\widetilde{T}(\mathbb{1}_A)]^2\,\mathrm{Tr}[\tilde{\rho}_E^2] \\ &\quad + \frac{1}{8^n}\sum_{\substack{\xi,\nu,\nu'\in\{0,1,2,3\}^n \\ /\{\nu,\nu'\}=\{0,0\}}}\mathrm{Tr}[\sigma_\xi\widetilde{T}(\sigma_\nu)]\,\mathrm{Tr}[\sigma_\xi\widetilde{T}(\sigma'_\nu)]\;\mathrm{Tr}_A[\sigma_\nu U_t\tilde{\rho}_{AE}U_t^\dagger]\,\mathrm{Tr}_A[\sigma_{\nu'}U_t\tilde{\rho}_{AE}U_t^\dagger] \\ &= \mathrm{Tr}[\tilde{\tau}_B^2]\,\mathrm{Tr}[\tilde{\rho}_E^2] + \frac{1}{8^n}\sum_{\substack{\nu,\nu'\in\{0,1,2,3\}^n \\ /\{\nu,\nu'\}=\{0,0\}}}T_{\nu,\nu'}\;\mathrm{Tr}_A[\sigma_\nu U_t\tilde{\rho}_{AE}U_t^\dagger]\,\mathrm{Tr}_A[\sigma_{\nu'}U_t\tilde{\rho}_{AE}U_t^\dagger],\end{aligned} \tag{5.65}$$

where one defines $T_{\nu,\nu'} \coloneqq \sum_\xi \mathrm{Tr}[\sigma_\xi\widetilde{T}(\sigma_\nu)]\,\mathrm{Tr}[\sigma_\xi\widetilde{T}(\sigma'_\nu)]$. The square root of the LHS of eq. (5.63) can hence be bounded by

$$\mathbb{E}\|\mathcal{T}(U_t\rho_{AE}U_t^\dagger) - \tau_B\otimes\rho_E\|_1 \leq \left(\mathbb{E}\,\frac{1}{8^n}\sum_{\substack{\nu,\nu'\in\{0,1,2,3\}^n \\ \{\nu,\nu'\}\neq\{0,0\}}}T_{\nu,\nu'}\;\mathrm{Tr}_A[\sigma_\nu U_t\tilde{\rho}_{AE}U_t^\dagger]\,\mathrm{Tr}_A[\sigma_{\nu'}U_t\tilde{\rho}_{AE}U_t^\dagger]\right)^{1/2} \tag{5.66}$$

using Jensen's inequality in addition to the previous calculations. In the next step, we are going to connect this expression with the second moment operator using the swap trick given in eq. (4.10), in a similar fashion as in the proof of Theorem 4.1. For any ν and ν',

$$\begin{aligned}&\mathbb{E}\,\mathrm{Tr}\left[\mathrm{Tr}_A[\sigma_\nu U_t\tilde{\rho}_{AE}]\,\mathrm{Tr}_A[\sigma_{\nu'}U_t\tilde{\rho}_{AE}]\right] = \mathbb{E}\,\mathrm{Tr}\left[\mathrm{Tr}_A[\sigma_\nu U_t\tilde{\rho}_{AE}]\otimes\mathrm{Tr}_{A'}[\sigma_{\nu'}U_t\tilde{\rho}_{A'E'}]\mathbb{F}_{EE'}\right] \\ &= \mathrm{Tr}\left[\mathrm{Tr}_{AA'}[(\sigma_\nu\otimes\sigma_{\nu'})(M^2_{\mathrm{circ}(\mathrm{Haar},t)}\otimes\mathcal{I}_{EE'})(\tilde{\rho}_{AE}\otimes\tilde{\rho}_{A'E'})\mathbb{F}_{EE'}\right] \\ &= \frac{1}{4^n}\sum_{\mu,\mu'\in\{0,1,2,3\}^n}\mathrm{Tr}[(\sigma_\nu\otimes\sigma_{\nu'})M^2_{\mathrm{circ}(\mathrm{Haar},t)}(\sigma_\mu\otimes\sigma_{\mu'})\otimes\mathrm{Tr}_A[\sigma_\mu\tilde{\rho}_{AE}]\otimes\mathrm{Tr}_{A'}[\sigma_{\mu'}\tilde{\rho}_{A'E'}]],\end{aligned} \tag{5.67}$$

where we used again a projection expansion on Pauli basis.
We now recall that $M^{k=2}_{\mathrm{circ(Haar},t)}(\sigma_\mu \otimes \sigma_{\mu'}) = 0$ when $\mu \neq \mu'$ because of eq. (3.7), and we denote the Pauli coefficient of σ_ν after t steps of the random circuit applied on Pauli string σ_μ with

$$Q^t_{\mathrm{circ}}(\mu,\nu) \coloneqq \frac{1}{4^n} \operatorname{Tr}\left[\sigma_\nu \otimes \sigma_\nu \, M^{k=2}_{\mathrm{circ(Haar},t)}(\sigma_\mu \otimes \sigma_\mu)\right]. \tag{5.68}$$

The term under the square root of the RHS of eq. (5.66) can then be re-formulated as

$$\frac{1}{4^n} \sum_{\mu \in \{0,1,2,3\}^n, \mu \neq 0} \operatorname{Tr}[\operatorname{Tr}_A[\sigma_\mu \tilde{\rho}_{AE}]^2] \sum_{\nu \in \{0,1,2,3\}^n, \nu \neq 0} \operatorname{Tr}[\widetilde{T}(\sigma_\nu)^2] Q^t_{\mathrm{circ}}(\mu,\nu). \tag{5.69}$$

One applies now the main technical result of the paper, namely,

Theorem 5.15 (Theorem 4.1 in ref. [24])**.** *For any constants $\delta \in (0, 1/16)$ and $\eta \in (0,1)$ there exists a constant c such that for any $t \geq c\, n \log^2 n$ and all Pauli strings σ_μ of weight ℓ and large enough n*

$$\sum_{\nu \in \{0,1,2,3\}^n, \nu \neq 0} \left| Q^t_{\mathrm{circ}}(\mu,\nu) n - p_\delta(\nu) \right| \leq \frac{1}{(3-\eta)^\ell \binom{n}{\ell}} \frac{1}{\mathrm{poly}(n)} \quad \forall \nu, \tag{5.70}$$

where p_δ is a (possibly subnormalized) distribution such that

$$p_\delta(\nu) \leq \frac{16^{\delta n}}{4^n - 1}. \tag{5.71}$$

We then rewrite eq. (5.69) in order to include this distribution and subsequently use the above result, that is,

$$(5.69) = \frac{1}{4^n} \sum_{\ell=1}^n \sum_{\mu : |\mu| = \ell} \operatorname{Tr}[Tr_A[\sigma_\mu \tilde{\rho}_{AE}]^2] \sum_{\nu \in \{0,1,2,3\}^n, \nu \neq 0} \operatorname{Tr}[\widetilde{\mathcal{T}}(\sigma_\nu)^2](p_\delta(\nu) + Q^t_{\mathrm{circ}}(\mu,\nu) - p_\delta(\nu)) \tag{5.72}$$

$$\leq \frac{1}{4^n} \sum_{\mu \neq 0} \operatorname{Tr}[Tr_A[\sigma_\mu \tilde{\rho}_{AE}]^2] \sum_{\nu \neq 0} \operatorname{Tr}[\widetilde{\mathcal{T}}(\sigma_\nu)^2] \frac{4^{\delta n}}{4^n - 1}$$
$$+ \frac{1}{4^n} \sum_{\ell=1}^n \sum_{\mu : |\mu| = \ell} \operatorname{Tr}[Tr_A[\sigma_\mu \tilde{\rho}_{AE}]^2] \frac{1}{(3-\eta)^\ell \binom{n}{\ell}} \frac{1}{\mathrm{poly}(n)} \max_\nu \operatorname{Tr}[\widetilde{\mathcal{T}}(\sigma_\nu)^2]. \tag{5.73}$$

Now, we recall that some of these terms are by definition linked to the quantum collision entropy, namely,

$$\sum_\mu \operatorname{Tr}[Tr_A[\sigma_\mu \tilde{\rho}_{AE}]^2] = 2^n \operatorname{Tr}[\tilde{\rho}^2_{AE}] = 2^n \, 2^{-H_2(A|E)_\rho} \tag{5.74}$$

and since $\Phi_{A'A} = \frac{1}{4^n} \sum_\nu \sigma_\nu \otimes \sigma_\nu$ one has

$$2^{-H_2(A|B)_\tau} = \frac{1}{8^n} \sum_\nu \operatorname{Tr}[\widetilde{\mathcal{T}}(\sigma_\nu)^2]. \tag{5.75}$$

We can then re-formulate the first term in eq. (5.73) as

$$\frac{1}{4^n} \sum_{\mu \neq 0} \operatorname{Tr}[Tr_A[\sigma_\mu \tilde{\rho}_{AE}]^2] \sum_{\nu \neq 0} \operatorname{Tr}[\widetilde{\mathcal{T}}(\sigma_\nu)^2] \frac{4^{\delta n}}{4^n - 1} = 4^{\delta n - n} \sum_{\nu \neq 0} \operatorname{Tr}[\widetilde{\mathcal{T}}(\sigma_\nu)^2] \frac{2^n \operatorname{Tr}[\tilde{\rho}^2_{AE}] - \operatorname{Tr}[\tilde{\rho}^2_E]}{4^n - 1}$$
$$\leq 4^{\delta n} \, 2^{-H_2(A|B)_\tau} \, 2^{-H_2(A|E)_\rho}. \tag{5.76}$$

Now, there is the second term in eq. (5.73) to be bounded by an inverse polynomial. To prove this, Brown and Fawzi showed first

$$\sum_{\nu:|\nu|=\ell} \mathrm{Tr}[Tr_A[\sigma_\nu \tilde{\rho}_{AE}]^2] \leq 12\, n^4 (3-\eta)^\ell \binom{n}{\ell} \tag{5.77}$$

leading to the desired bound, namely,

$$\frac{1}{4^n} \sum_{\ell=1}^{n} \sum_{\mu:|\mu|=\ell} \mathrm{Tr}[Tr_A[\sigma_\mu \tilde{\rho}_{AE}]^2] \frac{1}{(3-\eta)^\ell \binom{n}{\ell}} \frac{1}{\mathrm{poly}(n)} \max_\nu \mathrm{Tr}[\widetilde{\mathcal{T}}(\sigma_\nu)^2] \leq \tag{5.78}$$

$$\leq \frac{1}{4^n} \max_\nu \mathrm{Tr}[\widetilde{\mathcal{T}}(\sigma_\nu)^2] \frac{12\, n^5}{\mathrm{poly}(n)} \leq \frac{1}{\mathrm{poly}(n)} \tag{5.79}$$

and this concludes the proof of Theorem 5.14. □

As we pointed out before the proof, this calculations will also connect to a decoupling theorem our main technical results, Theorem 7.19. The latter is in turn an equivalent formulation of Theorem 5.15 in the framework of stochastically evolving continuous-time unitary evolutions and so will be employed in the same way leading to fast decoupling. Both technical results rely on a random walk on the zero chain in a similar fashion as in the work of Harrow and Low. In fact, the work of Brown and Fawzi is based once again on the zero chain with the same transition matrix (the one given in eq. (5.9)) since, while they look for a different property, they are using the same local RQC . As we will see in Section 7.2.1, the continuous-time stochastic evolution induces another random walk on the zero chain whose accelerated chain is exactly the one of eq. (5.11). This means that, in order to show Theorem 7.19, we are going to borrow from the work of Brown and Fawzi results regarding the accelerated chain that we are going to outline in the following paragraphs.
The principal result is a bound on Pauli coefficients after t steps of the zero chain, namely,

Lemma 5.16. *Let P be the Markov chain transition matrix (5.9). For any constants $\delta \in (0, 1/16)$ and η in$(0,1)$ there exists a constant c such that for $t \geq c\, n \log^2 n$ and all integers $1 \leq \ell \leq n$ and $1 \leq k \leq n$, we have for large enough n*

$$P^t(\ell, k) \leq 4^{\delta n} \frac{\binom{n}{k} 3^k}{4^n - 1} + \frac{1}{(3-\eta)^\ell \binom{n}{\ell}} \frac{1}{\mathrm{poly}(n)}. \tag{5.80}$$

Conceptually, this means that $t \geq c\, n \log^2 n$ steps of the transition matrix P lead to a distribution which is close to the stationary one, $\frac{\binom{n}{k} 3^k}{4^n - 1}$, up to an exponential factor $4^{\delta n}$, with failure probability $\frac{1}{(3-\eta)^\ell \binom{n}{\ell}} \frac{1}{\mathrm{poly}(n)}$.
To achieve this, one first divides the state space Ω into three regions: $[1, r_-), [r_-, r_+]$ and $(r_+, n]$ for $r_- = (3/4 - \delta)n$ and $r_+ = (3/4 + \delta)n$.
Since it stands

$$\binom{n}{r_-} 3^{r_-} \geq 4^{(1-\delta)n} \qquad \text{and} \qquad \binom{n}{r_+} 3^{r_+} \geq 4^{(1-\delta)n} \tag{5.81}$$

for sufficiently large n, one has for $r \in [r_-, r_+]$ and any $t \geq 1$

$$\begin{aligned}
\mathbb{P}^t(r,k) &= \frac{4^n-1}{\binom{n}{r}3^r}\frac{\binom{n}{r}3^r}{4^n-1}\mathbb{P}^t(r,k) \\
&\leq \frac{4^n-1}{\binom{n}{r}3^r}\frac{1}{4^n-1}\sum_{\ell=1}^{n} 3^\ell \binom{n}{\ell}\mathbb{P}^t(\ell,k) \\
&\leq \frac{4^n-1}{\binom{n}{r}}3^r\frac{\binom{n}{k}3^k}{4^n-1} \\
&\leq 4^{\delta n}\frac{\binom{n}{k}3^k}{4^n-1}
\end{aligned} \tag{5.82}$$

and so the condition (5.80) is immediately satisfied when starting within the region $[r_-, r_+]$. This means that we should investigate how many steps the chain needs in order to reach this subset of the state space when starting from an arbitrary position $\ell \in [1, r_-)$ or $\ell \in (r_+, n]$. In the former case, the probability for the steps number T_{r_-} needed to reach the region $[r_-, r_+]$ being larger than $O(n\log^2 n)$ can be upper bounded by the following lemma.

Lemma 5.17 (Lemma 4.3 in ref. [24]). *Let $\delta \in (0, 1/16)$ and $\eta \in (0,1)$. Then for a large enough constant c (depending on δ and η) and large enough n, we have for all $\ell \leq r_-$*

$$\mathbb{P}\left(T_{r_-} > c\, n\log^2 n\right) \leq 2^{-2n} + \frac{1}{(3-\eta)^\ell \binom{n}{\ell}}\frac{1}{\mathrm{poly}(n)}. \tag{5.83}$$

The proof can be split again in two lemmas, one giving a bound for exceeding a certain number of accelerated steps S and the other one providing a bound on the waiting time when the former event does not happen. More precisely, one has

$$\mathbb{P}\left(T > t+s\right) \leq \mathbb{P}\left(S > s\right) + \mathbb{P}\left(S \leq s, W_1 + W_2 + \cdots + W_S > t\right), \tag{5.84}$$

where W_j denotes the waiting time between the accelerated steps $j-1$ and j.
With the knowledge of the accelerated chain in eq. (5.11) and a Chernoff-type bound one obtains for $s > \frac{n}{3\delta}$

$$\mathbb{P}\left(S > s\right) \leq \exp\left(-\frac{\delta^2}{18}s\right), \tag{5.85}$$

so that one can bound this probability with an arbitrarily large exponential for some $O(n)$ accelerated steps.
Conversely, the waiting time can be bounded using a Gambler's ruin argument (in fact, the real problem arises when the walk, after reaching states with weights being at least a fraction of n, goes back to states with low weights, where the expected waiting time is large) so that we obtain

$$\mathbb{P}\left(S \leq s, W_1 + W_2 + \cdots + W_S > c\, n\log^2 n\right) \leq \frac{1}{(3-\eta)^\ell \binom{n}{\ell}}\frac{1}{\mathrm{poly}(n)}, \tag{5.86}$$

and this corresponds to the second term in the RHS of eq. (5.83).

The case for a walk starting from $\ell \in (r_+, n]$ is easier to bound, namely, as

$$\mathbb{P}\left(T_{r_+} > c\, n\log^2 n\right) \leq 2^{-2n} \tag{5.87}$$

since the bound on the waiting time does not require a specific argument: in the region for $k > r_+$, one has $P(k,k) \leq 4/5$ and so all waiting times $W_1 + W_2 + \cdots + W_S$ are stochastically dominated by a geometric distribution with parameter $4/5$, allowing for a direct application of a

Chernoff-type bound.

To go from Lemma 5.16 to the main result given in Theorem 5.15 an additional step is required. The complete and explicit argument is quite complex and long, hence we provide here only a summary thereof.

Let us go back to the full chain $Q(\mu,\nu)$, assuming for the rest of the argument that the waiting time does not exceed $c\,n\log^2 n$ (event bounded by eq. (5.86)), and rewrite it as

$$Q = \frac{2}{5}R + \frac{3}{5}M \tag{5.88}$$

where $R = \frac{1}{2}\tilde{R} + \frac{1}{2}\mathbb{F}R$, $\mathbb{F}$ is the operator that interchanges two random sites chosen at random, and $\tilde{R} = \frac{1}{n(n-1)}\sum_{i,j}\tilde{R}_{ij}$ and $M = \frac{1}{n(n-1)}\sum_{i,j}M_{ij}$ are given by

$$\tilde{R}_{ij}(\mu,\nu) = \begin{cases} 1 & \text{if } |\mu_i,\mu_j| = |\nu_i,\nu_j| = 0 \\ 1/3 & \text{if } |\mu_i,\mu_j| = |\nu_i,\nu_j| = 1, \mu_i = \nu_i = 0 \\ 1/3 & \text{if } |\mu_i,\mu_j| = |\nu_i,\nu_j| = 1, \mu_j = \nu_j = 0 \\ 1/9 & \text{if } |\mu_i,\mu_j| = |\nu_i,\nu_j| = 2 \end{cases} \tag{5.89}$$

and

$$M_{ij}(\mu,\nu) = \begin{cases} 1 & \text{if } |\mu_i,\mu_j| = |\nu_i,\nu_j| = 0 \\ 1/9 & \text{if } |\mu_i,\mu_j| = 1 \text{ and } |\nu_i,\nu_j| = 2 \\ 1/9 & \text{if } |\mu_i,\mu_j| = 2 \text{ and } |\nu_i,\nu_j| = 1 \\ 1/27 & \text{if } |\mu_i,\mu_j| = |\nu_i,\nu_j| = 2 \end{cases} \tag{5.90}$$

respectively.
There are two important remarks. R does not change the weight of the string μ, only performs permutations of the support and cycling of the Pauli labels 1,2,3. Additionally, R and M commute, and so one can write T steps of the full chain Q as

$$Q^T = \sum_{T_1+T_2=T}\left(\frac{3}{5}\right)^{T_1}\left(\frac{2}{5}\right)^{T_2}\binom{T}{T_1}M^{T_1}R^{T_2}. \tag{5.91}$$

Hence, one can first let the chain M run in order to reach a certain weight k, and then let R run to change the support without altering its size. Assuming that after T_1 steps of the chain M we reached the Pauli string ν^M with $|\nu^M| = k$, we analyze now how the support evolves by constructing a random walk on the intersection between the support of the Pauli string touched at time $T_1 + S$, that we shall call ν^S, and some target string ν, both having support size k. Note that if the support were completely random the intersection would then be made of k^2/n qubits. Let us call the chain I_S, then its transition matrix is given by

$$I_S(p,q) = \begin{cases} \frac{(k-p)^2}{n(n-1)} & \text{if } q = p+1 \\ \frac{p(n-2k+p)}{n(n-1)} & \text{if } q = p-1 \\ 1 - \frac{(k-p)^2}{n(n-1)} - \frac{p(n-2k+p)}{n(n-1)} & \text{if } p = q. \end{cases} \tag{5.92}$$

The stationary distribution is

$$\pi_I(k) = \frac{\binom{k}{k'}\binom{n-k}{k-k'}}{\binom{n}{k}} \tag{5.93}$$

for $k' \in \{0, 1, \ldots, k\}$. We want to bound the probability of reaching state k' when starting from some state r'. The closer is r' to k^2/n, the closer is this probability to the stationary distribution π_I; more precisely, whenever $r' \in [k^2/n - \delta_2 n, k^2/n + \delta_2 n]$ for some positive constant δ_2,

$$\sum_{|\nu'|=k : I_S = k'} Q^{T=T_1+S}(\mu, \nu') \leq n^2 2^{nh\left(\frac{\delta_2}{\delta_0}\right)} 4^{\delta n} \frac{\binom{k}{k'}\binom{n-k}{k-k'}}{\binom{n}{k}} \frac{\binom{n}{k} 3^k}{4^n - 1}, \tag{5.94}$$

where h is the *binary entropy function* $h(\alpha) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$. The probability to fail in reaching the region $r' \in [k^2/n - \delta_2 n, k^2/n + \delta_2 n]$ can be bounded by an arbitrarily exponentially decreasing function in n after $O(n)$ steps using a Chernoff-type bound, as already done previously, and can hence be neglected.
It remains to show that the probability in eq. (5.94) is more or less evenly split among all ν' with support size k. For any permutation $\pi \in S_n$ and any relabeling $\mathcal{L}$ of the 3 Pauli indices it stands that $R(\gamma \circ \mathcal{L}(\mu), \gamma \circ \mathcal{L}(\nu)) = R(\mu, \nu)$ and hence $R(\mu, \gamma \circ \mathcal{L}(\nu)) = R(\mu, \nu)$ for any γ and $\mathcal{L}$ that leave μ unchanged. Therefore all $3^{k-k'}\binom{n-k}{k-k'}$ permutations and relabelings of ν acting outside the support of ν^M satisfy the condition and are also equal in probability; in addition, all combinations of possible intersections of the support of ν^M and ν can be seen as a simple relabeling of the Pauli indices so that there are additional $\binom{k}{k'}$ equally distributed strings for each transformation acting outside the support of ν^M. Finally, all relabelings of each site in the intersection of the supports produce 3 additional equally distributed strings ν. One can show that a relabeling of at least $k' - \delta_1 n$ sites happens with very high probability after $s = O(n \log n)$ applications of the chain $\tilde{R}$ (i.e., the complementary event can be bounded by an arbitrarily exponentially decreasing function). Summarizing, a total of $3^{k-\delta_1 n}\binom{n-k}{k-k'}\binom{k}{k'}$ strings are equally distributed. Using eq. (5.94), we can finally write

$$\begin{aligned}
Q^T(\mu, \nu) &\leq \frac{1}{3^{k-\delta_1 n}\binom{n-k}{k-k'}\binom{k}{k'}} \sum_{|\nu'|=k} Q^T(\mu, \nu') && (5.95)\\
&\leq \frac{1}{3^{k-\delta_1 n}\binom{n-k}{k-k'}\binom{k}{k'}} n^2 2^{nh\left(\frac{\delta_2}{\delta_0}\right)} 4^{\delta n} \frac{\binom{k}{k'}\binom{n-k}{k-k'}}{\binom{n}{k}} \frac{\binom{n}{k} 3^k}{4^n - 1} && (5.96)\\
&\leq n^2 2^{nh\left(\frac{\delta_2}{\delta_0}\right)} 3^{\delta_1 n} \frac{4^{\delta n}}{4^n - 1} && (5.97)\\
&\leq \frac{16^{\delta n}}{4^n - 1} && (5.98)
\end{aligned}$$

for an adequate choice of δ_1 and δ_2 and large enough n. With this bound and the term coming from the bound for a waiting time exceeding $c\, n \log^2 n$ from eq. (5.86) we obtain Theorem 5.15.

CHAPTER 6

Randomized benchmarking

In order to realize a multi-qubit quantum computer it is necessary to estimate the accuracy of an experimental implementation of unitary operations with high precision. One of the most relevant methods to verify the error characterizing the incorrect implementation is quantum process tomography [102, 103], which allows for a complete description of the implemented unitary gate but suffers however from two major drawbacks. First of all, the method is not scalable in the number of qubits: for a system of n qubits with dimension $d = 2^n$, $\Theta(d^2)$ expectation values have to be estimated. Furthermore, this approach is sensitive to state preparation and measurement (SPAM) errors since it cannot distinguish and handle separately the one related to the quantum gate with the one coming from the generation of the initial state and the final measurement. This challenge is aggravated by the fact that gate errors suitable for fault tolerant quantum computing are extremely small, so that the characterization needs to be very precise.
An alternative and successful approach that we are going to illustrate in the following pages is given by *randomized benchmarking*. This method is based on the randomization of the noise channel associated to the unitary gate, which leads to a simpler extraction of information that remains unaltered by the protocol. Also, it allows to overcome the two issues related to quantum process tomography since it is robust against SPAM errors and, at least for protocols involving the Clifford group, scalable in the system size. This however comes at the cost of recovering less information: what we actually get is the *average gate fidelity* of the noise channel, that can then be related to more physically relevant objects such as the diamond norm. Furthermore, additional assumptions on the underlying model are required, in particular regarding the element to be characterized, that is, the noise channel describing the experimental unitary gate implementation. This may sound odd, but it is actually a general feature of estimation theory: one needs some prior knowledge regarding the quantity one tries to gauge in order to use nontrivial estimation methods [104]; for several protocols one assumes identical noise levels for each gate in the set of operators to be benchmarked, which is a rather strong assumption that we aim to overcome with our novel approach.

The literature on randomized benchmarking is quite extended, so we will focus on theoretical discoveries leaving out experimental investigations. We will start with the arguably most common and recognized randomized benchmarking method presented by Magesan, Gambetta and Emerson [45, 105] after providing some mathematical background on the subject. However, the basic idea has been generalized in several ways, using instead, for instance, the single-qubit

dihedral group [106]. Still, the known schemes extract information averaged over a full group and, except for protocols tied to the Clifford group [107], assume identical noise channel for all gates up to small gate-dependent perturbations. To partially address the problem and allow for extraction of the fidelity of a specific gate, a scheme called *interleaved randomized benchmarking* [108, 109] makes use of random sequences of Clifford gates interlaced by the particular gate whose noise is to be individually characterized, but it is still limited to the Clifford group itself or the T-gate. In several ways the commonly made assumptions on the underlying processes are very demanding. For example, refs. [45, 46, 105] require the 2-design property and so they are restricted to the Clifford group on n qubits, whose size scales as $\Theta(2^{O(n^2)})$. If one instead makes use of schemes involving different gate sets, then the assumption on uniform noise channel for each gate of the group may be overly strong. In our novel approach illustrated in Section 6.3 we will see that some of these requirements can be largely relaxed at little additional cost, as far as quantum resources are concerned.

6.1 The randomized benchmarking protocol with Clifford group

We are now going to illustrate the benchmarking protocol with Clifford gates. In few words, we apply multiple random Clifford operators in sequence to an initial state ρ and then, before measuring the final outcome with respect to some POVM E, we invert all gates with a single final operator. We repeat the procedure multiple times in order to obtain an average of the sample outcome which can be assumed to be close to the full average over the Clifford group. More precisely, we perform the following steps.

(1) Draw a random sequence $\mathbf{k}_\ell = (k_1, \dots, k_\ell) \in \mathbb{N}_{|\mathcal{C}|}^\ell$.

(2) Apply to an initial state ρ (e.g., $\rho = |0\rangle\langle 0|$) the sequence of random Clifford gates (labeled with $\mathbb{N}_{|\mathcal{C}|}^\ell$) generated according to $\mathbf{k}_\ell$,

$$\mathcal{S}_{\mathbf{k}_\ell} = \mathcal{C}_{\text{inv}} \circ \mathcal{C}_{k_\ell} \circ \cdots \circ \mathcal{C}_{k_2} \circ \mathcal{C}_{k_1}, \tag{6.1}$$

where $\mathcal{G}_{\text{inv}} := \mathcal{G}_{k_1}^\dagger \circ \cdots \circ \mathcal{G}_{k_\ell}^\dagger$ is the channel (still in the Clifford group) given by the composition of the inversion of all previous gates.

(3) Perform a POVM E to be defined later and measure the survival probability

$$\mathcal{Q}_{\mathbf{k}_\ell} = \text{Tr}[E\ \mathcal{S}_{\mathbf{k}_\ell}(\rho)]. \tag{6.2}$$

Repeat this step sufficiently many times in order to obtain a statistic that gives a reliable result outcome.

(4) Repeat the previous steps for sufficiently many (say N) random sequences $\mathbf{k}_{\ell,1}, \dots, \mathbf{k}_{\ell,N}$ of length ℓ. Then, calculate the *average survival probability*

$$\mathcal{Q}_{\text{seq}}(\ell) = \frac{1}{N} \sum_{\mathbf{k}_\ell} \mathcal{Q}_{\mathbf{k}_\ell}. \tag{6.3}$$

The number N of random sequences should be chosen such that

$$\mathcal{Q}_{\text{seq}}(\ell) \approx \mathcal{Q}_{\text{avg}}(\ell), \tag{6.4}$$

where $\mathcal{Q}_{\text{avg}}$ is the fidelity averaged over all possible sequences.

(5) Repeat the previous steps for different lengths ℓ and insert $\mathcal{Q}_{\text{seq}}(\ell)$ into the fitting model,

$$\mathcal{Q}^{(0)}_{\text{avg}}(\ell) = Ap^{\ell} + B \qquad \text{zero order model} \tag{6.5}$$

$$\mathcal{Q}^{(1)}_{\text{avg}}(\ell) = \mathcal{Q}^{(0)}_{\text{avg}}(\ell) + C(\ell-1)(w-p^2)p^{\ell-2} \qquad \text{first order model} \tag{6.6}$$

where $\mathcal{Q}^{(1)}_{\text{avg}}(\ell)$ is suitable when considering gate-dependent perturbation of the noise channel.
The parameters A, B, C absorb the SPAM errors and p is the polarization parameter that we want to obtain for the following equations expressing the average gate fidelity in terms of it.

Lemma 6.1 (Lemma 1 in ref. [54]). *For any quantum channel Λ and density operator ρ we have*

$$\text{twirl}(\Lambda(\rho)) \coloneqq \int_{\text{Haar}} U\Lambda(U^{\dagger}\rho U)U^{\dagger}\,\mathrm{d}U = p\rho + (1-p)\frac{\mathbb{1}}{d} =: \Lambda^{\text{dep}}(\rho), \tag{6.7}$$

where p is called depolarization parameter.

The depolarized channel can be interpreted as a quantum operator that with probability p returns the initial state and otherwise the fully mixed state.
It is straightforward to see that the average fidelity of the depolarized channel is related to the depolarization parameter by

$$\mathbb{E}(\mathcal{F}_{\Lambda^{dep},\mathcal{I}}) = p + \frac{1-p}{d}. \tag{6.8}$$

It is crucial to remark that both the average gate fidelity and the entanglement fidelity are invariant with respect to twirling, namely,

Lemma 6.2 (Lemma 2 in ref. [54]). *The average gate fidelity and the entanglement fidelity are both invariant with respect to the twirling over the Haar measure, that is,*

$$\mathbb{E}(\mathcal{F}_{\Lambda,\mathcal{I}}) = \mathbb{E}(\mathcal{F}_{\Lambda^{dep},\mathcal{I}}) \qquad \textit{and} \qquad F_{\text{ent}}(\Lambda) = F_{\text{ent}}(\Lambda^{\text{dep}}), \tag{6.9}$$

and thus the benchmarking does not alter the fidelity of the original noise channel. This condition uniquely defines the polarization parameter in eq. (6.7).

We now derive the fitting model in step (5). We notice that eq. (6.1) should return the identity channel, unless we take into account that the gates are not perfectly implemented. Namely, we assume that each Clifford gate is realized with an additional error channel Λ, identical for all gates in the group. The random sequence $\mathcal{S}_{\mathbf{k}_\ell}(\rho)$ then becomes

$$\mathcal{S}_{\mathbf{k}_\ell} = \Lambda \circ \mathcal{C}_{\text{inv}} \circ \Lambda \circ \mathcal{C}_{k_\ell} \circ \cdots \circ \Lambda \circ \mathcal{C}_{k_2} \circ \Lambda \circ \mathcal{C}_{k_1} = \Lambda \circ \mathcal{C}_{\text{inv}} \left(\bigcirc_{j=1}^{\ell} \Lambda \circ \mathcal{C}_{k_j} \right). \tag{6.10}$$

Let us now introduce the following reformulation with gates $\mathcal{D}_k$ again from the Clifford group.

(i) $\mathcal{D}_{k_1} \coloneqq \mathcal{C}_{k_1}$

(ii) define recursively $\mathcal{D}_{k_{j+1}} \coloneqq \mathcal{C}_{k_{j+1}} \circ \mathcal{D}_{k_j}$

(iii) from the two previous steps, follows $\mathcal{D}_{\text{inv}} = \mathcal{I}$.

Hence, we have now

$$\mathcal{S}_{\mathbf{k}_\ell} = \Lambda \circ \mathcal{D}_{k_\ell}^\dagger \circ \Lambda \circ \mathcal{D}_{k_\ell} \circ \cdots \circ \mathcal{D}_{k_2}^\dagger \circ \Lambda \circ \mathcal{D}_{k_2} \circ \mathcal{D}_{k_1}^\dagger \circ \Lambda \circ \mathcal{D}_{k_1} \tag{6.11}$$

and averaging over all possible sequences

$$\mathcal{S}_{\text{avg}}^{(0)}(\ell) = \frac{1}{|\mathcal{C}|^\ell} \sum_{\text{all sequences}} \Lambda \circ \mathcal{D}_{k_\ell}^\dagger \circ \Lambda \circ \mathcal{D}_{k_\ell} \circ \cdots \circ \mathcal{D}_{k_1}^\dagger \circ \Lambda \circ \mathcal{D}_{k_1} \tag{6.12}$$

$$= \Lambda \circ \left(\frac{1}{|\mathcal{C}|} \sum_k \mathcal{D}_k^\dagger \circ \Lambda \circ \mathcal{D}_k \right) \circ \cdots \circ \left(\frac{1}{|\mathcal{C}|} \sum_k \mathcal{D}_k^\dagger \circ \Lambda \circ \mathcal{D}_k \right) \tag{6.13}$$

$$= \Lambda \circ \underbrace{\Lambda^{\mathcal{C}} \circ \cdots \circ \Lambda^{\mathcal{C}}}_{\ell \text{ times}}, \tag{6.14}$$

where $\Lambda^{\mathcal{C}}$ is the noise channel Λ twirled over the (uniformly distributed) Clifford group. Recalling from eq. (3.9) the twirling condition over unitary 2-designs, we know that $\Lambda^{\mathcal{C}} = \Lambda^{\text{dep}}$. This implies, using eq.(6.7),

$$\mathcal{Q}_{\text{avg}}^{(0)} = \text{Tr}[E\, \Lambda \bigcirc^\ell \Lambda^{\text{dep}}(\rho)] = \text{Tr}[E\ \Lambda(\rho)]\, p^\ell + \text{Tr}\left[E\ \Lambda\left(\frac{\mathbb{1}}{d}\right)\right](1-p^\ell) \coloneqq Ap^\ell + B, \tag{6.15}$$

where

$$A = \text{Tr}\left[E\ \Lambda\left(\rho - \frac{\mathbb{1}}{d}\right)\right] \qquad \text{and} \qquad B = \text{Tr}\left[E\ \Lambda\left(\frac{\mathbb{1}}{d}\right)\right] \tag{6.16}$$

absorb state preparation and measurement errors, since an imperfect realization of E' of E and ρ' of ρ do not alter the fitting model expression.

One can slightly relax the condition for the noise channel, allowing for a small time- and gate-dependent perturbation term, namely, $\Lambda_{j,k} = \Lambda + \delta\Lambda_{j,k}$ and rewrite the first order random sequence as

$$\mathcal{S}_{\text{avg}}^{(1)}(\ell) = \mathcal{S}_{\text{avg}}^{(0)} + \sum_{j=1}^{\ell} \frac{1}{|\mathcal{C}|^\ell} \sum_{\text{all sequences}} \Lambda \circ \mathcal{D}_{k_\ell}^\dagger \circ \Lambda \circ \mathcal{D}_{k_\ell} \circ \cdots \circ \mathcal{D}_{k_j}^\dagger \circ \delta\Lambda_{j,k_j} \circ \mathcal{D}_{k_j} \circ \cdots \circ \mathcal{D}_{k_1}^\dagger \circ \Lambda \circ \mathcal{D}_{k_1}. \tag{6.17}$$

The trick now is to revert the notation $\mathcal{D}_{k_j} = \mathcal{C}_{k_j} \circ \mathcal{D}_{k_{j-1}}$ so that the perturbations of steps $1, \ldots, \ell$ lead to the expression

$$\sum_{j=2}^{\ell} \Lambda \circ \underbrace{\Lambda^{\text{dep}} \circ \cdots \circ \Lambda^{\text{dep}}}_{\ell - j \text{ times}} \circ (\mathcal{W}_j \circ \Lambda)^{\text{dep}} \circ \underbrace{\Lambda^{\text{dep}} \circ \cdots \circ \Lambda^{\text{dep}}}_{j-2 \text{ times}} + \Lambda \circ \left(\Lambda^{\text{dep}}\right)^{\circ(\ell-1)} \circ \mathcal{W}_1, \tag{6.18}$$

where $\mathcal{W}_j \coloneqq \frac{1}{|\mathcal{C}|} \sum_k \mathcal{C}_k^\dagger \circ \delta\Lambda_{j,k} \circ \mathcal{C}_k$. Using the commuting properties of depolarizing channels we can rewrite

$$\sum_{j=2}^{\ell} \Lambda \circ (\mathcal{W}_j \circ \Lambda)^{\text{dep}} \left(\Lambda^{\text{dep}}\right)^{\circ(\ell-2)} + \Lambda \circ \left(\Lambda^{\text{dep}}\right)^{\circ(\ell-1)} \circ \mathcal{W}_1. \tag{6.19}$$

One can also incorporate a perturbation of the last channel Λ at step $\ell + 1$, depending only on the last gate $\mathcal{D}_{k_\ell}$, defining the channel $\mathcal{R}_{\ell+1} \coloneqq \frac{1}{|\mathcal{C}|} \sum_k \delta\Lambda_{\ell+1,k} \mathcal{D}_k^\dagger \circ \Lambda \circ \mathcal{D}_k$. The first order of the random sequence can be then expressed as

$$\mathcal{S}_{\text{avg}}^{(1)}(\ell) = \mathcal{S}_{\text{avg}}^{(0)} + \sum_{j=2}^{\ell} \Lambda \circ (\mathcal{W}_j \circ \Lambda)^{\text{dep}} \circ \left(\Lambda^{\text{dep}}\right)^{\circ(\ell-2)} \tag{6.20}$$

$$+ \Lambda \circ \left(\Lambda^{\text{dep}}\right)^{\circ(\ell-1)} \circ \mathcal{W}_1 + \mathcal{R}_{\ell+1} \circ \left(\Lambda^{\text{dep}}\right)^{\circ \ell} \tag{6.21}$$

so that the fitting model for the average survival probability becomes

$$\mathcal{Q}_{\text{avg}}^{(1)} = \text{Tr}[E\,\mathcal{S}_{\text{avg}}^{(1)}(\ell,\rho)] = Ap^{\ell} + B + C(\ell-1)(w-p^2)p^{\ell-2} \tag{6.22}$$

with

$$A = \text{Tr}\left[E\,\Lambda\left(\frac{\mathcal{Q}_1(\rho)}{p} - \rho + \frac{(p-1)\mathbb{1}}{pd}\right)\right] + \text{Tr}\left[E\,\mathcal{R}_{\ell+1}\left(\frac{d\rho-\mathbb{1}}{pd}\right)\right] \tag{6.23}$$

$$B = \text{Tr}\left[E\,\mathcal{R}_{\ell+1}\left(\frac{\mathbb{1}}{d}\right)\right] \tag{6.24}$$

$$C = \text{Tr}\left[E\,\Lambda\left(\rho - \frac{\mathbb{1}}{d}\right)\right] \tag{6.25}$$

$$w = \sum_{j=2}^{\ell} w_j/(\ell-1), \tag{6.26}$$

where w_j is the depolarization parameter defined by

$$(\mathcal{W}_j \circ \Lambda)^{\text{dep}}(\rho) = w_j\rho + (1-w_j)\frac{\mathbb{1}}{d}. \tag{6.27}$$

According to ref. [45], second order terms can be neglected when the following conditions stand. Let us first define a norm to quantify the distance between superoperators as

$$\|\mathcal{E}\|_{1\to1}^{H} = \max_{X \text{ hermitian and } \|X\|_1\leq 1} \|\mathcal{E}(X)\|_1 \tag{6.28}$$

and the gate dependent perturbation averaged over all steps as

$$\gamma_j \coloneqq \frac{1}{|\mathcal{C}|}\sum_j \|\Lambda_{j,k} - \Lambda\|_{1\to1}^{H}. \tag{6.29}$$

Then, the second order perturbation terms,

$$\sum_{j_2>j_1}\frac{1}{|\mathcal{C}|^{\ell}}\sum_{\text{all sequences}} \Lambda\circ\mathcal{D}_{k_\ell}^{\dagger}\circ\Lambda\circ\mathcal{D}_{k_\ell}\circ\cdots\circ\mathcal{D}_{k_{j_2}}^{\dagger}\circ\delta\Lambda_{j,k_{j_2}}\circ\mathcal{D}_{k_{j_2}}$$
$$\circ\cdots\circ\mathcal{D}_{k_{j_1}}^{\dagger}\circ\delta\Lambda_{j,k_{j_1}}\circ\mathcal{D}_{k_{j_1}}\circ\cdots\circ\mathcal{D}_{k_1}^{\dagger}\circ\Lambda\circ\mathcal{D}_{k_1} \tag{6.30}$$

can be upper bounded in $\|\cdot\|_{1\to1}^{H}$ norm by $\sum_{j_2>j_1}\gamma_{j_2}\gamma_{j_1}$ and, when assuming that the noise is time-independent and so $\gamma_j = \gamma \quad \forall j$, by $\sum_{j_2>j_1}\gamma^2 = \frac{m(m+1)}{2}\gamma^2$. The argument can be extended to all higher order terms, so that

$$|\mathcal{Q}_{\text{avg}}^{(r+1)} - \mathcal{Q}_{\text{avg}}^{(r)}| \leq \sum_{j_r>\cdots>j_1}\gamma_{j_r}\cdots\gamma_{j_1} \tag{6.31}$$

and again under time-independent noise assumption

$$|\mathcal{Q}_{\text{avg}}^{(r+1)} - \mathcal{Q}_{\text{avg}}^{(r)}| \leq \binom{\ell+1}{r}\gamma^r. \tag{6.32}$$

The choice of the $\|\cdot\|_{1\to1}^{H}$ in place of the more usual diamond norm is motivated by the fact that

$$\|\mathcal{E}_2 - \mathcal{E}_1\|_{1\to1}^{H} \leq \|\mathcal{E}_2 - \mathcal{E}_1\|_{\diamond} \tag{6.33}$$

so that the bound provided by the former is tighter.

The argument just now presented is presumably flawed and constitutes the starting point of a new fitting [107] for the same Clifford gates protocol, fixing the problem and also lifting gate-independence noise assumption for for those protocols strictly tied to the Clifford group. Actually, the contribution of all higher order terms are (applying the binomial theorem)

$$h_1 = \sum_{q\geq 1} \binom{\ell+1}{q} \gamma^q = (1+\gamma)^\ell - 1 \approx \mathrm{e}^{\ell\gamma} - 1 \qquad \text{zero order residuum} \tag{6.34}$$

$$h_2 = \sum_{q\geq 2} \binom{\ell+1}{q} \gamma^q \approx \mathrm{e}^{\ell\gamma} - 1 - \ell\gamma \qquad \text{first order residuum} \tag{6.35}$$

for the zero order and second order model, respectively. To stay in the range of magnitude 0.01 with respect to the dominant terms, the larger sequence lengths are $\ell \approx 0.01/\gamma$ and $\ell \approx 0.1/\gamma$. On the other hand, numerics from ref. [110] suggests that, in order to fit the exponential decay, sequences of length $\ell \approx 1/r$ (where we recall r being the average error rate of the noise channel) are required. Hence, we conclude that negligible higher order terms impose the condition $\gamma \leq 0.01r$ or respectively $\gamma \leq 0.1r$. The first issue is to certify that the experimental implementation is indeed in this regime, since it is likely to require fully reconstruction of the process matrix for each noise channel. Secondly, such small deviations from the average noise channel Λ may be a too strong constraint for effective implementations, in particular for the zero order model. By showing that most of the gate-dependent perturbation noise terms cancel each others out, one can improve the bound of ref. [45] at the cost of an additional exponentially decaying term in the fitting model. That is [107, Theorem 4],

$$\mathcal{Q}_{\mathrm{avg}} = Ap^\ell + B + \varepsilon_\ell, \tag{6.36}$$

where the perturbation term ε_m^ℓ satisfies

$$|\varepsilon_\ell| \leq \delta_1 \delta_2^\ell \tag{6.37}$$

for some δ_1, δ_2 that quantifiy the amount of gate dependence.
While, as mentioned above, the protocol is now suitable to benchmark gates with a noise component that can be differentiated for each of them, one should remark that it still retrieves averaged quantities only and not the individual noise levels.

We now turn on another important point, namely, the scalability of the protocol. The Clifford group is large, in the sense that it scales as $2^{O(n^2)}$; consequently, the number of sequences of length ℓ is $2^{\ell\, O(n^2)}$. This implies that in order to obtain the full average over all sequences, the protocol does not scale either in system size nor in sequence length. The authors of ref. [45] provide a bound using Hoeffding inequality,

$$\mathbb{P}\left(|\frac{1}{N}\sum_{k=1}^{N} Q_k - \mathcal{Q}_{\mathrm{avg}}| \geq \varepsilon \right) \leq 2e^{\frac{-2k\varepsilon^2}{(b-2)^2}}, \tag{6.38}$$

where $[a,b] \subseteq [0,1]$ is the interval where the survival probabilities $\mathcal{Q}$'s can take value. For example, supposing that we want a fidelity within 0.99 accuracy with failing probability $\varepsilon = 10^{-3}$, and assuming that $b - a = 0.2$, we will need approximatively $k \approx 7 \cdot 10^4$ different sequences. While promising compared against quantum process tomography, this bounds seems to be significantly loose according to suggestions from numerical simulations and experimental realizations, where up to 100 sequences are used for each value ℓ [111], and with respect to the analytical investigation in ref. [112] based on the irreducible decomposition of the two-fold tensor representation of the Clifford group [113].

6.2 Interleaved randomized benchmarking and other protocols

As previously mentioned, one of the major drawbacks of randomized benchmarking is that it estimates only an error averaged over the gate set: gate-dependent components are not retrieved. To overcome the issue, ref. [108] propose a protocol – called *interleaved randomized benchmarking* – to characterize the error of an arbitrary individual Clifford gate $C^\star$ by slightly modifying the random sequence. After using the usual randomized benchmarking method described above and obtaining the depolarization parameter p, one should run a second protocol where the random gates are interlaced with the gate $C^\star$, that is,

$$\mathcal{S}^\star_{\mathbf{k}_\ell} = \Lambda_{\text{inv}} \circ \mathcal{C}_{\text{inv}} \circ \mathcal{C}^\star \circ \Lambda_{\mathcal{C}^\star} \circ \Lambda_{k_\ell} \circ \mathcal{C}_{k_\ell} \circ \cdots \circ \mathcal{C}^\star \circ \Lambda_{\mathcal{C}^\star} \circ \Lambda \circ \mathcal{C}_{k_2} \circ \mathcal{C}^\star \circ \Lambda_{\mathcal{C}^\star} \circ \Lambda \circ \mathcal{C}_{k_1} \tag{6.39}$$

$$= \Lambda_{\text{inv}} \circ \mathcal{C}_{\text{inv}} \left(\bigcirc_{j=1}^{\ell} \mathcal{C}^\star \circ \Lambda_{\mathcal{C}^\star} \circ \Lambda_{k_j} \circ \mathcal{C}_{k_j} \right), \tag{6.40}$$

where the inverse gate is given by the composition of all random gates and interleaved $\mathcal{C}^\star$ operators and where we consider the noisy implementation of $\Lambda_{\mathcal{C}^\star}$ of $\mathcal{C}^\star$ and $\Lambda_{k_j} = \Lambda + \delta\Lambda_{k_j}$ (i.e., considering a small perturbation) of the random gate $\mathcal{C}_{k_j}$. Defining $\mathcal{D}_{k_1} \coloneqq \mathcal{C}_{k_1}$ and recursively $\mathcal{D}_{k_{j+1}} \coloneqq \mathcal{C}_{k_{j+1}} \circ \mathcal{C}^\star \circ \mathcal{D}_{k_j}$, we can re-forumulate the last expression as

$$\mathcal{S}^\star_{\mathbf{k}_\ell} = \Lambda_{\text{inv}} \bigcirc_{j=1}^{\ell} \left(\Lambda_{\mathcal{C}^\star} \circ \Lambda_{k_j} \right)^{\text{dep}}. \tag{6.41}$$

Again, measuring the survival probability $\mathcal{Q}_{\mathbf{k}^\star_\ell}$ and averaging over a number $N^\star$ of different random sequences for different values of ℓ, one can fit the zero (6.34) or first order fitting model (6.35) depending whether one assumes the Λ_{k_j} to be the same for all random gates or having a slight gate-dependent perturbation, obtaining a second depolarization parameter $p_{\mathcal{C}^\star}$. Together with p, it is the possible to estimate the interval where $\mathbb{E}(\mathcal{F}_{\Lambda_{\mathcal{C}^\star}, \mathcal{I}})$ lies, which is within the two values

$$\mathbb{E}(\mathcal{F}_{\text{min,max}}) = \frac{(d-1)p_{\mathcal{C}^\star}/p}{d} \pm I \quad \text{where} \qquad I = \min \begin{cases} \frac{(d-1)(|p-p_{\mathcal{C}^\star}/p|+1-p)}{d} \\ \frac{2(d^2-1)(1-p)}{pd^2} + \frac{4\sqrt{1-p}\sqrt{d^2-1}}{p}. \end{cases} \tag{6.42}$$

A similar interleaved randomized benchmarking protocol was proposed in ref. [109], this time to individually benchmark the T-gate by alternating it with randomly selected gates from the Clifford and Pauli groups.

The Clifford group or more in general 2-design distributions are the common choice for randomized benchmarking since they satisfy the depolarizing condition given by Lemma 6.7 at the core of the fitting model derivation. However, a few proposal have been done to go past the two design requirement [114]; one of this is the *dihedral benchmarking* [106] using a small group generated by rotations around the $\hat{z}$-axis of the Bloch sphere coupled with the reflection on $\hat{x}\hat{z}$-plane, that is,

$$\mathbb{D}_j \coloneqq \langle R_j, X \rangle \qquad \text{where} \quad R_j \coloneqq \exp\{\mathrm{i}2\pi/j \; Z\}. \tag{6.43}$$

By choising $j = 8$, we obtain a group containing the T-gate which, as we mentioned in Section 3.2, together with the gate H and CNOT, or more generally together with the Clifford group, builds a universal gate set. While lifting the two-design requirement, the method still assumes the noise to be gate-independent for all gates in the group, and is limited to single-qubit implementations.

The protocol is given by the following steps.

(1) Choose at random two binary strings of length ℓ, $\mathbf{z} = (z_1, \ldots, z_\ell) \in \mathbb{Z}_j^\ell$ and $\mathbf{x} = (x_1, \ldots, x_\ell) \in \{0, 1\}^\ell$.

(2) Prepare a system in an arbitrary initial state ρ.

(3) Apply the sequence

$$\mathcal{S}_{\mathbf{x},\mathbf{z}} = \mathcal{G}_{\text{inv}} \circ \mathcal{R}_j^{z_\ell} \circ \mathcal{X}^{x_\ell} \circ \mathcal{R}_j^{z_{\ell-1}} \circ \mathcal{X}^{x_{\ell-1}} \circ \cdots \circ \mathcal{R}_j^{z_1} \circ \mathcal{X}^{x_1}, \tag{6.44}$$

where $\mathcal{G}_{\text{inv}} \coloneqq \mathcal{X}^{b_1} \circ \mathcal{Z}^{b_2} \bigcirc_{k=\ell}^{1} \left[\mathcal{R}_j^k \circ \mathcal{X}^k\right]^\dagger$, for a choice of $b_1, b_2 \in \mathbb{Z}_2$ explained below.

(4) Perform a POVM E.

(5) Repeat step (3) and (4) sufficiently many times in order to obtain a valid statistic of the survival probability to a desired precision.

(6) Repeat the previous steps for a large enough number of different random sequences $\mathbf{z}$ and $\mathbf{x}$ to obtain the average survival probability.

(7) Repeat the procedure for different value of ℓ and, for $b_1 = b_2 = 0$, fit the average survival probability data into the model

$$\mathcal{Q}_{\text{avg}}(\ell, b_1 = 0, b_2 = 0) = Ap_0^\ell + Bp_1^\ell + C \tag{6.45}$$

where A, B, C are constants absorbing SPAM errors and p_0, p_1 are the parameters characterizing the average gate fidelity through the formula given in eq. (2.30) and the connection between the fidelity of a channel and its twirled version in Lemma 6.2.

$$\mathbb{E}(\mathcal{F}_{\Lambda,\mathcal{I}}) = \frac{1}{2} + \frac{1}{6}(p_0 + 2p_1). \tag{6.46}$$

One can also exploit linear combinations of the average survival probabilities for different values of b_1 and b_2 to simplify the model into a linear fitting problem, that is,

$$\begin{aligned} &\mathcal{Q}_{\text{avg}}(\ell, b_1 = 0, b_2 = 0) + \mathcal{Q}_{\text{avg}}(\ell, b_1 = 0, b_2 = 1) \\ &- \mathcal{Q}_{\text{avg}}(\ell, b_1 = 1, b_2 = 0) - \mathcal{Q}_{\text{avg}}(\ell, b_1 = 1, b_2 = 1) \\ &= 4Ap_0^\ell \end{aligned} \tag{6.47}$$

and

$$\mathcal{Q}_{\text{avg}}(\ell, b_1 = 0, b_2 = 0) - \mathcal{Q}_{\text{avg}}(\ell, b_1 = 0, b_2 = 1) = 2Bp_1^\ell. \tag{6.48}$$

The derivation of the fitting model follows the same idea of the one with Clifford gates without perturbation terms. Accounting for a noise channel Λ – identical for all operators – representing the imperfect implementation of the ideal unitary gates, we write

$$\mathcal{S}_{\text{avg}} = \Lambda \circ \mathcal{X}^{b_1} \circ \mathcal{Z}^{b_2} \circ \left(\Lambda^{\mathbb{D}_j}\right)^{\circ \ell}, \tag{6.49}$$

where $\Lambda^{\mathbb{D}_j}$ is the twirling of the noise channel Λ with respect to the dihedral group $\mathbb{D}_j$.
Note that by construction $\Lambda^{\mathbb{D}_j}$ commutes with all elements of the diehedral group $\mathbb{D}_j$ and therefore, from Schur's Lemma, it can be decomposed into a block-diagonal form with respect to the irreducible representations of the dihedral group, given by

$$\begin{aligned} &R_j^k\, X^x \to 1 && \text{trivial} \\ &R_j^k\, X^x \to \begin{pmatrix} \cos(2\pi k/j) & (-1)^{x+1}\sin(2\pi k/j) \\ \sin(2\pi k/j) & (-1)^x \cos(2\pi k/j) \end{pmatrix} && \text{faithful} \\ &R_j^k\, X^x \to (-1)^x && \text{parity} \end{aligned}$$

for $k \in \mathbb{Z}_j,\ x \in \{0, 1\}$.
The matrix representation of the twirled channel includes each of these representations exactly once, and so can be written (using a basis transformation bringing the matrix group in block-diagonal form) as a diagonal matrix, i.e.,

$$\Lambda^{\mathbb{D}_j} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & p_1 & 0 & 0 \\ 0 & 0 & p_1 & 0 \\ 0 & 0 & 0 & p_0 \end{pmatrix}, \tag{6.50}$$

so that in Pauli-Liouville presentation the average gate sequence can be expressed as

$$\mathcal{S}_{\text{avg}} = \Lambda \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & (-1)^{b_2} p_1^{\ell} & 0 & 0 \\ 0 & 0 & (-1)^{b_1+b_2} p_1^{\ell} & 0 \\ 0 & 0 & 0 & (-1)^{b_1} p_0^{\ell} \end{pmatrix} \tag{6.51}$$

and hence the fitting model for the survival probability is

$$\mathcal{Q}_{\text{avg}}(\ell, b_1 = 0, b_2 = 0) = (-1)^{b_1} A p_0^{\ell} + ((-1)^{b_1+b_2} B_1 + (-1)^{b_2} B_2) p_1^{\ell} + C. \tag{6.52}$$

The protocol can also be used to estimate an upper bound of the error implementation relative to the T-gate, using the fact that the group $\mathbb{D}_8$ can be divided in two cosets, namely, $\mathbb{D}_4$, generated by the Pauli matrix X and the phase gate P, and $\mathrm{T} \cdot \mathbb{D}_4$. Precise details are given in ref. [106].

At the beginning of this chapter, we mentioned quantum process tomography as a non-scalable and SPAM sensitive method to fully characterize a quantum channel. In ref. [115], the authors describe a method to make use of randomized benchmarking to obtain tomographic information about the unital part of a quantum channel. The link between the two techniques allows to reconstruct the unital part of a CPTP map in a SPAM robust way, inheriting the advantages of the randomized benchmarking protocol, but at the same time permitting the characterization of arbitrarily large errors. The idea at the heart of this approach is that any unitary map can be written as a linear combination of elements in the Clifford group and at the same time this group spans the unital subspace of quantum CPTP maps (cfr. [115, Lemma IV.1]).

Using a modified version of interleaved randomized benchmarking protocol one can obtain the average gate fidelity between a Clifford gate $\mathcal{C}_j$ and an arbitrary quantum channel $\mathcal{E}$. By applying in sequence

$$\mathcal{S}_{\mathbf{k}_\ell} = \mathcal{C}_{k_\ell}^{\dagger} \circ \mathcal{C}_j^{\dagger} \circ \mathcal{E} \circ \mathcal{C}_{k_\ell} \circ \cdots \circ \mathcal{C}_{k_2}^{\dagger} \circ \mathcal{C}_j^{\dagger} \circ \mathcal{E} \circ \mathcal{C}_{k_2} \circ \mathcal{C}_{k_1}^{\dagger} \circ \mathcal{C}_j^{\dagger} \circ \mathcal{E} \circ \mathcal{C}_{k_1}, \tag{6.53}$$

according to the other protocols, this leads to an estimate of the average fidelity $\mathbb{E}(\mathcal{F}_{\mathcal{E},\mathcal{C}_j})$ and hence, using eq. (2.30), to $\mathrm{Tr}[\mathcal{E}\mathcal{C}_j^{\dagger}]$ for any Clifford gate. This allows to reconstruct the unital part of $\mathcal{E}$. Additionally, assuming that a unitary channel $\mathcal{U}$ can written as $\mathcal{U} = \sum_j \beta_j C_j$, then one can retrieve by linear combination of the randomized benchmarking data an estimate of $\mathrm{Tr}[\mathcal{E}\mathcal{U}^{\dagger}]$ (and hence again the average gate fidelity $\mathbb{E}(\mathcal{F}_{\mathcal{E},\mathcal{U}})$). For a single T-gate, this will require the measurement of three Clifford gates, since

$$\mathrm{T} = \frac{1}{2}\mathbb{1} + \frac{1-\sqrt{2}}{\sqrt{2}} Z + \frac{1}{\sqrt{2}} \mathrm{e}^{-\mathrm{i}\frac{\pi}{4}Z}. \tag{6.54}$$

To obtain $\mathbb{E}(\mathcal{F}_{\mathcal{E},\mathcal{U}})$ with precision ε with probability $1 - \delta$, one requires (cfr. [115, section VI B])

$$O\left(N_{\mathcal{U}} \left(\frac{\sum_j |\beta_j|}{\varepsilon} \right)^4 \log \frac{N_{\mathcal{U}}}{\delta} \right) \tag{6.55}$$

samples (without taking into account the number of measurements for the same sequence to obtain the survival probability), where $N_{\mathcal{U}}$ is the number of Clifford gates with non-zero linear coefficient β_j. Note that this number may scale as $O(d^{10})$ when $N_{\mathcal{U}} = O(d^2)$.

In a similar spirit, we propose a scheme aiming at loosing some of the strong assumptions regarding the noise channel while maintaining robustness against SPAM errors at the cost of increasing the classical computational effort.

6.3 Randomized benchmarking for individual quantum gates

In our work *Randomized benchmarking for individual quantum gates* [42], we present a new protocol that is expected to be significant in two ways: in the first place, it is suitable to benchmark quantum gates *individually*, including ones that are outside the Clifford group. Since schemes for universal quantum computing necessarily make use of such gates, this constitutes an important step forward. At the same time, we do not require twirling over the full Clifford group or a 2-design, but only over a relatively small local symmetry group coupled with transpositions gates. The novel idea in the present work is to exploit the symmetry of the quantum gate itself in an appropriate fashion and hence reduce the amount of necessary quantum resources and computational power. To achieve this goal, we harness advanced tools from representation theory to arrive at schemes that require similar physical operations, but not tied to the 2-designs property and which can make predictions beyond known prescriptions.
More conceptually speaking, and putting this contribution into a broader context, we show that one can interpolate between common assumptions made when characterising quantum processes: in other words, there is "room in the middle" between full quantum process tomography, which is largely assumption-free but comes along with daunting resource requirements, and conventional randomised benchmarking, which requires significantly less effort and is also robust against state preparation and measurement (SPAM) errors, although it makes strong assumptions. We believe that this conceptual insight into the ontology of assumptions when characterising quantum processes subject to unknown quantum noise is equally important.

The setting

In the following, we are going to describe a protocol that provides the *average gate fidelity* of the noisy channel Λ which characterizes the imperfect implementation $\widetilde{U}$ of a target ideal unitary gate U,

$$\mathbb{E}(\mathcal{F}_{\widetilde{\mathcal{U}},\mathcal{U}}) = \mathbb{E}(\mathcal{F}_{\Lambda,\mathcal{I}}) \coloneqq \int_{\text{Haar}} \text{Tr}\left[|\phi\rangle\langle\phi| \, \Lambda(|\phi\rangle\langle\phi|)\right] \, \mathrm{d}\phi. \tag{6.56}$$

It is key to our method to explicitly exploit the local and permutation symmetries of U, allowing for a drastic reduction of the fitting parameters and also inheriting robustness with respect to SPAM errors. In this way, fitting models of well-known randomized benchmarking protocols can be uplifted to this setting involving fewer assumptions.
Throughout this section, we consider quantum gates acting on n-qubit systems and we are interested in benchmarking the accuracy of their implementation in a quantum circuit making use of the protocol that we are going to explain later on. The method is particularly suitable for gates consisting of tensor products of local gates, hence admitting additional symmetries with respect to the exchange of qubit subsystems and it is applicable in those situations of single layers of local unitary quantum gates. For concreteness and as a guiding example, we shall put emphasis on single layers of circuits whose gates consist of tensor compositions of the T-gate with other "popular" gates belonging to the Clifford group $\mathcal{C}_n$, namely, H, P and CNOT that we

write once again

$$\mathrm{H} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad \mathrm{T} = \begin{pmatrix} \mathrm{e}^{-\mathrm{i}\pi/8} & 0 \\ 0 & \mathrm{e}^{\mathrm{i}\pi/8} \end{pmatrix},$$
$$\mathrm{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \qquad \mathrm{P} = \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{i} \end{pmatrix}, \tag{6.57}$$

leading to a universal set. More generally, the protocol is suitable to benchmark tensor products of local gates consisting in arbitrary rotations around $\hat{x}$,$\hat{y}$,$\hat{z}$-axes of the Bloch sphere. In the following, we will denote with $U = U_1 \otimes \cdots \otimes U_n$ the multi-qubit unitary operator – with U_j acting on qubit j – that we intend to benchmark, acting as a single layer of a unitary circuit. At the heart of the analysis will be its symmetry group, constructed from the symmetries of the local gates U_j composing U and the permutations of qubits which the local gate acts upon. More precisely, we choose the local symmetry group A_j of U_j as the subset of the single qubit Clifford group whose elements commute with U_j. As an example, for $U_j = \mathrm{T} = \exp(-\mathrm{i}\pi Z/8)$, the group of local symmetries is given by

$$A_{\mathrm{T}} \coloneqq \{\, U \in \mathcal{C}_n : U^\dagger Z\, U = Z \,\}\,. \tag{6.58}$$

This is an abelian group of 4 elements isomorphic to the cyclic group of order 4, $\mathbb{Z}_4$. The set of all possible permutations interchanging qubits affected by the same local gate is another symmetry group of the target unitary U; taking a pratical example, for the gate $U = \mathrm{T} \otimes \mathrm{H} \otimes \mathrm{T} \otimes \mathrm{H} \otimes \mathrm{T}$, this group is isomorphic to $S_3 \times S_2$, i.e., all permutations of the first, third and fifth subsystems combined with the transposition of the second and forth subsystems. The full symmetry group G is then obtained through the *semi-direct product* $A_n \rtimes \Pi$, where A_n is the *direct product* of the local symmetry groups A_j constructed by the Kronecker product of the respective elements, and Π is the representation of the subgroup of S_n consisting of all allowed permutations of the qubits subsystems.

Remark

In order to apply our full protocol and combine the group A_n with Π, all local symmetry groups A_j must be abelian. This is indeed a necessary condition to reconstruct the irreducible representations of the full group G with the sole knowledge of the composing groups, as we will discuss in a dedicated paragraph. Fortunately, this is the case for the symmetry of the gates in eq. (6.57) and all other rotations around Bloch axes. Should the local symmetry groups not be all abelian, the protocol is still valid setting $G = A_n$, i.e., without considering permutation symmetries.

Assumptions and physical motivation

We denote with calligraphic letters the channel acting by gate conjugation on density operators, i.e., $\mathcal{U}(\rho) \coloneqq U^\dagger \rho\, U$, and the noisy implementation of the idealized gate channel $\mathcal{U}$ as $\widetilde{\mathcal{U}} \coloneqq \Lambda_{\mathcal{U}} \circ \mathcal{U}$, i.e., we account for a gate-dependent error channel $\Lambda_{\mathcal{U}}$ whose average fidelity we want to characterize with the proposed protocol. Since randomized benchmarking can be interpreted as a trade-off between the level of characterization of the noise channel and the amount of physical and computational resources needed, we will make the following assumption: the Pauli-Liouville representation of the twirled noise channel, $\Lambda_{\mathcal{U}}^G \coloneqq |G|^{-1} \sum_{j \in \mathbb{N}_{|G|}} \mathcal{G}_j^\dagger \Lambda_{\mathcal{U}}\, \mathcal{G}_j$, is *almost jointly diagonalizable* with the target unitary channel $\mathcal{U}$ (where again we consider it as a matrix in

Pauli-Liouville representation), in the spirit of ref. [116]. This means that when the matrix $\mathcal{U}$ is brought to diagonal form by some unitary transformation V, the off-diagonal elements of $\Lambda_{\mathcal{U}}^{G}$ under the same transformation are small. This is true in two cases. The first possibility is that both $\mathcal{U}$ and the twirled noise $\Lambda_{\mathcal{U}}^{G}$ are diagonalizable simultaneously in a certain basis, e.g., when the decomposition of the representation of the symmetry group into irreducible representations has no multiplicity: in this case, both $\mathcal{U}$ and the twirled noise $\Lambda_{\mathcal{U}}^{G}$ are "forced" to assume a diagonal form with respect to the irreducible subspaces. If this is not the case, then $\Lambda_{\mathcal{U}}^{G}$ assumes a sparse form with some off-diagonal entries, which have to be small with respect to the diagonal elements. This is fulfilled whenever the original noise channel $\Lambda_{\mathcal{U}}$ related to the implementation of the gate $\mathcal{U}$ was almost jointly diagonalizable to begin with, or put in another perspective (see again ref. [116]), it is almost commuting. This assumption is valid when the gate U is generated by a Hamiltonian H applied for some run time t [117], i.e., $U = \mathrm{e}^{-\mathrm{i}Ht}$, which can be perturbed for a small fraction of the time, or be applied for too much or too little time. More precisely, assume that the gate $\widetilde{U}$, which is the physical realization of the ideal gate U, is obtained during the application of some Hamiltonian H, perturbed (we denote the perturbed Hamiltonian as R) for a fraction of time Δt, i.e., $\widetilde{U} = \mathrm{e}^{-\mathrm{i}(R\Delta t + Ht)}$. Using the *Zassenhaus formula* [118], we can rewrite the implemented gate as

$$\widetilde{U} = \mathrm{e}^{-\mathrm{i}HT}\mathrm{e}^{-\mathrm{i}R\Delta t}\prod_{n=2}^{\infty} e^{C_n(HT,R\Delta t)} \tag{6.59}$$

$$= \mathrm{e}^{-\mathrm{i}HT}\left(\mathbb{1} - \mathrm{i}R\Delta t + \S\Delta t\right) + O(\Delta t^2), \tag{6.60}$$

where

$$\S \coloneqq \sum_{n=2}^{\infty} c_n [\underbrace{H,[H,\ldots,[H}_{n-1 \text{ times}},R]\ldots]]T^{n-1}, \tag{6.61}$$

with the Zassenhaus coefficients c_n that can be recursively calculated for instance as in ref. [118]. This implies that the off-diagonal elements of the matrix $\widetilde{U}$ – computed in the eigenbasis of U – are of order Δt, justifying our assumption on the noise Λ. Furthermore, we ask the gates belonging to the symmetry group G to be implementable with high accuracy. These gates either perform a permutation of the subspaces of the system or belong to the Clifford group and so can be for instance benchmarked with the well-known protocols [45, 105, 107] to guarantee high fidelity.

The protocol

We propose a slightly modified version of the previous protocols. We apply in succession channels defined by the gate U after the one induced by a gate uniformly drawn at random from the symmetry group G. Note that, unlike previous protocols, the target gate U is not part of the twirling group G: this is one of the reason why one can benchmark arbitrarily small rotations over the Bloch axes with a relatively small number of gates. For a fixed sequence length ℓ, the protocol is constituted by the following steps:

(1) Prepare an initial state ρ (e.g., $\rho = |0\rangle\langle 0|$).

(2) Draw a random sequence $\mathbf{k}_\ell = (k_1, \ldots, k_\ell) \in \mathbb{N}_{|G|}^{\ell}$.

(3) Apply the following operation generated by the symmetry operations $\mathcal{G}_{k_i}$ to the initial state ρ

$$\mathcal{C}_{\mathbf{k}_\ell}(\rho) = \mathcal{G}_{\mathrm{inv}} \circ \mathcal{U} \circ \mathcal{G}_{k_\ell} \circ \cdots \circ \mathcal{U} \circ \mathcal{G}_{k_1}, \tag{6.62}$$

where $\mathcal{G}_{\mathrm{inv}} \coloneqq \mathcal{G}_{k_1}^{\dagger} \circ \cdots \circ \mathcal{G}_{k_\ell}^{\dagger}$ is the channel given by the composition of the inversion of all previous random gates channels.

(4) Perform a POVM E and measure the *survival probability* $\mathcal{Q}_{\mathbf{k}_\ell} = \mathrm{Tr}[E\,\mathcal{C}_{\mathbf{k}_\ell}(\rho)]$. To obtain an appropriate precision for $F_{\mathbf{k}_\ell}$, this step has to be repeated sufficiently many times.

(5) Repeat the previous step for sufficiently many (say N) random sequences $\mathbf{k}_{\ell,1}, \ldots, \mathbf{k}_{\ell,N}$ of length ℓ. Then, calculate the *sequences survival probability*

$$\mathcal{Q}_{\mathrm{seq}}(\ell,\rho) = \frac{1}{N}\sum_{\mathbf{k}_\ell} \mathcal{Q}_{\mathbf{k}_\ell} = \frac{1}{N}\sum_{\mathbf{k}_\ell} \mathrm{Tr}[E\,\mathcal{C}_{\mathbf{k}_\ell}(\rho)]. \tag{6.63}$$

The number K of random sequences should be chosen such that

$$\mathcal{Q}_{\mathrm{seq}} \approx \mathcal{Q}_{\mathrm{avg}}\ , \tag{6.64}$$

where $\mathcal{Q}_{\mathrm{avg}}$ is the survival probability averaged over all possible sequences. The choice can be motivated by an analysis on the variance of the random variable $\mathcal{Q}$, with $\mathcal{Q}_{\mathbf{k}_\ell}$ being a *realization* and $\mathcal{Q}_{\mathrm{avg}}$ the *mean* of the distribution.

(6) Repeat the previous steps for different lengths ℓ.

(7) Insert $\mathcal{Q}_{\mathrm{seq}}$ into the zero- or first-order fitting model,

$$\mathcal{Q}^{(0)}_{\mathrm{avg}}(\ell,\rho) = \sum_{j=1}^{4^n} (\lambda_j\, d_j)^\ell \xi_j \tag{6.65}$$

$$\mathcal{Q}^{(1)}_{\mathrm{avg}}(\ell,\rho) = \mathcal{Q}^{(0)}_{\mathrm{avg}}(\ell,\rho) + \sum_{i\neq j}\sum_{p=0}^{\ell-1} (\lambda_i d_i)^p\,(\lambda_j d_j)^{\ell-p-1}\,\zeta_{i,j} \tag{6.66}$$

where $\{\,d_j\,\}$ are the eigenvalues of the target matrix $\mathcal{U}$ and $\xi_j, \zeta_{i,j}$ are constants absorbing SPAM errors, and obtain the parameters $\{\,\lambda_j\,\}$ characterizing the average gate fidelity of $\widetilde{\mathcal{U}}$ with respect to $\mathcal{U}$.

The fitting model

Considering a noise channel Λ (where we now drop the subscript $\mathcal{U}$ to lighten notation) at each implementation of $\mathcal{U}\circ\mathcal{G}_k$, we can write

$$\mathcal{C}_{k_\ell} = \Lambda' \circ \mathcal{G}_{\mathrm{inv}} \bigcirc_{t=\ell}^{1} \Lambda \circ \mathcal{U} \circ \mathcal{G}_{k_t}. \tag{6.67}$$

Note that the error channel Λ' characterizing the implementation of $\mathcal{G}_{\mathrm{inv}}$ can be different from the error for the implementation of $\mathcal{U}\circ\mathcal{G}_{k_t}$. Now, defining recursively $\mathcal{B}_{k_t} := \mathcal{G}_{k_t}\circ\mathcal{B}_{k_{t-1}}$ with $\mathcal{B}_{k_1} = \mathcal{G}_{k_1}$, and using the invariance of U under the action of G, we can rewrite

$$\mathcal{C}_{k_\ell} = \Lambda' \bigcirc_{t=\ell}^{1} \mathcal{B}_{k_t}^\dagger \circ \Lambda \circ \mathcal{B}_{k_t} \circ \mathcal{U}. \tag{6.68}$$

When averaging over all possible sequences, we get $\mathcal{C}_{\mathrm{avg}} = \Lambda' \bigcirc_{t=\ell}^{1} \Lambda^G \circ \mathcal{U}$, where $\Lambda^G := |G|^{-1}\sum_{j\in\mathbb{N}_{|G|}} \mathcal{B}_j^\dagger \circ \Lambda \circ \mathcal{B}_j$ is now like U invariant with respect to the action of G. At this point we use Schur's Lemma to diagonalize the Pauli-Liouville representations of $\mathcal{U}$ and Λ^G. If for instance the decomposition of G into irreducible representations does not contain any multiplicity, the two matrices are simultaneously diagonalizable. If conversely multiple of the same irreducible representations occurs, in general there is no basis which brings both into a diagonal form and so,

when diagonalizing the matrix $\mathcal{U}$, the other will assume a block form, where each of these blocks corresponds to an irreducible representation. For the moment, we assume that both matrices are simultaneously diagonalizable. Writing $d_1, \ldots, d_{4^n}$ and $\lambda_1, \ldots, \lambda_{4^n}$ to denote the diagonal elements of $\mathcal{U}$ and Λ^G respectively, we have

$$\mathcal{Q}^{(0)}_{\text{avg}}(\ell, \rho) = \sum_{j=1}^{4^n} (\lambda_j \, d_j)^\ell \xi_j, \tag{6.69}$$

with $\xi_j \coloneqq \text{Tr}[E \, \Lambda'(v_j)] \, \langle \rho, v_j \rangle$ absorbing the state preparation and measurement errors and where $\{ \, v_j \, \}_j$ is the set of the symmetry adapted vectors diagonalizing $\mathcal{U}$.
We can now consider the case where also off-diagonal matrix entries are present. The main reason of the protocol is that, by twirling the error channel over the symmetry group G, we reduce the number of these off-diagonal matrix entries and so drastically decrease the amount of parameters in the fitting model. Let us write $\Lambda^G = \Lambda_0 + \Lambda_{\text{off}}$, with Λ_0 being jointly diagonalizable with $\mathcal{U}$. Provided $\Lambda_{\text{off}} = \{ \, \mu_{i,j} \, \}_{i \neq j}$ to be "small" (i.e., the second order perturbation being negligible), we can consider the first order model

$$\mathcal{Q}^{(1)}_{\text{avg}}(\ell, \rho) = \mathcal{Q}^{(0)}_{\text{avg}}(\ell, \rho) + \sum_{i \neq j} \sum_{p=0}^{\ell-1} (\lambda_i d_i)^p \, (\lambda_j d_j)^{\ell-p-1} \, \zeta_{i,j}, \tag{6.70}$$

with $\zeta_{i,j} \coloneqq \mu_{i,j} d_j \;\; \text{Tr}[E \, \Lambda'(v_i)] \langle \rho, v_j \rangle$. This expression may be re-formulated into a simpler form, e.g., using the geometric series formula we obtain

$$\mathcal{Q}^{(1)}_{\text{avg}}(\ell, \rho) = \mathcal{Q}^{(0)}_{\text{avg}}(\ell, \rho) + \sum_{i \neq j} \frac{(\lambda_j d_j)^\ell - (\lambda_i d_i)^\ell}{\lambda_j d_j - \lambda_i d_i} \zeta_{i,j}. \tag{6.71}$$

As already mentioned, since we twirled over the symmetry group and so Λ^G is block-diagonal, a number of $\mu_{i,j}$ can be set to 0 in advance. More precisely, when a representation of the symmetry group is written as a direct sum of irreducible representations as

$$\pi(g) = \bigoplus_{\alpha \, \text{irrep}} \mathbb{1}_{m_\alpha} \otimes \pi^\alpha(g), \tag{6.72}$$

where m_α is the multiplicity of the irreducible representation π^α, two matrices X and Y which are both commuting with $\pi(g)$ assume the form

$$X = \bigoplus_\alpha x^\alpha \otimes \mathbb{1}_{\dim \alpha} \qquad \text{and} \qquad Y = \bigoplus_\alpha y^\alpha \otimes \mathbb{1}_{\dim \alpha}, \tag{6.73}$$

where x^α, y^α are square matrices with $\dim x^\alpha = \dim y^\alpha = m_\alpha$. One can then choose a basis such that all x^α are diagonal (so that X will assume a diagonal form), while Y will maintain a similar form $Y = \bigoplus_\alpha \widetilde{y}^\alpha \otimes \mathbb{1}_{\dim \alpha}$. Hence, in our case, while diagonalizing $\mathcal{U}$ (from the Pauli-Liouville representation), Λ^G maintains a form as in eq. (6.73).

Construction of irreducible representations of semi-direct product groups

As we have discussed in the introduction of this section, it is possible to couple two groups to construct a new one using direct and semi-direct products. We can also obtain all irreducible representations of the latter using knowledge about irreducible representations of the original two groups alone. For direct product, the procedure is straightforward, namely,

Theorem 6.3 ([119], Theorem 10, Chapter 3)**.** *Each irreducible representation of a direct group $G_1 \times G_2$ is isomorphic to a representation $\pi_1 \otimes \pi_2$ with π_1 and π_2 being irreducible representations of group G_1 and G_2 respectively.*

For a group generated by a semi-direct product $N \rtimes H$, a more sophisticated machinery is needed (cfr. refs. [119, 120]), and works only if the normal subgroup N is also *abelian*, i.e., all elements commute with respect to the group operation. Assuming N to be abelian, its irreducible representations $\{\chi_i\}_i$ are 1-dimensional and carry an action of G by

$$g \cdot \chi_i(a) = \chi_i(g^{-1}ag) \qquad \forall a \in N \text{ and } g \in G. \tag{6.74}$$

Now, consider the orbits of the characters induced by the action of H and choose a set of representatives $\{\chi_r\}_r$. For each r, let H_r be the stabilizer subgroup of χ_r in H and then define $G_r = N \cdot H_r$. Now extend χ_r to G_r by

$$\chi_r(ah) = \chi(a) \qquad \forall a \in N \text{ and } h \in H_r. \tag{6.75}$$

Let θ be an irreducible representations of H_r and lift it to an irreducible representation $\widetilde{\theta}$ of G_r through the canonical projection $P : G_r \to G_r/N$. As a final step, compose the two representations and obtain a representation $\rho_{r,\widetilde{\theta}}$ of the group G by *induction*, i.e., $\rho_{r,\widetilde{\theta}} = \mathrm{Ind}_{G_r}^{G}(\chi_r \cdot \widetilde{\theta})$. From ref. [119, Proposition 25], we know that the so constructed representations $\rho_{r,\widetilde{\theta}}$ are irreducible and exhaust all irreducible representations of G. Since we will only need the characters $\chi_{\rho_{r,\widetilde{\theta}}}$ of the irreducible representations of G to apply Schur's Lemma in our protocol, we will not elaborate on what induced representations are. To obtain the sought characters, it suffices to make use of a Mackey-type formula

$$\chi_{\rho_{r,\widetilde{\theta}}}(s) = \frac{1}{|G_r|} \sum_{\substack{g \in G \\ g^{-1}sg \in G_r}} \chi_r \cdot \chi_{\widetilde{\theta}}\,(g^{-1}sg). \tag{6.76}$$

Connecting to the average gate fidelity

To retrieve from the fitted parameters the actual quantity that we aim to estimate, that is, the average gate fidelity, we recall eq. (2.30) in Section 2.1 for an orthonormal basis $\{V_j\}_j$ such that $\mathrm{Tr}[V_j V_k] = d$,

$$\mathbb{E}(\mathcal{F}_{\mathcal{E},\mathcal{I}}) = \frac{dF_{\mathrm{ent}}(\mathcal{E}) + 1}{d+1} = \frac{\sum_j \mathrm{Tr}[V_j^\dagger \mathcal{E}(V_j)] + d^2}{d^2(d+1)}, \tag{6.77}$$

so that the average gate fidelity of the twirled error channel Λ^G is related to the parameters $\{\lambda_j\}_j$ obtained in the fitting model in eqs. (6.65)-(6.71) by

$$\mathbb{E}(\mathcal{F}_{\Lambda^G,\mathcal{I}}) = \frac{\sum \lambda_j + d}{d(d+1)}. \tag{6.78}$$

Now the question is what information about the original noise channel we can extract from the twirled channeld Λ^G. In fact their average gate fidelities are the same, since the entanglement fidelity is invariant under twirling over the symmetry group G. Let us rewrite from eq. (2.29)

$$F_{\mathrm{ent}}(\Lambda^G) = d^{-3} \sum_j \mathrm{Tr}[V_j^\dagger \Lambda^G(V_j)] \tag{6.79}$$

$$= \frac{d^{-3}}{|G|} \sum_{k=1}^{|G|} \sum_j \mathrm{Tr}[V_j^\dagger g_k^\dagger \Lambda(g_k V_j g_k^\dagger) g_k] \tag{6.80}$$

$$= \frac{d^{-3}}{|G|} \sum_{k=1}^{|G|} \sum_j \mathrm{Tr}[(W_j^k)^\dagger \Lambda(W_j^k)], \tag{6.81}$$

where we denote $W_j^k = g_k V_j g_k^\dagger$ and used cyclicity of the trace. Since W_j^k is again an orthogonal basis with respect to the Hilbert-Schmidt inner product (i.e., $\mathrm{Tr}[(W_{j'}^k)^\dagger W_j^k] = d\,\delta_{j'j}\ \forall k$), then $d^{-3}\sum_j \mathrm{Tr}[(W_j^k)^\dagger \Lambda(W_j^k)] = F_{\mathrm{ent}}(\Lambda)$ so that

$$F_{\mathrm{ent}}(\Lambda^G) = \frac{1}{|G|}\sum_{k=1}^{|G|} F_{\mathrm{ent}}(\Lambda) = F_{\mathrm{ent}}(\Lambda) \tag{6.82}$$

and hence

$$\mathbb{E}(\mathcal{F}_{\Lambda,\mathcal{I}}) = \mathbb{E}(\mathcal{F}_{\Lambda^G,\mathcal{I}}). \tag{6.83}$$

Characterizing the error of the single gate U

In order to recover the fidelity of the gate U from the noise Λ, which originates from the composition of U and a unitary gate from the symmetry group G, we first consider the χ *matrix representation of* $\mathcal{E}$,

$$\mathcal{E}(\rho) = \sum_{i,j} \chi_{i,j} P_i \rho P_j\ . \tag{6.84}$$

We can characterize the error of the gate U distinguishing it from the one coming from the symmetry group G, that we consider to be $\mathcal{N}$ for all element in the group (which can be benchmarked separately using for instance the known methods to benchmark Clifford gates). Using the bound from ref. [115, Appendix D] (where we set $i = 0$), we have

$$\begin{aligned}|\chi_{0,0}^{\Lambda\circ\mathcal{N}} - \chi_{0,0}^{\Lambda}\chi_{0,0}^{\mathcal{N}}| \leq 2\left((1-\chi_{0,0}^{\Lambda})\chi_{0,0}^{\Lambda}(1-\chi_{0,0}^{\mathcal{N}})\chi_{0,0}^{\mathcal{N}}\right)^{1/2} \\ + (1-\chi_{0,0}^{\Lambda})(1-\chi_{0,0}^{\mathcal{N}}).\end{aligned} \tag{6.85}$$

For an arbitrary channel $\mathcal{E}$, we know that $\chi_{0,0}^{\mathcal{E}} = \mathrm{Tr}[\mathcal{E}]/d^2$ (cfr. [105, eq. 2.30] and eq. (6.78)), so that we can recover the fidelity of the gate U form the one of the gates belonging to G and from $\mathbb{E}(\mathcal{F}_{(\Lambda\circ\mathcal{N})^G,\mathcal{I}})$ obtained with our protocol. The bound is particularly valid in the regime $\chi_{0,0}^{\mathcal{N}} \approx 1$, i.e., when the gates of the symmetry group can be implemented with high fidelity.

Confidence interval

To asses the number of different random sequences that have to be sampled in order to justify $\mathcal{Q}_{\mathrm{seq}}(\ell) \approx \mathcal{Q}_{\mathrm{avg}}(\ell)$ for a given sequence length ℓ, Wallmann and Flammia in ref. [46] provided bounds on the variance for the Clifford randomized benchmarking protocol described in ref. [105]. Their results show that a relatively small number of random sample is needed. We want to prove a bound similar to that of ref. [46, Theorem 10] for the variance

$$\sigma_\ell^2 = \frac{1}{|G|^\ell}\sum_{\mathbf{k}_\ell} \mathcal{Q}_{\mathbf{k}}^2(\ell,\rho) - \mathcal{Q}_{\mathrm{avg}}(\ell,\rho)^2. \tag{6.86}$$

In Pauli-Liouville representation and using $(E|\mathcal{C}|\rho)^2 = (E^{\otimes 2}|\mathcal{C}^{\otimes 2}|\rho^{\otimes 2})$, this can be expressed in terms of a scalar product in the form

$$\sigma_\ell^2 = \frac{1}{|G|^\ell}\sum_{\mathbf{k}_\ell} (E^{\otimes 2}|\mathcal{C}_{\mathbf{k}_\ell}^{\otimes 2}|\rho^{\otimes 2}) - (E^{\otimes 2}|\mathcal{C}_{\mathrm{avg},\ell}^{\otimes 2}|\rho^{\otimes 2}). \tag{6.87}$$

Now, we assume to be in the regime $\Lambda = \mathbb{1} + W\Delta t$, where W is a bounded matrix under additional assumption $\mathrm{Tr}\, W = \Theta(d^2)$, and expand the expression for the variance up the second

order in Δt,

$$\sigma_\ell^2 = \Delta t^2 (E^{\otimes 2}| \sum_{j=1}^{\ell} \frac{1}{|G|} \sum_{B\in G} (\mathcal{U}\otimes\mathcal{U})^{\ell-j} (\mathcal{B}^\dagger W \mathcal{B} \otimes \mathcal{B}^\dagger W \mathcal{B})(\mathcal{U}\otimes\mathcal{U})^j |\rho^{\otimes 2}) \tag{6.88}$$

$$- (E^{\otimes 2}| \sum_{j=1}^{\ell} (\mathcal{U}\otimes\mathcal{U})^{\ell-j} (W^G \otimes W^G)(\mathcal{U}\otimes\mathcal{U})^j |\rho^{\otimes 2}) + O(\ell^2 r^2 d^4). \tag{6.89}$$

The first term can be bounded as in ref. [46] using diamond norm properties and ref. [46, Proposition 9] with $4d(d+1)\ell r$. Again following that argument, the terms of order $O(\Delta t^3 W^3)$ and $O(\Delta t^4 W^4)$ are $O(\ell^2 r^2 d^4)$. Knowing the structure of W^G from the analysis of the symmetry group G, we can upper bound the number of non-zero terms as

$$\sum_\alpha m_\alpha^2 \, d_\alpha \leq \max_\alpha m_\alpha \sum_\alpha m_\alpha \, d_\alpha = \max_\alpha m_\alpha \, d^2. \tag{6.90}$$

From now on, we denote $m = \max_\alpha m_\alpha$ and $q = \max_{i,j} q_{i,j}$, the largest matrix entry of W, that we assume being independend of d. The second term in expression (6.88) obeys to the inequality

$$(E^{\otimes 2}| \sum_{j=1}^{\ell} (\mathcal{U}\otimes\mathcal{U})^{\ell-j} (W^G \otimes W^G)(\mathcal{U}\otimes\mathcal{U})^j |\rho^{\otimes 2}) \leq \ell q^2 m^2 d^4. \tag{6.91}$$

Now using

$$\mathrm{Tr}[\Lambda] = d(d+1)\mathbb{E}(\mathcal{F}) - d \tag{6.92}$$

follows

$$\Delta t = -r \frac{d(d+1)}{\mathrm{Tr}[W]}, \tag{6.93}$$

and so $\Delta t = O(r)$ since we assumed $\mathrm{Tr}\, W = \Theta(d^2)$. Hence, the second term of eq. (6.88) is $O(m^2 \, \ell \, r^2 d^4)$. While we do not have an exact estimation for the scaling of m for the general case, in the illustrated example for tensor copies of T-gate this goes as $O(\log d)$. Summarizing gives a bound for the variance

$$\sigma_\ell^2 \leq 4d(d+1)\ell r + O(\ell^2 r^2 d^4) + O(m^2 \, \ell \, r^2 d^4), \tag{6.94}$$

where the second term dominates the third one for $\ell \gg m^2$, i.e., in this regime the bound is exactly equivalent to the one of ref. [46, Theorem 10]. This bound however is probably not tight, and we are interested whether a bound similar to the one provided in ref. [112] can be obtained.

6.3.1 Example: Tensor copies of T-gate

We present an example to assess our protocol on one of the most relevant quantum gates, the T-gate, which together with the Hadamard and CNOT gates gives rise to a universal quantum circuit (cfr. Section 3.2). We are going to benchmark tensor copies of this gate too, up to four, in order to get a feeling on the scalability and necessary resources for this method. We give in the following the step-by-step protocol.

[1] Produce the n-Kronecker product group, denoted by A_n, of the local abelian symmetry group

$$\left\{ \begin{pmatrix} 1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1 \end{pmatrix}, \begin{pmatrix} 1&0&0&0\\0&0&-1&0\\0&1&0&0\\0&0&0&1 \end{pmatrix}, \begin{pmatrix} 1&0&0&0\\0&0&1&0\\0&-1&0&0\\0&0&0&1 \end{pmatrix}, \begin{pmatrix} 1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&1 \end{pmatrix} \right\},$$

which is isomorphic to the *cyclic group* of order 4, $\mathbb{Z}_4$.

[2] Construct the representation of the symmetric group S_n permuting the local subsystems.

[3] Construct the full symmetry group G as a semi-direct product of A_n and S_n by multiplying the respective 4^n-dimensional matrix representations. Each $g \in G$ is given by $g = a.\sigma$, with $a \in A_n, \sigma \in S_n$.

[4] From the character table of $\mathbb{Z}_4$,

$\mathbb{Z}_4$	e	γ	γ^2	γ^3
χ^0	1	1	1	1
χ^1	1	i	-1	-i
χ^2	1	-1	1	-1
χ^3	1	-i	-1	i

construct the character table of A_n by taking the product of the respective characters

$$\chi^{c_1,c_2,\ldots,c_n}(\ell_1,\ell_2,\ldots,\ell_n) \coloneqq \chi^{c_1}(\ell_1)\chi^{c_2}(\ell_2)\ldots\chi^{c_n}(\ell_n)\ , \tag{6.95}$$

where $\ell_j \in \mathbb{Z}_4$ and c_j is the label representing the irreducible representation.

[5] Divide the characters of A_n into orbits with respect to the action of S_n given by

$$\sigma.\chi(a)^{c_1,c_2,\ldots,c_n} \coloneqq \chi(\sigma^{-1}a\sigma)^{c_1,c_2,\ldots,c_n}. \tag{6.96}$$

In this particular case, the action of S_n works as a permutation of the labels of the irreducible representations, i.e.,

$$\sigma.\chi(a)^{c_1,c_2,\ldots,c_n} = \chi(a)^{\sigma(c_1,c_2,\ldots,c_n)}. \tag{6.97}$$

Choose for each orbit a representative element, for instance $\chi(a)^{c_1,c_2,\ldots,c_n}$ with $c_1 \leq c_2 \leq \cdots \leq c_n$, building a set $\{\,\chi_j\,\}_j$.

[6] For each representative element χ_j, find the stabilizer group H_j as a subgroup of S_n.

[7] For each irreducible representation π of H_j, write an irreducible representation of the subgroup $G_j \coloneqq A_n \cdot H_j$ of G by

$$\tilde{\rho}^j_\pi(a, g_j) = \chi^j(a) \cdot \pi(g_j). \tag{6.98}$$

[8] Obtain the characters of the representation ρ^j_π of G induced by $\tilde{\rho}^j_\pi$ with the Mackey-type formula,

$$\chi_{\rho^j_\pi}(s) = \frac{1}{|G_j|} \sum_{\substack{t\in G \\ t^{-1}st\in G_j}} \chi_{\tilde{\rho}^j_\pi}(t^{-1}st), \tag{6.99}$$

and obtain the irreducible representation multiplicity m^j_π in the decomposition of the Pauli-Liouville representation of G by the formula

$$m^j_\pi = \frac{1}{|G|} \sum_{g\in G} \Big(\chi_{\rho^j_\pi}(g)\Big)^* \cdot \phi(g), \tag{6.100}$$

where $\phi(g)$ is the trace of g in Pauli-Liouville representation.

In case of $n = 4$, for instance, there are 256 different irreducible representations of A_4 and five stabilizer groups: the full permutation group S_4 for the irreducible representations of the form $\chi^{a,a,a,a}$, $a \in 0, 1, 2, 3$, giving rise to $4 \cdot 5 = 20$ irreducible representations for G, S_3 for the representative irreducible representations of the form $\chi^{a,a,a,b}$ and $\chi^{a,b,b,b}$ with $a < b$, giving rise to $12 \cdot 3 = 36$ new irreducible representations, $S_2 \times S_2$ (isomorphic to the Klein 4 group) for representative elements $\chi^{a,a,b,b}$ with $a < b$, so that a total of $6 \cdot 4 = 24$ irreducible representations of G are derived, again representative elements $\chi^{a,a,b,c}, \chi^{a,b,b,c}, \chi^{a,b,c,c}$ with $a < b < c$ have stabilizer S_2, producing additional $12 \cdot 2 = 24$ irreducible representations; finally, the single representative element $\chi^{0,1,2,3}$ is the representative element of the sole orbit with trivial stabilizer. Hence, we have in total 105 different induced irreducible representations of G whose characters are obtained using eq. (6.99). As one can see from Table 6.1(d), only 22 of these irreducible representations decompose the twirled noise matrix, and the trivial representation has the highest multiplicity.

Results for n $\leq$ 4 We have obtained the irreducible decompositions for up to four tensor copies of the T-gate (see the Appendix for the *Mathematica* code) and report in the next page, Table 6.1, the decomposition of each twirled noise matrix. The superscripts of χ label the irreducible representations of A_n, while after the semicolon we denote the irreducible representation of the stabilizer group, where e denotes the trivial representation, *sgn* the sign representation, *std* the standard representation for all subgroups of S_4, $\ker_a$ the *Kernel a* representation of the Klein 4 group isomorphic to $S_2 \times S_2$, while 2dim denotes the 2-dimensional representation of S_4. We note that χ^2 never appears in the decomposition, and that the highest multiplicity, being $n + 1$, is always related to the trivial representation of the full group G. Additionally, we note that exploiting Schur's Lemma and the above considerations, the number of λ_j to be fitted when benchmarking copies of the T-gate is $\sum_{\alpha \text{ irrep}} m_\alpha$; from 1 to 4 qubits, this number is $4, 11, 24, 46$. In Figure 6.1 we illustrate the Pauli-Liouville representation of the twirled noise channel. One can notice that it becomes more sparse by increasing the system size.

Irreducible representation	Multip.
χ^0	2
χ^1	1
χ^3	1

(a) One T-gate decomposition.

Irreducible representation	Multip.	Irreducible representation	Multip.
$\chi^0, \chi^0; e$	3	$\chi^0, \chi^1; e$	2
$\chi^1, \chi^1; e$	1	$\chi^0, \chi^3; e$	2
$\chi^3, \chi^3; e$	1	$\chi^1, \chi^3; e$	1
$\chi^0, \chi^0, \mathrm{sgn}$	1		

(b) Two T-gates decomposition.

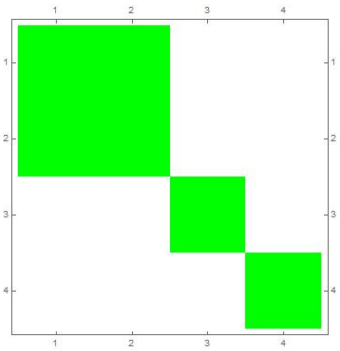
Irrep	Dim	Multip.	Irrep	Dim	Multip.	Irrep	Dim	Multip.
$\chi^0, \chi^0, \chi^0; e$	1	4	$\chi^0, \chi^0, \chi^3; e$	3	3	$\chi^0, \chi^0, \chi^1; \mathrm{sgn}$	3	1
$\chi^1, \chi^1, \chi^1; e$	1	1	$\chi^0, \chi^1, \chi^1; e$	3	2	$\chi^0, \chi^0, \chi^3; \mathrm{sgn}$	3	1
$\chi^3, \chi^3, \chi^3; e$	1	1	$\chi^1, \chi^1, \chi^3; e$	3	1	$\chi^0, \chi^1, \chi^3; e$	6	2
$\chi^0, \chi^0, \chi^0; \mathrm{std}$	2	2	$\chi^0, \chi^3, \chi^3; e$	3	2			
$\chi^0, \chi^0, \chi^1; e$	3	3	$\chi^1, \chi^3, \chi^3; e$	3	1			

(c) Three T-gates decomposition.

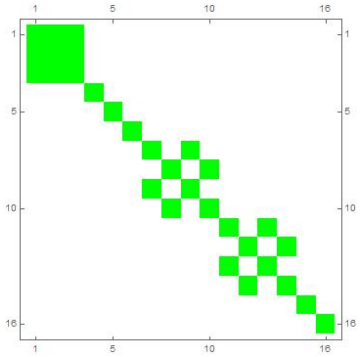
Irrep	Dim	M.	Irrep	Dim	M.	Irrep	Dim	M.
$\chi^0, \chi^0, \chi^0, \chi^0; e$	1	5	$\chi^1, \chi^1, \chi^1, \chi^3; e$	4	1	$\chi^1, \chi^1, \chi^3, \chi^3; e$	6	1
$\chi^1, \chi^1, \chi^1, \chi^1; e$	1	1	$\chi^0, \chi^3\chi^3, \chi^3; e$	4	2	$\chi^0, \chi^0, \chi^1, \chi^1; \ker_a$	6	1
$\chi^3, \chi^3, \chi^3, \chi^3; e$	1	1	$\chi^1, \chi^3\chi^3, \chi^3; e$	4	1	$\chi^0, \chi^0, \chi^3, \chi^3; \ker_a$	6	1
$\chi^0, \chi^0, \chi^0, \chi^0; 2\mathrm{dim}$	2	1	$\chi^0, \chi^0, \chi^0, \chi^1; \mathrm{std}$	8	2	$\chi^0, \chi^0, \chi^1, \chi^3; e$	12	3
$\chi^0, \chi^0, \chi^0, \chi^0; \mathrm{std}$	3	3	$\chi^0, \chi^0, \chi^0, \chi^3; \mathrm{std}$	8	2	$\chi^0, \chi^1, \chi^1, \chi^3; e$	12	2
$\chi^0, \chi^0, \chi^0, \chi^1; e$	4	4	$\chi^0, \chi^0, \chi^1, \chi^1; e$	6	3	$\chi^0, \chi^1, \chi^3, \chi^3; e$	12	2
$\chi^0, \chi^0, \chi^0, \chi^3; e$	4	4	$\chi^0, \chi^0, \chi^3, \chi^3; e$	6	3	$\chi^0, \chi^0, \chi^1, \chi^3; \mathrm{sgn}$	12	1
$\chi^0, \chi^1, \chi^1, \chi^1; e$	4	2						

(d) Four T-gates decomposition.

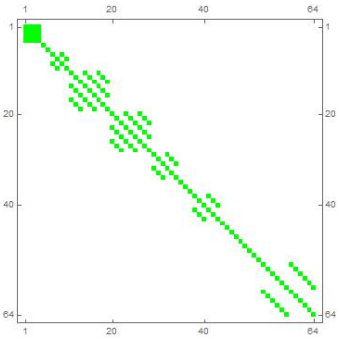
Table 6.1: Irreducible decomposition of the symmetry group G of multiple tensor copies of the T-gate channel. The superscripts of χ label the irreducible representations of A_n while the word after the semicolon the irreducible representation of the stabilizer group: e denotes the trivial representation, sgn the sign representation, std the standard representation, $\ker_\mathrm{a}$ the Kernel a representation of the Klein 4 group, 2dim the 2-dimensional representation of S_4.



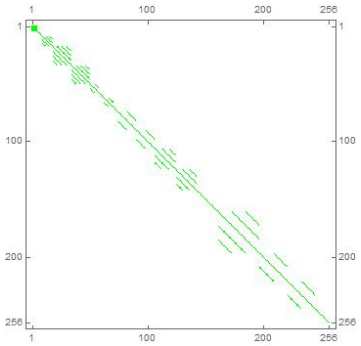
(a) Twirled noise matrix for one T-gate.



(b) Twirled noise matrix for two T-gates.



(c) Twirled noise matrix for three T-gates.



(d) Twirled noise matrix for four T-gates.

Figure 6.1: Matrix representation for the twirled noise of tensor copies of T-gates. The green area represents the non-zero matrix entries. For each matrix, the irreducible representations are sorted as in Table 6.1, reading from top to bottom, left to right.

CHAPTER 7

Brownian motion over the unitary group

In Chapter 5, we have described random quantum circuits and their link with random walks, a discrete stochastic process. One of the principal aim of this work is to embed random quantum circuits into a more general and continuous-time stochastic process over the unitary group, namely, *Brownian motion*, which as we shall see in the following is directly related to the context of *stochastically fluctuating local Hamiltonians* [64, 121]. In our article *Mixing properties of stochastic quantum Hamiltonians* [41] we have proven that these two settings indeed have in common many properties, from mixing results to the underlying random walks, as we shall illustrate in this chapter.

The study of continuous-time unitary evolution, besides its relevance from a purely mathematical perspective, can be of interest in order to understand spontaneous processes in Nature, such as dissipative ones, and are reminiscent in many ways and sometimes exactly model *thermalising dynamics* of interacting quantum systems with many constituents [122]. This link has prominently been explored in the context of *black hole thermalisation* (cfr. Section 7.3). This phenomenon is connected to the still unresolved puzzle asking how quickly black holes release information about their microscopic state. Based on considerations from string theory and gauge-gravity correspondences [123, 124], it is increasingly becoming clear that black holes do not destroy information when evaporating. This insight raises the question on what time scales this release of information precisely happens. It has been suggested that the time scale is set by the time it takes to *"scramble"* the microscopic degrees of freedom of the black hole, in a way that initial perturbations will be locally undetectable. In this regard, according to the famous *fast scrambling conjecture*, black holes should indeed be perfect scramblers, taking a time logarithmic in the number of degrees of freedom [37, 64].

More generally, any experimental setting in quantum mechanics will necessarily be interacting with a classical exterior in one way or another. Many decoherence mechanisms can well be approximated by a classical degree of freedom fluctuating randomly in time. In fact, effects like magnetic field fluctuations are of this type, and so are Gaussian noisy processes in condensed matter physics. This type of noise is usually seen as a detrimental type of *decoherence*, deteriorating the correlation present in the quantum mechanical system. This connection to *local dissipative dynamics* will be made clear below in Subsection 7.3 and in particular Proposition 7.28.

Again more technologically or pragmatically speaking, it should be highlighted that fluctuating Hamiltonians by no means have to reflect unwanted external noise. Quite to the contrary, in many applications in which random quantum circuits are envisioned, one can as well replace the quantum

circuit by the mere time evolution under such a fluctuating Hamiltonian. In many situations this can lead to a significantly simplified prescription, compared to implementing precisely controlled quantum gates that are designed according to samples of some suitable classical probability distribution. That is to say, in a number of instances, fluctuating Hamiltonians can be seen as being vastly more feasible than random circuits that require the accurate realization of quantum gates.

Brownian motion (also referred to as *Wiener process* in Probability theory) has a long history. The Roman poet and philosopher Titus Lucretius anticipated in 60 b.C. in the second volume of *De rerum natura* the description of this process,

> *For look closely, whenever rays are let in and pour the sun's light through the dark places in houses: for you will see many tiny bodies mingle in many ways all through the empty space right in the light of the rays, and as though in some everlasting strife wage war and battle, struggling troop against troop, nor ever crying a halt, harried with constant meetings and partings; so that you may guess from this what it means that the first-beginnings of things are for ever tossing in the great void. So far as may be, a little thing can give a picture of great things and afford traces of a concept. And for this reason it is the more right for you to give heed to these bodies, which you see jostling in the sun's rays, because such jostlings hint that there are movements of matter too beneath them, secret and unseen. For you will see many particles there stirred by unseen blows change their course and turn back, driven backwards on their path, now this way, now that, in every direction everywhere. You may know that this shifting movement comes to them all from the first-beginnings. For first the first-beginnings of things move of themselves; then those bodies which are formed of a tiny union, and are, as it were, nearest to the powers of the first-beginnings, are smitten and stirred by their unseen blows, and they in their turn, rouse up bodies a little larger. And so the movement passes upwards from the first-beginnings, and little by little comes forth to our senses, so that those bodies move too, which we can descry in the sun's light; yet it is not clearly seen by what blows they do it.*[125]

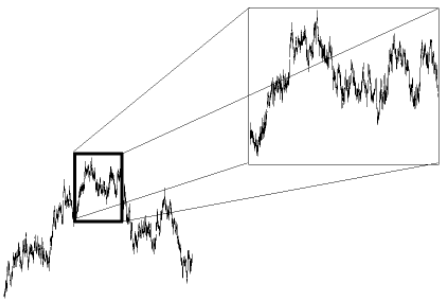
Many centuries later, in 1784, the physiologist and botanist Jan Ingenhousz described the scattered motion of a coal dust particle on a surface of alcohol similarly to the observations of the botanist Robert Brown in 1827 regarding the edgy motion of pollen grains suspended in water. Albert Einstein in 1905 made use of Brown's considerations and by formalizing Brownian motion as a stochastic process he gave an interpretation for particle diffusion that also served to indirectly confirm the existence of atoms and molecules. However, a completely mathematically rigorous construction of this process and existence thereof is due to the mathematician Norbert Wiener in 1923.

We are now going to approach a more formal description. Brownian motion over a topological group such as $\mathbb{U}(N)$ is quite different from its equivalent over an Euclidean space. However, recalling the definition and the main properties of the latter is certainly a useful exercise before diving into the theory of Brownian motion over Lie groups.

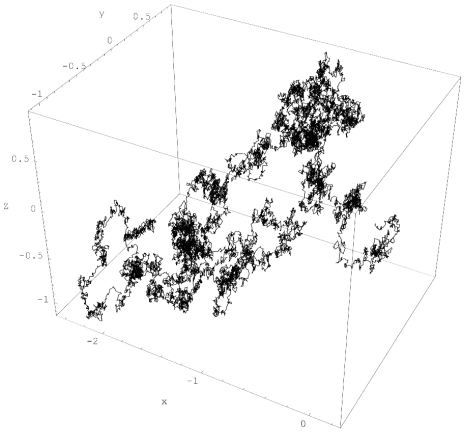
Definition 7.1 (Brownian motion, cfr. Definition 2.18 in ref. [126])**.** *A real-valued stochastic process B_t, $t \in [0, \infty)$ is called* Brownian motion *if*

(1) $B_0 = 0$,

(2) B *has* stationary *increments, i.e., for s, t with $s \leq t$, $B_t - B_s$ is equal in distribution to B_{t-s},*



(a) Scaling property of Brownian motion (picture taken from Multiscale Statistics for Evolving Complex Systems, by R. H. Riedi, http://www.stat.rice.edu/riedi/research-easy.html.)



(b) Brownian motion in 3 dimensions (picture taken from Brownian motion, Wikipedia, by T. J. Sullivan, https://en.wikipedia.org/wiki/Brownian motion.)

Figure 7.1

(3) B *has* independent *increments, that is, for all* $0 < t_1 < t_2 < \cdots < t_n$*, the increments* $B_{t_1} - B_0, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$ *are independent random variables,*

(4) for all $0 \leq s < t$*,* $B_t - B_s \sim \mathcal{N}(0, t-s)$*,*

(5) the paths $t \mapsto B_t$ *are continuous* almost surely*.*

Brownian motions exhibit self-similarity properties, namely,

Proposition 7.2. *Let* B_t *be a Brownian motion. Then* $\widetilde{B}^1_t = B_1 - B_{1-t}$ *for* $0 \leq t \leq 1$ *is equal in distribution to* B_t*, (time-shifting). Moreover,* $\widetilde{B}^2_t = \begin{cases} 0 & t = 0 \\ t\, B_{1/t} & t > 0 \end{cases}$ *(time-inversion) and* $\widetilde{B}^3_t = 1/c B_{c^2 t}$ *with* $c > 0$ *(Brownian scaling) are Brownian motions.*

The latter construction, $\widetilde{B}^3_t$, tells us that we can "zoom in" a Brownian motion and obtain again a Brownian motion, somehow in the same fashion as a fractal (see Fig. 7.1a).
One can also generalize the process introducing a *drift* term μ and re-scaling the variance by σ^2 writing $X_t = X_0 + \mu t + \sigma B_t$. The increments are then distributed as $\mathcal{N}(\mu(t-s), \sigma^2(t-s))$, and so we can describe any continuous-time (non-deterministic) *Lévy process* through Brownian motion.

While this stochastic process is, by definition, almost surely continuous, it is also nowhere differentiable (again almost surely), i.e., for all t

$$\limsup_{h \to 0} \frac{B(t+h) - B(t)}{h} = +\infty \qquad \text{and} \qquad \liminf_{h \to 0} \frac{B(t+h) - B(t)}{h} = -\infty. \tag{7.1}$$

This implies that the usual rules for differential and integration do not apply. Kiyoshi Itō introduced a new theory in order to treat these operations for Brownian motion, called *Itō calculus* [127, 128], nowadays widely applied in different contexts from pure mathematics to

econometrics.

Constructing Brownian motion (and hence proving existence thereof) is possible in different ways, e.g., with a Lévy construction, with harmonic functions or as a limit of a random walk. To briefly illustrate the last case, let us consider a sequence of independent and identically distributed random variables $\{ X_n \}_{n\geq 1}$ with $\mathbb{E}[X_n] = 0$ and $\mathrm{var}(X_n) = 1$. Note that we are not losing generality, since we can always consider the renormalization

$$\frac{X_n - \mathbb{E}[X_n]}{\sqrt{\mathrm{var}(X_n)}}. \tag{7.2}$$

Let us now construct a random walk $S_k = \sum_{n=1}^{k} X_n$ and interpolate between the points, namely, let us construct a random function in the space of continuous functions in $[0,\infty)$, $\mathbf{C}([0,\infty))$,

$$S(t) \coloneqq S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor)(S_{\lfloor t \rfloor + 1} - S_{\lfloor t \rfloor}). \tag{7.3}$$

We re-scale the interpolation on $\mathbf{C}([0,1])$ by $S_n^*(t) = S(nt)/\sqrt{n}$. Then the following theorem, called *Donsker's invariance principle* or alternatively *functional central limit theorem*, states that:

Theorem 7.3 (Donsker's invariance principle, cfr. [129])**.** *The distribution S_n^* converges weakly to a Brownian motion on $t \in [0,1]$.*

Note that the central limit theorem, extended by Donsker to the whole interval $[0,1]$, is the underlying reason for the distribution of the increments of Brownian motion to be strictly normal.

Brownian motion can of course be brought in Euclidean spaces of larger dimension (Fig. 7.1b). Then, for $B_t \in \mathbb{R}^d$, the increments $(B(t) - B(s))_1, \ldots, (B(t) - B(s))_d$ are independent random variables distributed according to $\mathcal{N}(0, t-s)$. Again, one can generalize the process adding a drift vector $\mu \in \mathbb{R}^d$ and a diffusion (symmetric, positive semi-definite) matrix $\Sigma \in \mathbb{R}^d \times \mathbb{R}^d$ and write $X_t = X_0 + \mu t + \Sigma B_t$ so that the increments are multivariate normal distributed with expectation $\mu(t-s)$ and covariance matrix $\Sigma\Sigma^T(t-s)$.

Brownian motion over a Lie group is however somehow different to construct. In particular, central limit theorem is no more valid and so the distribution of the increments should not be treated as normal distributed. Hence, we define Brownian motion on the unitary group $\mathbb{U}(n)$ as follows [34–36]:

Definition 7.4 (Brownian motion on the unitary group)**.** *A process U_t on the unitary group $\mathbb{U}(N)$ is called* Brownian motion *if the following conditions are satisfied.*

(1) $U_0 = \mathbb{1}$.

(2) For any time $t \geq 0$, the increments are stationary, i.e., for any $\Delta t > 0$ the increment $U_{t+\Delta t} U_t^\dagger$ is equal in distribution to $U_{\Delta t} U_0^\dagger$.

(3) For all $0 < t_1 < t_2 < \cdots < t_n$, the (left) increments $U_{t_1} U_0^\dagger, U_{t_2} U_{t_1}^\dagger, \ldots, U_{t_n} U_{t_{n-1}}^\dagger$ are independent.

(4) The paths $t \mapsto U_t$ are continuous almost surely.

Brownian motion U_t on the Lie group $\mathbb{U}(N)$ corresponds to Brownian motion W_t on the Lie algebra $\mathfrak{u}(N)$ through the exponential map, which for a matrix Lie group is given by the

series $\exp(X) = \sum_{n=0}^{\infty} X^n/n!$. More precisely, one can construct Brownian motion on $\mathbb{U}(N)$ by *injecting the differential* of a Brownian motion from $\mathfrak{u}(N)$ via the product integral of the exponential map [130, 131],

$$U_t = \lim_{\Delta t \to 0} \prod_{\ell=t/\Delta t}^{1} \exp\left\{W_{\ell\Delta t} - W_{(\ell-1)\Delta t}\right\} U_0\,. \tag{7.4}$$

We can turn to a more physical description through quantum Hamiltonians, whose increments are denoted by

$$H_{\ell,\Delta t} \coloneqq \mathrm{i}\,\Theta_{\ell,\Delta t} \tag{7.5}$$

with

$$\Theta_{\ell,\Delta t} \coloneqq \frac{1}{\Delta t}\left[W_{\ell\Delta t} - W_{(\ell-1)\Delta t}\right] \tag{7.6}$$

being the increments in the Lie algebra $\mathfrak{u}(N)$. More specifically, in this work we will consider *local* Hamiltonian increments on the physical quantum system consisting of n subsystems of dimension d, so that N becomes d^n. Those subsystems interact according to a pattern captured by an interaction graph with vertex set V and edge set E. In the special case of $d = 2$, this is referred to the *qubit case*, and the system is an *n-qubit system*.
We assume that $\Theta_{\ell,\Delta t}$ from eq. (7.6) is local with respect to an interaction graph (V, E), where each vertex in V corresponds to a d-level subsystem. Only qudits connected by an edge $e \in E$ may interact, i.e.,

$$\Theta_{\ell,\Delta t} = \sum_{e\in E} \theta^{(e)}_{\ell,\Delta t}\,, \tag{7.7}$$

where each local term $\theta^{(e)}_{\ell,\Delta t}$ is supported on e. The local terms are explicitly given by

$$\theta^{(e)}_{\ell,\Delta t} = -\mathrm{i}\,h^{(e)}_0 + \sum_{\mu} A^{(e)}_\mu\, \xi^{(e,\mu)}_{\ell,\Delta t}\,, \tag{7.8}$$

where we specify each term in this equation in the following. $h^{(e)}_0$ are deterministic Hermitian operators reflecting a constant drift in the evolution. Each noise operator $A^{(e)}_\mu$ acts on the two vertices connected by e as A_μ and as the identity elsewhere. $\{A_\mu\}_\mu$ is a basis of the real Lie algebra

$$\mathfrak{u}(d^2) \coloneqq \{X \in \mathbb{C}^{d^2\times d^2} :\ X = -X^\dagger\}. \tag{7.9}$$

$\xi^{(e,\mu)}_k$ are real random variables representing the noise. We assume that the noise satisfies

$$\mathbb{E}\left[\xi^{(e,\mu)}_{\ell,\Delta t}\right] = 0, \tag{7.10}$$

$$\mathbb{E}\left[\xi^{(e,\mu)}_{\ell,\Delta t}\,\xi^{(e',\mu')}_{\ell',\Delta t}\right] = -\frac{a}{\Delta t}\,\delta_{\ell,\ell'}\,\delta_{e,e'}\,\kappa^{-1}_{\mu,\mu'}, \tag{7.11}$$

where $a > 0$ is an arbitrary constant and the matrix κ is defined by

$$\kappa_{\mu,\nu} \coloneqq -2d^2\,\mathrm{Tr}(A^\dagger_\mu A_\nu). \tag{7.12}$$

As we will explain later, this matrix is in fact the *Killing metric tensor* associated with the basis $\{A_\mu\}_\mu$.

Remark 7.5 (Orthonormal basis)**.** *If the basis $\{A_\mu\}_\mu$ is orthonormal then our assumption (7.11) on the covariance simplifies to*

$$\mathbb{E}\left[\xi^{(e,\mu)}_{\ell,\Delta t}, \xi^{(e',\mu')}_{\ell',\Delta t}\right] = \frac{a}{2d^2\Delta t}\,\delta_{\ell,\ell'}\,\delta_{e,e'}\,\delta_{\mu,\mu'} \tag{7.13}$$

which represents white noise*. This happens, for instance, if we choose the Pauli basis (see the example processes in Example 7.1.5 and Section 7.2).*

Remark 7.6 (Overcomplete sets of operators)**.** *Additionally, we may consider an overcomplete set of operators* $\{A_\mu\}_\mu \in \mathfrak{u}(d^2)$ *as long as they give rise to a negative contribution to the generator* (7.23)*, since this will increase the gap of the moment operator induced by the stochastic evolution and hence make the convergence even faster.*

The above described Brownian motion with Hamiltonian increments as in eq. (7.7) with the specified $\theta_\ell^{(e)}$ induces a Brownian motion U_t on the unitary group. We denote the distribution of U_t at time t by $\mathrm{SLH}(t)$ and write the according expectation as $\mathbb{E}_{\mathrm{SLH}(t)}$.

7.1 Brownian motion on the unitary group is a *k*-design

In this section we investigate mixing properties of random quantum processes in quantum many-body systems. We show that the locally generated Brownian motion gives rise to an efficiently approximate unitary k-design of arbitrary order, i.e., each of its moment operators converges to the one of the Haar measure in polynomial run time. Furthermore, the convergence rate is compatible to that of a random quantum circuit in discrete time that we have discussed in Section 5.2. Our main technical contribution is a connection between the generator of the local diffusion and the Casimir element of the special unitary group. This allows us to obtain an explicit uniform lower bound on the gap of the local generator, i.e., independent of the order k. Hence, our contribution provides a new class of probability measures on the unitary group whose set of generated unitaries has the spectral gap property [98, 132], and moreover, with an explicitly known constant. This might be an unexpected result, as the convergence time of the k-th moment increases with k for many processes.

Hence, for any of the time-fluctuating local Hamiltonians fulfilling the description given in the previous part of the chapter, these results can be expressed in terms of quantum tensor product expanders or approximate unitary designs as follows.

Theorem 7.7 (Local Brownian motions on $\mathbb{U}(d^n)$ are quantum (λ, k)-tensor product expanders, Theorem 9 in ref. [41])**.** *Let* U_T *be a unitary Brownian motion with the increments given in eqs.* (7.7)-(7.12) *with the interaction graph* (V, E) *being either a complete graph or a* 1*D nearest neighbour graph. Then, for any run time*

$$T \geq 850 \lceil \log_d(4k) \rceil^2 d^2 k^5 k^{3.1/\ln(d)} \frac{\ln(1/\lambda)}{a}, \tag{7.14}$$

U_T *is a quantum* (λ, k)*-tensor product expander.*

Then, using Lemma 3.12, we immediately obtain the subsequent corollary.

Corollary 7.8 (Approximate unitary k-designs, Corollary 10 in ref. [41])**.** *For any run time*

$$T \geq 850 \lceil \log_d(4k) \rceil^2 d^2 k^5 k^{3.1/\ln(d)} \frac{nk\ln(d) + \ln(1/\varepsilon)}{a}, \tag{7.15}$$

U_T *is an* ε*-approximate unitary* k*-design.*

Theorem 7.7 can be seen as a unifying statement on random quantum processes. It extends the results on random local quantum circuits, as considered in ref. [27, Corollary 7], to continuous time dynamics under fluctuating Hamiltonians. Note that the scaling of the minimal run time required for the generating of a unitary k-design is by a factor of n smaller with respect to the circuit setting. This is due to the number of Hamiltonian interactions per time step growing linearly in the system size for a 1D graph (which is, as discussed in Lemma 7.16, the slowest

setting among all complete graphs), while for local random quantum circuits only one gate per step is applied. If we re-scale the stochastic Hamiltonian with a pre-factor $O(1/\sqrt{n})$ so that the k-th moment operator may be written in the same form as the one induced by a random quantum circuit, i.e.,

$$M^k = \frac{1}{n}\sum_j \left(m^k\right)^{j,j+1}, \tag{7.16}$$

where $\left(m^k\right)^{j,j+1}$ denotes the local moment operator applied on qubits j and $j+1$, then we would instead obtain the same scaling for the mixing time. Therefore, we can consider the two scenarios as perfectly compatible.

7.1.1 Proof of Theorem 7.7

In this section, we prove Theorem 7.7 bounding the time after which the stochastic time evolution becomes a tensor product expander. As a crucial step we investigate the gap of the local generator induced by the Hamiltonian increments as given in eqs. (7.7)-(7.12).
The proof will be structured as follows: we first derive in Lemma 7.9 the generator of the k-th moment operator and then describe how this allows us to express it as a tensor product expander using previous results on random quantum circuits. In Subsection 7.1.2 we provide the central mathematical result of this work, namely, a diagonalisation of the local generator by relating it to the Casimir element in the enveloping algebra of $\mathfrak{su}(d^2)$. Since only certain irreps are contained in the direct sum decomposition of the Casimir element, we will observe that no eigenvalue can assume a value in the interval (0,1), giving rise to a local gap.
Much of the developed machinery will build upon the representation theory of the special unitary group. It will also be helpful to use the identification of maps on matrices with matrices using the vectorisation isomorphism (cfr. Section 2.1) to express the k-th moment operator as

$$M_\mu^k = \mathbb{E}_\mu[\pi_{k,k}(U)], \tag{7.17}$$

where $\pi_{k,k}(U)$ is the (k,k)-mixed tensor representation of the group element $U \in \mathbb{SU}(N)$ given by

$$\pi_{k,k}(U) := U^{\otimes k} \otimes \overline{U}^{\otimes k}. \tag{7.18}$$

We also make use of the corresponding representation of the Lie algebra $\mathfrak{su}(N)$ which is also denoted by $\pi_{k,k}$ and satisfies the following for all $X \in \mathfrak{su}(N)$:

$$\pi_{k,k}(\exp(X)) = \exp(\pi_{k,k}(X)), \tag{7.19}$$

with

$$\pi_{k,k}(X) = \sum_{i=1}^{k} X \otimes \mathbb{1}_{\bar{i}} + \sum_{i=k+1}^{2k} \overline{X} \otimes \mathbb{1}_{\bar{i}} \tag{7.20}$$

and

$$X \otimes \mathbb{1}_{\bar{i}} := \mathbb{1}_1 \otimes \ldots \otimes \mathbb{1}_{i-1} \otimes X \otimes \mathbb{1}_{i+1} \otimes \ldots \otimes \mathbb{1}_{2k}. \tag{7.21}$$

This representation plays a central role in our analysis of the gap of the k-th moment operator of the stochastic time evolution.
The k-th moment operator $M^k_{\mathrm{SLH}(T)}$ has a generator that we explicitly calculate in the following. In fact, the lemma also holds for general Brownian motions on $\mathbb{U}(N)$, not only the locally generated ones considered in our theorems.

Lemma 7.9 (The generator of the k-th moment operator)**.** *Let M_T^k be the k-th moment operator of a unitary Brownian motion with increments $\Theta_{\Delta t}$ as in eq. (7.6) at time T. Then*

$$M_T^k = \mathrm{e}^{G^k\, T} \tag{7.22}$$

with

$$G^k = \lim_{t\to 0}\left(\mathbb{E}\left[\pi_{k,k}(\Theta_t)\right] + \frac{1}{2}\mathbb{E}\left[\pi_{k,k}(\Theta_t)^2\, t\right]\right). \tag{7.23}$$

Note that as Θ_t is anti-Hermitian, G^k is negative semidefinite. If the Brownian motion is universal then the kernel of G^k is the invariant subspace of M^k.
Most steps in the proof of this lemma will be also used again in the proof of Theorem 7.7.

Proof. As M_T^k is a Markov process, we have

$$M_T^k = (M_{\Delta t}^k)^{T/\Delta t}. \tag{7.24}$$

With the mixed tensor representation (7.18) and the definition of the k-th moment operator (3.4) we obtain for a single time step

$$\begin{aligned} M_{\Delta t}^k &= \mathbb{E}\left[\pi_{k,k}\left(U_{\Delta t}\right)\right] && (7.25)\\ &= \mathbb{E}\left[\pi_{k,k}\left(\exp\left\{\Theta_{\Delta t}\,\Delta t\right\}\right)\right] && (7.26)\end{aligned}$$

with $U_{\Delta t} = \mathrm{e}^{\Theta_{\Delta t}\,\Delta t}$ and the increments $\Theta_{\Delta t}$ from eq. (7.6). Using a Taylor expansion yields

$$\begin{aligned} M_{\Delta t}^k &= \mathbb{E}\left[\mathrm{e}^{\pi_{k,k}(\Theta_{\Delta t})\Delta t}\right] && (7.27)\\ &= \sum_{p=0}^{\infty}\frac{(\Delta t)^p}{p!}\mathbb{E}\left[\pi_{k,k}(\Theta_{\Delta t})^p\right] && (7.28)\\ &= \mathbb{1}_{\dim(\pi_{k,k})} + \mathbb{E}\left[\pi_{k,k}(\Theta_{\Delta t})\right]\Delta t + \mathbb{E}\left[\pi_{k,k}(\Theta_{\Delta t})^2\right]\frac{\Delta t^2}{2} + O(\Delta t^2). && (7.29)\end{aligned}$$

Composing the time steps as in eq. (7.24), we obtain

$$M_T^k = \lim_{\Delta t\to 0}\left(\mathbb{1}_{\dim(\pi_{k,k})} + \left(\mathbb{E}\left[\pi_{k,k}(\Theta_{\Delta t})\right] + \mathbb{E}\left[\frac{1}{2}\pi_{k,k}(\Theta_{\Delta t})^2\Delta t\right]\right)\Delta t\right)^{T/\Delta t} \tag{7.30}$$

and finishes the proof. □

Additionally, we remark the following lemma on random quantum circuits generated by general local distributions, which is implicitly contained in ref. [27, Corollary 7].

Lemma 7.10 (Relating global and local gaps)**.** *Let μ_{loc} be a distribution on $\mathbb{U}(d^2)$ and $\mathrm{circ}(\mu_{loc})$ be the distribution on $\mathbb{U}(d^n)$ that applies a unitary drawn according to μ_{loc} to a uniformly chosen edge of an interaction graph $i, i+1$. Then, its moment operator satisfies*

$$\|M^k_{\mathrm{circ}(\mu_{loc})} - M^k_{\mathrm{Haar}}\|_\infty \leq 1 - \left(1 - \|m^k_{\mu_{loc}} - m^k_{\mathrm{Haar}_{loc}}\|_\infty\right)\left(1 - \|M^k_{\mathrm{circ}(\mathrm{Haar}_{loc})} - M^k_{\mathrm{Haar}}\|_\infty\right), \tag{7.31}$$

where Haar_{loc} denotes the Haar measure on $\mathbb{U}(d^2)$.

Proof. Setting $\left(M^k_{\text{Haar}}\right)^\perp = \mathbb{1} - M^k_{\text{Haar}}$ we have the relation

$$\|M^k_{\text{circ}(\mu_{loc})} - M^k_{\text{Haar}}\|_\infty = \|\left(M^k_{\text{Haar}}\right)^\perp M^k_{\text{circ}(\mu_{loc})} \left(M^k_{\text{Haar}}\right)^\perp\|_\infty \tag{7.32}$$

$$= \|\frac{1}{|E|}\sum_{e\in E}^{n-1} \left(M^k_{\text{Haar}}\right)^\perp m^{k,(e)}_{\mu_{loc}} \left(M^k_{\text{Haar}}\right)^\perp\|_\infty \,. \tag{7.33}$$

By denoting $\gamma := \|m^k_{\mu_{loc}} - m^k_{\text{Haar}_{loc}}\|_\infty$, we find $m^{k,(e)}_{\mu_{loc}} \leq (1-\gamma)\, m^{k,(e)}_{\text{Haar}_{loc}} + \gamma\mathbb{1}$, which implies the operator inequality

$$\left(M^k_{\text{Haar}}\right)^\perp \left(\sum_{e\in E} m^{k,(e)}_{\mu_{loc}}\right) \left(M^k_{\text{Haar}}\right)^\perp \leq \gamma|E| \left(M^k_{\text{Haar}}\right)^\perp \tag{7.34}$$

$$+ (1-\gamma) \left(M^k_{\text{Haar}}\right)^\perp \left(\sum_{e\in E} m^{k,(e)}_{\text{Haar}_{loc}}\right) \left(M^k_{\text{Haar}}\right)^\perp \,. \tag{7.35}$$

Since $(1-\gamma)$ is positive, we can use the bound $A \leq \|A\|_\infty$ for the second summand on the right hand side which, together with

$$\|M^k_{\text{circ}(\text{Haar}_{loc})} - M^k_{\text{Haar}}\|_\infty = \|\left(M^k_{\text{Haar}}\right)^\perp M^k_{\text{circ}(\text{Haar}_{loc})} \left(M^k_{\text{Haar}}\right)^\perp\|_\infty \tag{7.36}$$

finishes the proof. □

Now we present the main proof of Theorem 7.7. Part of it will be completed with the lemmas stated and proved subsequently.

Proof of Theorem 7.7. Thanks to Lemma 3.14 it is enough to bound the gap of the k-th moment operator $M^k_{\text{SLH}(T)}$. According to Lemma 7.15, the time constant part of the Hamiltonian does not affect the invariant subspace nor the gap of $M^k_{\text{SLH}(T)}$. Hence, we can set without loss of generality $h^{(e)}_0 = 0 \quad \forall e$. Additionally, in Lemma 7.16 we prove that the gap of an interaction graph being a complete graph is larger than the one of a 1D graph. We hence consider only the latter case in the proof.
Using the approximation (7.29) and expressing Θ in terms of the local terms $\theta^{(e)}$ (as in eq. (7.8)) we obtain

$$M^k_{\text{SLH}(\Delta t)} = \mathbb{1}_{\dim(\pi_{k,k})} + \sum_{e\in E} \mathbb{E}\left[\pi_{k,k}(\theta^{(e)}_{\Delta t})^2\right] \frac{\Delta t^2}{2} + O(\Delta t^2) \tag{7.37}$$

Using another Taylor approximation yields

$$M^k_{\text{SLH}(\Delta t)} = \mathbb{1}_{\dim(\pi_{k,k})} + \frac{1}{n}\sum_{e\in E} \mathbb{E}\left[\pi_{k,k}\left(\sqrt{n}\,\theta^{(e)}_{\Delta t}\right)^2\right] \frac{\Delta t^2}{2} + O(\Delta t^2) \tag{7.38}$$

$$= \frac{1}{n}\sum_{e\in E} \mathbb{E}\left[\left(\exp\{\sqrt{n}\,\theta^{(e)}_{\Delta t}\,\Delta t\}\right)^{\otimes k,k}\right] + O(\Delta t^2)\,. \tag{7.39}$$

Next, we view $G(\theta^{(e)}_{\Delta t}) := \exp\{\sqrt{n}\,\theta^{(e)}_{\Delta t}\,\Delta t\}$ as a random gate in a *G-random quantum circuit* considered in ref. [27]. The (system size independent) local k-th moment operator on edge $e \in E$ is

$$m^{k,(e)}_{\Delta t} := \mathbb{E}\left[\left(\exp\{\sqrt{n}\,\theta^{(e)}_{\Delta t}\,\Delta t\}\right)^{\otimes k,k}\right]. \tag{7.40}$$

Note that this k-th moment operator also corresponds to a Brownian motion but with a variance re-scaled by a factor of n, cfr. also the parameter a in eq. (7.11). As its gap, i.e., the difference between the largest and second largest eigenvalue, does not depend on e we simply denote the gap of $m_{\Delta t}^{k,(e)}$ by $\Delta(m_{\Delta t}^k)$. Then the local gap lemma 7.10 yields directly

$$\left\| M^k_{\mathrm{SLH}(\Delta t)} - M^k_{\mathrm{Haar}} \right\|_\infty \leq 1 - \Delta\left(m^k_{\Delta t}\right)\left(1 - \|M^k_{\mathrm{circ(Haar)}} - M^k_{\mathrm{Haar}}\|_\infty\right), \tag{7.41}$$

where we recall that $M^k_{\mathrm{circ(Haar)}}$ is the k-th moment operator of single step of a local random quantum circuit whose gates are chosen from the Haar measure. The gap of $M^k_{\mathrm{circ(Haar)}}$ can be lower bounded with Theorem 5.8, that is,

$$\left(1 - \|M^k_{\mathrm{circ(Haar)}} - M^k_{\mathrm{Haar}}\|_\infty\right) \geq \frac{1}{425\, n \lceil \log_d(4t)\rceil^2\, d^2\, k^5\, k^{3.1/\ln(d)}}. \tag{7.42}$$

Together with eq. (7.41), these results imply

$$\|M^k_{\mathrm{SLH}(\Delta t)} - M^k_{\mathrm{Haar}}\|_\infty \leq 1 - \Delta(m^k_{\Delta t})\, s/n \tag{7.43}$$

with

$$s := \left(425 \lceil \log_d(4k)\rceil^2 d^2 k^5 k^{3.1/\ln(d)}\right)^{-1}. \tag{7.44}$$

In order to calculate $\Delta(m^k_{\Delta t})$ we use Lemma 7.9, eq. (7.40) and

$$\mathbb{E}\left[\pi_{k,k}(\theta_t^{(e)})\right] = 0 \tag{7.45}$$

so that we can express $m_{\Delta t}^{k,(e)}$ as

$$m_{\Delta t}^{k,(e)} = \exp\left(n\, g_k^{(e)}\, \Delta t\right) = \mathbb{1} + n\, g_k^{(e)}\, \Delta t + O(\Delta t^2) \tag{7.46}$$

with

$$n\, g_k^{(e)} = \frac{1}{2} \lim_{t\to 0} \mathbb{E}\left[\pi_{k,k}(\sqrt{n}\, \theta_t^{(e)})^2\, t\right] = \frac{n}{2} \lim_{t\to 0} \mathbb{E}\left[\pi_{k,k}(\theta_t^{(e)})^2\, t\right]. \tag{7.47}$$

Hence,

$$\Delta(m^k_{\Delta t}) = n\, \Delta(g_k)\, \Delta t + O(\Delta t^2), \tag{7.48}$$

where $\Delta(g_k)$ denotes again the spectral gap to the invariant subspace, i.e., minus the largest non-zero eigenvalue of $g_k^{(e)}$.
As both k-th moment operators have the same unit eigenvalue eingenspace according to Lemma 3.14, $\|M^k_{\mathrm{SLH}(\Delta t)} - M^k_{\mathrm{Haar}}\|_\infty$ is the second largest eigenvalue of $M^k_{\mathrm{SLH}(\Delta t)}$. Hence,

$$\begin{aligned}
\|M^k_{\mathrm{SLH}(T)} - M^k_{\mathrm{Haar}}\|_\infty &= \lim_{\Delta t\to 0} \|\left(M^k_{\mathrm{SLH}(\Delta t)}\right)^{T/\Delta t} - M^k_{\mathrm{Haar}}\|_\infty && (7.49)\\
&= \lim_{\Delta t\to 0} \left(1 - s\, \Delta(g_k)\right)^{T/\Delta t} && (7.50)\\
&= \exp\left(-T\, s\, \Delta(g_k)\right). && (7.51)
\end{aligned}$$

Observation 7.12 and Lemma 7.13 imply that the gap is the same as the variance (7.11) of the noise, $\Delta(g_k) = a/2$, which completes the proof. □

7.1.2 Local gap

In order to calculate the local gap $\Delta(g_k)$, the following representations for the algebra $\mathfrak{su}(N)$ will be used.

$$\text{Trivial rep.} \quad \pi_1 : \mathfrak{su}(N) \to \mathfrak{gl}(1,\mathbb{C}), \qquad X \mapsto 0, \tag{7.52}$$

$$\text{Fundamental rep.} \quad \pi_f : \mathfrak{su}(N) \to \mathfrak{gl}(N,\mathbb{C}), \qquad X \mapsto X, \tag{7.53}$$

$$\text{Adjoint rep.} \quad \pi_{\text{ad}} : \mathfrak{su}(N) \to \mathfrak{gl}\left(\mathfrak{su}(N)\right), \qquad X \mapsto \text{ad}_X, \tag{7.54}$$

where ad_X is defined by $\text{ad}_X(Y) \coloneqq [X,Y]$.

Observation 7.11 (Omitting the phase)**.** *From the mixed-tensor representation we note that we can restrict the analysis on the $\mathfrak{su}(N)$ algebra instead of $\mathfrak{u}(N)$.*
Let $\mathcal{C}(n)$ be the center of $\mathbb{U}(n)$, that is, the set of all scalar matrices $\lambda\mathbb{1}$ with λ element of the circle group $\mathbb{T} = \{\, \lambda : |\lambda| = 1 \,\}$. *Let*

$$\mathcal{Z}(n) = \mathbb{SU}(n) \cap \mathcal{C}(n) = \{\, e^{\mathrm{i}2\pi j/n}\mathbb{1} : j = 0, \dots, n-1 \,\}$$

be the center of $\mathbb{SU}(n)$, i.e., the subset of the $\mathcal{C}(n)$ with the roots of the unity.
It is known that [133]

$$\mathbb{U}(n) = \mathbb{SU}(n) \times \mathcal{C}(n)/\mathcal{Z}(n), \tag{7.55}$$

that is, the direct product $\mathbb{SU}(n) \times \mathcal{C}(n)$ is a n-fold cover of the unitary group $U(n)$; for any $U \in \mathbb{U}(n)$ with $\det U = \lambda$, its inverse image with respect to the cover map is given by n Cartesian pairs $\left\{ \left(S\, e^{-\mathrm{i}2\pi j/n}, \beta\, e^{\mathrm{i}2\pi j/n}\, \mathbb{1} \right) : j = 0, \dots, n-1 \right\}$, where β is an arbitrary n-root of λ and $S \in \mathbb{SU}(n)$ such that $U = \beta S$.
Let us define the interval in the unit circle $I(n) = \{\, \lambda = \mathrm{e}^{\mathrm{i}\phi} : \phi \in [0, 2\pi/n) \,\} \subset \mathbb{T}$. Then $\mathbb{U}(n)$ is homeomorphic to $\mathbb{SU}(n) \times I(n)$ and any matrix $U \in \mathbb{U}(n)$ can then be written uniquely as $U = \beta S$, with $\beta = \sqrt[n]{\det U} \in I(n)$ and $S \in \mathbb{SU}(n)$. Hence

$$\pi_{k,k}(U) = U^{\otimes k} \otimes \overline{U}^{\otimes k} = (\beta S)^{\otimes k} \otimes (\overline{\beta S})^{\otimes k} = \pi_{k,k}(S) \tag{7.56}$$

and so the in this representation the two algebras are indistinguishable.

The *Killing form* K in $\mathfrak{su}(N)$ is the symmetric bilinear form defined by

$$K(X,Y) \coloneqq \text{Tr}\left[\text{ad}_X \text{ad}_Y\right]. \tag{7.57}$$

Denoting the Hilbert-Schmidt inner product of X and Y (in the fundamental representation) by $\langle X, Y\rangle = \text{Tr}(X^\dagger Y)$, the Killing form of $\mathfrak{su}(N)$ can also be written as

$$K(X,Y) = -2N \langle X, Y\rangle. \tag{7.58}$$

In terms of a basis $\{X_\mu\}_{\mu=1}^{N^2-1}$ of $\mathfrak{su}(N)$ the *Killing metric tensor* κ is defined by

$$\kappa_{\mu,\nu} \coloneqq K(X_\mu, X_\nu), \tag{7.59}$$

as was already indicated in eq. (7.12). Then, the *Casimir element* in a matrix representation π is

$$C(\pi) \coloneqq \sum_{\mu,\nu} \kappa^{-1}_{\mu,\nu}\, \pi(X_\mu)\pi(X_\nu). \tag{7.60}$$

According to eqs. (7.47) and (7.8), the local generator $g_k^{(e)}$ of our unitary process with vanishing driving $h_0^{(e)} = 0$ is given by

$$g_k^{(e)} = \frac{1}{2} \lim_{\Delta t \to 0} \mathbb{E} \left[\pi_{k,k} \left(\sum_{\mu=1}^{N^2-1} A_\mu^{(e)} \, \xi^{(e,\mu)} \right)^2 \right] \Delta t = -\frac{a}{2} \sum_{\mu,\nu=1}^{N^2-1} \kappa_{\mu,\nu}^{-1} \, \pi_{k,k} \left(A_\mu^{(e)} \right) \pi_{k,k} \left(A_\nu^{(e)} \right) \tag{7.61}$$

(where $N \coloneqq d^2$). The second equality follows from our central assumption (7.11). All $g_k^{(e)}$ are tensor copies of a local operator g_k. Therefore, we will suppress the subscripts e in this section from now on.

Observation 7.12 (Casimir element)**.** *Let g_k be the generator of the local k-th moment operator in eq. (7.61). Then*

$$g_k = -\frac{a}{2} \, C(\pi_{k,k}) \, . \tag{7.62}$$

More generally, an overcomplete set $\{ A_\mu \}$ can also be admitted, as already mentioned in Remark 7.6. The final result about the convergence rate – up to a constant $O(1)$ – is still valid as long as the generator and the Casimir element are related by an equation of the form

$$g_k = -\frac{a'}{2} \, C(\pi_{k,k}) + g' \, , \tag{7.63}$$

where $a' > 0$ and g' is negative semidefinite so that it can only increase the gap.
In the following, we prove that the eigenvalues of the Casimir do not assume a value within the interval $(0, 1)$, for all k.

Lemma 7.13 (Casimir gap)**.** *Let $\mathcal{I}_k$ be the set of irreducible representations occurring in $\pi_{k,k}$ and let $m_k(\pi) \in \mathbb{N}$ denote the multiplicity of each such representation π. Then*

$$C(\pi_{k,k}) \simeq \bigoplus_{\pi \in \mathcal{I}_k} c_2(\pi) \, \mathbb{1}_{\dim(\pi)} \otimes \mathbb{1}_{m_k(\pi)}, \tag{7.64}$$

where

$$c_2(\pi) \begin{cases} = 0 & \text{if } \pi \simeq \pi_1, \\ = 1 & \text{if } \pi \simeq \pi_{\mathrm{ad}}, \\ > 1 & \text{otherwise.} \end{cases} \tag{7.65}$$

In particular, the spectral gap of $C(\pi_{k,k})$ is independent of k.

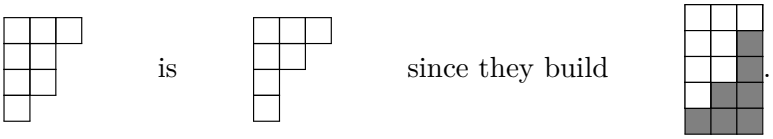
Proof. Since the Casimir element is an element of the center of the universal enveloping algebra, from Schur's Lemma follows that it acts as a multiple of the identity in each irreducible representation (see ref. [134, Chapter 12]), so that (7.64) is immediate. Now, since the tensor product between the fundamental representation and its conjugate are isomorphic to the direct sum of the trivial and the adjoint ones, this means that the representation $\pi_{k,k}$ is isomorphic to $(\pi_1 \oplus \pi_{\mathrm{ad}})^{\otimes k}$.
The trivial representation is guaranteed to occur in the decomposition of $\pi_{k,k}$ into irreducible representations (for example, via $\pi_1^{\otimes k}$) and leads to the eigenvalue $c_2(\pi_1) = 0$. The adjoint representation always occurs – for example, via $\pi_{\mathrm{ad}} \otimes \pi_1^{\otimes(k-1)}$ and permutations thereof – too, and leads to the eigenvalue $c_2(\pi_{\mathrm{ad}}) = 1$. If we can show that no other irreducible representation π with $c_2(\pi) \leq 1$ occurs, the proof is complete.
One might think that this requires rather detailed knowledge about how tensor product representations of the form $\pi_{\mathrm{ad}}^{\otimes l}$ decompose into irreducible representations. To follow the next argument, some basic knowledge regarding Young diagrams is necessary. These are arrays of boxes arranged

in $N-1$ left-justified rows whose length is non-increasing from top to bottom, each of them connected to an irreducible representation, e.g.,

In particular, the following holds true.

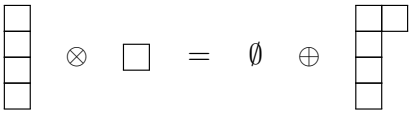
- The Young diagram of the fundamental representation is given by one single box  .
- The trivial representation does not have any box; we can denote it by $\emptyset$.
- The adjoint representation is given by a column of $N-1$ boxes and a second column made of a single box. For example, the adjoint representation of $\mathfrak{su}(5)$ is given by

- The conjugate representation of a Young diagram whose first row contains ℓ boxes is given by the complementary diagram (rotated by 180 degrees) shaping the rectangle of N rows and ℓ columns. For example, for $\mathfrak{su}(5)$ the conjugate representation of



Note that the conjugate diagram of the fundamental representation is given by a single column of $N-1$ boxes, while the adjoint representation is self-conjugate.

Young diagrams are particularly helpful when decomposing the tensor product of two representations into a direct sum of irreducible representations. Here, one follows two steps: first, one combines the boxes of the two diagrams by adding, one at a time, all boxes in the first row of the second diagram to the first one, respecting the condition of non-increasing length from top to bottom for the rows of the newly created diagrams and remembering that each of them can have at most N rows. One repeats the procedure for all rows in the second diagram. As a second step, one discards all diagrams which do not satisfy specific rules that we are not going to mention here; for a full description, see ref. [135]. Furthermore, for the algebra $\mathfrak{su}(N)$ all columns with N boxes occurring in a diagram can be deleted.
Recalling that the tensor product of the fundamental representation and its conjugate can be decomposed as a direct sum of the trivial and the adjoint representation and taking again $\mathfrak{su}(5)$ as an example, we have

$$\overline{U}\otimes U=\overline{\pi}_f\otimes\pi_f=$$  $$=\pi_1\oplus\pi_{\rm ad}\,, \qquad (7.66)$$

since a diagram with a column of $N=5$ boxes is equivalent to the trivial representation.
An alternative way to express an irreducible representation of $\mathfrak{su}(N)$ is to associate a *Dynkin label* $(\lambda_1,\lambda_2,\ldots,\lambda_{N-1})$, where λ_b gives the number of columns made of b boxes. For instance,

the fundamental representation is given by the label $(1,0,\ldots,0)$ and the adjoint representation by $(1,0,\ldots,0,1)$.
It is in fact sufficient to exploit a remarkably basic property which is shared by all the irreducible $\mathfrak{su}(d^2)$ representations occurring in $\pi_{k,k}$: their Young diagrams must have a number of boxes which is divisible by d^2. This can be seen for instance by induction: (π_1 and π_{ad}) are two representations made of 0 and d^2 boxes respectively. Now consider a representation π whose number of boxes is divisible by d^2; $\pi \otimes (\pi_1 \oplus \pi_{\mathrm{ad}})$ is again a direct sum of representation divisible by d^2, since tensoring with the trivial one does nothing and tensoring with the adjoint adds d^2 boxes to the Young diagram of π. According to Young calculus only d^2 boxes can be cancelled at once. Hence, if the statement is true for $(\pi_1 \oplus \pi_{\mathrm{ad}})^{\otimes k-1}$, then it holds for $(\pi_1 \oplus \pi_{\mathrm{ad}})^{\otimes k}$.
Indeed, all such representations π other than the trivial and the adjoint one satisfy $c_2(\pi) > 1$ as we will show in Lemma 7.14 below. □

Let $\lambda := (\lambda_1, \ldots, \lambda_{N-1})$ with $\lambda_i \in \mathbb{N}_0$ denote the Dynkin label of an irreducible representation π of $\mathfrak{su}(N)$. The eigenvalue of the Casimir element in the irreducible representation π is

$$c_2(\pi) = \frac{1}{2N} \sum_{i,j=1}^{N-1} (\lambda_i + 2)(A^{-1})_{i,j} \lambda_j \ , \tag{7.67}$$

where A is the Cartan matrix of $\mathfrak{su}(N)$ [55, §21.3]. The *inverse* Cartan matrix is directly given by

$$(A^{-1})_{i,j} = \frac{1}{N} \begin{cases} i\,(N-j) & \text{if } i \leq j \\ j\,(N-i) & \text{if } i > j \end{cases} , \tag{7.68}$$

and is symmetric. We now show the following lemma.

Lemma 7.14 (Young diagrams)**.** *Let $N > 2$ and π be an irreducible representation of $\mathfrak{su}(N)$ such that the number of boxes in its Young diagram is divisible by N. If π is* not *isomorphic to the trivial or adjoint representation, then $c_2(\pi) > 1$.*

Proof. First observe that we can immediately rule out all irreducible representations whose Young diagrams consist of a single column because the maximal column height for $\mathfrak{su}(N)$ is $N-1$ (i.e., Dynkin labels having a single entry 1 and 0 everywhere else). In the following we will analyse the growth behavior of the quadratic form (7.67) as we move from one irreducible representation (i.e., Dynkin label) to the next one.
It will turn out very helpful to know the column sums of the inverse Cartan matrix A^{-1}. Clearly, the sum of the first (or equally the last) column is $(N-1)/2$. The sum of any other column is strictly greater than this value. Indeed, pick a column j and denote its sum by a_j. One can easily convince oneself that $a_j = j\,(N-j)/2$.
Now we compare the quadratic Casimir eigenvalues of different irreducible representations, i.e., Dynkin labels λ. As it turns out, adding 1 to any component of any Dynkin label λ always increases this eigenvalue at least by almost $1/2$,

$$c_2(\lambda + e_i) - c_2(\lambda) \geq \frac{N^2-1}{2N^2} =: \Delta_N. \tag{7.69}$$

Here e_i is the i-th canonical basis vector of $\mathbb{R}^{N-1}$. So, starting from the trivial representation with $c_2(0,0,\ldots,0) = 0$ we immediately obtain the crude lower bound

$$c_2(\lambda_1, \ldots, \lambda_{N-1}) \geq \Delta_N \sum_{i=1}^{N-1} \lambda_i = \Delta_N \|\lambda\|_1. \tag{7.70}$$

Observe that $2\Delta_N < 1 < 3\Delta_N$. Thus we are guaranteed to obtain a quadratic Casimir eigenvalue strictly greater than 1 whenever we add at least three arbitrary columns to the (empty!) Young diagram of the trivial representation.
This leaves us with those irreducible representations whose Young diagrams have exactly two columns, i.e., with the Dynkin labels $(1,1,0,\dots,0)$, $(2,0,0,\dots,0)$ and all permutations thereof. As is well-known (and can be checked easily with the explicit formula below) the quadratic Casimir eigenvalue of the adjoint representation $(1,0,\dots,0,1)$ is exactly 1. We would like to show that any other placement of the two ones yields a strictly greater eigenvalue. Suppose these occur in positions $1 \le \alpha < \beta < N$. Then,

$$\begin{aligned} c_2(\lambda) &= \frac{1}{2N}\Big((A^{-1})_{\alpha,\alpha} + 2(A^{-1})_{\alpha,\beta} + (A^{-1})_{\beta,\beta} + 2a_\alpha + 2a_\beta\Big) \\ &\ge \frac{1}{2N}\Big((A^{-1})_{1,1} + 2(A^{-1})_{1,N-1} + (A^{-1})_{N-1,N-1} + 2a_1 + 2a_{N-1}\Big) \\ &= c_2(1,0,\dots,0,1) \\ &= 1. \end{aligned} \tag{7.71}$$

It is easy to see that this inequality turns into a strict one if either of the two ones is *not* at the first or last position. Finally consider a Dynkin label λ with a single non-vanishing component $\lambda_\alpha = 2$ at position α (i.e., a Young diagram with exactly two columns of height α),

$$c_2(\lambda) = \frac{2}{N}\Big((A^{-1})_{\alpha,\alpha} + a_\alpha\Big) = \frac{N+2}{N^2}\,\alpha\,(N-\alpha). \tag{7.72}$$

From the global minimum of the quadratic function $\alpha\,(N-\alpha)$ we easily obtain the lower bound

$$c_2(\lambda) \ge 1 + \frac{N-2}{N^2} \tag{7.73}$$

and thus $c_2(\lambda) > 1$ for all $N > 2$ as claimed. □

7.1.3 Hamiltonian driving

We now show that a time constant part in a stochastic Hamiltonian cannot affect the gap of the k-th moment operator.

Lemma 7.15 (Hamiltonian driving)**.** *Let M_T^k be the k-th moment operator of a universal Brownian motion with increments $\Theta_{\Delta t}$ as in eq. (7.6). Write $\Theta_{\Delta t}$ as*

$$\Theta_{\Delta t} = -\mathrm{i}\,H_0 + F_{\Delta t}\,, \tag{7.74}$$

where $-\mathrm{i}\,H_0$ and $F_{\Delta t}$ are its anti-Hermitian time constant and fluctating parts, respectively, with

$$F_{\Delta t} = \sum_\mu B_\mu\,\xi^\mu_{\Delta t}, \qquad B_\mu^\dagger = -B_\mu\,, \qquad \mathbb{E}[\xi^\mu_{\Delta t}] = 0\,, \qquad \text{and} \qquad \mathbb{E}[\xi^\mu_{\Delta t}\,\xi^\nu_{\Delta t}] = -\frac{a}{\Delta t}\,\delta_{\mu,\nu}\,.$$

Let $\tilde{M}_T^k$ be defined similarly but without driving, i.e., with $H_0 = 0$. Then $\tilde{M}_T^k$ and M_T^k have the same gap, i.e.,

$$\|\tilde{M}_T^k - M^k_{\mathrm{Haar}}\|_\infty = \|M_T^k - M^k_{\mathrm{Haar}}\|_\infty. \tag{7.75}$$

Proof. Lemma 3.14 implies that the gap of $M_{\Delta t}^k$ is $\|M_{\Delta t}^k - M^k_{\mathrm{Haar}}\|_\infty$. Hence,

$$\|M_T^k - M^k_{\mathrm{Haar}}\|_\infty = \lim_{\Delta t\to 0} \|M_{\Delta t}^k - M^k_{\mathrm{Haar}}\|_\infty^{T/\Delta t} \tag{7.76}$$

is the gap of M_T^k and, similarly, for $\tilde{M}_T^k$.

Using the connection between Brownian motion and its increments (7.4) and a Trotter-Suzuki approximation we obtain

$$\begin{aligned} M_{\Delta t}^k &= \mathbb{E}\left[\exp\{\pi_{k,k}(-\mathrm{i}H_0 + F_{\Delta t})\,\Delta t\}\right] + O(\Delta t^2) \\ &= \mathbb{E}\left[\exp\{\pi_{k,k}(F_{\Delta t})\,\Delta t\}\right]\,\exp\{\pi_{k,k}(-\mathrm{i}H_0)\,\Delta t\} + O(\Delta t^2) \\ &= \tilde{M}_{\Delta t}^k \exp\{\pi_{k,k}(-\mathrm{i}H_0)\,\Delta t\} + O(\Delta t^2)\,. \end{aligned} \tag{7.77}$$

As $\exp\{\pi_{k,k}(-\mathrm{i}H_0)\,\Delta t\}$ is a fixed unitary, up to an error of order $O(\Delta t^2)$, the gap of $M_{\Delta t}^k$ and $\tilde{M}_{\Delta t}^k$ are the same. This finishes the proof. □

7.1.4 More general interaction graphs

The generator from Lemma 7.9 of the k-th moment operator of the unitary Brownian motion inherits the locality structure from the increments (7.7). Hence, it can be written as

$$G^k = \sum_{e \in E} g_k^{(e)}\,, \tag{7.78}$$

where G^k is the generator associated to $\Theta_{\Delta t}$ and $g_k^{(e)}$ to $\theta_{\Delta t}^{(e)}$ according to eq. (7.23). Presumably, among all connected graphs, the gap of G^k could have a minimum for $1D$ nearest neighbour graphs. Here, we show that adding edges to this graph can only increase the gap, which can only lead to a faster mixing in Theorem 7.7.
In the following lemma, the spectral gap $\Delta(G)$ of an operator G is the difference of the second smallest and smallest singular value.

Lemma 7.16 (The spectral gap of the generator is concave)**.** *Let $(G_i)_i$ be a finite set of negative semidefinite and Hermitian operators with common non-trivial kernel and p be a probability vector. Then*

$$\Delta\left(\sum_i p_i\,G_i\right) \geq \sum_i p_i\,\Delta(G_i)\,. \tag{7.79}$$

This lemma implies that the gap of the generator (7.78) can only become smaller when one removes edges from E, while keeping E connected. Hence, the gap in the case of a one dimensional graph can also only be smaller as the gap in case of a complete graph.

Proof. Let K denote the common kernel of $(G_i)_i$. Then it is also the kernel of any operator in the convex hull of $(G_i)_i$. The gap of G_i is the smallest singular value of G_i restricted to the orthogonal complement of K and similarly for $G := \sum_i p_i\,G_i$. Hence, it is enough to show that the smallest singular value as the function

$$G \mapsto \min_{\langle x|x\rangle=1} |\,\langle x|\,G\,|x\rangle\,| \tag{7.80}$$

is concave. But this follows from the smallest singular value being the minimum of the linear functions $G \mapsto \langle x|\,G\,|x\rangle$. □

Remark 7.17 (Frustration free Hamiltonians)**.** *The same argument applies when the operators are all positive semidefinite. Hence, the gap of frustration free Hamiltonians, as considered in ref. [27], is also a concave function, i.e., it can only increase under taking convex combinations.*

7.1.5 Example: White noise in the Pauli basis

We conclude the discussion on approximate unitary designs with an example involving the specific setting in eqs. (7.85) and (7.86), and see that the choice of the Pauli matrices as a basis precisely matches, under the representation theoretic approach, the assumption on the covariance for the variables ξ.
Consider $n = 2$ qubits (thus $N = 4$) and the Hamiltonian increments

$$\Theta_{\Delta t} := -\mathrm{i} \sum_{\alpha,\beta=0}^{3} \sigma_\alpha \otimes \sigma_\beta \, \xi_{\Delta t}^{(\alpha,\beta)} \,, \tag{7.81}$$

where $\xi_{\Delta t}^{(\alpha,\beta)}$ are i.i.d. real random variables with zero mean and covariance

$$\mathrm{cov}[\xi_{\Delta t}^{(\alpha,\beta)} \xi_{\Delta t}^{(\alpha',\beta')}] = \delta_{\alpha,\alpha'}\delta_{\beta,\beta'} \frac{1}{\Delta t} \,, \qquad \forall \alpha, \beta. \tag{7.82}$$

Leaving out the term $\sigma_0 \otimes \sigma_0 \, \xi_{\Delta t}^{(0,0)}$ we can easily restrict $\Theta_{\Delta t}$ to its traceless part

$$\Theta_{0,\Delta t} = \sum_{\mu=1}^{15} \tau_\mu \, \xi_{\Delta t}^{\mu} \,, \tag{7.83}$$

where we defined the anti-Hermitian operators $\tau_\mu := -\mathrm{i}\, \sigma_{\mu_1} \otimes \sigma_{\mu_2}$ so that $\{ \tau_1, \tau_2, \dots, \tau_{15} \} = \{ \tau_{(0,1)}, \tau_{(0,2)}, \dots, \tau_{(3,3)} \}$ form a basis of the fundamental representation of $\mathfrak{su}(4)$. From eq. (7.12) we compute the Killing metric tensor (7.59) with respect to this basis as

$$\kappa_{\mu,\nu} = -8 \, \mathrm{Tr}(\tau_\mu^\dagger \tau_\nu) = -32 \delta_{\mu,\nu}. \tag{7.84}$$

From eq. (7.11) and the assumption in eq. (7.82) immediately follows $a = 32$. Observation 7.12 tells us then $g_2 = -16\, C(\pi_{2,2}) =$ and hence the second moment operator $M^{k=2}_{\mathrm{SLH}(\Delta t)}$ has a gap of $16\Delta t$, matching eq. (7.110) in the decoupling section that we are going to discuss.

With this, we conclude the analysis of the first principal results emerging from the study of the diffusion over the unitary group induced by stochastic local Hamiltonians. In the next section we will investigate a second mixing property for such a setting, namely, fast decoupling, again linked to the analogous result for random quantum circuit illustrated in Section 5.3.

7.2 Fast decoupling induced by Brownian motion on the unitary group

We show in the following decoupling with almost linear scaling in the system size under a unitary evolution obeying to Brownian motion laws. We interpret the time-fluctuating Hamiltonian in the framework of a *continuous-time random walk*, relating it with the discrete random walk induced by random quantum circuits with Haar distribution according to the description given in Chapter 5. The continuous-time walk has been first formalised by Montroll and Weiss [136] as a sequence of random transitions (jumps) spaced out by waiting times and been object of successive study [137], being applied to a wide range of fields of physics [138–140].
The exact correspondence between the accelerated steps of the random walk induced by random quantum circuits given in ref. [28] and the jumps of the continuous-time random walk generated by the fluctuating Hamiltonian infers a close similarity between the discrete circuit and the continuous process and can be hence used to relate results from these two settings. This fact is again to be seen in the task of bringing discrete and continuous-time processes involving the

unitary group under the same umbrella.

We restrict ourself into a more specific formulation of the fluctuating quantum Hamiltonian, considering an n-qubit system and taking the Pauli matrices together with the identity (times the imaginary unit) as a basis of the Lie algebra $\mathfrak{u}(2^n)$. Inspired by the Hamiltonian given in ref. [64], for which a reading of the *fast scrambling conjecture* has been studied, we set the increments in eqs. (7.7) and (7.8) of the Brownian motion to be

$$\Theta_{n,\ell,\Delta t} = -\mathrm{i}\, H_{n,\ell,\Delta t} \coloneqq -\mathrm{i} \left(\frac{2}{n(n-1)}\right)^{1/2} \sum_{j<k} \sum_{\alpha,\beta=0}^{3} \sigma_\alpha^j \otimes \sigma_\beta^k \, \xi_{\ell,\Delta t}^{(j,k,\alpha,\beta)} \,, \tag{7.85}$$

where $\sigma_\alpha^j \otimes \sigma_\beta^k$ means that $\sigma_\alpha \otimes \sigma_\beta$ is applied on qubits labeled j and k, respectively. We recall that $\xi_{\ell,\Delta t}^{(j,k,\alpha,\beta)}$ are i.i.d. real random variables with zero mean and covariance

$$\mathbb{E}\left[\xi_{\ell,\Delta t}^{(j,k,\alpha,\beta)} \, \xi_{\ell',\Delta t}^{(j',k',\alpha',\beta')}\right] = \frac{1}{\Delta t}\, \delta_{\ell,\ell'} \delta_{k,k'} \delta_{j,j'} \delta_{\alpha,\alpha'} \delta_{\beta,\beta'} \qquad \forall j,k,\alpha,\beta \tag{7.86}$$

which is obtained from eq. (7.13) by choosing $a = 8$. The pre-factor of $(2/(n(n-1)))^{1/2}$ is chosen so that the initial rate of diffusion of a local operator scales as $O(1/n)$. This is to normalize the time scale for the diffusion process in order to compare it with the random quantum circuit model in refs. [24, 28], where the probability that a local operator experiences a random gate is $2/n$ per discrete time step.
For our results, we need to define the permutation invariance property. This condition is required to deduce a dominant probability distribution on the final Pauli coefficients when starting with an analysis of the evolution of the Pauli weights. Indeed, the random walk on Pauli weights does not distinguish among strings having same support size but different support, hence it provides the probability distribution for each set of strings with the same support size, but not on Pauli strings taken singularly.

The permutation invariance property has already been debated in the proof of ref. [28] showing that random quantum circuits with Haar measure are approximate unitary 2-designs. In ref. [95] it has been discussed that this essential condition in the proof had not been granted and an argument making use on random transpositions based on the work of Diaconis (see refs. [38, 39]) has been put forward solving this issue. Since we cannot prove that permutation invariance is achieved with sufficiently high probability by the stochastic Hamiltonian evolution itself within a run time scaling almost linearly in n, we impose it as a pre-condition for the initial state. Actually, we can relax the condition and ask for a "large portion" of the qubits, but not necessarily all, to be invariant with respect to an arbitrary permutation. This allows us to apply our result to a larger family of states, for instance those whose support is very small. More formally, we define the permutation invariance property as follows.

Definition 7.18 (Permutation invariance property)**.** *Let $0 \leq \gamma < 1$. Let $\sigma_{\pi(\mu)}$ denote a Pauli string whose label is given by interchanging the sub-indices of μ according to the permutation π. Then, for an arbitrary quantum state ρ of a n-qubit system, we say that it satisfies the* γ-permutation invariance property *if there exists a subset of $(1-\gamma)n$ qubits which is invariant with respect to any permutation, i.e.,*

$$\mathrm{Tr}[\sigma_\mu \rho] = \mathrm{Tr}[\sigma_{\pi(\mu)} \rho] \tag{7.87}$$

for every Pauli string σ_μ and every permutation π on this subset of qubits.

Note that any state ρ with $|\mathrm{supp}(\rho)| \leq \gamma n$ is permutation invariant with respect to this definition.

Our second main result states that, under a unitary evolution describing Brownian motion, decoupling is achieved with a run time scaling almost linear in system size. First of all we need an upper bound on the distance between the distribution of these Pauli coefficients and a distribution which is close to the uniform one.
We denote the Pauli basis coefficients after a continuous-time evolution with run time T as

$$Q^T(\mu,\nu) := \frac{1}{4^n}\,\mathrm{Tr}\left[\sigma_\nu \otimes \sigma_\nu\, M^{k=2}_{n,\mathrm{SLH}(T)}(\sigma_\mu \otimes \sigma_\mu)\right]. \tag{7.88}$$

Then we formulate the following condition.

Theorem 7.19 (Mixing condition for Pauli coefficients, Theorem 12 in ref. [41])**.** *For any constants $\delta \in (0, 1/16)$, $\eta \in (0,1)$ there exist constants $\varsigma > 0$ and $0 < \gamma_0 \leq 1/2$ such that for a total run time $T \geq \varsigma\, n \log^2 n$ and large enough n*

$$\sum_{\nu\in\{0,1,2,3\}^n,\nu\neq 0} \left|Q^T(\mu,\nu) - p_\delta(\nu)\right| \leq \frac{1}{(3-\eta)^\ell \binom{n}{\ell}} \frac{1}{\mathrm{poly}(n)} \tag{7.89}$$

where σ_μ is an arbitrary string whose support has size ℓ and has a subset of $(1-\gamma)n$ qubits, with $\gamma < \gamma_0$, which is invariant with respect to any permutation, and p_δ is a (possibly sub-normalised) distribution on Pauli strings such that

$$p_\delta(\nu) \leq \frac{5^{\delta n}}{4^n - 1} \quad \forall \nu. \tag{7.90}$$

From Theorem 7.19 we obtain the final result on decoupling. Theorem 7.20 is achieved by arguments analogous to the random quantum circuit case given in ref. [24] so that we do not need any additional work to prove it.

Theorem 7.20 (Fast decoupling, Theorem 13 in ref. [41])**.** *Consider a bipartite quantum state $\rho_{AE} \in \mathcal{S}_{AE}$ of an n-qubit system A coupled with some other system E. Let then ρ_{AE} undergo a unitary evolution U_t induced by stochastic local Hamiltonian increments as in eq. (7.85) acting upon system A, followed by a completely positive trace preserving map $\mathcal{T} : \mathcal{S}_A \to \mathcal{S}_B$ which maps from A to another system B. Let $\tau_{A'B}$ denote the Choi-Jamiolkowski isomorph of $\mathcal{T}$. Then, for any $\delta \in (0, 1/16)$ there exist $\varsigma > 0$ and $0 < \gamma_0 \leq 1/2$ such that for all γ-permutation invariant states with $\gamma < \gamma_0$ and total run times $T \geq \varsigma\, n \log^2 n$ and for large enough n*

$$\mathbb{E}_{\mathrm{SLH}(T)}\left\{\left\|\mathcal{T}\left(U_T\,\rho_{AE}U_T^\dagger\right) - \tau_B \otimes \rho_E\right\|_1\right\} \leq \left(\frac{1}{\mathrm{poly}(n)} + 5^{\delta n}\cdot 2^{-H_2(A|B)_\tau - H_2(A|E)_\rho}\right)^{1/2}, \tag{7.91}$$

where $\mathrm{SLH}(T)$ denotes the distribution over the unitary group induced by the Brownian motion with run time T.

7.2.1 Proof of Theorem 7.19

The section is devoted to the proof Theorem 7.19. To show our result, we consider a fluctuating Hamiltonian on a complete graph whose increments are given in eq. (7.85), in the limit of $\Delta t \to 0$. As already mentioned, this result then implies Theorem 7.20 by application of the same proof technique used for the random quantum circuit case in ref. [24].
First, we analyse how the support size of an initial Pauli string evolves during the process, then we observe how the qubits are made invariant under relabelling of the Pauli elements; this, together with the permutation invariance condition, leads to the desired result. Decoupling of an

arbitrary n-qubit system A is mainly described by the second moment operator induced by the evolution. The expansion coefficients in the Pauli basis are given in eq. (7.88). We recall that, since the Brownian motion on $\mathbb{U}(2^n)$ is Markovian, the second moment operator at time T on X is given by concatenating $T/\Delta t$ times the operator $M^{k=2}_{n,\mathrm{SLH}(\Delta t)}$, i.e.,

$$M^{k=2}_{n,\mathrm{SLH}(T)}(X) = \lim_{\Delta t\to 0} \underbrace{M^{k=2}_{n,\mathrm{SLH}(\Delta t)} \circ \cdots \circ M^{k=2}_{n,\mathrm{SLH}(\Delta t)}}_{T/\Delta t \text{ times}} (X) \tag{7.92}$$

$$=: \lim_{\Delta t\to 0} \bigcirc_{s=1}^{T/\Delta t} M^{k=2}_{n,\mathrm{SLH}(\Delta t)} (X) \ . \tag{7.93}$$

Note that, since the Hamiltonian in eq. (7.85) generating Brownian motion is dependent on system size, we must include an additional subscript.
In Taylor approximation, up to an error $O(\Delta t^2)$, $M^{k=2}_{n,\mathrm{SLH}(\Delta t)}$ results from the sum of two-qubit moment operators acting on any possible qubit pair j,k, i.e.,

$$M^{k=2}_{n,\mathrm{SLH}(\Delta t)} = \frac{2}{n(n-1)} \sum_{j<k} \left(M^{k=2}_{2,\mathrm{SLH}(\Delta t)}\right)^{j,k} + O(\Delta t^2) \ . \tag{7.94}$$

This can be seen through calculations analogous to the ones from eqs. (7.37)-(7.39). We can hence interpret this process as a qubit pair being uniformly randomly chosen at every time step $(\ell-1)\Delta t$ and a two-qubit unitary $U_{2,\ell,\Delta t} := \exp\{-\mathrm{i}\, H_{2,\ell,\Delta t}\, \Delta t\}$ being applied. Therefore, in the following subsection we first consider the restricted two-qubit case, which provides useful results and insights to be used for the investigation of the general case with n qubits.

Two-qubit analysis of the second moment operator

Considering a two-qubit system, here we would like to understand the evolution of $M^{k=2}_{2,\mathrm{SLH}(T)}$ through $M^{k=2}_{2,\mathrm{SLH}(\Delta t)}$ and show the following lemma, which is compatible to the analysis of the local gap discussed in the previous section (as showed in Example 7.1.5) .

Lemma 7.21 (Two-qubit case)**.** *Then the local second moment operator associated to the Hamiltonian increments* (7.85) *converges exponentially to the second moment operator of the uniform distribution, i.e.,*

$$\|M^{k=2}_{2,\mathrm{SLH}(T)} - M^{k=2}_{2,\mathrm{Haar}}\|_\infty \leq \mathrm{e}^{-16\,T}. \tag{7.95}$$

Proof. To prove the convergence rate, we want to express $M^{k=2}_{n,\mathrm{SLH}(\Delta t)}$ in terms of the Pauli basis and compute the gap. We can see directly that the identity on 4 qubits is an eigenvector with unit eigenvalue

$$M^{k=2}_{2,\mathrm{SLH}(\Delta t)}(\mathbb{1}_4) = \mathbb{E}\left[U^{\otimes 2}_{2,\ell,\Delta t} \mathbb{1}_4 (U^\dagger_{2,\ell,\Delta t})^{\otimes 2}\right] = \mathbb{1}_4. \tag{7.96}$$

We then observe the unitary evolution acting on a Pauli element $\sigma_\mu \otimes \sigma_\nu$, with $\mu,\ \nu \in \{0,1,2,3\}^2$ and calculate its expectation with a Taylor expansion for the unitary, taking into account terms with leading order in Δt (and suppressing subscripts for H),

$$\begin{aligned} M^{k=2}_{2,\mathrm{SLH}(\Delta t)}(\sigma_\mu \otimes \sigma_\nu) &= \mathbb{E}\left[U_{2,\ell,\Delta t}\ (\sigma_{\mu_1}\otimes\sigma_{\mu_2})\ U^\dagger_{2,\ell,\Delta t} \otimes U_{2,\ell,\Delta t}\ (\sigma_{\nu_1}\otimes\sigma_{\nu_2})\ U^\dagger_{2,\ell,\Delta t}\right] \\ &= \mathbb{E}\left[\left(\mathbb{1}_2 - \mathrm{i}H\Delta t - \frac{1}{2}H^2\Delta t^2\right)\ (\sigma_{\mu_1}\otimes\sigma_{\mu_2})\ \left(\mathbb{1}_2 + \mathrm{i}H\Delta t - \frac{1}{2}H^2\Delta t^2\right)\right. \\ &\quad \left. \otimes \left(\mathbb{1}_2 - \mathrm{i}H\Delta t - \frac{1}{2}H^2\Delta t^2\right)\ (\sigma_{\nu_1}\otimes\sigma_{\nu_2})\ \left(\mathbb{1}_2 + \mathrm{i}H\Delta t - \frac{1}{2}H^2\Delta t^2\right)\right] \\ &\quad + O(\Delta t^2). \end{aligned} \tag{7.97}$$

We now recall that the ξ white noise variables are i.i.d. with zero mean and covariance as in eq. (7.86). Considering only the non-vanishing linear terms in Δt in the expectation, we have

$$\begin{aligned} M^{k=2}_{2,\mathrm{SLH}(\Delta t)}(\sigma_\mu \otimes \sigma_\nu) = \sigma_\mu \otimes \sigma_\nu + \Delta t^2 \mathbb{E}\Big[H\sigma_\mu H \otimes \sigma_\nu + \sigma_\mu \otimes H\sigma_\nu H\Big] \\ - \frac{\Delta t^2}{2}\mathbb{E}\left[H^2\sigma_\mu \otimes \sigma_\nu + \sigma_\mu H^2 \otimes \sigma_\nu + \sigma_\mu \otimes H^2\sigma_\nu + \sigma_\mu \otimes \sigma_\nu H^2\right] \\ - \Delta t^2 \mathbb{E}\Big[[H,\sigma_\mu] \otimes [H,\sigma_\nu]\Big] + O(\Delta t^2). \end{aligned} \tag{7.98}$$

Let us consider the second term, in particular

$$\mathbb{E}\left[H\sigma_\mu H \otimes \sigma_\nu\right] = \tfrac{1}{\Delta t}\left(\textstyle\sum_{\alpha,\beta}(\sigma_\alpha \otimes \sigma_\beta)(\sigma_{\mu_1} \otimes \sigma_{\mu_2})(\sigma_\alpha \otimes \sigma_\beta)\right) \otimes (\sigma_{\nu_1} \otimes \sigma_{\nu_2}). \tag{7.99}$$

If $\mu = 0$, then

$$\mathbb{E}\left[H\,\mathbb{1}_2\,H \otimes \sigma_\nu\right] = \mathbb{E}\left[H^2 \otimes \sigma_\nu\right] = \frac{16}{\Delta t}\,\mathbb{1}_2 \otimes \sigma_\nu. \tag{7.100}$$

Otherwise, for $\mu \neq 0$, at least one among μ_1 and μ_2 is not 0. Let us assume $\mu_1 \neq 0$. Then, $\forall \beta$, $\sigma_\alpha\sigma_{\mu_1}\sigma_\alpha \otimes \sigma_\beta\sigma_{\mu_2}\sigma_\beta$ equals $\sigma_{\mu_1} \otimes \sigma_\beta\sigma_{\mu_2}\sigma_\beta$ for $\alpha = 0, \mu_1$ and $-\sigma_{\mu_1} \otimes \sigma_\beta\sigma_{\mu_2}\sigma_\beta$ for the other two indices of α. Thus, summing over α gives 0. The same applies for μ_1 arbitrary, $\mu_2 \neq 0$. We conclude that the second term in the expression for $M^{k=2}_{2,\mathrm{SLH}(\Delta t)}$ vanishes if both μ and ν are different from $\{0,0\}$.
Now we look at the first part of the third term and we get that

$$\mathbb{E}\left[H^2\sigma_\mu \otimes \sigma_\nu\right] = \frac{1}{\Delta t}\sum_{\alpha,\beta}(\sigma_\alpha \otimes \sigma_\beta)^2\sigma_\mu \otimes \sigma_\nu = \frac{16}{\Delta t}\,\sigma_\mu \otimes \sigma_\nu. \tag{7.101}$$

Hence, keeping terms to leading order in Δt we have

$$M^{k=2}_{2,\mathrm{SLH}(\Delta t)}(\sigma_\mu \otimes \sigma_\nu) = (1 - 32\Delta t)\sigma_\mu \otimes \sigma_\nu - \Delta t^2 \mathbb{E}\left[[H,\sigma_\mu] \otimes [H,\sigma_\nu]\right] \tag{7.102}$$

$$= (1 - 32\Delta t)\sigma_\mu \otimes \sigma_\nu - \Delta t \sum_{\alpha,\beta}[\sigma_\alpha \otimes \sigma_\beta, \sigma_{\mu_1} \otimes \sigma_{\mu_2}] \otimes [\sigma_\alpha \otimes \sigma_\beta, \sigma_{\nu_1} \otimes \sigma_{\nu_2}], \tag{7.103}$$

when both μ and ν are different from $\{0,0\}$, and conversely

$$M^{k=2}_{2,\mathrm{SLH}(\Delta t)}(\mathbb{1}_2 \otimes \sigma_\nu) = (1 - 16\Delta t)\mathbb{1}_2 \otimes \sigma_\nu, \tag{7.104}$$

$$M^{k=2}_{2,\mathrm{SLH}(\Delta t)}(\sigma_\mu \otimes \mathbb{1}_2) = (1 - 16\Delta t)\sigma_\mu \otimes \mathbb{1}_2. \tag{7.105}$$

We now divide the set of all possible strings $\sigma_\mu \otimes \sigma_\nu$ in three parts: the identity $\mathbb{1}_4$, the set of strings of the form $\sigma_\mu \otimes \sigma_\mu$, and all remaining strings of the form $\sigma_\mu \otimes \sigma_\nu$ with $\mu \neq \nu$. We can then make use of the matrix representation of the operator $M^{k=2}_{2,\mathrm{SLH}(\Delta t)}$ as a matrix with respect to Pauli basis, which gives

$$M^{k=2}_{2,\mathrm{SLH}(\Delta t)} = \begin{pmatrix} 1 & & \\ & A & \\ & & B \end{pmatrix}, \tag{7.106}$$

where A is a 15×15 matrix related to the set of $\sigma_\mu \otimes \sigma_\mu$ elements (without the identity $\mathbb{1}_{16}$) and B is a 240×240 matrix for $\sigma_\mu \otimes \sigma_\nu$ elements. The detailed proof of this finding is laid out in the separate subsequent Lemma 7.22.
We now consider the matrix A; we compute the action of $M^{k=2}_{2,\mathrm{SLH}(\Delta t)}$ over all possible $\sigma_\mu \otimes \sigma_\mu$ and look for eigenvalues. We obtain a non-degenerate eigenvalue 1 whose eigenvector is the uniform sum over all non-identity Pauli matrices

$$\mathbb{F} = \frac{1}{15}\sum_{\gamma \neq 0}\sigma_\gamma \otimes \sigma_\gamma. \tag{7.107}$$

We then have a 9-fold degenerate eigenvalue $1-40\Delta t$ and a 5-fold degenerate eigenvalue $1-24\Delta t$. We are free to bound all these eigenvalues with $1-16\Delta t$. We now deal with the action of the second moment operator on terms of the form $\sigma_\mu \otimes \sigma_\nu$ with $\mu, \nu \neq 0$ and $\mu \neq \nu$. Only four choices of $\sigma_\alpha \otimes \sigma_\beta$ do not commute for a given pair μ, ν, i.e.,

$$M^{k=2}_{2,\mathrm{SLH}(\Delta t)}(\sigma_\mu \otimes \sigma_\nu) = (1-32\Delta t)\sigma_\mu \otimes \sigma_\nu - 4\Delta t\{\pm\sigma_{\gamma_1} \otimes \sigma_{d_1} \pm \sigma_{\gamma_2} \otimes \sigma_{d_2} \pm \sigma_{\gamma_3} \otimes \sigma_{d_3} \pm \sigma_{\gamma_4} \otimes \sigma_{d_4}\} \tag{7.108}$$

with $\gamma_i \neq d_i$, for each $\sigma_\mu \otimes \sigma_\nu$. This means that each column of the matrix B has one entry $(1-32\Delta t)$ (in the diagonal element) and four entries $\pm 4\Delta t$, and 0 otherwise. Hence,

$$\|B\|_1 = \max_j \sum_i |a_{i,j}| = 1 - 16\Delta t. \tag{7.109}$$

By the Gershgorin circle theorem, and taking also into account (7.104) and (7.105), we can upper bound the highest eigenvalue of B with $1-16\Delta t$. For a single time step, the two-qubit second moment operator can be upper bounded by the following diagonal matrix

$$M^{k=2}_{2,\mathrm{SLH}(\Delta t)} \leq \begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1-16\Delta t & & & \\ & & & \ddots & & \\ & & & & 1-16\Delta t \end{pmatrix}, \tag{7.110}$$

where we recall that the 2-fold degenerate eigenvalue 1 corresponds to the identity and $\mathbb{F}$. □

Lemma 7.22 (Local second moment operator)**.** *$M^{k=2}_{2,\mathrm{SLH}(\Delta t)}$ is Hermitian, maps elements of the set of strings of the form $\sigma_\mu \otimes \sigma_\mu$ to a linear combination of elements of the same set and elements of the set of strings of the form $\sigma_\mu \otimes \sigma_\nu$ with $\mu \neq \nu$ again to a linear combination of elements of the same set, such that there is no mixing between the two sets. Hence, we can represent the operator $M^{k=2}_{2,\mathrm{SLH}(\Delta t)}$ as a matrix with respect to Pauli basis in the following form*

$$M^{k=2}_{2,\mathrm{SLH}(\Delta t)} = \begin{pmatrix} 1 & & \\ & A & \\ & & B \end{pmatrix}, \tag{7.111}$$

where A is a 15×15 matrix related to the set of $\sigma_\mu \otimes \sigma_\mu$ elements (without the identity $\mathbb{1}_{16}$) and B is a 240×240 matrix for $\sigma_\mu \otimes \sigma_\nu$ elements.

Proof. From eq. (7.102) and (7.103), follows directly that $M^{k=2}_{2,\mathrm{SLH}(\Delta t)}$ is Hermitian. Moreover we see, again from eq. (7.103), that elements of the set $\sigma_\mu \otimes \sigma_\mu$ are mapped to a linear combination of elements of the same set. This, in addition to the fact that $M^{k=2}_{2,\mathrm{SLH}(\Delta t)}$ is Hermitian, implies that elements of the set $\sigma_\mu \otimes \sigma_\nu$ with $\mu \neq \nu$ are mapped again to a linear combination of elements of the same set. □

Next, we make use of this analysis to understand how n-qubit Pauli strings evolve during the continuous-time process. As already mentioned, the continuous-time random walk induced by the Hamiltonian increments can be interpreted as a sequence of jumps defining a discrete random walk spaced out by i.i.d. waiting times.

Markov chain analysis on weights

The proof strategy for Lemma 7.19 begins with the analysis of the evolution of the coefficients: we observe how the support size behaves during the process, inferring a probability that, for a given initial string σ_μ with support size ℓ, after run time T the string has support size k. Conditioned on some specific event E_W that we will discuss later, this probability can be upper bounded as

$$\mathbb{P}\left(\{T,\ell,k\} \;\middle|\; E_W\right) := \sum_{|\nu|=k} Q^T_{E_W}(\mu,\nu) \leq \binom{n}{k} 3^k \frac{4^{\delta n}}{4^n - 1}. \tag{7.112}$$

Having a total of $\binom{n}{k}3^k$ strings with support size k, we then show that almost all of them have the same probability.
Considering the analysis in the previous section on the two-qubit case and that, the local structure of $M^{k=2}_{n,\mathrm{SLH}(\Delta t)}$ given in eq. (7.94) we introduce a Markov chain over the weights of the string similarly to ref. [28] (where this projected chain is called *zero chain*). The chain runs over the state space $\Omega = \{1,2,\ldots,n\}$ and the transition probability from ℓ at time t to k at time $t+\Delta t$ is described by the matrix element

$$P(\ell,k) := \sum_{\nu:|\nu|=k} \frac{1}{4^n} \operatorname{Tr}\left[\sigma_\nu \otimes \sigma_\nu \, M^{k=2}_{n,\mathrm{SLH}(\Delta t)} (\sigma_\mu \otimes \sigma_\mu)\right] \tag{7.113}$$

for any choice of μ with support size ℓ.

Lemma 7.23 (Transition matrix of the zero chain)**.** *The zero chain has transition matrix P on state space $\Omega = \{1,2,\ldots,n\}$,*

$$P(\ell,k) = \begin{cases} 1 - \frac{16\ell(3n-2\ell-1)}{n(n-1)}\Delta t & k = \ell \\ \frac{16\ell(\ell-1)}{n(n-1)}\Delta t & k = \ell - 1 \\ \frac{48\ell(n-\ell)}{n(n-1)}\Delta t & k = \ell + 1 \\ 0 & \textit{otherwise} \end{cases} \tag{7.114}$$

for $1 \leq \ell, k \leq n$.

Proof. We consider the analysis of the two-qubit second moment operator in Section 7.2.1. It is straightforward to note that, after application of $M^{k=2}_{n,\mathrm{SLH}(\Delta t)}$, the weight of the string can only vary by 1 or stay the same. The weight decreases if a pair of two non-identity terms $\sigma \otimes \sigma$ is chosen and is transformed in a pair with one identity element (namely, $\sigma \otimes \mathbb{1}$ or $\mathbb{1} \otimes \sigma$); there are in total four choices for $\sigma_\alpha \otimes \sigma_\beta$ which produce such a transition. According to the two-qubit case, the probability that one of these Pauli operators is chosen is $4 \cdot 4\Delta t = 16\Delta t$ and since the probability of choosing a pair with weight 2 is $\ell(\ell-1)/(n(n-1))$, we have

$$P(\ell,\ell-1) = \frac{16\ell(\ell-1)}{n(n-1)}\Delta t \ . \tag{7.115}$$

The weight of the string can be increased if an identity term paired with a non-identity term is chosen (i.e., $\sigma \otimes \mathbb{1}$ or $\mathbb{1} \otimes \sigma$) and transformed into a pair of two non-identity terms $\sigma \otimes \sigma$. The probability of obtaining such a result (conditioned on choosing such a pair) after application of the two-qubit second moment operator is $24\Delta t$, since there are in total 6 choices for $\sigma_\alpha \otimes \sigma_\beta$ to produce such a transition. Furthermore, the probability of choosing an identity and non-identity pair is given by $2\ell(n-\ell)/(n(n-1))$; hence

$$P(\ell,\ell+1) = \frac{48\ell(n-\ell)}{n(n-1)}\Delta t. \tag{7.116}$$

Finally, the probability of staying at the same weight is obtained by simply requiring the total probability to sum to unity. □

It is therefore possible to reach each state of the chain, meaning that it is *irreducible*. Moreover, the chain contains self loops, being hence *aperiodic*. From these two properties follows that the chain is also *ergodic*, thus converging to a unique stationary distribution.

Lemma 7.24 (Stationary distribution of zero chain)**.** *The stationary distribution of the zero chain is*

$$\omega_0(k) = \frac{3^k \binom{n}{k}}{4^n - 1}. \tag{7.117}$$

Proof. This follows from straightforward calculation. □

The stationary distribution is actually analogous to the one of the chain induced by a random quantum circuit under the Haar measure (see ref. [28, Lemma 5.3]). Another crucial analogy is the exact equivalence of the *accelerated chain* (i.e., the chain conditioned on moving) of the two different settings. This means that, when moving, the random walk on weights is identically biased for both random quantum circuits under Haar distribution and the stochastic Hamiltonian process. From the description of Montroll and Weiss, the jumps of the random quantum circuit are contained in the fluctuating Hamiltonian evolution, spaced out by i.i.d. waiting times. Concretely, the accelerated chain is given by

$$P_{\text{accel}}(\ell, k) = \begin{cases} 0 & k = \ell \\ \frac{\ell - 1}{3n - 2\ell - 1} & k = \ell - 1 \\ \frac{3(n-\ell)}{3n-2\ell-1} & k = \ell + 1 \\ 0 & \text{otherwise.} \end{cases} \tag{7.118}$$

With these analogies, we can prove the next theorem using results from the proof of ref. [24, Theorem 4.2], many of which are illustrated in Section 5.3. We should take care of the parts of the proof involving the waiting time, because it is where the two walks differ. Now, we re-formulate the result for the continuous-time case.

Lemma 7.25 (Mixing condition on support size)**.** *Let P be the Markov chain transition matrix defined in Lemma 7.23. For any constants $\delta \in (0, 1/16), \eta \in (0, 1)$ there exists a constant $\varsigma > 0$ such that for $T \geq \varsigma\, n \log^2 n$ and all integers $1 \leq \ell \leq n$ and $1 \leq k \leq n$, we have for large enough n*

$$\mathbb{P}\left(\{T, \ell, k\}\right) = \sum_{\nu : |\nu| = k} Q^T(\ell, k) \leq \binom{n}{k} 3^k \frac{4^{\delta n}}{4^n - 1} + \frac{1}{(3-\eta)^\ell \binom{n}{\ell}} \frac{1}{\text{poly}(n)}\ , \tag{7.119}$$

where $\{T, \ell, k\}$ is the event that an initial Pauli string with support size ℓ, after a run time T, has weight equal to k.

Proof. We start by defining the following points,

$$r_- := \left(\frac{3}{4} - \delta\right) n \quad \textit{and} \quad r_+ := \left(\frac{3}{4} + \delta\right) n. \tag{7.120}$$

Then, considering eq. (5.82), it follows that for an initial weight of $\ell \in [r_-, r_+]$

$$\mathbb{P}(\{T, \ell, k\}) \leq \binom{n}{k} 3^k \frac{4^{\delta n}}{4^n - 1} \tag{7.121}$$

for any $T > 0$.

To deal with the case $\ell \in [1, r_-)$, for random quantum circuits it has been shown that the probability that the interval $[r_-, r_+]$ of the state space has been reached is very high for a number of gates $O(n \log^2 n)$. Here we prove the same scaling result for the run time of the continuous-time process, that is, the total waiting time between the jumps, can be bound with the following lemma.

Lemma 7.26 (Waiting time)**.**

$$\mathbb{P}(E_W^c) \coloneqq \mathbb{P}(W_{r_-} > \varsigma\, n \log^2 n) \leq \frac{1}{(3-\eta)^\ell \binom{n}{\ell}} \frac{1}{\mathrm{poly}(n)} \tag{7.122}$$

for some sufficiently large ς*.*

Proof of Lemma 7.26. To prove this result on the waiting time, we first assume that we reach the region $[r_-, r_+]$ within $S \leq s$ accelerated steps for some $s = O(n)$ and we bound the probability that the waiting time exceeds $\varsigma\, n \log^2 n$. We will deal with the case of $\mathbb{P}(S > s)$ afterwards. Now, let M be the smallest site visited during the walk, and let $\{y_i\}_{i=1}^S$ be a sequence of accelerated steps where $S \leq s$, with waiting times $\{W_i\}_{i=1}^S$ respectively, satisfying the event

$$H = \bigcap_{j=1}^{n} \left[\sum_{k=1}^{S} \mathbb{I}(X_k \leq j) \leq zj/\mu \right], \tag{7.123}$$

where $\mathbb{I}$ is the indicator function and X_k is the random variable assuming values in $\Omega = \{1, 2, \ldots, n\}$ describing the state of the chain at step k and z chosen as $O(\log n)$. In words, this means that, if H occurs, then no site has been visited "too often". This is a useful event, since the smaller is the value of the current state of the chain, the smaller is the parameter of the exponential distribution dominating the waiting time. Namely, we have

$$1 - P(k,k) = \frac{16k(3n-2k-1)}{n(n-1)} \Delta t \geq \frac{16k}{n} \Delta t. \tag{7.124}$$

So, dealing with three events, we consider the bound

$$\begin{aligned}
\mathbb{P}(W > t) &= \mathbb{P}(W > t \ \cap \ H \ \cap \ S \leq s) + \mathbb{P}(W > t \ \cap \ H \ \cap \ S > s) \\
&\quad + \mathbb{P}(W > t \ \cap \ H^c \ \cap \ S \leq s) + \mathbb{P}(W > t \ \cap \ H^c \ \cap \ S > s) \\
&\leq \mathbb{P}(W > t \mid H \ \cap \ S \leq s) + \mathbb{P}(H \ \cap \ S > s) \\
&\quad + \mathbb{P}(H^c \ \cap \ S \leq s) + \mathbb{P}(H^c \ \cap \ S > s) \\
&\leq \mathbb{P}(W > t \mid H \ \cap \ S \leq s) + \mathbb{P}(H^c \mid S \leq s) + \mathbb{P}(S > s).
\end{aligned} \tag{7.125}$$

Conditioning on the two previous event and setting $M = m$ for arbitrary $m \in \{1, \ldots, \ell\}$, we have to find an upper bound for the waiting time being too large; more precisely for a given run time t, we show:

Lemma 7.27 (Waiting time conditioning on event H)**.**

$$\max_{\{y_i\}} \mathbb{P}\left(W(y_1) + \cdots + W(y_S) \geq t \;\middle|\; M = m \ , \ H \right) \leq \mathrm{e}^{-\frac{8k}{n} t}\, 2^{zm/\mu} \mathrm{e}^{zm/(2\mu) \log n}. \tag{7.126}$$

Proof of Lemma 7.27. We recall that this is the exactly the sequence visiting m for zm/μ (for simplicity, we assume it to be an integer) times and all other $j > m$ sites for z/μ times, hence

$$W(y_1) + \cdots + W(y_S) \leq \sum_{i=1}^{zm/\mu} E_{m,i} + \sum_{i=1}^{z/\mu} \sum_{k=m+1}^{r} E_{k,i}, \tag{7.127}$$

where $E_{k,i}$ are i.i.d. exponential distributions with parameter $p(k) = 16k/n$. Now applying Markov's inequality we obtain

$$\mathbb{P}\left(\sum_{i=1}^{zm/\mu} E_{m,i} + \sum_{i=1}^{z/\mu}\sum_{k=m+1}^{r} E_{k,i} > t\right) \leq \frac{\mathbb{E}\left[\exp\left\{\alpha\left(\sum_{i=1}^{zm/\mu} E_{m,i} + \sum_{i=1}^{z/\mu}\sum_{k=m+1}^{r} E_{k,i}\right)\right\}\right]}{\mathrm{e}^{\alpha t}} \tag{7.128}$$

$$= \mathrm{e}^{-\alpha t}\left(\frac{p(m)}{p(m)-\alpha}\right)^{zm/\mu}\prod_{k=m+1}^{r}\left(\frac{p(k)}{p(k)-\alpha}\right)^{z/\mu}$$

for $\alpha < p(m)$. Let us choose $\alpha = p(m)/2$, then we have

$$\mathbb{P}\left(\sum_{i=1}^{zm/\mu} E_{m,i} + \sum_{i=1}^{z/\mu}\sum_{k=m+1}^{r} E_{k,i} > t\right) \leq \mathrm{e}^{-\frac{8m}{n}t}\, 2^{zm/\mu}\left(\prod_{k=m+1}^{r}\frac{2k}{2k-m}\right)^{z/\mu} \tag{7.129}$$

$$\leq \mathrm{e}^{-\frac{8m}{n}t}\, 2^{zm/\mu}\mathrm{e}^{zm/(2\mu)\,\log n}.$$

□

With this lemma we obtain an equivalent result for the waiting time as in ref. [24] up to the prefactor of t. Hence, for $t > \varsigma\, n\log^2 n$ with ς sufficiently large, applying the bounds on the probabilities $\mathbb{P}(M = m)$ for each value of $m \in \{1,\ldots,\ell\}$ proved for the random quantum circuit case, we have

$$\mathbb{P}\left(W_{r_-} > t \mid H \cap S \leq s\right) = \sum_{m=1}^{\ell}\mathbb{P}(M=m)\max_{\{y_i\}}\mathbb{P}\left(W(y_1)+\cdots+W(y_S) \geq t \mid M=m\right) \tag{7.130}$$

$$\leq \frac{1}{(3-\eta)^{\ell}\binom{n}{\ell}}\frac{1}{\mathrm{poly}(n)}.$$

The last two probability terms in eq. (7.125) depend only on the path of the accelerated random walk before reaching the interval $[r_-, r_+]$. Looking at the accelerated chain and considering ℓ being in the region $[1, (3/4-\delta)n)$, we have $3(n-\ell)/(3n-2l-1) \geq 1/2+\delta$ for any n. So, constructing a random walk X'_k starting at the origin moving forward with probability $1/2+\delta$ and backward with $1/2-\delta$, it follows

$$\mathbb{P}\left(S > s\right) \leq \mathbb{P}\left(X'_s < r_- - \ell\right) \tag{7.131}$$

$$= \mathbb{P}\left(X'_s < 2\delta\, s - (2\delta\, s + \ell - r_-)\right)$$

$$\leq \exp\left(-\frac{(2\delta s + \ell - r_-)^2}{2s}\right),$$

where in the last inequality we have used the Chernoff-Hoeffding bound in ref. [28, Lemma A.3] assuming $2\delta s + \ell - r_- > 0$. We conclude that the probability for the waiting time to be larger than $s \geq \phi n$ is exponentially decreasing in n for large enough ϕ. The last remaining term in eq. (7.125) can instead be bounded by

$$\mathbb{P}(H^c \mid S \leq s) \leq \frac{1}{(3-\eta)^{\ell}\binom{n}{\ell}}\frac{1}{\mathrm{poly}(n)} \tag{7.132}$$

so that the proof of Lemma 7.26 is now complete. □

The case that remains to be discussed is the one of an initial Pauli string with support size $\ell \in (r_+, n]$ to reach $[r_-, r_+]$; again the analysis is divided on accelerated steps and waiting times. Regarding the former, the probability of going backward is larger than the one of moving forward starting from point z with

$$P(z, z+1) \stackrel{!}{=} P(z, z-1) \qquad \Leftrightarrow \tag{7.133}$$

$$\frac{z-1}{3n-2z-1} = 3\frac{n-z}{3n-2z-1}, \tag{7.134}$$

from which follows that

$$z = \frac{3}{4}n + \frac{1}{4}. \tag{7.135}$$

This means that for any $n > 1/(4\delta)$ the probability of moving backward at each site of region $(r_+, n]$ is at least $1/2 + \epsilon$ for some $\epsilon > 0$, and again using the argument for the case with $\ell < r_-$ the probability of not reaching r_+ in $S \leq s$ steps is upper bounded by an exponential decreasing function for $s \geq \phi' n$ for sufficiently large ϕ'. In this instance, all waiting times are stochastically dominated by parameter $p(3n/4) = 12$, hence there is no necessity to define an event equivalent to H. For $S \leq s$ accelerated steps, using again a Chernoff-Hoeffding inequality, the bound on the total waiting time is exponentially decreasing in s for a run time $W_{r_+} > (\log 2/6)s$. The proof of Lemma 7.25 is then complete. □

From the zero chain to the full distribution

Once the weight distribution has reached an equilibrium such that the condition in eq. (7.112) is fulfilled, we need to show that all Pauli strings sharing the same weight have a similar probability. To prove this, we need to show that almost all Pauli strings with the same support but different Pauli labels $\{1, 2, 3\}$ are equivalent in probability. This, together with the permutation invariance property assumed for the initial state, which is conserved during the whole stochastic Hamiltonian process, will bring us to the desired result.

Let M be the Markov chain on the first n-qubits induced by $M^2_{n,\Delta t}$, and define an accelerated version as

$$A := \frac{1}{36\Delta t}(M - (1 - 36\Delta t)\mathcal{I}). \tag{7.136}$$

If we define an operator

$$R = \frac{2}{n(n-1)} \sum_{j<k} R_{j,k} \ , \tag{7.137}$$

where $R_{j,k}$ randomises one qubit site in the following way,

$$R_{j,k}(\sigma^j_\mu \otimes \sigma^k_\nu) = \begin{cases} \frac{1}{3}\sum_{\alpha=1,2,3} \sigma^j_\alpha \otimes \mathbb{1}^k & \text{if } \mu \neq 0, \nu = 0, \\ \frac{1}{3}\sum_{\alpha=1,2,3} \mathbb{1}^j \otimes \sigma^k_\alpha & \text{if } \mu = 0, \nu \neq 0, \\ \frac{1}{6}\sum_{\alpha=1,2,3} \sigma^j_\alpha \otimes \sigma^k_\nu + \frac{1}{6}\sum_{\alpha=1,2,3} \sigma^j_\mu \otimes \sigma^k_\alpha & \text{if } \mu \neq 0, \nu \neq 0, \\ \mathbb{1}^j \otimes \mathbb{1}^k & \text{if } \mu = \nu = 0, \end{cases} \tag{7.138}$$

then according to Section 7.2.1, the accelerated chain can be written as

$$A = \frac{1}{3}R + \frac{2}{3}L, \tag{7.139}$$

where

$$L = \frac{2}{n(n-1)} \sum_{j<k} L_{j,k} \tag{7.140}$$

and

$$L_{j,k}(\sigma_\mu^j \otimes \sigma_\nu^k) = \begin{cases} \frac{1}{6}\sum_{\alpha=1,2,3} \sigma_{\mu+1}^j \otimes \sigma_\alpha^k + \frac{1}{6}\sum_{\alpha=1,2,3} \sigma_{\mu+2}^j \otimes \sigma_\alpha^k & \text{if } \mu \neq 0, \nu = 0, \\ \frac{1}{6}\sum_{\alpha=1,2,3} \sigma_\mu^j \otimes \sigma_{\nu+1}^k + \frac{1}{6}\sum_{\alpha=1,2,3} \sigma_\mu^j \otimes \sigma_{\nu+2}^k & \text{if } \mu = 0, \nu \neq 0, \\ \frac{1}{12}\left(\sigma_{\mu+1}^j \otimes \sigma_\nu^k + \sigma_{\mu+2}^j \otimes \sigma_\nu^k + \sigma_\mu^j \otimes \sigma_{\nu+1}^k + \sigma_\mu^j \otimes \sigma_{\nu+2}^k\right) & \\ +\frac{1}{6}\left(\sigma_{\mu+1}^j \otimes \mathbb{1}^k + \sigma_{\mu+2}^j \otimes \mathbb{1}^k + \mathbb{1}^j \otimes \sigma_{\nu+1}^k + \mathbb{1}^j \otimes \sigma_{\nu+2}^k\right) & \text{if } \mu \neq 0, \nu \neq 0, \\ \mathbb{1}^j \otimes \mathbb{1}^k & \text{if } \mu = \nu = 0, \end{cases} \tag{7.141}$$

with the notation $\sigma_{3+1} = \sigma_{2+2} = \sigma_1$ and $\sigma_{3+2} = \sigma_2$. Note that R does not produce any change in the weight or transpositions between identities and non-identity elements, it solely performs a local randomisation of the Pauli labels. This means that only the chain L is responsible for the random walk on the weights.
We would like to upper bound the probability that more than βn sites have not been randomised after s steps of chain R (we denote the complement of this event as E_R). Knowing that there are $\binom{n}{\beta n}$ such regions, this is given by union bound

$$\mathbb{P}(E_R^c) \leq \binom{n}{\beta n}(1-\beta)^s \leq 2^{h(\beta)n}\, \mathrm{e}^{-\beta s}, \tag{7.142}$$

where $h : [0,1] \to [0,1]$ is the binary entropy function. This probability can then be upper bounded by an arbitrarily exponentially decreasing function in n for some $s = O(n)$. Hence, to ensure that s randomisations have been performed to fulfill the event E_R with sufficiently large probability, given eq. (7.139) and by application of an Hoeffding's inequality follows that it is again sufficient to apply $O(n)$ steps of the accelerated chain A. Since the waiting time is dominated by an exponential distribution with parameter 36, the bound on the probability for the waiting time of this process to exceed $W_R = \varsigma_R\, n$ can be bounded by an arbitrarily exponentially decreasing function in n for a sufficiently large ς_R with the same argument used for the random walk on weights when starting from $\ell > r_+$.
In conclusion, assuming that event E_W and E_R have been satisfied, we have for $\gamma < \gamma_0 \leq 1/2$:

1. For strings ν with support size $k \leq \gamma_0\, n$,

$$Q^T(\mu,\nu) \leq \sum_{|\nu|=k} Q^T(\mu,\nu) \leq \binom{n}{\gamma_0\, n} 3^{\gamma_0\, n} \frac{4^{\delta n}}{4^n - 1} \leq 2^{n\, h(\gamma_0)}\, 3^{\gamma_0\, n} \frac{4^{\delta n}}{4^n - 1}. \tag{7.143}$$

2. For strings ν with support size $k \geq (1-\gamma_0)n$, given event E_R at least $(1-\beta)n$ sites of the support have been uniformly randomised, hence

$$Q^T(\mu,\nu) \leq \frac{1}{3^{k-\beta\, n}} \sum_{|\nu|=k} Q^T(\mu,\nu) \leq 2^{n\, h(\gamma_0)}\, 3^{\beta\, n} \frac{4^{\delta n}}{4^n - 1}. \tag{7.144}$$

3. For strings ν with support size $\gamma_0\, n < k = \kappa n < (1-\gamma_0)n$ such that $\kappa - \gamma_0 = O(1)$ (otherwise, we can apply slightly modified versions of the bounds in the two previous cases),

given event E_R at least $(1-\beta)n$ sites of the support have been uniformly randomised. In addition, if we assume the γ-permutation invariance property for the initial string σ_μ, we obtain

$$Q^T(\mu,\nu) \leq \frac{1}{3^{k-\beta\, n}} \frac{1}{\binom{(1-\gamma)n}{k-\gamma n}} \sum_{|\nu|=k} Q^T(\mu,\nu) \leq 3^{\beta\, n} \left[\frac{1}{\kappa-\gamma_0}\right]^{\gamma_0\, n} \frac{4^{\delta n}}{4^n-1}. \tag{7.145}$$

Now, for an appropriate choice of β and γ_0,

$$Q^T(\mu,\nu) \leq \frac{5^{\delta n}}{4^n-1} \tag{7.146}$$

for all μ and ν.
Also, having proven that there exists ς such that, for all $T \geq \varsigma n \log^2 n$, $\mathbb{P}(E_R^c)$ is bounded by an exponentially decreasing function in n and that

$$\mathbb{P}(E_W^c) \leq \frac{1}{(3-\eta)^\ell \binom{n}{\ell}} \frac{1}{\mathrm{poly}(n)} \tag{7.147}$$

and having proven that, if both event have been satisfied and the permutation invariance property is assumed, we have

$$Q^T(\mu,\nu) \leq \frac{5^{\delta n}}{4^n-1} \tag{7.148}$$

for all μ and ν, we conclude the proof for the main Lemma 7.19.

As mentioned in the main result section, the decoupling theorem is valid for all states which are invariant with respect to any permutation on $(1-\gamma)n$ qubits, in the sense of Definition 7.18, and not only for Pauli strings taken singularly. Consider a set of $\min\left\{\binom{n-\gamma n}{\ell-\gamma n}, \binom{n-\gamma n}{\ell}\right\} \leq b_{\ell,\gamma} \leq \max\left\{\binom{n-\gamma n}{\ell-\gamma n}, \binom{n-\gamma n}{\ell}\right\}$ Pauli strings $\{\sigma_\mu\}_\mu$ with support size ℓ which is invariant with respect to any of such permutations. Assuming that the above events have been satisfied, at least the same number of qubits in the final Pauli strings $\{\sigma_\nu\}_\nu$ is invariant with respect to permutations since the stochastic evolution preserves this property. Hence, for the argument from the previous subsection, we have:

$$\sum_\mu Q^T(\mu,\nu) \leq b_{\ell,\gamma} \frac{5^{\delta n}}{4^n-1} \tag{7.149}$$

This, together with the fact that $\mathrm{Tr}[\sigma_\mu \rho]$ is the same for all strings related by these permutations, allows to apply the proof in ref. [24] for the decoupling Theorem for all density states ρ composed by permutation invariant sets of Pauli strings.

7.3 Fast scrambling and other applications

We discuss in the following two interesting applications for Brownian motion on the unitary group, that is, black holes scrambling and dissipative dynamics, and hint at a third one making use of fluctuating Hamiltonian dynamics in quantum information processing.

Fast scrambling In the last decade, black holes have been considered from a quantum information perspective, providing toy models and fresh insights to the field. In particular, it has been conjectured that they are *fast scramblers* [37, 141, 142]. A system is *scrambled* when any previous perturbation has been thoroughly spread among the degrees of freedom so that to recover the information contained in the original perturbation one should access simultaneously

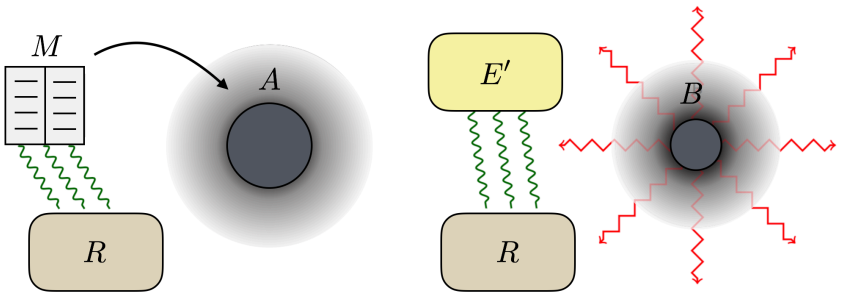


Figure 7.2: A quantum memory system M is initially entangled with a reference system denoted by R and is subsequently thrown into a black hole A (left picture). As the black holes leaks out Hawking radiation, it shrinks into a smaller system B. When a controlled subystem of the irradiated environment E', has become maximally entangled with the reference system R, the initial information M has been mirrored (right picture).

a large fraction of the entire system. The minimum time for mixing information is then called *scrambling time*. More specifically, in ref. [37] three hypotheses have been outlined: the most rapid scramblers take logarithmic time in the degrees of freedom, the bound is saturated for matrix quantum mechanics, i.e., systems whose degrees of freedom are $n \times n$ matrices, black holes are the fastest scramblers in Nature. The authors of ref. [64] brought evidence about the conjectures regarding scrambling in logarithmic time by investigating Brownian quantum circuit and Ising model on sparse random graphs. There are two related mixing conditions for unitary dynamics that satisfy the requirements for scrambling, as discussed in refs. [37, 141], or ref. [142], respectively. The relation between our results and both of these conditions will be discussed in the following.

In ref. [141], one considers the black hole's internal system A and the radiated environment E. Furthermore, one defines an additional reference system R, initially maximally entangled with a quantum memory system M that is subsequently thrown into the black hole. As the Hawking radiation leaks out, we would like R to become maximally entangled with a subsystem of E over which we can have control, hence having recovered the initial state of M, and so interpreting the black hole as a mirror (see Fig. 7.2). This may be translated into a scrambling condition through a decoupling theorem. As the black hole evaporates, A shrinks into a smaller system B which decouples from R. More formally, this means that

$$\mathbb{E}_{\text{Haar}} \left\{ \left\| \text{Tr}_{A\setminus B} \left(U_A \, \rho_{AR} U_A^\dagger \right) - \frac{\mathbb{1}_B}{|B|} \otimes \rho_R \right\|_1 \right\} \leq 2^{-\gamma}, \tag{7.150}$$

where ρ_{AR} is a quantum state where subsystem E shares m Bell pairs with A, and A is otherwise mixed, and γ is the difference between the number of qubits emitted as Hawking radiation and the number of qubits of system M. The approximate statement

$$\mathbb{E}_{\omega} \left\{ \left\| \text{Tr}_{A\setminus B} \left(U_A \, \rho_{AR} U_A^\dagger \right) - \frac{\mathbb{1}_B}{|B|} \otimes \rho_R \right\|_1 \right\} \leq \sqrt{4^{-\gamma} + 4^m \varepsilon} \tag{7.151}$$

is satisfied in expectation for an ensemble of unitary transformations ω being an approximate 2-design in the sense that the Pauli coefficients are close to the uniform distribution, i.e.,

$$\sum_{\nu \neq 0} |q_\omega(\mu, \nu) - q_u(\mu, \nu)| \leq \varepsilon \qquad \forall \mu \ , \tag{7.152}$$

where

$$q_u(\mu,\nu) = \frac{1}{4^n-1} \quad \forall \mu,\nu \qquad \text{and} \qquad q_\omega(\mu,\nu) = \frac{1}{4^n} \operatorname{Tr}\left[\sigma_\nu \otimes \sigma_\nu \, M_\omega^{k=2}(\sigma_\mu \otimes \sigma_\mu)\right] . \tag{7.153}$$

This condition was shown in ref. [28] to be satisfied by a random quantum circuit of size $O(n \log n)$ (when $\epsilon = 1/\mathrm{poly}(n)$) and analogously by a stochastic local Hamiltonian, according to the analysis on random walk in Section 7.2.1 and following the same reasoning as in ref. [28], with a run time $T = O(n \log n)$. However, in order to compare time scales with ref. [64], we take the same convention and divide the global scrambling time by the time it takes to scramble a single subsystem; in this case we obtain a scrambling time of $\tau_* = O(\log n)$. Hence, our work also provides an alternative proof for the scaling of the scrambling time in ref. [64], although our argument does not involve any intermediate conjecture, such as the final statements of ref. [64, Appendix B].
In ref. [142], a slightly different scrambling condition is required for the unitarity of black hole evaporation to hold, given postselection on the final state at the singularity inside the black hole. One considers the composite system $\mathcal{H}_M \otimes \mathcal{H}_{\text{in}} \otimes \mathcal{H}_{\text{out}}$ representing the infalling matter, the infalling negative energy Hawking radiation behind the event horizon and the outgoing positive energy Hawking radiation outside the horizon, respectively. Again, one defines a reference system S which is maximally entangled with a subsystem $M_1 \subset M$. After the application of a random unitary transformation U on $\mathcal{H}_M \otimes \mathcal{H}_{\text{in}}$ and subsequently tracing out the complement subsystem of S, we have (cfr. [142, eq.(3)])

$$\mathbb{E}_{\text{Haar}} \left\{ \left\| \operatorname{Tr}_{\overline{S}} \left(U \, \rho \, U^\dagger \right) - \frac{\mathbb{1}_S}{|S|} \right\|_1 \right\} \leq \sqrt{\frac{|\mathcal{H}_{M_1}|}{|\mathcal{H}_{in}|}}. \tag{7.154}$$

A relaxed version of this bound, namely,

$$\mathbb{E}_{\omega} \left\{ \left\| \operatorname{Tr}_{\overline{S}} \left(U \, \rho \, U^\dagger \right) - \frac{\mathbb{1}_S}{|S|} \right\|_1 \right\} \leq \sqrt{5^{\delta n} \frac{|\mathcal{H}_{M_1}|}{|\mathcal{H}_{in}|} + \frac{1}{\mathrm{poly}(n)}} \, , \tag{7.155}$$

where $n = \log_2 \left(|\mathcal{H}_{in}||\mathcal{H}_M| \right)$, follows from the condition

$$\sum_{\nu \neq 0} |q_\omega(\mu,\nu) - 4^{\delta n} q_u(\mu,\nu)| \leq \frac{1}{(3-\eta)^\ell \binom{n}{\ell}} \frac{1}{\mathrm{poly}(n)} \, , \tag{7.156}$$

for every Pauli string σ_μ with support size ℓ .

The above condition was shown to hold in ref. [24] for random quantum circuits of size $O(n \log^2 n)$. Applying the equivalence established Section 7.2.1, it follows from Theorem 7.19 that this is fulfilled by a stochastic local Hamiltonians in time $\tau'_* = O(\log^2 n)$, when again we take the convention of ref. [64] and divide global scrambling time by the time to scramble a single subsystem.

Dissipative dynamics arising from fluctuations As pointed at the beginning of this chapter, there is an intimate relationship between time-fluctuating dynamics and *Markovian dissipative evolution*, a connection that we will now make explicit. Brownian motion U_t on the unitary group yields an average dynamics given by

$$\rho(t) \coloneqq \mathbb{E}[U_t \rho U_t^\dagger], \tag{7.157}$$

which describes a dissipative quantum Markovian evolution of the state ρ. In this sense, *time-fluctuating classical noise* is precisely a specific source of *dissipation*. Indeed, the use of *controlled*

dissipative Markovian dynamics has received much interest in recent years [143–145].

The generator of the dynamical semi-group given by the evolution (7.157) has been calculated in Lemma 7.9. For $k = 1$, the generator is immediately expressed in the following proposition.

Proposition 7.28 (Fluctuations as dissipative processes)**.** *Let U_t be a Brownian motion with increments $\Theta_{\Delta t}$ as in eq. (7.6). Write $\Theta_{\Delta t}$ as*

$$\Theta_{\Delta t} = -\mathrm{i}\, H_0 + F_{\Delta t}\,, \tag{7.158}$$

where $-\mathrm{i}\, H_0$ and $F_{\Delta t}$ are its anti-Hermitian time constant and fluctuating parts, respectively, with

$$F_{\Delta t} = \sum_\mu B_\mu\, \xi^\mu_{\Delta t}, \qquad B_\mu^\dagger = -B_\mu\,, \qquad \mathbb{E}[\xi^\mu_{\Delta t}] = 0\,, \qquad \text{and} \qquad \mathbb{E}[\xi^\mu_{\Delta t}\, \xi^\nu_{\Delta t}] = -\frac{a}{\Delta t}\, \delta_{\mu,\nu}\,.$$

Then $\rho(t) = \mathbb{E}[U_t \rho U_t^\dagger]$ gives rise to a quantum dynamical semi-group and evolves according to the Lindblad equation

$$\frac{d}{dt}\rho(t) = -\mathrm{i}\,[H_0, \rho(t)] - a \sum_\mu \left(B_\mu \rho B_\mu^\dagger - \frac{1}{2}\left(B_\mu^\dagger B_\mu \rho + \rho B_\mu^\dagger B_\mu \right) \right) \tag{7.159}$$

with $\rho(0) = \rho$.

Proof. According to Lemma 7.9 the evolution has a generator

$$G^1 = \lim_{t\to 0} \mathbb{E}\left[\Theta_t \otimes \mathbb{1} + \mathbb{1} \otimes \overline{\Theta}_t\right] + \frac{1}{2} \lim_{t\to 0} \mathbb{E}[(\Theta_t \otimes \mathbb{1} + \mathbb{1} \otimes \overline{\Theta}_t)^2]\, t\,, \tag{7.160}$$

where the mixed tensor representation $\pi_{1,1}$ from eq. (7.20) is used. It remains to show that the generator is of Lindblad form. The proposition's hypothesis yields

$$G^1 = -i\, H_0 \otimes \mathbb{1} + \mathrm{i}\, \mathbb{1} \otimes \overline{H}_0 + \frac{1}{2} \lim_{t\to 0} \mathbb{E}[(F_t^2 \otimes \mathbb{1} + \mathbb{1} \otimes \overline{F}_t^2 + 2F_t \otimes \overline{F}_t)]\, t \tag{7.161}$$

$$= -i\, H_0 \otimes \mathbb{1} + \mathrm{i}\, \mathbb{1} \otimes H_0^T - \frac{a}{2} \sum_\mu (B_\mu^2 \otimes \mathbb{1} + \mathbb{1} \otimes B_\mu^{2\,T} - 2B_\mu \otimes B_\mu^T)\,, \tag{7.162}$$

where we have used that $\overline{B}_\mu = -B_\mu^T$. The identification $\mathrm{vec}(XYZ) = (X \otimes Z^T)\,\mathrm{vec}(Y)$ and $B_\mu^\dagger = -B_\mu$ finish the proof. □

Applications in quantum information processing We finally mentioned a third, immediate, application, which seems yet particularly important when having potential technological applications in quantum information processing in mind. It should be now clear that whenever the aim is to realize an approximate unitary design, the evolution under a fluctuating Hamiltonian constitutes a valuable option. We have illustrated a number of domains of quantum information where unitary designs play a prominent role, and seen that with a suitable random circuit one can generate approximate unitary designs. Such a scheme, however, requires the precise implementation of a deep quantum circuit consisting of a large number of local quantum gates. The above results tell that, instead of implementing a quantum circuit, a suitably stochastic Hamiltonian evolution can be used to give rise to exactly the same dynamics and hence be chosen whenever such an approach is more feasible.

CHAPTER 8

Conclusions and outlook

In this thesis we investigated distributions over the unitary group and properties thereof from a quantum information theoretic perspective. We motivated the fundamental importance of the group by discussing applications in a number of circumstances and settings. After the preliminaries, in Chapter 3 we introduced the notion of distributions of unitaries, in particular highlighting the concept of universality and the construction of the Haar measure as a uniform distribution. We formalized in the same chapter a notion whose importance we have extensively motivated throughout the work: suitable distributions that constitute so-called unitary designs which approximate the Haar measure. As a first novel result, we discussed how quantum cryptography in the Private Quantum Channel formalism is more robust than the classical counterpart in presence of an imperfect source of randomness by expressing the encryption through a random unitary operator drawn from a unitary design. We then illustrated in Chapter 4 another particularly relevant topic in quantum information theory strictly tied to unitary 2-designs, namely, decoupling. This is involved in a number of applications, from state merging to thermodynamic equilibration and again, as we discussed in the last chapter, as a black hole scrambling condition. We have collected different formulations for decoupling theorems, in terms of different entropy measures such as the quantum collision entropy, the min-entropy and the hypothesis-testing entropy. In Chapter 5 we connected these two topics, unitary distributions and decoupling, with a prominent setting of quantum information, that is, random quantum circuits. Indeed, with this construction one is able to implement efficiently unitary designs and also decouple rapidly a system from another. In Chapter 6 we illustrated randomized benchmarking, one of the most relevant emerging applications for random quantum circuits. By twirling noise channels over suitable groups of unitary operators, we can estimate the fidelity of experimental gates implementations. Previous literature focuses on twirls over 2-designs, such as the Clifford group, to characterize this quantity. While those protocols rely on the depolarization of the noise channel, we exploit conversely symmetries of the associated target unitary gate: the underlying mathematical tools of representation theory and in particular the Schur's Lemma force the twirled noise matrix representation into a block-diagonal form, whose matrix entries can be fitted as parameters. This, at the cost of higher classical computational resources, allows to benchmark individually quantum gates outside the Clifford group. In Chapter 7 we eventually switched to a continuous-time framework with the final goal of unifying random processes over the unitary group under the same umbrella. More precisely, we showed that results on unitary designs and decoupling obtained with the implementation of a random quantum circuit can

be extended to Brownian motions on the manifold of the unitary group. We formalized the stochastic process through differential Brownian increments on the Lie algebra subsequently injected through the exponential map into the group. Two applications of this diffusion model are presented: dissipation and black holes scrambling. One of the first outlook of this work is to find other settings where the presented results can be applied; we conjecture that this is the case of experimental implementations of noisy Hamiltonians for quantum optics purposes.
Again regarding the continuous-time scheme, one should investigate a possible improvement in the scaling of the moments convergence with respect to their degree, perhaps to the point of making the result completely independent of it, as already argued in ref. [27] for the random quantum circuit counterpart. Another important improvement would be to remove the permutation invariant condition for the initial state considered for fast decoupling. We tried several shuffling and mixing techniques (cfr. refs. [38, 39, 48]) without success, but we are still of the opinion that this assumption can be eliminated.
Now moving to the novel benchmarking protocol, one of the first follow-up projects should certainly involve a sharper bound for the confidence interval in the fashion of the one provided in ref. [112] for Clifford group twirling. An approach in the fashion of ref. [113], studying the tensor representations of the involved symmetry groups, could in principle lead to similar bounds.

This work does not however aim at providing new results and bounds in terms of new frameworks only, but also wants to stimulate mathematical applications in quantum information theory. In order to develop a novel randomized benchmarking protocol, we extensively applied group and representation theoretic tools to exploit systems and operators symmetries. Furthermore, we extended the discrete-time setting of random quantum circuits to one of the most relevant topics of probability theory, namely, Brownian motion. To prove equivalent results, we shifted to a Lie algebra formulation and made use of tools such as Young diagrams and Dynkin labels to establish a gap for the local generator of the moment operator which is entirely independent of its degree. This is by itself a remarkable result, in the same fashion of the one that Bourgain and Gamburd have shown in ref. [98] for gate sets with algebraic entries, where in our case we also provide an explicit value. Again, to show our version of decoupling theorem with random unitaries chosen accordingly to a diffusion model and to estimate convergence times, we combined random walk on Pauli strings results from refs. [28] and [24] regarding stationary distributions, projections of chains and barrier absorptions. This author hence hopes that both results as well as mathematical techniques will be of interest in this and other fields.

Acknowledgements and final words

Almost four years have passed since the start of my PhD journey. While preparing for this adventure I read, somewhere, that the PhD period resemble in some way a monastic experience, at the end of which the worthy candidate is awarded with a degree and the academic dress. Never been a monk myself, I still believe that the metaphor may be very accurate, at least if I consider my path. Devotion to knowledge and will of breaking its boundaries are the leading principles of the PhD student; reading new articles on the arXiv is is morning ritual (although true monks probably start the day a bit earlier!); parting form family and friends and isolating from the world for long hours of science contemplation his trial of endurance and perseverance; the pleasure of discovery and of interacting with very interesting and capable scientists his reward.

Now leaving the metaphor, let me thank the persons who accompanied and guided me during my PhD studies. First of all, I would like to thank my supervisor, Jens Eisert, a thousand of times, even though this wouldn't be nearly remotely sufficient to express my gratitude to him. He opened to me the doors of the fabulous (not always, but most of the time!) academic research world, assigned me extremely engaging and intriguing projects, inspired me with his ideas, wisdom and teaching. Thank you SO much for all of this, Jens!
During this years I had the immense pleasure of being part of the *Randomness Team*, together with other four brilliant scientists. I express my greatest gratitude to Albert H. Werner, Martin Kliesch and Oliver Buerschaper for their collaboration and mentoring, and to Winton Brown for the same reasons, but even more importantly for his invaluable friendship.
I also would like to say thank you to all members of our group: everyone contributed to create a splendid and pleasant atmosphere and a stimulating research environment, every single day.

Now my family. First of all comes my mother Claudia, for all the love, support and protection she has given to me. No words will ever be able to express her a grain of my love and gratitude. Thanks to my father Giuseppe, for transmitting to me the pure love for mathematics, Nature and its laws; my journey in science started from his teaching, ranging from Greek mythology to algebra. Thanks to my aunt Marina and my uncle Stöffi for all their affection, support and for the beautiful holidays in Japan. Finally, thanks to our Lancillotto, the valiant rabbit. Thanks to you all, with all my heart.

Bibliography

[1] Albert Einstein. Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik*, 322(6):132–148, 1905.

[2] Max Born. Zur Quantenmechanik der Stoßvorgänge. *Zeitschrift für Physik*, 37(12):863–867, 1926.

[3] Werner Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43(3):172–198, 1927.

[4] Erwin Schrödinger. Quantisierung als Eigenwertproblem. *Annalen der Physik*, 384(4): 361–376, 1926.

[5] Erwin Schrödinger. Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften*, 23(48):807–812, 1935.

[6] Hugh Everett. *Theory of the Universal Wavefunction*. PhD thesis, Princeton University, 1956.

[7] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, 1935.

[8] John Stewart Bell. On the Einstein-Podolsky-Rosen paradox. *Physics Physique Fizika*, 1: 195–200, 1964.

[9] David Hilbert. Probleme der Grundlegung der Mathematik. *Mathematische Annalen*, 102: 1–9, 1929. Lecture given at the International Congress of Mathematicians, 3 September 1928.

[10] David Hilbert and Wilhelm Ackermann. *Grundzüge der Theoretischen Logik*. Cambridge Studies in Advanced Mathematics. Springer-Verlag Berlin Heidelberg, 1959.

[11] Kurt Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I. *Monatshefte für Mathematik und Physik*, 38(1):173–198, 1931.

[12] Alonzo Church. A note on the Entscheidungsproblem. *Journal of Symbolic Logic*, 1:40–41, 1936.

[13] Alan Turing. On computable numbers, with an application to the Entscheidungsproblem. In *Proceedings of the London Mathematical Society*, volume 42 of *2*, pages 230–265, 1936.

[14] Claude E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27:379–423, 1948.

[15] Gordon E. Moore. Cramming more components onto integrated circuits. *Electronics*, 38: 114, 1965.

[16] Aram W. Harrow and Ashley Motanaro. Quantum computational supremacy. *Nature*, 549: 203–209, 2017.

[17] Peter W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, SFCS '94, pages 124–134. IEEE Computer Society, 1994. ISBN 0-8186-6580-7.

[18] R. L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM*, 21(2):120–126, 1978.

[19] Michael J. Bremner, Ashley Montanaro, and Dan J. Shepherd. Average-case complexity versus approximate simulation of commuting quantum computations. *Phys. Rev. Lett.*, 117:080501, 2016.

[20] Sergio Boixo and et al. Characterizing quantum supremacy in near-term devices. *Nature Physics*, 14:595 – 600, 2018.

[21] C. Neill and et al. A blueprint for demonstrating quantum supremacy with superconducting qubits. *Science*, 360(6385):195–199, 2018.

[22] Micheal A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 10th anniversary edition, 2010.

[23] Renato Renner. Quantum information theory, 2011. Lecture notes ETH.

[24] Winton Brown and Omar Fawzi. Decoupling with random quantum circuits. *Comm. Math. Phys.*, 340:867–900, 2015.

[25] Winton Brown and Lorenza Viola. Convergence rates for arbitrary statistical moments of random quantum circuits. *Phys. Rev. Lett.*, 104:250501, 2010.

[26] Yaakov S. Weinstein, Winton Brown, and Lorenza Viola. Parameters of pseudorandom quantum circuits. *Phys. Rev. A*, 78:052332, 2008.

[27] F. G. S. L. Brandao, A. W. Harrow, and M. Horodecki. Local random quantum circuits are approximate polynomial-designs. *Comm. Math. Phys.*, 346:397–434, 2016.

[28] Aram W. Harrow and Richard A. Low. Random quantum circuits are approximate 2-designs. *Comm. Math. Phys.*, 291:257, 2009.

[29] Michael A. Nielsen, Mark R. Dowling, Mile Gu, and Andrew C. Doherty. Quantum computation as geometry. *Science*, 311(5764):1133–1135, 2006.

[30] Michael A. Nielsen, Mark R. Dowling, Mile Gu, and Andrew C. Doherty. Optimal control, geometry, and quantum computing. *Phys. Rev. A*, 73:062323, 2006.

[31] Mark R. Dowling and Michael A. Nielsen. The geometry of quantum computation, 2006. arXiv:quant-ph/0701004.

[32] David Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, second edition, 2009.

[33] David Applebaum. *Probability on Compact Lie Groups*. Probability Theory and Stochastic Modelling. Springer International Publishing, first edition, 2014.

[34] S. Itō. Brownian motions in a topological group and in its covering group. *Rend. Circ. Mat. Palermo*, 1:40–48, 1952.

[35] Boris Tsirelson. Unitary Brownian motions are linearisable, 1998. arXiv:math/9806112.

[36] Ming Liao. *Lévy processes in Lie groups*, volume 162. Cambridge university press, 2004.

[37] Yasuhiro Sekino and Leonard Susskind. Fast scramblers. *JHEP*, 10:65, 2008.

[38] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Probability theory and related fields*, 57(2):159–179, 1981.

[39] David Aldous and Persi Diaconis. Shuffling cards and stopping time. *Am. Math. Soc. Mon.*, 93(5):333–348, 1986.

[40] Persi Diaconis. The cutoff phenomenon in finite Markov chains. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 93, pages 1659–1664, 1996.

[41] E. Onorati, O. Buerschaper, M. Kliesch, W. Brown, A. H. Werner, and J. Eisert. Mixing properties of stochastic quantum Hamiltonians. *Communications in Mathematical Physics*, 355(3):905–947, 2017.

[42] E. Onorati, A. H. Werner, and J. Eisert. Randomized benchmarking for individual quantum gates. *Phys. Rev. Lett.*, 123:060501, 2019.

[43] Michael A. Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. *Phys. Rev. A*, 303:249–252, 2002.

[44] John Watrous. Semidefinite programs for completely bounded norms. *Th. Comp.*, 5(11), 2009.

[45] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson. Robust randomized benchmarking of quantum processes. *Phys. Rev. Lett.*, 106:042311, 2011.

[46] Joel J. Wallman and Steven T. Flammia. Randomized benchmarking with confidence. *New. J. Phys.*, 16:103032, 2014.

[47] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[48] Persi Diaconis and Laurent Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *The Annals of Applied Probability*, 6(3):695–750, 1996.

[49] Brian C. Hall. *Lie Groups, Lie Algebras, and Representations.* Springer International Publishing, second edition, 2015.

[50] Andrew Knighlty and Charles Li. *Traces of Hecke Operators*, volume 133. American Mathematical Society, 2006.

[51] André Weil. *L'intégration dans les groupes topologiques et ses applications*, volume 1145. Hermann, 1965.

[52] Frédéric Dupuis, Mario Berta, Jürg Wullschleger, and Renato Renner. One-shot decoupling. *Communications in Mathematical Physics*, 328(1):251–284, 2014.

[53] Oleg Szehr, Frédéric Dupuis, Marco Tomamichel, and Renato Renner. Decoupling with unitary approximate two-designs. *New Journal of Physics*, 15(5):053022, 2013.

[54] P. Horodecki, M. Horodecki, and R. Horodecki. General teleportation channel, singlet fraction and quasidistillation. *Phys. Rev. A*, 60, 1999.

[55] William Fulton and Joe Harris. *Representation theory: A first course*. Springer, Heidelberg, 1991.

[56] Benoit Collins and Piotr Sniady. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Comm. Math. Phys.*, 264:773–795, 2006.

[57] F. G. S. L. Brandao and M. Horodecki. Exponential quantum speed-ups are generic. *Q. Inf. Comp.*, 13:0901, 2013.

[58] S. Hallgren and A. W. Harrow. Superpolynomial speedups based on almost any quantum circuit. *Proc. of the 35th Int. Coll. Aut. Lang. Prog. LNCS*, 5125:782, 2008.

[59] V. Dunjko and H. J. Briegel. Quantum mixing of Markov chains for special distributions. *New Journal of Physics*, 17(7):073004, 2015.

[60] M. Kabanava, R. Kueng, H. Rauhut, and U. Terstiege. Stable low-rank matrix recovery via null space properties. *Information and Inference: A Journal of the IMA*, 5(4):405–441, 2016.

[61] M. Ohliger, V. Nesme, and J. Eisert. Efficient and feasible state tomography of quantum many-body systems. *New J. Phys.*, 15:015024, 2013.

[62] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland. Randomized benchmarking of quantum gates. *Phys. Rev. A*, 77:012307, 2008.

[63] Linxi Zhang, Chuanghua Zhu, and Changxing Pei. Randomized benchmarking using unitary t-design for fidelity estimation of practical quantum circuit, 2017. arXiv:1711.08098.

[64] N. Lashkari, D. Stanford, M. Hastings, T. J. Osborne, and P. Hayden. Towards the fast scrambling conjecture. *JHEP*, 2013:22, 2013.

[65] Winton Brown and Omar Fawzi. Scrambling speed of random quantum circuits, 2012. arXiv:1210.6644.

[66] Richard Low. *Pseudo-randomness and learning in quantum computation*. PhD thesis, university of Bristol, 2010.

[67] S. Bravyi and D. Gosset. Improved classical simulation of quantum circuits dominated by Clifford gates. *Phys Rev. Lett.*, 116:250501, 2016.

[68] Peter Selinger. Generators and relations for n-qubit Clifford operators. *Logical Methods in Computer Science*, 11, 2015.

[69] M. Ozols. Clifford group, 2008. Essays at University of Waterloo.

[70] Daniel Gottesman. *Stabilizer Codes and Quantum Error Correction*. PhD thesis, Caltech, 1997.

[71] Daniel Gottesman. The Heisenberg representation of quantum computers. In *Proceedings of the XXII International Colloquium on Group Theoretical Methods in Physics*, pages 32–43, 1998.

[72] Zak Webb. The Clifford group forms a unitary 3-design. *Quantum Information and Computation*, 16, 2016.

[73] Huangjun Zhu. Multiqubit Clifford groups are unitary 3-designs, 2015. arXiv:1510.02619.

[74] C. Dankert, R. Cleve, J. Emerson, and E. Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A*, 80, 2009.

[75] J. Emerson, R. Alicki, and K. Zyczkowski. Scalable noise estimation with random unitary operators. *J. of Optics B*, 7, 2005.

[76] Claude E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28(4):656–715, 1949.

[77] James L. McInnes and Benny Pinkas. On the impossibility of private key cryptography with weakly random keys. In *Proceedings of the 10th Annual International Cryptology Conference on Advances in Cryptology*, CRYPTO '90, pages 421–435, London, UK, 1991. Springer-Verlag. ISBN 3-540-54508-5.

[78] Miklos Santha and Umesh V. Vazirani. Generating quasi-random sequences from semi-random sources. *Journal of Computer and System Sciences*, 33(1):75–87, 1986.

[79] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.*, 17(2):230–261, 1988.

[80] A. Ambainis, M. Mosca, A. Tapp, and R. De Wolf. Private quantum channels. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 547–553, 2000.

[81] M. Horodecki, J. Oppenheim, and A. Winter. Quantum state merging and negative information. *Commun. Math. Phys.*, 269:107, 2007.

[82] Michal Horodecki, Jonathan Oppenheim, and Andreas Winter. Partial quantum information. *Nature*, 436:673–676, 2005.

[83] Berry Groisman, Sandu Popescu, and Andreas Winter. On the quantum, classical and total amount of correlations in a quantum state. *Physical Review A*, 72:032317, 2005.

[84] Patrick Hayden, Michal Horodecki, Jon Yard, and Andreas Winter. A decoupling approach to the quantum capacity. *Open Systems and Information Dynamics*, 15:7–19, 2008.

[85] Noah Linden, Sandu Popescu, Anthony J. Short, and Andreas Winter. Quantum mechanical evolution towards thermal equilibrium. *Physical Review E*, 79:061103, 2009.

[86] Philippe Faist. Thermodynamic state preparation. Master thesis, ETH Zürich, 2011.

[87] Lidia del Rio, Johan Aberg, Renato Renner, Oscar Dahlsten, and Vlatko Vedral. The thermodynamic meaning of negative entropy. *Nature*, 474:61–63, 2011.

[88] Renato Renner. *Security of Quantum Key Distribution.* PhD thesis, ETH Zürich, 2005.

[89] Emilio Onorati. Linking decoupling theorem and hypothesis-testing entropy. Master thesis, ETH Zürich, 2013.

[90] Dorit Aharonov, Alexei Kitaev, and Noam Nisan. Quantum circuits with mixed states. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, pages 20–30. ACM Press, 1998.

[91] Dorit Aharonov. Quantum computing. *Annual Reviews of Computational Physics*, 6: 259–346, 1999.

[92] Barbara M. Terhal and David P. Di Vincenzo. Adaptive quantum computation, constant depth quantum circuits and Arthur-Merlin games. *Quant. Inf. Comp.*, 4:134–145, 2004.

[93] R. Oliveira, O. C. O. Dahlsten, and M. B. Plenio. Efficient generation of generic entanglement. *Phys. Rev. Lett.*, 98:130502, 2007.

[94] Persi Diaconis and Laurent Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. App. Prob.*, 6(3):695–750, 1996.

[95] I. T. Diniz and D. Jonathan. Comment on the paper "Random quantum circuits are approximate 2-designs". *Comm. Math. Phys.*, 304, 2011.

[96] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times.* American Mathematical Society, 2008.

[97] Donald Knuth. *The art of computer programming*, volume 2. Addison-Wesley, third edition, 1997.

[98] Jean Bourgain and Alex Gamburd. A spectral gap theorem in SU(d). *Journal of the European Mathematical Society*, 14(5):1455–1511, 2012.

[99] Bruno Nachtergaele. The spectral gap for some spin chains with discrete symmetry breaking. *Comm. Math. Phys.*, 175:565–606, 1996.

[100] Roberto I. Oliveira. On the convergence to equilibrium of Kac's random walk on matrices. *The Annals of Applied Probability*, 19:1200–1231, 2009.

[101] Russ Bubley and Martin Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. In *Proceedings 38th annual Symposium on Foundations of Computer Scicence*, pages 223–231, 1997.

[102] I. L. Chuang and M. A. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *J. Mod. Opt.*, 44:2455–2467, 1997.

[103] M. Mohseni, A. T. Rezakhani, and D. A. Lidar. Quantum-process tomography: Resource analysis of different strategies. *Phys. Rev. A*, 77:032322, 2008.

[104] D. W. T. Rippin. Parameter estimation in engineering and science, by James V. Beck and Kenneth J. Arnold, published by John Wiley and Sons, 1977. *AIChE Journal*, 24(2): 367–367, 1978.

[105] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. Lett.*, 85:042311, 2012.

[106] C. Arnaud Dugas, Joel Wallman, and Joseph Emerson. Characterizing universal gate sets via dihedral benchmarking. *Phys. Rev. A*, 92:060302, 2015.

[107] Joel J. Wallman. Randomized benchmarking with gate-dependent noise. *Quantum*, 2:47, 2018.

[108] Easwar Magesan, Jay M. Gambetta, B. R. Johnson, Colm A. Ryan, Jerry M. Chow, Seth T. Merkel, Marcus P. da Silva, George A. Keefe, Mary B. Rothwell, Thomas A. Ohki, Mark B. Ketchen, and M. Steffen. Efficient measurement of quantum gate error by interleaved randomized benchmarking. *Phys. Rev. Lett.*, 109:080505, 2012.

[109] Robin Harper and Steven T Flammia. Estimating the fidelity of T gates using standard interleaved randomized benchmarking. *Quantum Science and Technology*, 2(1):015008, 2017.

[110] C. Granade, C. Ferrie, and D. G. Cory. Accelerated randomized benchmarking. *New J. Phys.*, 17, 2015.

[111] K. R. Brown, A. C. Wilson, Y. Colombe, C. Ospelkaus, A. M. Meier, E. Knill, D. Leibfried, and D. J. Wineland. Single-qubit-gate error below 10^{-4} in a trapped ion. *Phys. Rev. A*, 84, 2011.

[112] J. Helsen, J. Wallman, S. T. Flammia, and S. Wehner. Multi-qubit randomized benchmarking using few samples, 2017. arXiv:1701.04299.

[113] Jonas Helsen, Joel J. Wallman, and Stephanie Wehner. Representations of the multi-qubit Clifford group. *Journal of Mathematical Physics*, 59(7):072201, 2018.

[114] Andrew W. Cross, Easwar Magesan, Lev S. Bishop, John A. Smolin, and Jay M. Gambetta. Scalable randomized benchmarking of non-Clifford gates. *npj Quantum Information*, 2, 2016.

[115] Shelby Kimmel, Marcus P. da Silva, Colm A. Ryan, Blake R. Johnson, and Thomas Ohki. Robust extraction of tomographic information via randomized benchmarking. *Phys. Rev. X*, 4:011050, 2014.

[116] K. Glashoff and M. M. Bronstein. Almost-commuting matrices are almost jointly diagonalizable, 2013. arXiv:1305.2135.

[117] B. E. Anderson, H. Sosa-Martinez, C. A. Riofrio, I. H. Deutsch, and P. S. Jessen. Accurate and robust unitary transformations of a high-dimensional quantum system. *Phys. Rev. Lett.*, 114:240401, 2015.

[118] Fernando Casas, Ander Murua, and Madlen Nadinic. Efficient computation of the Zassenhaus formula. *Comp. Phys. Comm.*, 183:2386–2391, 2012.

[119] Jean-Pierre Serre. *Linear Representations of Finite Groups*. Springer, 1996.

[120] Rolf Berndt. *Representations of Linear Groups*. Vieweg, first edition, 2007.

[121] A. C. R. Belton, M. Gnacik, and J. M. Lindsay. The convergence of unitary quantum random walks, 2014.

[122] C. Gogolin and J. Eisert. Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems. *Rep. Prog. Phys.*, 79:56001, 2016.

[123] T. Banks, W. Fischler, S. Shenker, and L. Susskind. M theory as a matrix model: A conjecture. *Phys. Rev. D*, 55:5112, 1997.

[124] Juan Maldacena. The large n limit of super-conformal field theories and supergravity. *Adv. Th. Math. Phys.*, 2:213, 1998.

[125] Cyril Bailey. Translation of de rerum natura. Oxford at the Clarendon Press. `http://oll.libertyfund.org/titles/carus-on-the-nature-of-things`, accessed: 09.08.2017.

[126] Achim Klenke. *Probability Theory: A comprehensive course*. Springer, second edition, 2014.

[127] Kiyoshi Itō. On a stochastic integral equation. *Proc. Japan Acad.*, 22(2):32–35, 1946.

[128] Kiyoshi Itō. Stochastic integral. *Proc. Imp. Acad.*, 20(8):519–524, 1944.

[129] Monroe D. Donsker. An invariance principle for certain probability limit theorems. Memoirs of the American Mathematical Society, 1951.

[130] Henry P. McKean. *Stochastic integrals*. Academic Press, 1969.

[131] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales*, volume 2. Cambridge Mathematical Library, second edition, 2000.

[132] Yves Benoist and Nicolas de Saxcé. A spectral gap theorem in simple Lie groups. *Inventiones mathematicae*, 205(2):337–361, 2016.

[133] Wolfgang Ziller. Lie groups. representation theory and symmetric spaces. Lecture notes, 2010.

[134] D. H. Sattinger and O. L. Weaver. *Lie groups and algebras with applications to physics, geometry and mechanics*. Springer-Verlag Berlin Heidelberg, 1986.

[135] Howard Georgi. *Lie algebras in particle physics*. Westview Press, second edition, 1999.

[136] E. W. Montroll and G. H. Weiss. Random walks on lattices II. *J. Math. Phys.*, 6:167–181, 1965.

[137] George H. Weiss. Aspects and applications of the random walk. *J. Stat. Phys.*, 79(1): 497–500, 1995.

[138] V. Zaburdaev, S. Denisov, and P. Hanggi. Perturbation spreading in many-particle systems: A random walk approach. *Phys. Rev. Lett.*, 106, 2011.

[139] J. H. P. Schulz and E. Barkai. Fluctuations around equilibrium laws in ergodic continuous-time random walks. *Phys. Rev. E*, 91, 2015.

[140] P. Chaudhuri, Y. Gao, L. Berthier, M. Kilfoil, and W. Kob. A random walk description of the heterogeneous glassy dynamics of attracting colloids. *J. Phys. Cond. Mat.*, 20, 2008.

[141] Patrick Hayden and John Preskill. Black holes as mirrors: quantum information in random subsystems. *Journal of High Energy Physics*, 2007(09):120–120, 2007.

[142] Seth Lloyd and John Preskill. Unitarity of black hole evaporation in final-state projection models. *Journal of High Energy Physics*, 2014(8):126, 2014.

[143] S. Diehl, A. Micheli, A. Kantian, B. Kraus, H. P. Buechler, and P. Zoller. Quantum states and phases in driven open quantum systems with cold atoms. *Nature Phys.*, 4:878, 2008.

[144] F. Verstraete, M. M. Wolf, and J. I. Cirac. Quantum computation and quantum-state engineering driven by dissipation. *Nature Phys.*, 5(9):633, 2009.

[145] M. J. Kastoryano, M. M. Wolf, and J. Eisert. Precisely timing dissipative quantum information processing. *Phys. Rev. Lett.*, 110:110501, 2013.

Appendix

Mathematica notebook for randomized benchmarking on 3-fold tensor copies of T-gate

Construct representation of S3

In[1]:=
```
F12 = ConstantArray[0, {64, 64}];
Do[
 Do[
  Do[
   j = 16 * c + 4 * b + a + 1; k = 16 * c + 4 * a + b + 1;
   F12[[j, k]] = 1,
   {a, 0, 3}], {b, 0, 3}], {c, 0, 3}]
```

In[3]:=
```
F23 = ConstantArray[0, {64, 64}];
Do[
 Do[
  Do[
   j = 16 * c + 4 * b + a + 1; k = 16 * b + 4 * c + a + 1;
   F23[[j, k]] = 1,
   {a, 0, 3}], {b, 0, 3}], {c, 0, 3}]
```

In[5]:=
```
F13 = ConstantArray[0, {64, 64}];
Do[
 Do[
  Do[
   j = 16 * c + 4 * b + a + 1; k = 16 * a + 4 * b + c + 1;
   F13[[j, k]] = 1,
   {a, 0, 3}], {b, 0, 3}], {c, 0, 3}]
```

In[7]:=
```
F123 = F13.F12;
F132 = F12.F13;
```

In[9]:=
```
S3Group = {IdentityMatrix[64], F12, F13, F23, F123, F132};
```

Character table of S3

In[10]:=
```
trivialrep = {1, 1, 1, 1, 1, 1};
signrep = {1, -1, -1, -1, 1, 1};
standardrep = {2, 0, 0, 0, -1, -1};
S3irreps = {trivialrep, signrep, standardrep};
```

Construct symmetry group of 3 tensor copies of T-gate and its irreps

In[14]:=
$$\text{LPTgate} = \left\{\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}, \begin{pmatrix}1&0&0&0\\0&0&-1&0\\0&1&0&0\\0&0&0&1\end{pmatrix}, \begin{pmatrix}1&0&0&0\\0&0&1&0\\0&-1&0&0\\0&0&0&1\end{pmatrix}, \begin{pmatrix}1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&1\end{pmatrix}\right\};$$

In[15]:=
```
Cyc4Trivial = {1, 1, 1, 1};
Cyc1 = {1, i, -i, -1};
Cyc2 = {1, -1, -1, 1};
Cyc3 = {1, -i, i, -1};
Cyc4Characters = {Cyc4Trivial, Cyc1, Cyc2, Cyc3};
```

In[20]:=
```
DirectProductGroup = Flatten[Table[
    KroneckerProduct[LPTgate[[p]], KroneckerProduct[LPTgate[[j]], LPTgate[[k]]]],
    {p, 1, 4}, {j, 1, 4}, {k, 1, 4}], 2];
```

In[21]:=
```
FullGroup =
  Flatten[Table[DirectProductGroup[[j]].S3Group[[k]], {k, 1, 6}, {j, 1, 64}], 1];
```

In[22]:=
```
TraceFullGroup = Map[Tr, FullGroup];
```

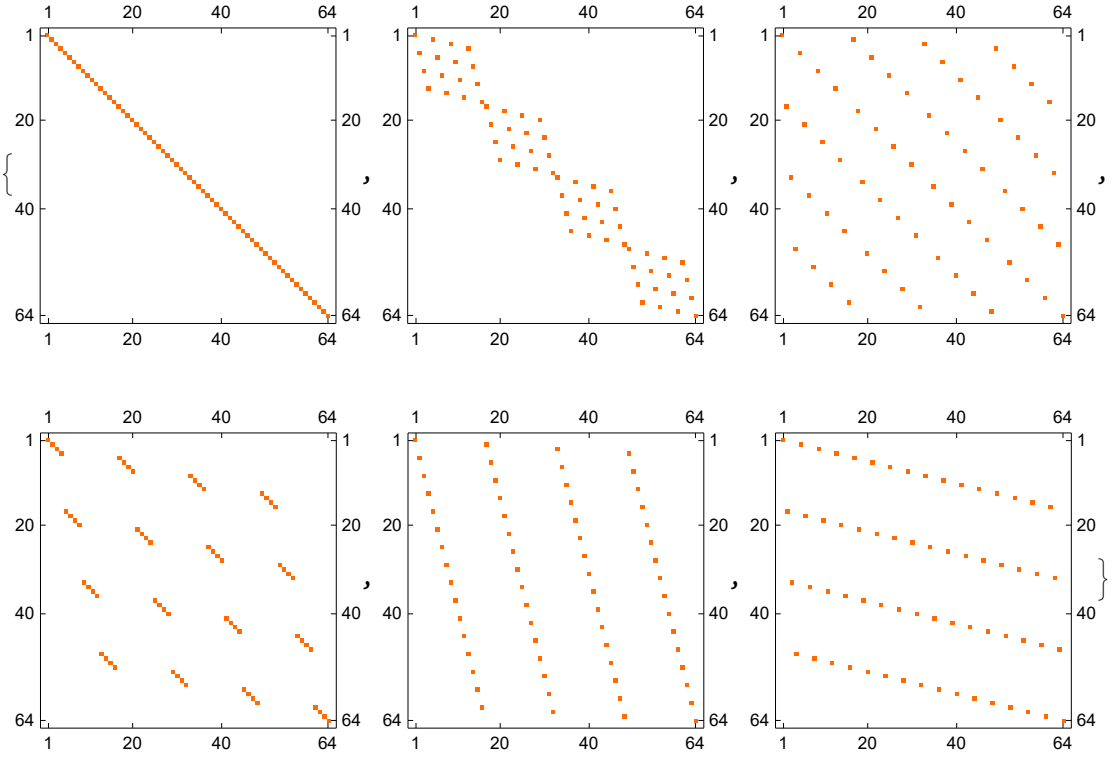
In[23]:=
```
DirectProductCharacters =
  Table[Flatten@Table[Cyc4Characters[[a, p]] * Cyc4Characters[[b, q]] *
      Cyc4Characters[[c, r]], {p, 1, 4}, {q, 1, 4}, {r, 1, 4}],
   {a, 1, 4}, {b, 1, 4}, {c, 1, 4}];
(*the labels a, b, c represent which irrep is chosen from the k4 Group,
the labels p,q,r the group element*)
```

Action of S3 on characters

In[24]:=
```
OneOrbits = Table[If[Inverse[S3Group[[ll]]].DirectProductGroup[[j]].S3Group[[ll]] ==
     DirectProductGroup[[k]], 1, 0], {ll, 1, 6}, {j, 1, 64}, {k, 1, 64}];
```

In[25]:=
```
MatrixPlot /@ OneOrbits
```

Out[25]=



In[26]:=
```
CharacterOrbits =
  Table[Table[OneOrbits[[j]].DirectProductCharacters[[a, b, c]], {j, 1, 6}],
   {a, 1, 4}, {b, 1, 4}, {c, 1, 4}];
(*first three entries label the representation,
last entry label the permutation group element*)
```

In[27]:=
```
CharacterOrbitsFlat = Flatten[CharacterOrbits, 2];
DirectProductCharactersFlat = Flatten[DirectProductCharacters, 2];
```

Sort orbits and select representative characters

In[29]:=
```
OrbitsLists = Flatten /@
  Map[Position[DirectProductCharactersFlat, #][[1]] &, CharacterOrbitsFlat, {2}]
```

Out[29]=
```
{{1, 1, 1, 1, 1, 1}, {2, 5, 17, 2, 5, 17}, {3, 9, 33, 3, 9, 33}, {4, 13, 49, 4, 13, 49},
 {5, 2, 5, 17, 17, 2}, {6, 6, 21, 18, 21, 18}, {7, 10, 37, 19, 25, 34},
 {8, 14, 53, 20, 29, 50}, {9, 3, 9, 33, 33, 3}, {10, 7, 25, 34, 37, 19},
 {11, 11, 41, 35, 41, 35}, {12, 15, 57, 36, 45, 51}, {13, 4, 13, 49, 49, 4},
 {14, 8, 29, 50, 53, 20}, {15, 12, 45, 51, 57, 36}, {16, 16, 61, 52, 61, 52},
 {17, 17, 2, 5, 2, 5}, {18, 21, 18, 6, 6, 21}, {19, 25, 34, 7, 10, 37},
 {20, 29, 50, 8, 14, 53}, {21, 18, 6, 21, 18, 6}, {22, 22, 22, 22, 22, 22},
 {23, 26, 38, 23, 26, 38}, {24, 30, 54, 24, 30, 54}, {25, 19, 10, 37, 34, 7},
 {26, 23, 26, 38, 38, 23}, {27, 27, 42, 39, 42, 39}, {28, 31, 58, 40, 46, 55},
 {29, 20, 14, 53, 50, 8}, {30, 24, 30, 54, 54, 24}, {31, 28, 46, 55, 58, 40},
 {32, 32, 62, 56, 62, 56}, {33, 33, 3, 9, 3, 9}, {34, 37, 19, 10, 7, 25},
 {35, 41, 35, 11, 11, 41}, {36, 45, 51, 12, 15, 57}, {37, 34, 7, 25, 19, 10},
 {38, 38, 23, 26, 23, 26}, {39, 42, 39, 27, 27, 42}, {40, 46, 55, 28, 31, 58},
 {41, 35, 11, 41, 35, 11}, {42, 39, 27, 42, 39, 27}, {43, 43, 43, 43, 43, 43},
 {44, 47, 59, 44, 47, 59}, {45, 36, 15, 57, 51, 12}, {46, 40, 31, 58, 55, 28},
 {47, 44, 47, 59, 59, 44}, {48, 48, 63, 60, 63, 60}, {49, 49, 4, 13, 4, 13},
 {50, 53, 20, 14, 8, 29}, {51, 57, 36, 15, 12, 45}, {52, 61, 52, 16, 16, 61},
 {53, 50, 8, 29, 20, 14}, {54, 54, 24, 30, 24, 30}, {55, 58, 40, 31, 28, 46},
 {56, 62, 56, 32, 32, 62}, {57, 51, 12, 45, 36, 15}, {58, 55, 28, 46, 40, 31},
 {59, 59, 44, 47, 44, 47}, {60, 63, 60, 48, 48, 63}, {61, 52, 16, 61, 52, 16},
 {62, 56, 32, 62, 56, 32}, {63, 60, 48, 63, 60, 48}, {64, 64, 64, 64, 64, 64}}
```

In[30]:=
```
uniqueOrbits = Union[Sort /@ OrbitsLists]
```

Out[30]=
```
{{1, 1, 1, 1, 1, 1}, {2, 2, 5, 5, 17, 17},
 {3, 3, 9, 9, 33, 33}, {4, 4, 13, 13, 49, 49}, {6, 6, 18, 18, 21, 21},
 {7, 10, 19, 25, 34, 37}, {8, 14, 20, 29, 50, 53}, {11, 11, 35, 35, 41, 41},
 {12, 15, 36, 45, 51, 57}, {16, 16, 52, 52, 61, 61}, {22, 22, 22, 22, 22, 22},
 {23, 23, 26, 26, 38, 38}, {24, 24, 30, 30, 54, 54}, {27, 27, 39, 39, 42, 42},
 {28, 31, 40, 46, 55, 58}, {32, 32, 56, 56, 62, 62}, {43, 43, 43, 43, 43, 43},
 {44, 44, 47, 47, 59, 59}, {48, 48, 60, 60, 63, 63}, {64, 64, 64, 64, 64, 64}}
```

In[31]:=
```
groupedOrbits = Sort[Sort /@ Union /@ uniqueOrbits, Length[#1] ≤ Length[#2] &]
```

Out[31]=
```
{{1}, {22}, {43}, {64}, {2, 5, 17}, {3, 9, 33}, {4, 13, 49}, {6, 18, 21},
 {11, 35, 41}, {16, 52, 61}, {23, 26, 38}, {24, 30, 54}, {27, 39, 42},
 {32, 56, 62}, {44, 47, 59}, {48, 60, 63}, {7, 10, 19, 25, 34, 37},
 {8, 14, 20, 29, 50, 53}, {12, 15, 36, 45, 51, 57}, {28, 31, 40, 46, 55, 58}}
```

In[32]:=
```
OrbitRepresentative = Min /@ groupedOrbits
```

Out[32]=
```
{1, 22, 43, 64, 2, 3, 4, 6, 11, 16, 23, 24, 27, 32, 44, 48, 7, 8, 12, 28}
```

In[33]:=
```
OrbitRepresentativeChar = DirectProductCharactersFlat[[#]] & /@ OrbitRepresentative;
```

Construct irreps with stabilizer S3

In[34]:=
```
(*trivial rep on S3*)
charact1 = Flatten[#, 1] &@Table[#, {xx, 6}] & /@ OrbitRepresentativeChar[[1 ;; 4]];
```

In[35]:=
```
(*sign rep on S3*)
charact2 = Flatten[#, 1] &@{#, -#, -#, -#, #, #} & /@ OrbitRepresentativeChar[[1 ;; 4]];
```

In[36]:=
```
(*standard rep on S3*)
charact3 =
  Flatten[#, 1] &@{2 #, 0 #, 0 #, 0 #, -#, -#} & /@ OrbitRepresentativeChar[[1 ;; 4]];
```

Construct irreps with stabilizer {Identity,F12}

In[37]:=
```
threeorbitssubgroup =
  Flatten[Table[DirectProductGroup[[j]].S3Group[[k]], {k, 1, 2}, {j, 1, 64}], 1];
```

In[38]:=
```
threeorbitssubgroupirreps = Table[
   {Flatten[{DirectProductCharacters[[a, b, b]], DirectProductCharacters[[a, b, b]]},
     1], Flatten[{DirectProductCharacters[[a, b, b]],
      -1 * DirectProductCharacters[[a, b, b]]}, 1]}, {a, 1, 4}, {b, 1, 4}];
```

Construct matrices representing the conjugate action t^(-1) s t

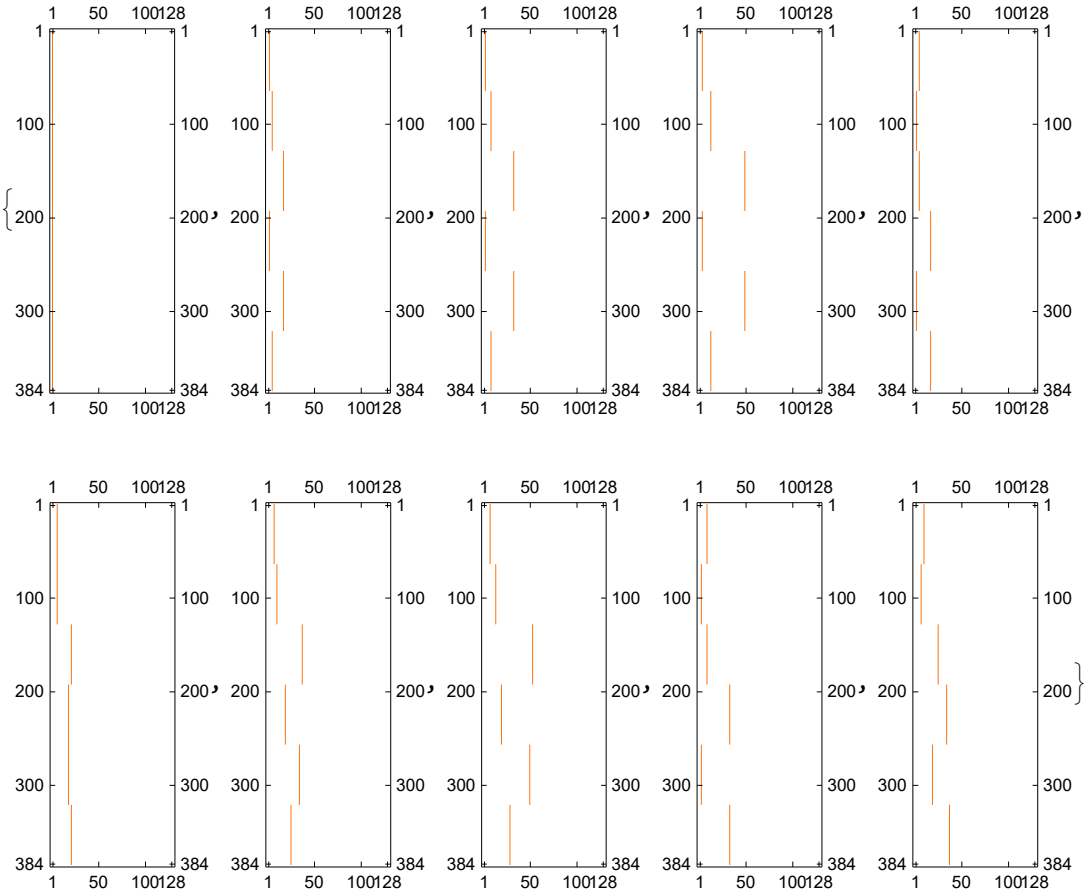
In[39]:=
```
conjugatematrices = Table[
   Inverse[FullGroup[[t]]].FullGroup[[s]].FullGroup[[t]], {s, 1, 384}, {t, 1, 384}];
```

Construct induced characters from irreps having orbits with three elements each

In[40]:=
```
inducedcharactermatrices =
  Table[If[conjugatematrices[[s, t]] == threeorbitssubgroup[[r]], 1, 0],
   {s, 1, 384}, {t, 1, 384}, {r, 1, 128}];
```

In[41]:=
```
MatrixPlot /@ inducedcharactermatrices[[1 ;; 10]]
```

Out[41]=



In[42]:=

In[43]:=
```
inducedcharactervectors =
  Table[Map[#.threeorbitssubgroupirreps[[a, b, t]] &, inducedcharactermatrices],
   {a, 1, 4}, {b, 1, 4}, {t, 1, 2}];
```

In[44]:=
```
threeorbitsinducedcharacters =
  Table[1 / 128 * Map[Total, inducedcharactervectors[[a, b, t]]],
   {a, 1, 4}, {b, 1, 4}, {t, 1, 2}];
```

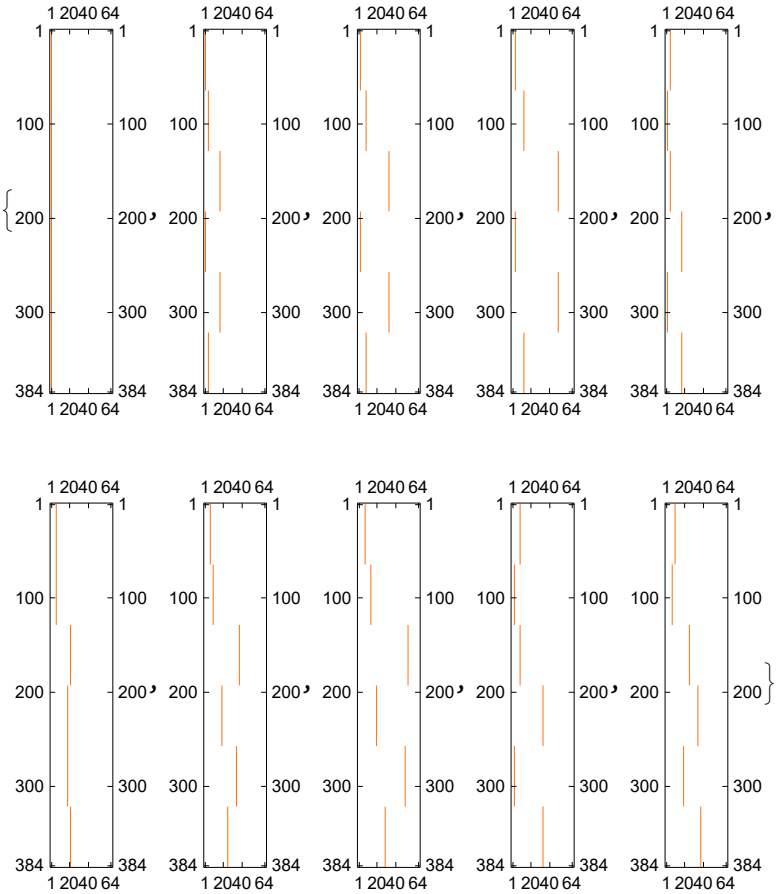
Construct induced characters from irreps having orbits with six elements each

Note that the subgroup given by the trivial stabilizer is the direct product group itself

In[45]:=
```
inducedcharactermatrices2 =
  Table[If[conjugatematrices[[s, t]] == DirectProductGroup[[r]], 1, 0],
   {s, 1, 384}, {t, 1, 384}, {r, 1, 64}];
```

In[46]:=
```
MatrixPlot /@ inducedcharactermatrices2[[1 ;; 10]]
```

Out[46]=



In[47]:=
```
inducedcharactervectors2 =
  Table[Map[#.DirectProductCharacters[[a, b, c]] &, inducedcharactermatrices2],
   {a, 1, 4}, {b, 1, 4}, {c, 1, 4}];
```

In[48]:=
```
sixorbitsinducedcharacters =
  Table[1 / 64 * Map[Total, inducedcharactervectors2[[a, b, c]]],
   {a, 1, 4}, {b, 1, 4}, {c, 1, 4}];
```

Sort irreps

There are in total 40 irreps, 12 from the one orbit characters with stabilizer S3, 12x2 from the three-orbit characters with stabilizer S2, 4 from the six-orbit characters with trivial stabilizer

In[49]:=
```
FullGroupirreps = ConstantArray[0, {40, 384}];
Do[FullGroupirreps[[a]] = charact1[[a]], {a, 1, 4}];
Do[FullGroupirreps[[a + 4]] = charact2[[a]], {a, 1, 4}];
Do[FullGroupirreps[[a + 8]] = charact3[[a]], {a, 1, 4}];
Do[FullGroupirreps[[12 + a - 1]] = threeorbitsinducedcharacters[[a, 1, 1]], {a, 2, 4}];
```

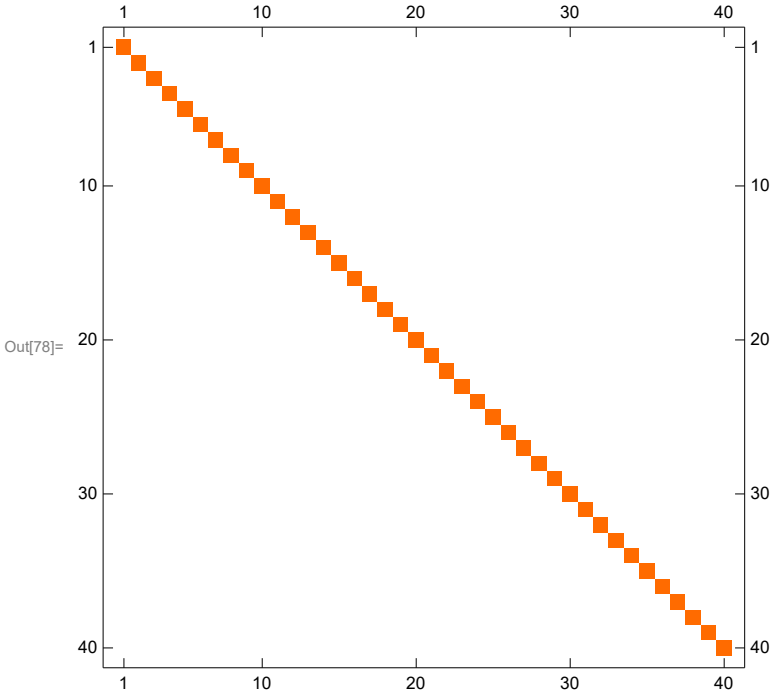
In[54]:=
```
FullGroupirreps[[16]] = threeorbitsinducedcharacters[[1, 2, 1]];
FullGroupirreps[[17]] = threeorbitsinducedcharacters[[3, 2, 1]];
FullGroupirreps[[18]] = threeorbitsinducedcharacters[[4, 2, 1]];
FullGroupirreps[[19]] = threeorbitsinducedcharacters[[1, 3, 1]];
FullGroupirreps[[20]] = threeorbitsinducedcharacters[[2, 3, 1]];
FullGroupirreps[[21]] = threeorbitsinducedcharacters[[4, 3, 1]];
FullGroupirreps[[22]] = threeorbitsinducedcharacters[[1, 4, 1]];
FullGroupirreps[[23]] = threeorbitsinducedcharacters[[2, 4, 1]];
FullGroupirreps[[24]] = threeorbitsinducedcharacters[[3, 4, 1]];
Do[FullGroupirreps[[24 + a - 1]] = threeorbitsinducedcharacters[[a, 1, 2]], {a, 2, 4}];
FullGroupirreps[[28]] = threeorbitsinducedcharacters[[1, 2, 2]];
FullGroupirreps[[29]] = threeorbitsinducedcharacters[[3, 2, 2]];
FullGroupirreps[[30]] = threeorbitsinducedcharacters[[4, 2, 2]];
FullGroupirreps[[31]] = threeorbitsinducedcharacters[[1, 3, 2]];
FullGroupirreps[[32]] = threeorbitsinducedcharacters[[2, 3, 2]];
FullGroupirreps[[33]] = threeorbitsinducedcharacters[[4, 3, 2]];
FullGroupirreps[[34]] = threeorbitsinducedcharacters[[1, 4, 2]];
FullGroupirreps[[35]] = threeorbitsinducedcharacters[[2, 4, 2]];
FullGroupirreps[[36]] = threeorbitsinducedcharacters[[3, 4, 2]];
FullGroupirreps[[37]] = sixorbitsinducedcharacters[[1, 2, 3]];
FullGroupirreps[[38]] = sixorbitsinducedcharacters[[1, 2, 4]];
FullGroupirreps[[39]] = sixorbitsinducedcharacters[[1, 3, 4]];
FullGroupirreps[[40]] = sixorbitsinducedcharacters[[2, 3, 4]];
```

Checking whether the irreps are correct...

In[77]:=
```
CheckOrthogonality =
  Table[If[1 / 384 * Sum[Conjugate[FullGroupirreps[[a, h]]] * FullGroupirreps[[b, h]],
      {h, 1, 384}] == 1, 1, 0], {a, 1, 40}, {b, 1, 40}];
```

In[78]:=
```
MatrixPlot[CheckOrthogonality]
```

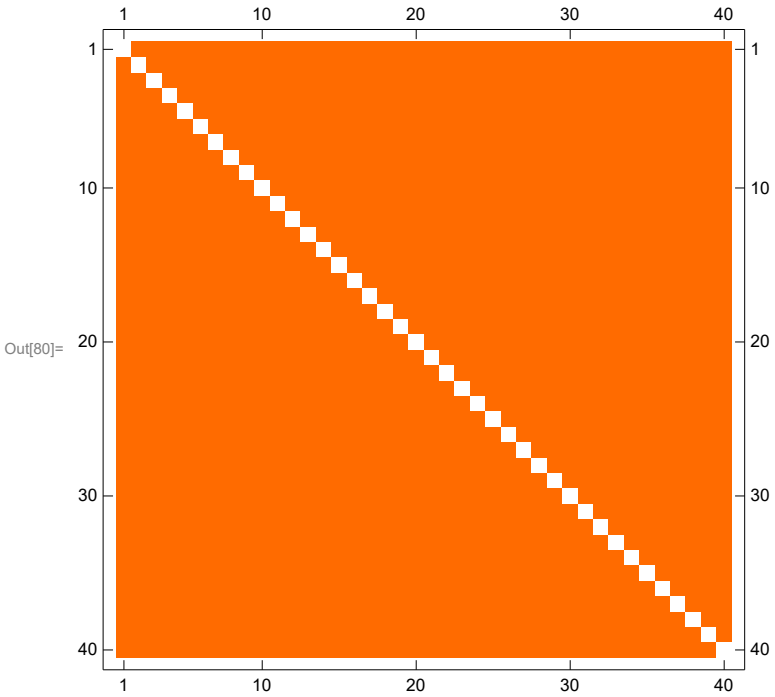
Out[78]=



In[79]:=
```
CheckOrthogonalityConverse =
  Table[If[1 / 384 * Sum[Conjugate[FullGroupirreps[[a, h]]] * FullGroupirreps[[b, h]],
      {h, 1, 384}] == 0, 1, 0], {a, 1, 40}, {b, 1, 40}];
```

In[80]:=
```
MatrixPlot[CheckOrthogonalityConverse]
```

Out[80]=



In[81]:=
```
Sum[FullGroupirreps[[a, 1]]^2, {a, 1, 40}] == 384 (*Dimension Formula*)
```

Out[81]=
```
True
```

All orthogonality and dimension formulae are correct!

Irreducible decomposition

In[82]:=
```
Do[
 Print[
  1 / 384 * Sum[Conjugate[FullGroupirreps[[m, j]]] * TraceFullGroup[[j]], {j, 1, 384}],
  " multiplicity of irrep ", m], {m, 1, 40}]
```

```
4 multiplicity of irrep 1
1 multiplicity of irrep 2
0 multiplicity of irrep 3
1 multiplicity of irrep 4
0 multiplicity of irrep 5
0 multiplicity of irrep 6
0 multiplicity of irrep 7
0 multiplicity of irrep 8
2 multiplicity of irrep 9
0 multiplicity of irrep 10
0 multiplicity of irrep 11
0 multiplicity of irrep 12
3 multiplicity of irrep 13
0 multiplicity of irrep 14
3 multiplicity of irrep 15
2 multiplicity of irrep 16
0 multiplicity of irrep 17
1 multiplicity of irrep 18
0 multiplicity of irrep 19
0 multiplicity of irrep 20
0 multiplicity of irrep 21
2 multiplicity of irrep 22
1 multiplicity of irrep 23
0 multiplicity of irrep 24
1 multiplicity of irrep 25
0 multiplicity of irrep 26
1 multiplicity of irrep 27
0 multiplicity of irrep 28
0 multiplicity of irrep 29
0 multiplicity of irrep 30
0 multiplicity of irrep 31
0 multiplicity of irrep 32
0 multiplicity of irrep 33
0 multiplicity of irrep 34
```

```
0 multiplicity of irrep 35
0 multiplicity of irrep 36
0 multiplicity of irrep 37
2 multiplicity of irrep 38
0 multiplicity of irrep 39
0 multiplicity of irrep 40
```

Zusammenfassung

Nach den Axiomen der Quantenmechanik wird die Zeitentwicklung quanten-mechanischer Systeme durch unitäre Operatoren beschrieben. Aus diesem Grund ist das Studium und Verständnis der entsprechenden Gruppe von großem Interesse. Ein Teilbereich der Forschung an dieser Gruppe mit besonders breitem Anwendungsgebiet in der Quanteninformationstheorie sind stochastische Zufallsprozesse über die unitäre Gruppe. Hier spielen sogenannte *unitäre Designs* eine zentrale Rolle. Dies sind besondere Dichteverteilungen, die die Momente des uniformen Haar-Maßes reproduzieren, und somit eng verknüpft sind mit Konzepten wie Quanten-Tomographie, Äquilibrierung, Thermalisierung, Verschlüsselung und Scrambling. Die bisherige Forschung konnte zeigen, dass unitäre Designs effizient angenähert werden können durch randomisierte Quanten-Schaltkreise mit lokalen unitären Gattern. Mithilfe von diesen Schaltkreisen lassen sich sowohl die Präzision experimenteller Implementierungen unitärer Gatter anhand von randomisierten Benchmarking-Protokollen auswerten, als auch Systeme die äußerst schnell von ihrer Umfeld entkoppeln.

Der hauptsächliche Beitrag dieser Arbeit ist die Entwicklung eines Formalismus, mit dem die oben beschriebenen Ergebnisse auf kontinuierliche Zeitentwicklungen angewendet werden können, das heisst, eine Untersuchung der Brown'sche Bewegung über die unitäre Gruppe. Um dies zu erreichen, entwickeln wir zum einen eine darstellungstheoretische Formulierung unitärer Designs und nutzen die daraus resultierenden Werkzeuge um einen Ausdruck für den Abstand zwischen lokalen Generatoren zu errechnen, die zu den verschiedenen Momenten einer Verteilung bei einem Diffusionsprozess in Zusammenhang stehen. Hieraus lässt sich weiter berechnen, in welchen Zeitskalen die Konvergenz dieser Momente zu denen des Haar-Maßes garantiert werden kann. Zum anderen beschreiben wir die stochastische Zeitentwicklung als eine Projektion auf einen Random Walk über Pauli-Matrizen und verknüpfen diese Formulierung des Problems mit randomisierten Quanten-Schaltkreisen. Hierdurch lässt sich eine Entkopplung von System und Umfeld erreichen, die nahezu linear in der Systemgröße skaliert.

Zusätzlich zu der Entwicklung eines einheitlichen Formalismus für Zufallsprozess über die unitäre Gruppe präsentiert diese Arbeit also neue mathematische Ergebnisse und Techniken im Bereich der Quanteninformationtheorie. Nicht zuletzt werden auch Anwendungen dieser Ergebnisse im Bereich der Dynamik schwarzer Löcher diskutiert, nämlich mit Hinblick auf das Informationsparadoxon.

Ein weiteres, neues Ergebnis dieser Arbeit, unabhängig von Brown'scher Bewegung, ist ein randomisiertes Benchmarking-Protokoll, welches existierende Protokolle zum Schätzen der Genauigkeit von Gattern verbessert, indem es Symmetrien in diesen Gattern ausnutzt. Auch dieser Fortschritt basiert auf Darstellungstheorie und nutzt diese, um den rechnerischen Aufwand und die Menge verwendeter Quanten-Ressourcen zu reduzieren.

Abstract

According to one of the fundamental axioms of quantum mechanics, unitary operators rule the evolution of any quantum system; it is thus of prominent importance to investigate the corresponding group and discover its properties. Random processes over the unitary group have indeed a wide range of applications in the context of quantum information; in particular, they are involved in the construction of peculiar distributions – called *unitary designs* – which mimic the uniform Haar measure by matching its moments and are thus deeply connected to the description of phenomena such as quantum tomography, equilibration, thermalization, encryption and scrambling. Previous approaches have shown that unitary designs are efficiently approximated by random quantum circuits with local unitary operators. Moreover, these circuits are used to characterize the precision of experimental implementations of unitary gates via *randomized benchmarking* protocols and to rapidly decouple a system from the environment.
The main novel contribution of this work is to extend these mixing properties to a continuous-time framework, namely, Brownian motion over the unitary group induced by stochastic local Hamiltonians. In order to achieve these new results, on the one hand we move to a representation theoretic formulation and make use of its tools to establish the gap of local generators linked to the moments of the distribution induced by the diffusion process; from this we then derive an expression for the length of time it takes to ensure convergence toward the moments of the Haar measure. On the other hand we project the stochastic evolution onto a random walk on Pauli matrices and tie this description to the analogous one for random quantum circuits to achieve decoupling in a run time scaling almost linearly with respect to the system size. In addition to providing a unifying framework for random processes over the unitary group, we hence aim at presenting new mathematical results and techniques for quantum information. We furthermore discuss applications to black holes dynamics in the perspective of the information paradox.
As an additional novel result not related to Brownian motion, we propose a randomized benchmarking protocol exploiting the symmetry of the gate whose accuracy has to be estimated, in order to overcome current shortcomings afflicting the known schemes. We again rely on representation theory to reduce the computational effort and the amount of employed quantum resources.

Selbständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbstständig und nur mit Hilfe der angegebenen Hilfsmittel angefertigt habe. Alle Stellen der Arbeit, die wörtlich oder sinngemäß aus Veröffentlichungen oder aus anderweitigen fremden Quellen entnommen wurden, sind als solche kenntlich gemacht. Diese Arbeit ist nicht schon einmal in einem früheren Promotionsverfahren eingereicht worden.

Berlin, den 20. Februar 2018 _____

Emilio Onorati