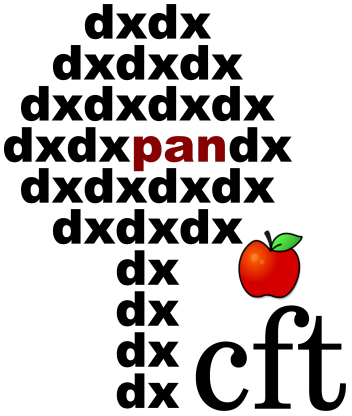
Spectral gaps, t-designs and ϵ-nets in quantum computing



Author: Oskar Szymon Słowik

Supervisor: prof. dr hab. Adam Sawicki

Co-supervisor: dr Oliver Reardon-Smith

Centrum Fizyki Teoretycznej Polskiej Akademii Nauk

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Physics

October 2025

Dedicated to my family—especially my grandfather Alfred, who sparked my interest in mathematics a long time ago

Abstract

Since the introduction of the quantum Turing machine almost half a century ago, quantum computing has emerged as a promising technology and an active interdisciplinary field of research. The quantum circuit model, which is polynomially equivalent to a quantum Turing machine, arose as the dominant formalism in quantum computation. Interestingly, such a gate-based model proved useful not only in describing quantum information processing in quantum computing hardware but also in applications such as many-body quantum systems, quantum complexity, and black hole physics. One of the reasons for such a diverse set of applications is that quantum circuits can be used to model the complexity of quantum states supported on discrete quantum physical systems. The question about the complexity of a specific unitary operation, which prepares the quantum state of interest up to a given precision, is known to be hard. However, the question about the joint upper bounds on such complexities is tangible and can be understood as a question about the computational efficiency of elementary quantum operations used to prepare the states. Indeed, the seminal Solovay-Kitaev (SK) theorem provides such bounds for any discrete universal gate set. However, it is known that certain gate sets enjoy the optimal scaling, which is better than the SK bound.

In this thesis, we explore the bounds on the efficiency of universal gate sets based on the spectral gap of the corresponding t-moment (or averaging) operators. We primarily focus on the finite-scale spectral gaps, which can, in principle, be computed.

Such an approach allows one to derive the non-constructive Solovay-Kitaev-like (SKL) theorems. We demonstrate how to obtain the SKL theorem using a construction based on the correspondence between δ-approximate t-designs and ϵ-nets, which are ubiquitous constructs, widely used in quantum information theory. We achieve such a correspondence by constructing polynomial approximations of the Dirac delta based on heat kernels, which are well-known and natural objects that find many applications in mathematical physics.

Using such an approach, we were able to improve the scaling of δ compared to the state of the art, while essentially retaining the scaling of t.

Aside from deriving the mentioned SKL theorem, we provide a relatively simple proof for the poly-logarithmic decay of the spectral gap with calculable constants and an alternative proof for the (global) spectral gap SKL theorem.

Finally, we introduce the notion of the Quantum Circuit Overhead (QCO) and the related notion of T-Quantum Circuit Overhead (T-QCO), which we believe are suitable measures to compare the efficiency of various gate sets and can be upper-bounded via simple formulas and numerical simulations. Aside from its direct relation to computational efficiency, the notion of the overhead can be used as a reasonable proxy for the actual cost-effectiveness of gate sets in certain NISQ and fault-tolerant architectures. We use this approach to gain insight into the efficiency of various random ensembles of single-qubit gate sets. Moreover, we numerically analyse several specific choices, such as Clifford+T and Super-Golden gate sets, obtaining interesting results concerning the efficiency of the famous T gate.

Crucially, we put a great emphasis on obtaining formulas where all constants are known or can be calculated in principle via numerical simulations on supercomputing clusters. Such an approach differs from some of the more mathematical works, in which the values of specific constants are not provided.

Streszczenie

Od czasu przedstawienia kwantowej maszyny Turinga prawie pół wieku temu, obliczenia kwantowe wyłoniły się jako obiecująca technologia i aktywny interdyscyplinarny obszar badań. Model obwodu kwantowego, który jest wielomianowo równoważny z kwantową maszyną Turinga, stał się dominującym formalizmem w obliczeniach kwantowych. Co interesujące, taki bramkowy model okazał się użyteczny nie tylko do opisu przetwarzania informacji kwantowej w sprzęcie do obliczeń kwantowych, lecz również znalazł zastosowania w obszarach takich jak wielociałowe układy kwantowe, złożoność kwantowa i fizyka czarnych dziur. Jednym z powodów tak różnorodnego wachlarza zastosowań jest fakt, że obwody kwantowe mogą zostać użyte do modelowania złożoności stanów kwantowych realizowanych przez dyskretne układy kwantowe. Wiadomo, że pytanie o złożoność konkretnej operacji unitarnej, która przygotowuje zadany stan kwantowy z zadaną dokładnością, jest trudne. Jednakże pytanie o wspólne ograniczenia górne na takie złożoności jest namacalne i może być rozumiane jako pytanie o efektywność obliczeniową elementarnych operacji kwantowych używanych do przygotowania stanów. W rzeczy samej, słynne twierdzenie Solovaya-Kitaeva (SK) wyznacza takie ograniczenia dla dowolnego dyskretnego zestawu bramek uniwersalnych. Jednakże wiadomo, że istnieją pewne zestawy bramek cechujące się optymalnym skalowaniem, które jest lepsze od ograniczenia wynikającego z twierdzenia SK.

W tej rozprawie badamy ograniczenia na efektywność uniwersalnych zestawów bramek opierając się na przerwie spektralnej odpowiadających im operatorów t-momentów (bądź uśredniania). Skupiamy się przede wszystkim na przerwach spektralnych na skończonej skali, które mogą być, co do zasady, wyliczone.

Takie podejście pozwala na wyprowadzenie niekonstruktywnych twierdzeń podobnych do twierdzenia Solovaya-Kitaeva (ang. Solovay-Kitaev-like; SKL). Demonstrujemy w jaki sposób można uzyskać twierdzenie SKL używając konstrukcji opartej na odpowiedniości

między δ-przybliżonymi t-designami oraz ϵ-netami, które są wszechobecnymi konstruktami, stosowanymi szeroko w teorii informacji kwantowej. Uzyskujemy taką odpowiedniość poprzez konstrukcję wielomianowych przybliżeń delty Diraca opartych na jądrach ciepła, które są dobrze znanymi i naturalnymi obiektami, znajdującymi wiele zastosowań w fizyce matematycznej. Używając takiego podejścia, byliśmy w stanie poprawić skalowanie δ względem najlepszego istniejącego wyniku, zasadniczo zachowując przy tym skalowanie t.

Poza wyprowadzeniem wspomnianego twierdzenia SKL, prezentujemy względnie prosty dowód na wielologarytmiczny zanik przerwy spektralnej ze stałymi, które są obliczalne oraz alternatywny dowód twierdzenia SKL z (globalną) przerwą spektralną.

Ostatecznie, wprowadzamy pojęcie Quantum Circuit Overhead (QCO) i powiązane z nim pojęcie T-Quantum Circuit Overhead, który naszym zdaniem mogą być odpowiednią miarą do porównywania efektywności różnych zestawów bramek kwantowych i może być ograniczony od góry za pomocą prostych wzorów i symulacji numerycznych. Poza bezpośrednim związkiem z efektywnością obliczeniową, pojęcie overheadu może być użyte jako rozsądna miara pośrednia (proxy) efektywności kosztowej w niektórych architekturach NISQ oraz z pełną korekcją błędów. Używamy tego podejścia, aby uzyskać wgląd w efektywność różnych losowych zespołów zestawów bramek jednokubitowych. Ponadto analizujemy numerycznie kilka konkretnych wyborów, takich jak zestaw bramek Clifford+T i Super-Golden, uzyskując interesujące wyniki dotyczące efektywności słynnej bramki T.

Co warte podkreślenia, kładziemy duży nacisk na to, aby wszystkie stałe w uzyskanych wzorach były znane bądź, co do zasady, obliczalne za pomocą symulacji numerycznych na klastrach superkomputerowych. Takie podejście różni się od niektórych bardziej matematycznych prac, w których wartości określonych stałych nie są podane.

Declaration

This dissertation is a result of my work as a PhD student at the Center for Theoretical Physics, Polish Academy of Sciences (CTP PAS), from October 2020 to September 2025. The work was supervised by prof. Adam Sawicki and later also by dr Oliver Reardon-Smith.

I declare that this thesis has not been submitted, in whole or in part, to obtain any other degree at the same institute or any other scientific institution.

The thesis is based on the following articles:

I. O. Słowik, A. Sawicki, "Calculable lower bounds on the efficiency of universal sets of quantum gates", *J. Phys. A: Math. Theor. 56 115304 (2023).*

II. O. Słowik, O. Reardon-Smith, A. Sawicki, "Fundamental solutions of the heat equation on unitary groups establish an improved relation between ϵ-nets and approximate unitary t-designs", *J. Phys. A: Math. Theor. 58 445301 (2025).*

III. O. Słowik, P. Dulian, A. Sawicki, "Quantum Circuit Overhead", *arXiv:2505.00683 (2025).*

which are an integral part of the thesis (see List of Publications for a complete list of the author's publications).

The thesis comprises six chapters, as well as a separate chapter that includes a complete list of the author's publications. Chapter 1 is an introduction to the thesis, which extends the abstract. Chapter 2 is a theoretical introduction with background information and the formulation of research problems. Chapters 3 to 5 are devoted to Papers I–III, respectively. Each chapter contains a summary of the paper and the contribution statement. Additionally, the relevant paper is attached at the end of the chapter. Chapter 6 contains the thesis summary and future research directions. A complete list of publications is provided in List of Publications.

Acknowledgements

I would like to thank Adam Sawicki and Oliver Reardon-Smith for their supervision, support, and our joint work during my PhD studies. I also thank Piotr Dulian for our collaboration, especially for providing his code. Moreover, I would like to express my gratitude to Péter Varjú for sharing some insights into the results he obtained.

Tools such as Grammarly and ChatGPT were used for language editing, editorial suggestions, assistance in generating certain figures, and initial fact-checking; the author reviewed all outputs and is solely responsible for any remaining errors.

I acknowledge the support from the Polish National Science Center (NCN), Grant No. 2015/18/E/ST1/00200 and No. 2020/37/B/ST2/02478, and I am grateful to the Director of the CTP PAS, Krzysztof Pawłowski, for offering me the position of Assistant, which secured my funding for an additional period.

Last but not least, I would like to thank everyone else who supported me during my studies — my cat Yuki, family, friends, and coworkers.

Contents

1 Introduction **1**

2 Preliminaries **7**

2.1 Basic notions . . . 7

2.1.1 Relevant groups . . . 8

2.1.2 Measures and functions . . . 9

2.1.3 Norms, balls and volumes . . . 12

2.2 Elements of group theory and analysis . . . 15

2.2.1 Elements of representation theory . . . 15

2.2.2 Peter-Weyl theorem and Fourier transform . . . 23

2.2.3 Averaging operators and spectral gap . . . 27

2.2.4 Balanced polynomials, t-moment operators and finite-scale spectral gap . . . 30

2.2.5 Casimir elements and Laplacian . . . 33

2.2.6 Approximate identities, heat and Fejér kernels . . . 36

2.3 Elements of quantum computation and information . . . 40

2.3.1 Unitary channels and ε-nets . . . 40

2.3.2 Quantum circuit model . . . 41

2.3.3 Words, complexity and universality . . . 44

2.3.4 Efficiency and cost . . . 48

2.3.5 Approximate unitary t-designs . . . 49

2.4 State of the art and research problems . . . 52

2.4.1 Spectral gap, efficient and optimal gate sets . . . 52

2.4.2 Decay of the spectral gap . . . 53

2.4.3 Solovay-Kitaev-like theorems . . . 53

2.4.4 Random walks, circuits and t-designs . . . 56

2.4.5 Research problems . . . 58

3 Paper I: Calculable lower bounds on the efficiency of universal sets of quantum gates **61**

3.1 Overview 61
3.2 Contribution statement 63
3.3 Errata 63

4 Paper II: Fundamental solutions of the heat equation on unitary groups establish an improved relation between ϵ-nets and approximate unitary t-designs 85
4.1 Overview 85
4.2 Contribution statement 87

5 Paper III: Quantum Circuit Overhead 134
5.1 Overview 134
5.2 Contribution statement 136

6 Summary and future directions 151
6.1 Summary 151
6.2 Future directions 153

Bibliography 159

Chapter 1

Introduction

The 20th century was rich in scientific breakthroughs. One of the most important advances in physics at that time was the birth of quantum mechanics. While Max Planck is often considered the "father" of quantum theory, the first mathematical framework of quantum mechanics was formulated in the mid-1920s by Werner Heisenberg and later by Erwin Schrödinger [1], marking the beginning of a new era in theoretical physics. Around ten years later, in 1936, Alan Turing laid the foundations of modern computer science by publishing a paper introducing a theoretical model of computation that can be made universal, later known as the Turing machine. Initially, quantum theory and computer science were seemingly unrelated branches of science, developing independently. However, at some point, physicists started to consider computational problems in the language of quantum theory. A key early step in this direction was made by Paul Benioff in 1980, with the introduction of the quantum Turing machine [1]. With the arrival of more advanced digital computers, scientists attempted to model quantum systems using computer simulations. However, physicists started to notice that such simulations may be inefficient due to the potential exponential scaling of required resources. This was articulated by Richard Feynman and Yuri Manin, who proposed using machines based on quantum phenomena to run potentially more efficient simulations [2, 3], thereby essentially establishing the concept of a quantum computer. The notion of a quantum computer was formalized by David Deutsch in 1985, who was also considering the applications of quantum computers to problems beyond quantum physics. Since then, numerous quantum algorithms [4] have been proposed. The first notable examples are the oracular algorithms of Deutsch (1985) and Simon (1994) [5, 6], which did not turn out to be particularly useful, but illustrated

[1] Heisenberg's "matrix mechanics" and Schrödinger's "wave mechanics" turned out to be mathematically equivalent.

query-complexity separations for some abstract problems. However, soon after that and inspired by Simon's work, Shor's algorithm (1994) was introduced, demonstrating an exponential quantum speed-up for the factoring and discrete logarithm problem. This sparked a surge in interest in quantum computing. Two years later, Grover's algorithm (1996) was introduced, demonstrating a generic quadratic quantum speed-up for the unstructured search problem [7, 8]. The more recent algorithms with potential applications include HHL (2008) [9] and more hardware-ready VQE and QAOA (2014) [10, 11].

Nowadays, quantum computing is an emerging technology and a multidisciplinary area of science, at the intersection of theoretical and experimental physics, computer science, mathematics, and engineering. Quantum computers are a type of computers that utilize quantum mechanical phenomena to solve computational problems and, as such, work in a different paradigm of computation than classical computers.

The quantum circuit model [4, 12], which is polynomially equivalent to the quantum Turing machine, is the universal and most widely used model of quantum computation. In this model, the quantum information stored in the quantum bit (qubit) register is processed through a series of elementary quantum operations, known as quantum gates. This is analogous to the classical circuit model, which can be used to describe the classical information processing occurring inside classical computers. Similarly to the classical circuits, which can be implemented using a finite set of elementary logic gates [2], a finite universal quantum gate set $\mathcal{S}$ is sufficient to realize an arbitrary n-qubit global unitary operation by a quantum circuit made out of the quantum gates from such a gate set. However, since there is a continuum of possible quantum operations, contrary to the classical case, the generic global quantum operation can be implemented only up to some finite error ϵ [3]. Although the gate-based quantum computers offer a potential advantage over classical computers, their current level of development makes such improvements debatable at best. This is mainly due to the moderate number of qubits, which, together with the quantum gates, state preparation, and measurement, are noisy enough to render the fault-tolerant quantum error correction schemes impossible. Executing potentially powerful quantum algorithms that require fault-tolerance, such as Shor's algorithm, is practically impossible in this scenario. Such devices are called noisy intermediate-scale quantum (NISQ) devices.

In the case of NISQ quantum hardware, reducing circuit length (i.e. the number of quantum gates used) and depth, especially the number of costly gates, such as the noisy entangling

[2] In fact, a single logical gate is needed - e.g. both NAND and NOR gates are universal.

[3] If the gate set is universal, then the error ϵ can be made arbitrarily small, for any target operation, by allowing long enough circuits.

gates, is necessary to keep the fidelities at an acceptable range and make the computations feasible [13, 14, 15]. Such circuit optimization is performed by the quantum compiler [16, 17, 4], which additionally decomposes quantum logical gates into the native gates used in a target quantum computer. Moreover, implementing quantum error correction schemes is associated with a significant overhead, which may cancel out potential speed-ups. Thus, even fault-tolerant quantum machines require a certain level of optimization, e.g., due to the high cost of the fault-tolerant implementation of specific gates used, e.g., the non-Clifford T gates in the case of Clifford+T quantum gates [18, 19, 20, 21, 22, 23, 24].

Such considerations naturally lead to questions like: What is the shortest circuit length ℓ, constructed from gates in $\mathcal{S}$, needed to implement an arbitrary global unitary quantum operation on an n-qubit register, up to some finite precision ϵ?; How does ℓ depend on the universal gate set $\mathcal{S}$ used? Such questions can be understood as questions about the efficiency of various gate sets $\mathcal{S}$ or, in the language of the complexities of unitary operations, about the joint upper bounds on the complexity of all global unitary operations with respect to $\mathcal{S}$ and for a certain precision ϵ.

Importantly, the questions about the efficiency of quantum gate sets are not only crucial to impose bounds on the compilation of quantum circuits executable on quantum computers, which are artificial, human-made machines. Although the statement that we all live inside a quantum computer is highly debatable [25], the dynamics of various naturally occurring objects can be described, or at least approximated, as a unitary dynamics of discrete quantum systems. Consequently, such dynamics naturally involve quantum information processing, which can be expressed in the universal language of quantum circuits. Such an approach has been recently used to study the dynamics of quantum many-body systems and, due to the dualities between physical theories and conjectures such as those of Brown and Susskind [26], to gain insight into the properties of black holes [27, 28, 29, 30, 31].

Although the seminal Solovay-Kitaev theorem [32, 12, 4] states that all universal gate sets are essentially efficient, namely $\ell = \mathcal{O}(\log^c(1/\epsilon))$, where c is a constant larger than one [4], examples of gate sets with the optimal asymptotic scaling $\ell = \Theta(\log(1/\epsilon))$ are known [34, 35, 36, 37]. In particular, the gate sets $\mathcal{S}$ with a so-called spectral gap enjoy such scaling. Nowadays, it is believed that every universal gate set has a spectral gap. Although the proof of such a conjecture (dating back to Sarnak) eluded mathematicians for years, the proof without additional assumptions on the gates [5] is still not known. In

[4] The value of the constant c depends on the proof and varies between $3+\alpha$, for any $\alpha > 0$ and ≈ 3.97. Recently, the cubic barrier has been broken with $c \approx 1.44$ [33].

[5] Such as the algebraicity of matrix entries, which is a property not satisfied generically.

fact, the history of the spectral gap results is longer and can be traced back to Kazhdan's property (T) [38].

However, the knowledge of the (global) spectral gap is not necessary to obtain poly-logarithmic bounds on ℓ. Indeed, there exist certain non-constructive theorems, similar to the Solovay-Kitaev theorem, that are based on the knowledge of the spectral gap at finite ϵ-dependent scales [39, 40, 41]. We refer to all constructive and non-constructive bounds on ℓ, which are poly-logarithmic in $1/\epsilon$, as Solovay-Kitaev-like (SKL) theorems. Contrary to the (global) spectral gap, the finite-scale spectral gap is, in principle, computable. Moreover, there exist poly-logarithmic bounds on the decay of the finite-scale spectral gap, which can be used to obtain SKL theorems using only the knowledge of the spectral gap at some group-dependent constant scale [40]. Unfortunately, such bounds often contain unknown constants, limiting their practical applications.

The main objective of this thesis is to study the calculable bounds on the efficiency of universal quantum gate sets.

In this thesis, we focus on the derivation of the bounds on the efficiency of quantum gate sets using the SKL theorems based on the finite-scale spectral gap. Crucially, we are interested not in the asymptotic scaling but in explicit bounds with all constants known or computable. Since such bounds already exist, our goal was to improve them, use more natural objects, or provide alternative, preferably simplified, proofs. Our strategy is mainly based on the establishment of the unitary δ-approximate t-design and ϵ-net correspondence using polynomial approximate identities stemming from the heat kernels, which are well-known and natural mathematical objects. This allows us to derive the SKL theorems based on the finite-scale spectral gap. Moreover, we introduce a new notion of the computational efficiency of the quantum gate sets - the Quantum Circuit Overhead (QCO), which we believe to be a good way for comparing the efficiency of different gate sets. Additionally, we introduce the related notion of the T-Quantum Circuit Overhead, which is more suitable for practical quantum hardware considerations. We also show how to upper-bound such overheads using the finite-scale spectral gap. Additionally, we supplement our theoretical considerations with numerical simulations on supercomputing clusters, demonstrating how our bounds can be used in practice. Such numerical simulations allow us to analyse various quantum gate sets, including the single-qubit Clifford+T, and shed light on the efficiency of the famous T gate and so-called Super-Golden Gates.

The thesis is organised as follows: Chapter 2 contains a common theoretical background needed to understand the results presented in the papers. It briefly covers the topics such as: basic notions, elements of group theory and analysis (including heat kernels), probability

theory, quantum computation and information (including ϵ-nets, δ-approximate t-designs and spectral gaps), and the formulation of the relevant state-of-the-art (SOTA) results, such as the SKL theorems and spectral gaps. Chapters 3, 4, and 5 are related to the papers I, II, and III, respectively. Each chapter includes information such as a short overview of the paper, the authors' contributions, and the full text of the paper. In Chapter 6, we summarize the thesis conclusions and discuss the future research directions, including the open problems. Finally, a complete list of authors' publications is given in List of Publications.

Chapter 2

Preliminaries

In this Chapter, we introduce the theoretical preliminaries underpinning the research work presented in the papers. In particular, we define the objects referred to in the thesis title - ε-nets, t-designs, and explain their relation to quantum computing. Finally, we present the chosen state-of-the-art (SOTA) results related to the thesis topic and formulate the research problems, indicating the ones addressed in the thesis.

2.1 Basic notions

For the matrix A we denote by $\bar{A}$ the matrix with complex conjugated entries, by A^T its transpose and by $A^\dagger$ its Hermitian conjugate, i.e. $A^\dagger = \overline{A^T}$. We say A is Hermitian if $A = A^\dagger$. For the square matrix A, we denote its trace as $\mathrm{Tr}(A)$ and its determinant as $\det(A)$. We denote the square identity matrix of dimension n as I_d. Similarly, for the matrix of zeros, we write 0_d.

By $M_d(\mathbb{C})$, we denote the set of $n \times n$ complex matrices. By $M_d^0(\mathbb{C})$, we denote the subset of $M_d(\mathbb{C})$ with matrices having trace zero.

The normalizer of a subset S in a group G is the set $N_G(S) \coloneqq \{g \in G| \quad gSg^{-1} = S\}$. The centralizer of S in G is $C_G(S) \coloneqq \{g \in G| \quad \forall_{s \in S}\, gs = sg\} \subseteq N_G(S)$. The center of the group G is $\mathcal{Z}(G) \coloneqq C_G(G)$.

2.1.1 Relevant groups

The three most widely used continuous groups in quantum information and computation theory are: the unitary group $\mathrm{U}(d)$, the special unitary group $\mathrm{SU}(d)$, and the projective group $\mathrm{PU}(d)$. We recall the definitions of such groups together with other relevant groups.

The general linear group,

$$\mathrm{GL}(d,\mathbb{C}) \coloneqq \{A \in M_d(\mathbb{C}) | \quad \det(\mathrm{A}) \neq 0\}. \tag{2.1}$$

The special linear group,

$$\mathrm{SL}(d,\mathbb{C}) \coloneqq \{A \in M_d(\mathbb{C}) | \quad \det(\mathrm{A}) = 1\}. \tag{2.2}$$

The unitary group,

$$\mathrm{U}(d) \coloneqq \{U \in M_d(\mathbb{C}) | \quad U^\dagger U = I_d\}. \tag{2.3}$$

The special unitary group,

$$\mathrm{SU}(d) \coloneqq U(d) \cap SL(d,\mathbb{C}). \tag{2.4}$$

We have the following inclusions as closed subgroups, $\mathrm{SU}(d) \subseteq \mathrm{SL}(d,\mathbb{C}) \subseteq \mathrm{GL}(d,\mathbb{C})$, and additionally $\mathrm{SU}(d) \subseteq \mathrm{U}(d) \subseteq \mathrm{GL}(d,\mathbb{C})$.

To define the projective groups, we use the notion of a group center. We have,

$$\mathcal{Z}(\mathrm{U}(d)) \coloneqq \{cI_d | \quad |c| = 1\} \cong \mathrm{U}(1), \tag{2.5}$$

and

$$\mathcal{Z}(\mathrm{SU}(d)) \coloneqq \{cI_d | \quad c^d = 1\} \cong \mathbb{Z}_d. \tag{2.6}$$

We define the projective unitary group $\mathrm{PU}(d)$ as the quotient of $\mathrm{U}(d)$ by the (right multiplication) of its center. Similarly, by taking the quotient of the $\mathrm{SU}(d)$ by its center, we obtain the projective special unitary group $\mathrm{PSU}(d)$, which is isomorphic to $\mathrm{PU}(d)$, i.e.

$$\mathrm{PU}(d) = \mathrm{U}(d)/\mathrm{U}(1) \cong \mathrm{SU}(d)/\mathbb{Z}_d = \mathrm{PSU}(d). \tag{2.7}$$

The groups $\mathrm{GL}(d,\mathbb{C})$, $\mathrm{GL}(d,\mathbb{C})$, $\mathrm{U}(d)$, $\mathrm{SU}(d)$ and $\mathrm{PU}(d)$ are connected matrix Lie groups. Out of them, $\mathrm{U}(d)$, $\mathrm{SU}(d)$ and $\mathrm{PU}(d)$ are compact real Lie groups and the others are non-compact complex Lie groups. Moreover, the groups $\mathrm{SL}(d,\mathbb{C})$ and $\mathrm{SU}(d)$ are simply connected [1].

Since the group $\mathrm{PU}(d)$ is defined as the quotient of either $\mathrm{U}(d)$ or $\mathrm{SU}(d)$, it is the image of the corresponding canonical projections. Since $\mathrm{SU}(d)$ is easier to work with for our purposes, we can define various structures on $\mathrm{SU}(d)$ and use the canonical projection

$$\pi : \mathrm{SU}(d) \to \mathrm{PU}(d) \tag{2.8}$$

to transport them to $\mathrm{PU}(d)$ or lift the objects of interest from $\mathrm{PU}(d)$ to $\mathrm{SU}(d)$. For example, every function f on $\mathrm{PU}(d)$ can be pulled back to the (unique) function $\tilde{f} = f \circ \pi$ on $\mathrm{SU}(d)$ which is constant on the fibres of π.

2.1.2 Measures and functions

For a compact group G, we denote its unique (normalised) Haar measure by μ. For compact groups [2], left and right Haar measure coincide so that μ is bi-invariant, i.e., for any Borel subset $A \subset G$ and element $g \in G$, we have

$$\mu(gA) = \mu(Ag) = \mu(A). \tag{2.9}$$

Such translation invariance of μ is a property very useful in calculations.

The Haar measure μ_S on $\mathrm{SU}(d)$ can be pushed forward to the Haar measure μ_P on $\mathrm{PU}(d)$, i.e. $\mu_P(A) = \mu_S(\pi^{-1}(A))$, whenever $\pi^{-1}(A)$ is μ_S-measurable. This way, the integration on $\mathrm{PU}(d)$ can be obtained by means of integration on $\mathrm{SU}(d)$ (a variant of the change of variable formula)

$$\int_X f \, d\mu_P = \int_{\tilde{X}} \tilde{f} \, d\mu_S, \tag{2.10}$$

where $X \subseteq \mathrm{PU}(d)$ is some Haar-measurable set and $\tilde{X} = \pi^{-1}(X)$.

[1] Excluding the degenerate case of $\mathrm{SL}(1,\mathbb{C}) = \mathrm{PU}(1) = \{1\}$, which is compact and complex and simply-connected. In fact, any connected, compact and complex Lie group is abelian.

[2] or more generally - unimodular groups

The Haar measure on G is an example of a finite, regular Borel measure. We denote general finite Borel measures on compact groups as ν.

For a chosen subset $A \subset G$, by $\mathbb{1}_A : X \to \{0, 1\}$ we denote the indicator function of the set A, i.e. $\mathbb{1}_A$ has value 1 at points of A and 0 at points of its compliment $G \setminus A$.

We say ν is a probability measure if it is real, non-negative, and normalized to 1.

A prominent example of ν, which is important in applications, is a discrete probability measure supported on a finite subset $\mathcal{S} \subset G$, especially its uniform version, which we denote as $\nu_{\mathcal{S}}$

$$\nu_{\mathcal{S}} \coloneqq \frac{1}{|\mathcal{S}|} \sum_{g \in \mathcal{S}} \delta_g, \tag{2.11}$$

where δ_g is a Dirac measure of g, i.e. $\delta_g(A) = \mathbb{1}_A(g)$ for any measurable set $A \subset G$. In general, we denote the support of ν as $\mathrm{supp}(\nu)$.

We say a probability measure ν is symmetric if $\nu(A) = \nu(A^{-1})$, for every measurable set A. In particular, this corresponds to $\nu_{\mathcal{S}}$ with symmetric (inverse-closed) set $\mathcal{S} = \{g_1, g_2, \ldots, g_k, g_1^{-1}, g_2^{-1}, \ldots, g_k^{-1}\}$.

For two finite Borel measures ν_1 and ν_2 on G and the measurable set A, their convolution measure $\nu_1 * \nu_2$,

$$(\nu_1 * \nu_2)(A) = \iint_{G \times G} \mathbb{1}_A(gh) d\nu_1(g) d\nu_2(h), \tag{2.12}$$

is also a finite measure.

Example 2.1. *The calculation of the ℓ self-convolution of $\nu_{\mathcal{S}}$ is particularly simple. We have*

$$\nu_{\mathcal{S}}^{*(\ell)} = \frac{1}{|\mathcal{S}|^\ell} \sum_{\omega \in \mathcal{S}^\ell} \delta_\omega \tag{2.13}$$

where $\mathcal{S}^\ell = \{g_1 g_2 \ldots g_\ell | \quad g_i \in \mathcal{S}\}$ is the set of all words of length ℓ over the alphabet $\mathcal{S}$.

In practice, one may prefer to work with the corresponding functions on G instead of the measures. For "sufficiently regular" measures, this can be made precise through the Radon-Nikodym derivative. Indeed, if ν is absolutely continuous with respect to μ (denoted $\nu \ll \mu$), i.e. for every μ-measurable set A, $\mu(A) = 0$ implies $\nu(A) = 0$, then there exist a measurable function $\frac{d\nu}{d\mu}$ [3], called the Radon-Nikodym derivative, such that for any measurable set A

[3] Unique up to a μ-null set.

$$\nu(A) = \int_A \frac{d\nu}{d\mu}(g)d\mu(g). \tag{2.14}$$

In such a case, we say ν has a density $k = d\nu/d\mu \in L^1(G)$ with respect to μ, and $||k||_1 \leq |\nu|(G)$ (see (2.15)), where $|\nu|(G)$ is the total variation of ν on G.

Moreover, the derivative of the convolution is the convolution of derivatives (as functions; see (2.18)). The measure $\nu_{\mathcal{S}}$ does not have the corresponding probability density as it is not absolutely continuous with respect to μ. However, by taking its convolution with appropriate approximations of identity (aka the mollifiers), we can consider the corresponding approximate densities.

By $L^p(G)$, $1 \leq p < \infty$, we denote the space of L^p-integrable complex functions on G, which is a Banach space under the norm [4]

$$||f||_p \coloneqq \left(\int_G |f(g)|^p d\mu(g)\right)^{1/p}. \tag{2.15}$$

In particular, for $p = 2$, the space $L^2(G)$ is a Hilbert space with a scalar product

$$\langle f, g\rangle \coloneqq \int_G f(x)\overline{g(x)}d\mu(x), \tag{2.16}$$

where $\overline{g}$ denotes the complex conjugate of a function g.

By $C(G)$ we denote the set of continuous complex functions on G, which is a Banach space under the supremum norm $||f||_\infty = \sup_{g\in G}|f(g)|$. The space $C(G) \subset L^p(G)$ is dense in any $L^p(G)$ with $1 \leq p < \infty$.

The following inequality turns out to be very useful for the study of $L^p(G)$-spaces [5].

Fact 2.1 (Hölder's inequality - specialised)**.** *Let $p, q \in [1, \infty]$ with $1/p + 1/q = 1$* [6]*. Then for f, g being the μ-measurable functions on G*

$$||fg||_1 \leq ||f||_p||g||_q. \tag{2.17}$$

[4] Formally, one should consider the equivalence classes of functions, where functions which agree μ-almost everywhere (i.e. up to a set of μ-measure zero) are identified.

[5] This inequality is valid for any measure space and any measurable real or complex functions.

[6] The ∞ case corresponds to the space $L^\infty(G)$ of essentially bounded functions on G. On a compact group we have $C(G) \subseteq L^\infty(G)$.

Moreover if additionally $p, q \in (1, \infty)$ *and* $f \in L^p(G)$ *and* $g \in L^q(G)$*, then one obtains an equality if and only if* $\alpha|f|^p = \beta|g|^q$ μ*-almost everywhere for some real numbers* $\alpha, \beta \geq 0$*, not both of them zero.*

A simple consequence of (2.1) and the finiteness of the Haar measure ($\mu(G) < \infty$) is that we have the inclusions $L^q(G) \subseteq L^p(G)$, for $p < q$ with the norms satisfying $||f||_p \leq \mu(G)^{\frac{1}{p}-\frac{1}{q}}||f||_q$. In particular, for the normalized Haar measure and $f \in L^2(G)$ being the probability density, we have $||f||_2 \geq 1$ with equality if and only if $f = 1$ μ-almost everywhere [7]. Thus, the discrepancy between the L^2-norm of a probability density on G and one can be understood as a measure of its non-uniformity.

We define the convolution of functions $f, g \in L^1(G)$ as

$$(f * g)(x) \coloneqq \int_G f(y)g(y^{-1}x)d\mu(y), \tag{2.18}$$

and the convolution of a finite Borel measure ν and $f \in L^1(G)$

$$(\nu * f)(x) \coloneqq \int_G f(y^{-1}x)d\nu(y). \tag{2.19}$$

If ν has density $k = d\nu/d\mu$, then $\nu * f = k * f$.

2.1.3 Norms, balls and volumes

In this subsection, we introduce the relevant norms and metrics. By G we mean any of the groups $\mathrm{U}(d)$, $\mathrm{SU}(d)$ and $\mathrm{PU}(d)$. Since we are restricted to the matrix case, we refrain from providing the more general versions of stated facts.

The default norm we use for the operators is the operator norm, denoted $||\cdot||_\infty$ [8]. In particular, for the matrix A, $||A||_\infty$ is the square root of the largest eigenvalue of the matrix $A^\dagger A$.

We denote the Hilbert-Schmidt norm of the operator as $||\cdot||_{HS}$. In particular, for the matrix $A = (a_{ij})$, $||A||_{HS} = \sqrt{\sum_i \sum_j |a_{ij}|^2} = \sqrt{\mathrm{Tr}\,(A^\dagger A)}$ and the corresponding scalar product is $\langle A, B\rangle_{HS} \coloneqq \mathrm{Tr}(AB^\dagger)$.

[7] Probability density is the non-negative real function normalized/integrating to one

[8] In some of our papers, we use $||\cdot||_{op}$ instead.

For matrices A and B with defined product, we have $||AB||_{HS} \leq ||A||_\infty \cdot ||B||_{HS}$ and $||BA||_{HS} \leq ||B||_{HS} \cdot ||A||_\infty$. In particular $||A||_\infty \leq ||A||_{HS}$.

A more unified way to look at $||\cdot||_\infty$ and $||\cdot||_{HS}$ is provided by the Schatten norm. For a $m \times n$ matrix A, we define

$$|A| \coloneqq \sqrt{A^\dagger A}, \tag{2.20}$$

and the Schatten p-norm

$$||A||_p \coloneqq (\mathrm{Tr}(|A|^p))^{1/p} = \left(\sum_{j=1}^{\min\{m,n\}} \sigma_j^p(A) \right)^{1/p}, \tag{2.21}$$

where $\sigma_j(A)$ denotes the j-th singular value, i.e. the j-th eigenvalue of $|A|$.

This agrees with the definition of the operator norm and the Schatten 2-norm corresponds to the Hilbert-Schmidt/Frobenius norm. For $p = 1$ we obtain the trace/nuclear norm $||A||_1 = \mathrm{Tr}(|A|)$. By combining von Neumann's trace inequality with Hölder's inequality for Euclidean spaces, we obtain Hölder's inequality for Schatten norms

$$|\langle A, B\rangle_{HS}| \leq ||A||_p ||B||_q. \tag{2.22}$$

All Schatten norms are sub-multiplicative, meaning that $||AB||_p \leq ||A||_p \cdot ||B||_p$. Moreover, they are unitarily invariant, which means $||UAV||_p = ||A||_p$, for all matrices A and all unitary matrices U and V.

Finally, we note that the Schatten norm can be defined in the same way for bounded linear operators on Hilbert spaces, possibly by specifying the $p = \infty$ case separately as the operator norm (e.g for non-compact operators), and the same properties hold.

Using the operator norm, we equip $\mathrm{U}(d)$ with the induced metric $d_\infty(\cdot, \cdot)$

$$d_\infty(U, V) \coloneqq ||U - V||_\infty. \tag{2.23}$$

By restricting $d_\infty(\cdot, \cdot)$ we define the metric on $\mathrm{SU}(d)$, denoting it using the same symbol.

To obtain the corresponding metric on $\mathrm{PU}(d)$, we minimize over the relative global phase:

$$d_P(\mathbf{U},\mathbf{V}) \coloneqq \min_{\varphi} ||U - e^{i\varphi}V||_\infty, \tag{2.24}$$

where $\mathbf{U},\mathbf{V} \in \mathrm{PU}(d)$ and $U,V \in \mathrm{U}(d)$ are the unitary representatives of the elements $\mathbf{U}$ and $\mathbf{V}$ respectively. The metric d_P can be also defined using the metric on $\mathrm{SU}(d)$:

$$d_P(\mathbf{U},\mathbf{V}) = \min_{\gamma \in \mathcal{Z}(\mathrm{SU}(d))} d_\infty(U,\gamma V), \tag{2.25}$$

where $U = \pi(\mathbf{U})$ and $V = \pi(\mathbf{V})$.

Due to the unitary invariance of the operator norm, the metrics $d(\cdot,\cdot)$ and $d_P(\cdot,\cdot)$ are translation-invariant.

By B_ε we denote the closed ε-ball in G centered at the group identity, with respect to the appropriate metric. The closed ε-ball centered at $g \in G$ is then a translation $B_\varepsilon(g) = gB_\varepsilon = B_\varepsilon g$.

The Haar volume of $B_\varepsilon \subset G$ [9] can be bounded as

$$(a_v\varepsilon)^{d^2-1} \leq \mu(B_\varepsilon) \leq (A_v\varepsilon)^{d^2-1}, \tag{2.26}$$

These constants can be obtained using methods from [42], where the balls in homogeneous spaces of $\mathrm{U}(d)$ are studied (see also [41])

TABLE 2.1: Common groups G used in quantum computing together with the constants providing the bounds of the volumes of balls in respective metrics.

G	a_v	A_v	dim$G$	metric
$\mathrm{U}(d)$	$\frac{1}{4\pi+2}$	$\frac{10}{\pi}$	d^2	d_∞ (2.23)
$\mathrm{SU}(d)$	$\frac{1}{8\pi+2}$	$\frac{10}{\pi}$	d^2-1	d_∞ (2.23)
$\mathrm{PU}(d)$	$\frac{1}{8\pi+2}$	87	d^2-1	d_P (2.24)

[9] Due to translational invariance of Haar measure and the metric, the volume of a ball does not depend on its origin.

2.2 Elements of group theory and analysis

In this thesis, we consider representations over the field of complex numbers. We always assume the groups and Lie algebras are finite-dimensional. We assume the reader is familiar with the basic facts from representation theory; we redirect readers without any exposition to standard books, e.g., [43, 44, 45, 46].

2.2.1 Elements of representation theory

We are interested in representations (reps) of compact groups. Due to the standard Haar-averaging "unitary trick", every finite-dimensional rep of a compact group is equivalent to a unitary rep. Moreover, every finite-dimensional rep of a compact group is completely reducible, i.e., decomposes into a direct sum of irreducible representations (irreps). As outlined in Subsection 2.2.2, due to the Peter-Weyl theorem, this is also the case for the regular representations of compact groups, which are infinite-dimensional unitary reps. In fact, every unitary irrep of a compact group is finite-dimensional, and every unitary rep of a compact group is completely reducible [47].

For completeness, we recall some basic definitions for unitary representations.

A unitary representation of a group G, as a continuous homomorphism $\pi : G \to \mathrm{U}(V_\pi)$, where $\mathrm{U}(V_\pi)$ denotes the unitary group on a complex Hilbert space V_π, equipped with the strong operator topology. The continuity means then that $g \to \pi(g)v$ is continuous as a mapping from G to V_π, for every fixed $v \in V_\pi$ (strong continuity). We call V_π the rep space, and in case of finite-dimensional V_π refer to $d_\pi \coloneqq \dim(V_\pi)$ as the dimension of π.

Two unitary reps π_1 and π_2 are said to be equivalent (isomorphic) if there is a unitary operator $U : V_{\pi_1} \to V_{\pi_2}$, such that for all $g \in G$, $\pi_1(g) = U\pi_2(g)U^\dagger$. We then write $\pi_1 \cong \pi_2$ (or $V_1 \cong V_2$, unless it may lead to confusion). In other words, they are the same up to the unitary change of basis.

For a rep π, a subspace $W \subseteq V_\pi$ is called invariant if $\pi(g)W \subseteq W$ for all $g \in G$.

A rep is called irreducible (irrep) if it is non-zero and V_π has no non-trivial closed invariant subspaces. By $\mathbf{1}$, we denote the irreducible trivial representation (which is one-dimensional).

Finite-dimensional (continuous) reps of Lie groups are automatically smooth.

A function f on G is central (or a class function), if $f(g) = f(hgh^{-1})$, for any $g, h \in G$.

A character of a finite-dimensional rep $\pi : G \to \mathrm{GL}(V)$ is a central function defined as

$$\chi_\pi(g) = \mathrm{Tr}(\pi(g)). \tag{2.27}$$

The finite-dimensional rep $\pi : G \to \mathrm{GL}(V_\pi)$ of a Lie group G with Lie algebra $\mathfrak{g}$ gives rise to a rep of $\mathfrak{g}$, $d\pi : \mathfrak{g} \to \mathfrak{gl}(V)$, called the derived rep

$$d\pi(X) \coloneqq \left.\frac{d}{dt}\right|_{t=0} \pi(e^{tX}). \tag{2.28}$$

Moreover, if G is connected, then π is an irrep if and only if $d\pi$ is an irrep.

To cover the infinite-dimensional unitary reps, notice that $t \mapsto \pi(e^{tX})$ is a strongly continuous one-parameter subgroup, so $d\pi$ can be defined using Stone's theorem with (2.28) replaced by strong derivative. The operator $d\pi$ is then skew-adjoint on its natural dense domain. Additionally, we obtain

$$\pi(e^{tX}) = e^{td\pi(X)}, \quad t \in \mathbb{R}, \tag{2.29}$$

which for $t = 0$ reproduces a well-known formula for finite-dimensional reps.

Example 2.2. *The derived rep of the adjoint representation of a group* $\mathrm{Ad} : G \to \mathfrak{gl}(\mathfrak{g})$, $\mathrm{Ad}(g) : X \mapsto gXg^{-1}$ *is called the adjoint representation of its Lie algebra* $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$, $\mathrm{ad}(X) : Y \mapsto [X, Y]$.

The adjoint representation of $\mathfrak{g}$ gives rise to asymmetric bilinear form on $\mathfrak{g}$, called the Killing form,

$$B(X, Y) \coloneqq \mathrm{Tr}\left(\mathrm{ad}(X) \circ \mathrm{ad}(Y)\right), \tag{2.30}$$

which is ad-invariant, i.e. $B([X, Y], Z) = B(X, [Y, Z])$. By changing ad to any finite-dimensional rep of $\mathfrak{g}$, we define the so-called trace form of π, which has the same properties as specified.

We say a Lie algebra $\mathfrak{g}$ is simple if it is non-abelian and has no non-zero proper ideals. A Lie algebra $\mathfrak{g}$ is semisimple if it is a direct sum of simple Lie algebras. Equivalently, $\mathfrak{g}$

has no non-zero abelian ideals. Moreover, this is equivalent to the Killing form being non-degenerate and also implies $\mathfrak{g} = [\mathfrak{g}, \mathfrak{g}]$. We say a Lie algebra $\mathfrak{g}$ is reductive if it decomposes (as Lie algebras) as $\mathfrak{g} = \mathfrak{a} \oplus [\mathfrak{g}, \mathfrak{g}]$, where $\mathfrak{a}$ is abelian and $[\mathfrak{g}, \mathfrak{g}]$ is semisimple. Lie algebras of compact Lie groups are reductive, with the semisimple part having a negative-definite Killing form (for real Lie groups).

The complex representation theory of a real Lie algebra $\mathfrak{g}$ is equivalent to the complex representation theory of its complexification $\mathfrak{g}_{\mathbb{C}} \coloneqq \mathfrak{g} \otimes_{\mathbb{R}} \mathbb{C}$.

The complex (finite-dimensional) representation theory of (finite-dimensional) complex semisimple Lie algebras is particularly simple and elegant e.g. due to the results such as the complete reducibility, weight and root space decomposition, highest weight theorem and Weyl character formula. Finally, such algebras themselves are classified, up to isomorphism, by the (finite disjoint sums of) Dynkin diagrams. However, many of these good properties also apply in the reductive case, with simple modifications stemming from the abelian component. Finally, the semisimple theory is closely related to the representation theory of compact Lie groups.

We now specialize to the representation theory of compact connected (real) Lie groups U(d) and SU(d), hence by G we denote one of such groups. We denote the (real) Lie algebra of G by $\mathfrak{g}$ and use $\mathfrak{g}_{\mathbb{C}} \coloneqq \mathfrak{g} + i\mathfrak{g}$ for its complexification.

The Cartan subalgebra (CSA) of $\mathfrak{g}_{\mathbb{C}}$ is an abelian and diagonalisable subalgebra of $\mathfrak{g}_{\mathbb{C}}$ which is maximal under set inclusion. Choose the maximal torus in $\mathbb{T} \subset G$ with Lie algebra $\mathfrak{t}$. Then the corresponding Cartan subalgebra of $\mathfrak{g}_{\mathbb{C}}$ is $\mathfrak{h} \coloneqq \mathfrak{t} + i\mathfrak{t}$. The maximal torus is unique up to conjugacy.

Consider a finite-dimensional rep (π, V) of $\mathfrak{g}_{\mathbb{C}}$. From the definition of the CSA, the family of operators $\{\pi(H) | \quad H \in \mathfrak{h}\}$ is simultaneously diagonalizable. Hence, we can organize the eigenvalues of their common eigenvectors using linear functionals on $\mathfrak{h}$, i.e. the elements from the dual space $\mathfrak{h}^*$.

We define a weight of V as an element $\lambda \in \mathfrak{h}^*$, such that the corresponding weight space

$$V_\lambda \coloneqq \{v \in V | \quad \pi(H)v = \lambda(H)v, \forall H \in \mathfrak{h}\} \tag{2.31}$$

is not zero. The vector space V decomposes into weight spaces,

$$V = \bigoplus_{\lambda \in w(\pi)} V_\lambda, \tag{2.32}$$

where by $w(\pi)$ we denote the set of weights of π.

By considering the specific case of the adjoint representation $\mathrm{ad} : \mathfrak{g}_{\mathbb{C}} \to \mathfrak{gl}(\mathfrak{g}_{\mathbb{C}})$, we arrive at the notion of a root. The root of $\mathfrak{g}_{\mathbb{C}}$ is the non-zero element $\alpha \in \mathfrak{h}^*$, such that the corresponding root space

$$\mathfrak{g}_\alpha := \{X_\alpha \in \mathfrak{g}_{\mathbb{C}} | \quad [H, X_\alpha] = \alpha(H)X_\alpha, \forall H \in \mathfrak{h}\} \tag{2.33}$$

is not zero. We denote the set of roots of $\mathfrak{g}$ as Φ. We have the following root space decomposition,

$$\mathfrak{g}_{\mathbb{C}} = \mathfrak{h} \oplus \bigoplus_{\alpha \in \Phi} \mathfrak{g}_\alpha. \tag{2.34}$$

The weight spaces (2.32) are not invariant. Indeed, the root vectors act on weight vectors by changing their weight spaces $\pi(X_\alpha) : V_\lambda \to V_{\lambda+\alpha}$, i.e. for $v_\lambda \in V_\lambda$

$$\pi(H)\pi(X_\alpha)v_\lambda = (\lambda + \alpha)(H)\pi(X_\alpha)v_\lambda. \tag{2.35}$$

We note that for the compact (real) Lie groups, the weights and roots take purely imaginary values in $\mathfrak{t}$. Hence, it is customary to use the notion of real weights $\lambda_{\mathbb{R}}$ and real roots $\alpha_{\mathbb{R}}$, which are the elements of $\mathfrak{t}^*$. The definitions are analogous to (2.31) and (2.33) with $\lambda(H)$ and $\alpha(H)$ replaced with $i\lambda_{\mathbb{R}}(H)$ and $i\alpha_{\mathbb{R}}(H)$ respectively and $H \in \mathfrak{t}$.

By picking a nondegenerate symmetric bilinear form B on $\mathfrak{h}$, we can identify $\mathfrak{h} \simeq \mathfrak{h}^*$ via Riesz isomorphism, i.e. for $\mathfrak{h} \to \mathfrak{h}^*$ we put $H \mapsto H^\flat := B(H, \cdot)$ and for $\mathfrak{h}^* \to \mathfrak{h}$ we put $\alpha \mapsto \alpha^\sharp$, where $B(\alpha^\sharp, H) = \alpha(H)$ for any $H \in \mathfrak{h}$. This defines the corresponding form on $\mathfrak{h}^*$, which we denote by the same symbol, given by $B(\alpha, \beta) = B(\alpha^\sharp, \beta^\sharp)$, for any $\alpha, \beta \in \mathfrak{h}^*$.

To form (an abstract) root system $\Phi \subset V$, which lives in a real span of roots $V = \mathrm{span}_{\mathbb{R}}\Phi$, we require B to induce an inner product on V. Additionally, the inner product needs to be invariant under a specific group of reflections, called the Weyl group, which can be ensured by picking ad-invariant B. We then denote $B(\cdot, \cdot)$ and its restrictions as $(\cdot, \cdot)$.

For $\alpha \in \mathfrak{h}^*$ we define $\alpha^\vee := \frac{2\alpha}{(\alpha,\alpha)} \in \mathfrak{h}^*$.

The Weyl group is then generated as $W := \langle s_\alpha | \quad \alpha \in \Phi \rangle$, where s_α is a reflection about a hyperplane perpendicular to α

$$s_\alpha(\lambda) = \lambda - (\lambda, \alpha^\vee)\alpha, \quad \lambda \in \mathfrak{h}^*, \tag{2.36}$$

which leaves Φ invariant.

From the properties of a root system, by choosing any hyperplane in V that does not intersect Φ, we obtain two sets of roots with equal cardinality. Each set is closed under addition and contains either α or -α, for all $\alpha \in \Phi$. The set of positive roots Φ^+ is then the set of roots lying at the chosen side of the hyperplane we decided. Additionally, we distinguish a set of simple roots Δ, which are the elements of Φ^+ that cannot be decomposed as a sum $\alpha + \beta$, for $\alpha, \beta \in \Phi^+$. Then, every $\alpha \in \Phi$ is a linear combination of simple roots with all the coefficients being either positive or negative integers.

For $\alpha, \beta \in \mathfrak{h}^*$, we say that α is higher than β, denoted $\alpha > \beta$, if $\alpha - \beta$ is a linear combination of simple roots with non-negative coefficients. Analogously we define $\alpha < \beta$. We say a weight $\lambda \in \mathfrak{h}^*$ is the highest, if there is no weight higher than λ.

We define the integral lattice in $\mathfrak{t}$

$$\mathfrak{t}_\mathbb{Z} \coloneqq \{X \in \mathfrak{t} | \quad e^{i2\pi X} = I\}. \tag{2.37}$$

We say $\lambda \in \mathfrak{h}^*$ is integral, if $(\lambda, \alpha^\vee) \in \mathbb{Z}$ for all $\alpha \in \Phi$. We say λ is analytically integral if $\lambda(X) \in \mathbb{Z}$ for all $X \in \mathfrak{t}_\mathbb{Z}$. Finally, we say λ is dominant if $(\lambda, \alpha^\vee) \geq 0$ for all $\alpha \in \Delta$.

The analytic integrality condition is needed to identify the irreps of the Lie algebra that integrate to the irreps of its Lie group.

We are now ready to characterize the irreps of G using the Theorem of the highest weights. The Theorem states that every finite-dimensional complex rep of G has a unique highest weight λ and such λ is a dominant element. Moreover, λ is the same for isomorphic reps. Finally, if the highest weight λ is dominant and analytically integral, there exists a finite-dimensional complex irrep with such a highest weight. It follows that the equivalence classes of finite-dimensional complex representations of G are bijectively labelled by the corresponding dominant and analytically integral highest weights.

From now on, by highest weights we will understand such highest weights bijectively labelling the equivalence classes of irreps.

Finally, we define the Weyl vector as

$$\delta \coloneqq \frac{1}{2} \sum_{\alpha \in \Phi^+} \alpha, \tag{2.38}$$

and we proceed to the group-specific computations. For $\mathrm{U}(d)$ we have the following real Lie algebra

$$\mathfrak{u}(d) = \{X \in M_d(\mathbb{C}) | \quad X^\dagger = -X\}. \tag{2.39}$$

The complexification reads $\mathfrak{u}(d)_\mathbb{C} \cong \mathfrak{gl}(d, \mathbb{C}) = M_d(\mathbb{C})$. We pick the maximal torus

$$\mathbb{T} \coloneqq \{\mathrm{diag}(e^{i\phi_1}, e^{i\phi_2}, \ldots, e^{i\phi_d}) | \quad \phi \in \mathbb{R}\}, \tag{2.40}$$

so that the toral algebra reads

$$\mathfrak{t} \coloneqq \{\mathrm{diag}(i\phi_1, i\phi_2, \ldots, i\phi_d) | \quad \phi \in \mathbb{R}\}, \tag{2.41}$$

and the CSA is

$$\mathfrak{h} \coloneqq \{\mathrm{diag}(z_1, z_2, \ldots, z_d) | \quad z \in \mathbb{C}\}. \tag{2.42}$$

For $\mathrm{SU}(d)$, we have $\mathfrak{su}(d) = \mathfrak{u}(d) \cap M_d^0(\mathbb{C})$ and $\mathfrak{su}(d)_\mathbb{C} \cong \mathfrak{sl}(d, \mathbb{C}) = M_d^0(\mathbb{C})$. The Lie algebra $\mathfrak{sl}(d, \mathbb{C})$ is semisimple.

The corresponding maximal torus in $\mathrm{SU}(d)$ is $\mathbb{T}_0 \coloneqq \mathbb{T} \cap \mathrm{SU}(d)$, with the toral algebra $\mathfrak{t}_0 \coloneqq \mathfrak{t} \cap M_d^0(\mathbb{C})$ and the CSA $\mathfrak{h}_0 \coloneqq \mathfrak{h} \cap M_d^0(\mathbb{C})$.

The Lie algebra $\mathfrak{u}(d)_\mathbb{C}$ is not semisimple, but it is reductive, with $\mathfrak{u}(d)_\mathbb{C} = \mathbb{C} I_d \oplus \mathfrak{sl}(d, \mathbb{C})$.

We introduce the linear functionals on $\mathfrak{h}$,

$$L_j : \begin{pmatrix} z_1 & & & \\ & z_2 & & \\ & & \ddots & \\ & & & z_d \end{pmatrix} \mapsto z_j. \tag{2.43}$$

so that $\{L_j | \quad 1 \leq j \leq d\}$ is the basis of $\mathfrak{h}^*$. Restricting to $\mathfrak{h}_0^*$, we can use $\{L_j | \quad 1 \leq j \leq d\}$ to span $\mathfrak{h}_0^*$, keeping in mind that (i.e. quotienting by) $L_1 + L_2 + \ldots + L_d = 0$.

The set of roots, $\Phi \coloneqq \{\alpha_{i,j} | \quad 1 \leq i, j \leq d\}$, where $\alpha_{i,j} \coloneqq L_i - L_j$, is the same for $\mathfrak{u}(d)_\mathbb{C}$ and $\mathfrak{su}(d)_\mathbb{C}$. This is not a coincidence, as it is easy to see that the roots need to vanish

on the abelian part of the reductive algebra and can be defined as the functionals on the semisimple part.

We choose the positive roots as $\Phi^+ \coloneqq \{\alpha_{i,j} | \quad 1 \leq i < j \leq d\}$ and the set of simple roots $\Delta \coloneqq \{\alpha_{i,i+1} | \quad 1 \leq i \leq d-1\}$.

The inner product on $\mathfrak{h}^*$ can be chosen to be $(L_i, L_j) = \delta_{ij}$, so that the (long) roots have squared length 2, which is a standard normalisation. The restriction to $\mathfrak{h}_0^*$ preserves the standard normalisation and yields $(L_i, L_j) = \delta_{ij} - \frac{1}{d}$. Such a restriction is proportional to the inner product on $\mathfrak{h}_0^*$ induced by the (unnormalised) negative Killing form on $\mathrm{sl}(d, \mathbb{C})$, $(L_i, L_j) = \frac{1}{2d}\left(\delta_{ij} - \frac{1}{d}\right)$, with squared root length $(\alpha_{i,j}, \alpha_{i,j}) = 1/d$, and restricts to an inner product on the root system V.

The root system $\Phi \subset V$ is of type A_{d-1} in Dynkin classification. The Weyl group isomorphic to the group of permutations S_d, acting as $\sigma \cdot L_j = L_{\sigma(j)}$, where $\sigma \in S_d \cong W$.

Expressing the weights as $\lambda = \sum_{i=1}^{d} \lambda_i L_i$, the highest weights of $\mathrm{U}(d)$ form a set

$$\Lambda = \{(\lambda_1, \lambda_2, \ldots, \lambda_d) \in \mathbb{Z}^d | \quad \lambda_j \geq \lambda_{j+1}, 1 \leq j \leq d-1\}, \tag{2.44}$$

which is 1-1 with the irreps of $\mathrm{U}(d)$.

One can check that any irrep of $\mathrm{U}(d)$ restricts to an irrep of $\mathrm{SU}(d)$, while any irrep of $\mathrm{SU}(d)$ extends to the irreps of $\mathrm{U}(d)$ (see e.g. [48]).

However, this mapping is not one-to-one. Since on $\mathfrak{sl}(d, \mathbb{C})$, $\sum_{j=1}^{d} L_j = 0$, any irrep of $\mathrm{U}(d)$ labelled by vectors which differ by a constant vector $(n, n, \ldots, n) \in \mathbb{Z}^d$ corresponds to the same irrep of $\mathrm{SU}(d)$. This $\mathbb{Z}$-gauge can be fixed by subtracting a constant vector with $n = \lambda_d$. Defining $\lambda_j^s \coloneqq \lambda_j - \lambda_d$, the set of highest weights for $\mathrm{SU}(d)$ reads

$$\Lambda_0 = \{(\lambda_1^s, \lambda_2^s, \ldots, \lambda_{d-1}^s) \in \mathbb{Z}_{\geq 0}^{d-1} | \quad \lambda_j^s \geq \lambda_{j+1}^s, 1 \leq j \leq d-2\}, \tag{2.45}$$

which can be interpreted as a description using Young diagrams $\lambda^s = (\lambda_1^s, \lambda_2^s, \ldots, \lambda_{d-1}^s)$.

In terms of the irreps of $\mathrm{PU}(d)$, which consists of equivalence classes of members of $\mathrm{U}(d)$ under the equivalence relation $U \sim e^{i\phi} U$, any irrep of $\mathrm{PU}(d)$ extends to an irrep of $\mathrm{U}(d)$ by choosing it to be constant on equivalence classes. An irrep of $\mathrm{U}(d)$ corresponds to an irrep of $\mathrm{PU}(d)$ exactly when it is constant on equivalence classes. By checking the action

of the center U(1), this happens when the highest weight vector satisfies $\sum_j \lambda_j = 0$. Thus, the irreps of PU(d) are 1-1 with the highest weights

$$\Lambda_P = \{(\lambda_1, \lambda_2, \ldots, \lambda_d) \in \mathbb{Z}^d | \quad \lambda_j \geq \lambda_{j+1},\, 1 \leq j \leq d-1\},\, \sum_j \lambda_j = 0\}. \tag{2.46}$$

Theorem 2.2 (Weyl Character Formula)**.** *Let $\mathfrak{g}$ be a complex semisimple Lie algebra and let π be an irrep of $\mathfrak{g}$ with highest weight λ, then*

$$\chi_\pi(e^H) = \frac{\sum_{w \in W} \det(w) e^{(w \cdot (\lambda + \delta))(H)}}{\prod_{\alpha \in \Phi^+} (e^{\alpha(H)/2} - e^{-\alpha(H)/2})},$$

for all $H \in \mathfrak{h}$ for which the denominator is non-zero. By $\det(w)$ we denote the determinant of a linear map $w : V \to V$ corresponding to $w \in W$.

For root systems of type A_{d-1}, $\det(w)$ is just the sign of the permutation $w \in S_d$. The denominator in Theorem 2.2 is called the Weyl denominator. Using Theorem 2.2 and calculating the limit $\lim_{H \to 0} \chi_\pi(e^H)$, we obtain the following Corollary.

Corollary 2.3 (Weyl dimension formula)**.** *Let $\mathfrak{g}$ be a complex semisimple Lie algebra and let π_λ be an irrep of $\mathfrak{g}$ with highest weight λ. Then,*

$$d_\pi = \frac{\prod_{\alpha \in \Phi^+} (\alpha, \lambda + \delta)}{\prod_{\alpha \in \Phi^+} (\alpha, \delta)}.$$

Theorem 2.2 holds verbatim for compact connected groups G and characters $\chi_\pi(t)$, where $t = e^H$ with $H \in \mathfrak{t}$.

Theorem 2.4 (Weyl Integration formula)**.** *Let G be a compact connected Lie group with a maximal torus $\mathbb{T} \subset G$. For any class function f*

$$\int_G f(g) d\mu(g) = \frac{1}{|W|} \int_T f(t) |\Delta(t)|^2 d\mu_{\mathbb{T}}(t), \tag{2.47}$$

where $d\mu_{\mathbb{T}}$ is the Haar measure on $\mathbb{T}$ and writing $t = \exp(H)$ for $H \in \mathfrak{t}$,

$$\Delta(t) \coloneqq \prod_{\alpha \in \Phi^+} (e^{\alpha(H)/2} - e^{-\alpha(H)/2}),$$

with α being roots relative to $\mathbb{T}$.

Example 2.3 (Volume of a ball in SU(2))**.** *Due to the invariance of the Haar measure, the volume of the ball is independent of the origin. For $SU(2)$ and $H = \mathrm{diag}(i\phi, -i\phi)$ we have $\alpha(H) = 2i\phi$, so $\Delta(t) = 2\sin(\phi)$. Since $|W| = |S_2| = 2$ and $d\mu_{\mathbb{T}}(t) = d\phi/2\pi$,*

$$\int_G f(g)d\mu(g) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin^2(\phi)d\phi. \tag{2.48}$$

In particular,

$$\mu(B_r) = \int_G \mathbb{1}_{B_r}(g)d\mu(g) = \frac{2}{\pi}\int_0^{\phi_r} \sin^2(\phi)d\phi, \tag{2.49}$$

where $\phi_r \coloneqq 2\cdot\arcsin(r/2)$, since $||e-g||_\infty = ||e-t||_\infty = 2\,|\sin(\phi/2)|$.

Thus,

$$\mu(B_r) = \frac{1}{\pi}\left(\phi_r - \frac{\sin(2\cdot\phi_r)}{2}\right), \tag{2.50}$$

Note that the diameter of $SU(d)$ is 2 and $\mu(B_2) = 1$. Taylor series expansion at $r = 0$ yields

$$\mu(B_r) = \frac{1}{6\pi}r^3 + \frac{1}{80\pi}r^5 + \frac{3}{1792\pi}r^7 + \mathcal{O}(r^9) \geq \frac{1}{6\pi}r^3. \tag{2.51}$$

2.2.2 Peter-Weyl theorem and Fourier transform

In this subsection, we discuss chosen parts of the Peter-Weyl theorem [49, 47], a fundamental result in harmonic analysis, and introduce the concept of the Fourier transform on a compact group G. All the representations considered in this subsection are unitary. We start with introducing the necessary definitions.

By $\widehat{G}$ we denote the set of equivalence classes of finite-dimensional unitary irreps of G. By a slight abuse of notation, we will use π to represent the elements of $\hat{G}$.

For an finite-dimensional rep π of G, we define its matrix coefficients as $\pi_{ij}(g) = \langle e_i, \pi(g)e_j, \rangle$, where $\{e_1, e_2, \ldots, e_{d_\pi}\}$ is the orthonormal basis of the rep space V_π. Matrix coefficients $\pi_{ij}(g)$ are smooth functions.

The left-regular rep of G, denoted λ, is the unitary rep on the Hilbert space $L^2(G)$ given by

$$\lambda(h)f(g) \coloneqq f(h^{-1}g). \tag{2.52}$$

The Peter-Weyl theorem states that λ decomposes into an orthogonal direct sum of all unitary irreps of G as follows

$$L^2(G) \cong \widehat{\bigoplus_{\pi \in \widehat{G}}} V_\pi^{\oplus d_\pi}, \tag{2.53}$$

where hat symbol denotes the closure.

The direct sum of all the copies of V_π in $L^2(G)$, $V_\pi^{\oplus d_\pi}$, is also called the π-isotypic component. Moreover, the theorem asserts that the set of elements $\pi_{ij}(g)$ is dense in the space of continuous complex functions $C(G)$, equipped with the supremum norm (hence also in $L^p(G)$ for $1 \leq p < \infty$). The theorem also specifies the following orthonormal basis of $L^2(G)$,

$$\left\{ \sqrt{d_\pi} \pi_{ij} | \pi \in \widehat{G}, 1 \leq i, j \leq d_\pi \right\}. \tag{2.54}$$

In particular, the space of smooth functions is dense in $L^2(G)$. It is informative to see how the irreps V_π are realized inside the function space $L^2(G)$.

We define $\mathcal{E}_\pi^{(j)} \coloneqq \text{span}\{\pi_{ij} | 1 \leq i \leq d_\pi\}$ as the span of the j-th column and span of all the matrix coefficients $\mathcal{E}_\pi \coloneqq \text{span}\{\pi_{ij} | 1 \leq i, j \leq d_\pi\}$. Such spans depend only on the isomorphism classes $\pi \in \hat{G}$. Any $f_\pi \in \mathcal{E}_\pi$ can be expressed as

$$f_\pi(g) = \text{Tr}(A\pi(g)), \tag{2.55}$$

where A is any complex $d_\pi \times d_\pi$ matrix. The spaces $\mathcal{E}_\pi^{(j)}$ are invariant under λ so that the restriction of λ to them form subrepresentations.

One can check that the mappings $\psi_j : V_\pi \to \mathcal{E}_\pi^{(j)}$

$$\psi_j : \sum_{i=1}^{d_\pi} c_i e_i \mapsto \sum_{i=1}^{d_\pi} c_i \pi_{ij}(g), \tag{2.56}$$

where $1 \leq j \leq d_\pi$, are the intertwiners providing the isomorphisms of V_π with the corresponding subspaces of $L^2(G)$. We have

$$\mathcal{E}_\pi = \bigoplus_{j=1}^{d_\pi} \mathcal{E}_\pi^{(j)}. \tag{2.57}$$

and for each $1 \leq j \leq d_\pi$, we can write $\lambda|_{\mathcal{E}_\pi^{(j)}} \cong \pi$ and the whole π-isotypic component corresponds to the restriction to $\mathcal{E}_\pi$. We can then write alternatively

$$L^2(G) = \widehat{\bigoplus_{\pi \in \widehat{G}}} \mathcal{E}_\pi. \tag{2.58}$$

Finally, we note that an analogous result can be obtained for the right-regular rep $(\rho(h)f)(g) \coloneqq f(gh)$, when the spans of the rows are considered instead of $\mathcal{E}_\pi^{(j)}$ and the matrix coefficients are replaced with their complex conjugates.

We now move to the definition of the Fourier coefficients of functions on G.

The Fourier coefficient of a function $f \in L^1(G)$ at a rep π, denoted $\widehat{f}(\pi)$, is the operator in $\mathrm{End}(V_\pi)$ defined via

$$\widehat{f}(\pi) \coloneqq \int_G \pi(g^{-1}) f(g) d\mu(g), \tag{2.59}$$

where by V_π we denote the representation space of irrep π. We turn $\mathrm{End}(V_{\pi_\lambda})$ into a Hilbert space $\mathrm{HS}(V_\pi)$ with inner product $d_\pi \langle \cdot, \cdot \rangle_{HS}$. Such a defined Fourier transform behaves under convolution as expected $\widehat{(f * g)}(\pi) = \widehat{f}(\pi)\widehat{g}(\pi)$.

Then one can show that the Fourier transform is an isomorphism of such Hilbert spaces

$$L^2(G) \cong \widehat{\bigoplus_{\pi \in \widehat{G}}} \mathrm{HS}(V_\pi). \tag{2.60}$$

In light of (2.60), one can think of $\hat{G}$ as the "spectrum" of G with "frequencies" π appearing with multiplicities d_π.

Using the Peter-Weyl theorem, we obtain the Fourier inversion formula for the functions $f \in L^2(G)$,

$$f(g) = \sum_{\pi \in \widehat{G}} d_\pi \mathrm{Tr}\left(\widehat{f}(\pi)\pi(g)\right), \tag{2.61}$$

where the convergence is in L^2-norm. If f is additionally continuous and

$$\sum_{\pi \in \hat{G}} d_\pi^{3/2} ||\hat{f}(\pi)||_{HS} < \infty, \tag{2.62}$$

then the Fourier series converges absolutely and uniformly. If G is additionally connected, the condition (2.62) follows from f being k-differentiable with $k > \frac{1}{2}\dim(G)$. Moreover, in this case, the Fourier coefficients of smooth functions are rapidly decreasing with $\lim_{|\lambda|\to\infty} ||\lambda||^k \hat{f}(\pi_\lambda) = 0$, for every $k \in \mathbb{Z}_+$ (see [50] for details).

As a consequence, we have the Plancherel identity

$$||f||_2^2 = \int_G |f(g)|^2 d\mu(g) = \sum_{\pi\in\widehat{G}} d_\pi ||\widehat{f}(\pi)||_{HS}^2. \tag{2.63}$$

The π-isotypic components $f_\pi(g) = d_\pi \mathrm{Tr}\left(\widehat{f}(\pi)\pi(g)\right)$ can be found via the orthogonal projection $P_\pi : L^2(G) \to \mathcal{E}_\pi$, given by the convolution with $d_\pi\chi_\pi$:

$$f_\pi = P_\pi f \coloneqq f * (d_\pi \chi_\pi). \tag{2.64}$$

This can be seen from the fact that

$$\widehat{\chi_\pi}(\pi') = \frac{\delta_{\pi\pi'}}{d_\pi} I_{d_\pi}, \tag{2.65}$$

which follows e.g. from the orthogonality of the matrix elements (Peter-Weyl theorem).

By applying the Peter-Weyl theorem to (a Hilbert space of) square-integrable central functions $L^2(G)^G$, we obtain a well-known fact that the characters of irreps of G form an orthonormal Hilbert basis for $L^2(G)^G$. The formula (2.61) then simplifies to

$$h = \sum_{\pi\in\hat{G}} \langle f, \chi_\pi\rangle \chi_\pi, \tag{2.66}$$

with $\hat{h}(\pi) = \frac{\langle f,\chi_\pi\rangle}{d_\pi} \mathbb{1}_{d_\pi}$, and (2.61) becomes

$$L^2(G) \cong \widehat{\bigoplus_{\pi\in\widehat{G}}} \mathbb{C}. \tag{2.67}$$

Example 2.4 (Abelian group; Fourier series)**.** *Consider a 1-dimensional torus* $\mathbb{T} \cong \mathrm{U}(1)$*. The irreps are one-dimensional with* $\chi_n(e^{i\phi}) = e^{in\phi}$*,* $n \in \mathbb{Z}$*. Then each* $f \in L^2(\mathbb{T})$ *has a form*

$$f(\phi) = \sum_{n=-k}^{k} \hat{f}(n) e^{in\phi} \tag{2.68}$$

with $\hat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\phi)e^{-in\phi}d\phi$.

Example (2.4) justifies naming (2.59) as Fourier coefficients and the interpretation of the approach presented in this Subsection as nonabelian Fourier analysis. Moreover, by expressing $e^{in\phi} = \cos(n\phi) + i\sin(n\phi)$, one can understand why the functions π_{ij} are often referred to as trigonometric polynomials (even in the nonabelian case).

Writing the central function as $h = \sum_{\pi\in\hat{G}} m_\pi d_\pi \chi_\pi$, the convolution with h acts as a "frequency filter", i.e. for a function $f \in L^2(G)$

$$\widehat{(f * h)}(\pi) = \widehat{(h * f)}(\pi) = m_\pi \hat{f}(\pi), \tag{2.69}$$

uniquely characterized by its frequency response function/multipliers: $m_h : \hat{G} \to \mathbb{C}$, $m_h(\pi) = m_\pi$.

2.2.3 Averaging operators and spectral gap

In this subsection, we introduce the notion of the averaging operator on the compact group G, also known as the mixing operators.

For a rep π and the finite Borel measure ν, we define the operator-valued integral

$$\pi(\nu) \coloneqq \int_G \pi(g)d\nu(g) \in \mathrm{End}(V_\pi) \tag{2.70}$$

by means of Bochner integration [10]. Similarly, we can define $\pi(f)$ for functions.

The averaging operator given by ν, is the operator $T_\nu : L^2(G) \to L^2(G)$,

$$(T_\nu f)(h) \coloneqq \lambda(\nu) = \int_G f(g^{-1}h)d\nu(g). \tag{2.71}$$

We have $T_\nu f = \nu * f$ and $T_{\nu_1 * \nu_2} = T_{\nu_1}T_{\nu_2}$. The operator T_ν is bounded. For ν being the probability measure, $||T_\nu||_\infty = 1$ since the norm is attained on constant functions. In such a case, T_ν is also called the Markov/mixing operator (see also 2.4.4). If additionally ν is symmetric, then T_ν is self-adjoint with the spectrum $\sigma(T_\nu) \subset [-1, 1]$.

[10] If π is finite-dimensional, one can use entry-wise integration.

From now on, we assume ν is a probability measure.

Due to Peter-Weyl, the operator T_ν can be made block-diagonal with finite blocks.

$$T_\nu \cong I_1 \oplus \bigoplus_{\pi \in \hat{G}_0} \pi(\nu)^{\oplus d_\pi}, \tag{2.72}$$

where by $\hat{G}_0$ we denote the set $\hat{G}$ without the trivial rep.

The operator T_μ is the orthogonal projector onto the space of constant functions on G. Hence, in the same basis as (2.72), it reads

$$T_\mu \cong I_1 \oplus \bigoplus_{\pi \in \hat{G}_0} 0_{d_\pi}^{\oplus d_\pi} \tag{2.73}$$

By $L_0^2(G)$ we denote the orthogonal complement of this space (i.e. the space of functions with average 0). We define the (uniform) spectral gap of T_ν as

$$\mathrm{gap}(\nu) \coloneqq 1 - ||T_\nu|_{L_0^2(G)}||_\infty \in [0, 1]. \tag{2.74}$$

We have

$$||T_\nu|_{L_0^2(G)}||_\infty = ||T_\nu - T_\mu||_\infty = \sup_{\pi \in \hat{G}_0} ||\pi(\nu)||_\infty. \tag{2.75}$$

We say ν has spectral gap if $\mathrm{gap}(\nu) > 0$.

Since $T_\mu = T_\mu T_\nu = T_\nu T_\mu$, one can show that

$$||T_{\nu^{*\ell}} - T_\mu||_\infty \leq (1 - \mathrm{gap}(\nu))^\ell \leq e^{-\ell \cdot \mathrm{gap}(\nu)}, \tag{2.76}$$

which essentially says that the existence of a gap implies exponential convergence $T_{\nu^{*\ell}} \to T_\mu$ in operator norm with growing ℓ. Since the Haar measure μ is our model uniform density, this corresponds to a measure ν^* becoming equidistributed.

The spectral gap at a finite scale is defined by restricting T_ν to a finite number of subspaces with "frequencies" bounded by a certain parameter r. For example,

$$\mathrm{gap}_r(\nu) \coloneqq 1 - \max_{0 < ||\lambda|| \leq r} ||\pi_\lambda(\nu)||_\infty, \tag{2.77}$$

where $||\lambda||$ is the chosen norm. Intuitively, such finite-scale information should be sufficient to characterize the equidistribution at scales $\sim 1/r$.

By excluding the space of constant functions, the question about the existence of a (uniform) gap becomes potentially non-trivial. To make it truly non-trivial we need to assume that ν is finitely-supported, i.e. $|\mathrm{supp}(\nu)| < \infty$ (or at least it is discrete or does not have the density; see below) and that ν is universal, i.e. $\overline{\langle \mathrm{supp}(\nu)\rangle} = G$. Indeed, ν is not universal iff there exists an invariant vector (fixed point) in $L^2_0(G)$, so that $||T_\nu||_\infty = 1$ and there is no gap. In fact, $\mathrm{Fix}(T_\nu) \cong L^2(G/H)$, where $H = \overline{\langle \mathrm{supp}(\nu)\rangle}$ and G/H is the set of left cosets of H in G ([51]; see also [52]).

However, even if ν is universal, so for each non-trivial irrep block $||\pi(\nu)||_\infty < 1$ for any $\pi \in \hat{G}$, the gap may be zero. Indeed, what is needed is the uniform bound $||\pi(\nu)||_\infty \leq 1-c$ for some universal constant c. Using the notion of almost invariant vectors, the question about the (uniform) gap can be related to Kazhdan's property (T), hence also to the notions such as expander graphs and strong ergodicity [38, 53].

Example 2.5 (No gap on circle)**.** *Consider a one-dimensional Torus* $\mathbb{T} \cong \mathrm{U}(1)$*. We pick any* $q \notin \mathbb{Q}$ *and choose* $S = \{e^{iq\pi}, e^{-iq\pi}\}$ *as the symmetric universal set with the corresponding symmetric probability measure* $\nu_S = \frac{1}{2}\left(\delta_{e^{iq\pi}} + \delta_{e^{-iq\pi}}\right)$*, so that* $\sigma(T_{\nu_S}) \in [-1, 1]$*. Evaluating on characters* $\chi_n : e^{i\phi} \mapsto e^{in\phi}$*,* $n \in \mathbb{Z}$*, we have*

$$(T_{\nu_S}\chi_n)(e^{i\phi}) = \frac{1}{2}\left(e^{in(\phi+q)} + e^{in(\phi-q)}\right) = \cos(nq\pi)\chi_n(e^{i\phi}). \tag{2.78}$$

We have $|\cos(nq)| < 1$*, but* $\sup_{n\neq 0}|\cos(nq)| = 1$ *so* $||T_{\nu_S}\big|_{L^2_0(\mathbb{T})}||_\infty = 1$*. If* $q \in \mathbb{Q}$ *then* $\mathcal{S}$ *is not universal (a finite set) and the norm is attained on all non-trivial characters* χ_n *with* $nq \in \mathbb{Z}$*.*

However, if ν has a density $k = d\nu/d\mu$, then $T_\nu f = k * f$, so T_ν is a block-diagonal compact operator with finite blocks. Thus, if additionally ν is universal, then it has a spectral gap. Moreover, if the density $k \in L^2(G)$, then T_ν is the Hilbert-Schmidt operator and one can use (2.63) to bound

$$||\pi(\nu)||_\infty \leq ||\pi(\nu)||_{HS} \leq \frac{||k||_{L^2}}{\sqrt{d_\pi}}. \tag{2.79}$$

This simple observation suggests studying the behaviour of finite-scale spectral gaps of finitely-supported measures, such as ν_S and its convolutions, using tools from harmonic

analysis by transforming them into measures with L^2-densities via convolution with appropriately chosen approximate identities/mollifiers.

Focusing now on the case of a finite universal set $\mathcal{S}$, it is interesting to ask what can be said about the upper bounds on $\mathrm{gap}(\nu_{\mathcal{S}})$, or equivalently on lower bounds on $||T_{\nu_{\mathcal{S}}}|_{L^2_0(G)}||_\infty$.

Following [54], one can obtain a bound for continuous groups from the, so-called, almost covering property by the balls of Haar volume $\frac{1}{2|\mathcal{S}|}$,

$$1 - \mathrm{gap}(\nu_{\mathcal{S}}) \geq \frac{1}{2\sqrt{|\mathcal{S}|}}. \tag{2.80}$$

A tighter bound (Kesten bound) can be obtained from random walk considerations (see Section 2.4.4).

2.2.4 Balanced polynomials, t-moment operators and finite-scale spectral gap

In this subsection, we introduce the concept of balanced polynomials and the related notion of t-moment operators. We explain the close relation between t-moment and averaging operators, introduced in subsection (2.2.3). Finally, we present a more natural approach to defining finite-scale spectral gaps.

Thinking of a matrix $U \in \mathrm{U}(d)$ as its image under the defining representation $U \mapsto U$ of $\mathrm{U}(d)$, we denote the matrix elements of the defining representations as $u_{i,j} = (U)_{i,j}$ (and their complex conjugates as $\overline{u}_{i,j}$). With a slight abuse of notation, we denote the defining representation as U. A balanced polynomial of degree t is a homogeneous polynomial with degree t in $u_{i,j}$ and t in $\overline{u}_{i,j}$. We denote the space of all such polynomials of degree t as $\mathcal{H}_t$. Such polynomials are linear combinations of the matrix elements of the representation $U \mapsto (U \otimes \bar{U})^{\otimes t}$ and as such, due to $(U \otimes \bar{U})^{\otimes t} \cong U^{\otimes t} \otimes \bar{U}^{\otimes t}$, every $f_t \in \mathcal{H}_t$ can be expressed as

$$f_t(U) = \mathrm{Tr}\left(A\left(U^{\otimes t} \otimes \bar{U}^{\otimes t}\right)\right), \tag{2.81}$$

for some $d^{2t} \times d^{2t}$ complex matrix A. Due to the global phase invariance, such balanced polynomials are well-defined on $\mathrm{SU}(d)$ and $\mathrm{PU}(d)$.

Since $U \otimes \bar{U} \cong \mathrm{Ad} \oplus \mathbf{1}$, where Ad is the adjoint representation of $\mathrm{U}(d)$, for $s \leq t$ the irreps of $U^{\otimes s} \otimes \bar{U}^{\otimes s}$ appear in $U^{\otimes t} \otimes \bar{U}^{\otimes t}$ and $\mathcal{H}_s \subseteq \mathcal{H}_t$. By Λ_t, we denote the set of highest

weights enumerating all (equivalence classes of) non-trivial irreps appearing in $U^{\otimes t} \otimes \bar{U}^{\otimes t}$. Then we can decompose the tensor product in two ways

$$\mathbf{1}^{\oplus m_0} \oplus \bigoplus_{\lambda \in \Lambda_t} \pi_\lambda^{\oplus m_\lambda} \cong U^{\otimes t} \otimes \bar{U}^{\otimes t} \cong \bigoplus_{N=0}^{t} \binom{t}{N} \mathrm{Ad}^{\otimes N}, \tag{2.82}$$

and m_0, m_λ are some natural numbers. On the right-hand side of (2.82), the binomial coefficient indicates taking the direct sum of that many copies of $\mathrm{Ad}^{\otimes N}$. The left-hand side of (2.82) is the decomposition into irreps.

Then (see [55]),

$$\Lambda_t = \{(\lambda_1, \lambda_2, \ldots, \lambda_d) \in \mathbb{Z}^d | \forall_{1 \leq i \leq d-1} \lambda_i \geq \lambda_{i+1}, |\lambda| = 0, |\lambda_+| \leq t\}, \tag{2.83}$$

where by $|\lambda|$ we denote the sum of elements of λ and by λ_+ we denote a subsequence of positive elements of λ.

We now specialize to the case of $\mathrm{PU}(d)$. For a fixed faithful (i.e. the homomorphism is injective) rep V of a compact group G, every irrep of G is contained in a tensor product $V^{\otimes k} \otimes \bar{V}^{\otimes l}$ [46]. The Ad descends to a well-defined and faithful irrep of $\mathrm{PU}(d)$ and $\mathrm{Ad} \cong \overline{\mathrm{Ad}}$. Hence, every irrep of $\mathrm{PU}(d)$ appears in some power $\mathrm{Ad}^{\otimes N}$ and in consequence in (2.82) for t larger than a certain constant. This observation is consistent with (2.83), since the condition $|\lambda| = 0$ simply restricts the weights of $\mathrm{U}(d)$ to the ones of $\mathrm{PU}(d)$.

For a probability measure ν on $\mathrm{PU}(d)$, we define its t-moment operator , $T_{\nu,t} : \mathcal{H}_t \to \mathcal{H}_t$ as

$$T_{\nu,t} \coloneqq \int_G d\nu(U) U^{t,t}, \tag{2.84}$$

where

$$U^{t,t} \coloneqq U^{\otimes s} \otimes \bar{U}^{\otimes t}, \tag{2.85}$$

and we introduce a quantity

$$\delta(\nu, t) \coloneqq \|T_{\nu,t} - T_{\mu,t}\|_\infty \in [0, 1]. \tag{2.86}$$

Using the decomposition into irreps (2.82), we can express $T_{\nu,t}$ as a block-diagonal matrix

$$T_{\nu,t} \cong \begin{bmatrix} I_{m_0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \pi_{\lambda_1}(\nu) \otimes I_{m_{\lambda_1}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \pi_{\lambda_M}(\nu) \otimes I_{m_{\lambda_M}} \end{bmatrix}, \tag{2.87}$$

for some M, with λ_i going through the set Λ_t in some fixed order. Comparing with the matrix for T_μ in the same basis, the trivial block cancels out and we can write

$$\delta(\nu, t) = \max_{\lambda \in \Lambda_t} ||\pi_\lambda(\nu)||_\infty, \tag{2.88}$$

since the multiplicities are irrelevant. We define $\delta(\nu) \coloneqq \sup_t \delta(\nu, t)$.

The irreps Λ_t are only of real or complex types, i.e. there are no quaternionic representations. Moreover, in practical calculations of (2.88), the set of complex type representations can be restricted due to unnecessary reflections of weights (see e.g. [48]).

On the other hand, using the averaging operator (2.71), we can define the gap at a scale t as

$$\mathrm{gap}(\nu, t) \coloneqq 1 - ||T_\nu\big|_{\mathcal{H}_t} - T_\mu\big|_{\mathcal{H}_t}||_\infty \tag{2.89}$$

However, one can check that the two notions coincide, i.e.

$$\mathrm{gap}(\nu, t) = 1 - \delta(\nu, t). \tag{2.90}$$

Indeed, since $\mathcal{H}_t$ is spanned by the matrix entries of $U \mapsto U^{\otimes t} \otimes \bar{U}^{\otimes t}$, it is also spanned by the matrix entries of the irreps from the decomposition (2.82). Thus, $\mathcal{H}_t$ as a subrepresentation of the left-regular representation λ, decomposes (reducibly) into

$$\mathcal{H}_t = 1 \oplus \bigoplus_{\lambda \in \Lambda_t} \mathcal{E}_{\pi_\lambda}, \tag{2.91}$$

and $T_\nu\big|_{\mathcal{E}_{\pi_\lambda}} \cong \pi_\lambda(\nu) \otimes d_{\pi_\lambda}$.

Hence, we can express $T_\nu\big|_{\mathcal{H}_t}$ in the same block-diagonal form as (2.87) with multiplicities $m_0 = 1$ and $m_{\lambda_i} = d_{\pi_{\lambda_i}}$ for $1 \leq i \leq M$ (see also [41]).

As a consequence,

$$\mathrm{gap}(\nu) = 1 - \delta(\nu). \tag{2.92}$$

2.2.5 Casimir elements and Laplacian

In this subsection, we briefly explain the relationship between the Casimir element of a Lie algebra $\mathfrak{g}$ and the Laplacian on a compact connected group G (see also [49]).

We pick an ad-invariant nondegenerate positive-definite symmetric bilinear form $(\cdot,\cdot)$ on $\mathfrak{g} \simeq T_eG$. This gives rise to a Riemannian bi-invariant metric on G (by translations).

For example, in the case of compact semisimple Lie groups, one may obtain a Riemannian metric by setting $(X, Y) = -B(X, Y)$, where B is the Killing form on $\mathfrak{g}$.

Let $\{X_i\}_{i=1}^n$ be a basis of $\mathfrak{g}$. Then the Casimir operator Ω of $\mathfrak{g}$ is an element of the universal enveloping algebra $U(\mathfrak{g})$ [44] defined as

$$\Omega := \sum_{i,j=1}^{n} X_i X^i = \sum_{i,j=1}^{n} g^{ij} X_i X_j, \tag{2.93}$$

where $g_{ij} = (X_i, X_j)$ and $(g^{ij}) = (g_{ij})^{-1}$ and by X_i we denote the generators of $U(\mathfrak{g})$ corresponding to the basis elements of $\mathfrak{g}$ and we skip the tensor product symbols.

The operator Ω does not depend on the choice of the basis and belongs to the center of $U(\mathfrak{g})$, i.e. the set of the elements of $U(\mathfrak{g})$ that commute with every element of $U(\mathfrak{g})$. Representations of $\mathfrak{g}$ give rise to the reps of the algebra $U(\mathfrak{g})$, which we will denote by the same symbol.

The elements $X \in \mathfrak{g} \simeq T_eG$ correspond to the left (right)-invariant vector fields, X^L (X^R), which are left (right) first order differential operators on G. The action of such fields on smooth functions $f \in \mathcal{C}^\infty(G)$ can be expressed using derived left (right) representations as

$$(d\lambda(X)f)(g) = -(X^L f)(g) = \frac{d}{dt}\Big|_{t=0} f(e^{-tX}g), \quad (d\rho(X)f)(g) = (X^R f)(g) = \frac{d}{dt}\Big|_{t=0} f(ge^{tX}). \tag{2.94}$$

Similarly, the elements of $U(\mathfrak{g})$ correspond to left (right) invariant differential operators on G, which can be expressed as polynomials in X_i^L (X_i^R).

Since Ω is central, one may check that $d\lambda(\Omega) = d\rho(\Omega)$, and we call such an operator the Laplacian on G, denoted $\Delta \coloneqq d\lambda(\Omega) = d\rho(\Omega)$.

The Laplacian is then a second-order differential operator,

$$\Delta = \sum_{i,j} g^{i,j} X_i^L X_j^L = \sum_{i,j} g^{i,j} X_i^R X_j^R, \tag{2.95}$$

which is symmetric and non-positive on $C^\infty(G)$.

The presented construction involving the Casimir element Ω can also be used when $(\cdot,\cdot)$ is indefinite (pseudo-Riemannian manifolds). One may also verify that Δ corresponds to the usual (pseudo-)Laplace-Beltrami operator (2.106) on a manifold G with a bi-invariant (pseudo-)Riemannian metric. For Riemannian manifolds, $(\cdot,\cdot)$ is positive definite and Δ is elliptic, which corresponds to heat diffusion. On the other hand, for Lorentzian manifolds, which are examples of pseudo-Riemannian manifolds with indefinite $(\cdot,\cdot)$, Δ is hyperbolic and corresponds to "wave-like" solutions.

We use the inner product $(\cdot,\cdot)$ on $\mathfrak{g}$ to identify $\mathfrak{h} \cong \mathfrak{h}^*$.

For a rep π of G, we define

$$\Omega_\pi \coloneqq d\pi(\Omega) = \sum_{i,j=1}^{n} g^{ij} d\pi(X_i) d\pi(X_j). \tag{2.96}$$

Then, one may show (see e.g. [50, 49]) that for an irrep π_λ of G, corresponding to the highest weight λ,

$$\Omega_\pi = -(\lambda, \lambda + 2\delta) \cdot I_{d_\pi} \tag{2.97}$$

and regarding Δ as a differential operator on G,

$$\Delta(\pi_\lambda)_{ij} = -(\lambda, \lambda + 2\delta)(\pi_\lambda)_{ij}. \tag{2.98}$$

Indeed, by Schur's lemma and from the fact Ω is central, $\Omega_\pi = c \cdot I_{d_\pi}$. The value of the scalar c can be found by expressing Ω in the Weyl basis $\{E_\alpha, H_i | \quad \alpha \in \Phi, 1 \leq i \leq \dim(T)\}$ of $\mathfrak{g}_\mathbb{C}$, chosen so that $(E_\alpha, E_{-\alpha}) = 1$, $(H_i, H_j) = \delta_{ij}$ and $E_\alpha + E_{-\alpha}, i(E_\alpha - E_{-\alpha}), H_i \in \mathfrak{g}$. The evaluation on the highest weight vector $v \in V_\pi$ yields $\Omega_\pi v = -(\lambda, \lambda + 2\delta)v$, which proves (2.97).

To prove (2.98), for $X \in \mathfrak{g}$ we compute

$$(X^R(\pi_\lambda)_{ij})(g) = \frac{d}{dt}\Big|_{t=0} (\pi_\lambda)_{ij}(ge^{tX}) = \sum_{k=1}^{d_\pi} (\pi_\lambda)_{ik}(g) \frac{d}{dt}\Big|_{t=0} (\pi_\lambda)_{kj}(e^{tX}). \tag{2.99}$$

Then, thinking of $\pi_\lambda(g)$ as matrix of functions, we can express (2.99) compactly as

$$(X^R \pi_\lambda)(g) = \pi_\lambda(g) d\pi_\lambda(X). \tag{2.100}$$

Extending this observation to $\mathrm{U}(\mathfrak{g})$, we apply it to Δ and use (2.97) to obtain

$$(\Delta\pi_\lambda)(g) = \pi_\lambda(g)\Omega_{\pi_\lambda} = -(\lambda, \lambda + 2\delta)\pi_\lambda(g). \tag{2.101}$$

Thus, the π_λ-isotypic components $\mathcal{E}_{\pi_\lambda}$ are the eigenspaces of Δ with eigenvalues $-k_\lambda$, where

$$k_\lambda \coloneqq (\lambda, \lambda + 2\delta). \tag{2.102}$$

In particular,

$$\Delta\chi_\lambda = -k_\lambda \chi_\lambda. \tag{2.103}$$

Finally, we note that for simplicity, we defined the Laplacian acting on $C^\infty(G)$. However, such a Laplacian is essentially self-adjoint and it can be extended to a (unique) self-adjoint Laplacian on $L^2(G)$ with the domain $H^2(G)$, i.e. the L^2 Sobolev space of order 2.

2.2.6 Approximate identities, heat and Fejér kernels

In this subsection, we introduce the notion of approximate identity on a compact (metrizable) group G with normalized Haar measure μ and provide some canonical examples.

Specializing to metric spaces, we define an approximate identity as follows [56].

Definition 2.5. An approximate identity on G with respect to a chosen metric is a net $\{\varphi_\alpha\}_{\alpha>0} \subset L^1(G)$ so that

1. (normalization) $\int_G \varphi_\alpha d\mu = 1$, for any $\alpha > 0$.
2. (bounded L^1-norm) $\sup_{\alpha>0}||\varphi_\alpha|| < \infty$
3. (vanishing outside ε-balls) $\lim_{\alpha\to 0} \int_{G\setminus B_\varepsilon} |\varphi_\alpha| d\mu = 0$, for any $\varepsilon > 0$.

Such conditions guarantee that the convolution $f * \varphi_\alpha$ reconstructs the function f in the limit $\alpha \to 0$, in terms of L^p convergence (with $1 \leq p < \infty$) and uniform or pointwise convergence (for continuous f) - see [56] for a precise statement.

The basic example of an approximate identity on G is the family $\{P_r\}_{r>0}$ of normalized characteristic functions of a ball

$$P_r(g) = \frac{\mathbb{1}_{B_r}(g)}{\mu(B_r)}, \tag{2.104}$$

The vanishing support is often a valuable property; however, P_r is discontinuous across ∂P_r and has an unbounded spectrum with frequency response $m_{P_r}(\pi) = \frac{1}{d_\pi \mu(B_r)} \int_{B_r} \chi_\pi(g) d\mu(g)$.

We now introduce the notion of the heat kernel. Recall that the heat kernel on $\mathbb{R}^N$ is an integral kernel, which is the fundamental solution to the heat equation. Thinking of convolutions, we may express the heat kernel as

$$H_t(x) = \frac{1}{(4\pi t)^{N/2}} e^{-||x||_2^2/4t}, \tag{2.105}$$

where $||\cdot||_2$ is the Euclidean norm on $\mathbb{R}^N$ and $t > 0$,

The heat equation on $\mathbb{R}^N$ can be generalized to other spaces, such as Riemannian manifolds (M, g), by replacing the standard Laplacian with the Laplace-Beltrami operator, which is given in local coordinates by

$$\Delta f = \frac{1}{\sqrt{|g|}} \partial_i \left(\sqrt{|g|} g^{ij} \partial_j f \right). \tag{2.106}$$

The initial value problem for the heat diffusion on G reads

$$\partial_t u(g,t) = \Delta u(g,t), \quad t > 0,\ g \in G \tag{2.107}$$

with some initial datum $u(g,0) = f(g)$, where e.g. $f \in C^\infty(G)$.

Equipping a compact group G with the bi-invariant Riemannian structure, stemming from the negative Killing form, we can use the results from Section 2.2.5. Similarly to the case of a flat N-dimensional torus $\mathbb{R}^N/\mathbb{Z}^N$, by fixing t and calculating the Fourier coefficients, we have $\hat{u}(\lambda,t) = e^{-k_\lambda t}\hat{f}(\lambda)$, where k_λ is the eigenvalue of Δ on the π_λ-isotypic component $\mathcal{E}_\pi$. Fourier inversion can be expressed as

$$u(g,t) = (e^{t\Delta} f)(g) = \sum_\lambda d_\lambda e^{-k_\lambda t} \mathrm{Tr}(\hat{f}(\pi_\lambda)\pi_\lambda(g)), \tag{2.108}$$

where $e^{t\Delta}$ forms so-called heat semigroup, which can be defined by means of the spectral theorem, and for $t > 0$ the series converges uniformly for any $f \in L^2(G)$.

Recalling that convolution with $d_\pi \chi_\pi$ is the projector on the π-isotypic component, we have

$$e^{t\Delta} f = f * H_t, \tag{2.109}$$

where the fundamental solution reads

$$H_t(g) = \sum_{\lambda \in \hat{G}} d_\lambda e^{-k_\lambda t} \chi_\lambda(g), \tag{2.110}$$

and the (negative) eigenvalues are $k_\lambda := (\lambda + 2\delta, \lambda)$. We call a function $H_t(g)$ the heat kernel. The solution given by convolution (2.109) is smooth for all $t > 0$. The Heat kernel family $\{H_t\}_{t>0}$ is a non-negative approximate identity on G. This can be proven, e.g., using the properties of a heat semigroup. Alternatively, quantitative bounds on the vanishing outside ε-balls can be obtained using the Gaussian bounds valid for compact Riemannian manifolds with non-negative Ricci curvature. Namely, for all $\delta > 0$ there exists $a_\delta > 0$ such that

$$H_t(g) \leq \frac{a_\delta}{\mathrm{Vol}_R(B_{R,\sqrt{t}})} \exp\left(-\frac{d_R^2(g,e)}{4(1+\delta)t}\right),\ t>0,\ g\in G, \tag{2.111}$$

where d_R is the bi-invariant Riemannian metric and $\mathrm{Vol}_R(B_{R,\sqrt{t}})$ is the Riemannian volume of the geodesic ball with radius $\sqrt{t}$ [57].

Additionally, note that the solution via the convolution (2.109) can have meaning for f being a finite measure or even a distribution.

Note that the eigenvalues k_λ depend on the choice/normalization of $(\cdot,\cdot)$.

Example 2.6. *For $SU(d)$ with $(X,Y) = -2d\mathrm{Tr}(XY)$, i.e. the negative Killing form, we have*

$$k_\lambda = \frac{1}{2d}\left(\sum_{j=1}^{d}\lambda_j^2 + \sum_{j=1}^{d}(d-2j+1)\lambda_j - \frac{|\lambda|^2}{d}\right). \tag{2.112}$$

and the roots have squared length $1/d$.

In particular, for $d=2$,

$$H_t(\phi) = \sum_{m=0}^{2t}(m+1)e^{-\frac{1}{8}(m^2+2m)t}\chi_m(\phi), \tag{2.113}$$

For the normalized Killing form, $(X,Y) = -\mathrm{Tr}(XY)$, so that the roots have squared length 2 and the right-hand side of (2.112) is rescaled by $2d$ factor.

The heat kernel on compact, semi-simple, simply-connected Lie groups can also be expressed using the Poisson summation form, which resembles (2.105)

$$H_t(\exp(X)) \sim j(\exp(X))^{-1}e^{||\delta||^2 t}(4\pi t)^{-N/2}\sum_{\gamma\in\Gamma}\pi(X^\flat+\gamma)e^{-\frac{1}{4t}||X^\flat+\gamma||^2}, \tag{2.114}$$

where j is the Weyl denominator, δ is a certain group constant and $\pi(X) \coloneqq \prod_{\alpha\in\Phi^+}\alpha(X)$ - see [58] for a precise statement and missing definitions.

The heat kernel is a low-pass filter with an unbounded spectrum but an exponentially decaying frequency response $m_{H_t}(\pi_\lambda) = e^{-k_\lambda t}$, with a tunable effective window in $||\lambda|| \lesssim 1/\sqrt{t}$.

Finally, we note that (shifted) heat kernels on compact Lie groups can be obtained by transporting the Euclidean heat kernel using the Dooley-Wildberger wrapping construction [59].

We now provide another example of a "nice" approximate identity on G. Denoting the norm induced by the Weyl-invariant inner product as $||\lambda||$, wee define the Dirichlet kernel with cutoff N as a sum of projectors (in the sense of convolution)

$$D_T(g) = \sum_{\lambda:\, ||\lambda|| \leq T} d_\lambda \chi_\lambda, \tag{2.115}$$

We can then define the Fejér kernel with cutoff T as an average of Dirichlet kernels

$$F_T(g) \coloneqq \frac{1}{T}\int_0^T D_t(g)dt = \sum_{\lambda: ||\lambda||_1 \leq T} \left(1 - \frac{||\lambda||}{T}\right)_+ d_\lambda \chi_\lambda(g), \tag{2.116}$$

where $(x)_+ \coloneqq \max\{x, 0\}$.

The Fejér kernel family $\{F_T\}_{1/T}$ is a non-negative approximate identity on G with net parameter $1/T$ [60].

The Fejér kernel is a low-pass filter with a bounded spectrum and a decreasing ramp frequency response $m_{F_T}(\pi_\lambda) = \left(1 - \frac{||\lambda||}{T}\right)_+$. The window has a cutoff at $||\lambda|| = T$.

The norm $||\lambda||$ can be replaced with other functions of λ defining the scale. For example, the natural scale can come from the (absolute values) of the eigenvalue of the Laplacian, k_λ, or the 1-norm $||\lambda||_1 = \sum_j |\lambda_j|$.

Finally, we note that the function χ_r

$$\chi_r = \frac{1}{r^{\dim(\mathbb{T})}} \left(\sum_{||\lambda|| \leq r} \chi_\lambda\right)^2 \tag{2.117}$$

introduced in [61] is an example of a non-negative approximate identity $\{\chi_r\}_{1/r}$ normalized between some group constants c, C

$$c \leq \int_G \chi_r(g)d\mu(g) = \frac{\#\{\lambda| \quad ||\lambda|| \leq r\}}{r^{\dim(\mathbb{T})}} \leq C. \tag{2.118}$$

It is a low-pass filter with a cutoff $||\lambda|| \sim r \leq 2r$, and its frequency response depends on the tensor product decomposition rules (Richardson-Littlewood/Clebsch-Gordan coefficients $c^{\lambda}_{\mu,\nu}$)

$$m_{\chi_r}(\pi_\lambda) = \frac{1}{d_\lambda r^{\dim(\mathbb{T})}} \sum_{||\mu||,||\nu|| \leq r} c^{\lambda}_{\mu,\nu}, \tag{2.119}$$

where $\chi_\mu \chi_\nu = \sum_{\lambda \in \hat{G}} c^{\lambda}_{\mu,\nu} \chi_\lambda$.

2.3 Elements of quantum computation and information

2.3.1 Unitary channels and ε-nets

The pure quantum state of a d-dimensional quantum system is a ray in the d-dimensional complex Hilbert space [11] $\mathcal{H} \cong \mathbb{C}^d$, i.e. the equivalence class of unit vectors from $\mathcal{H}$ modulo the global phase. Although formally the pure quantum states are the elements of the projectivisation $\mathbb{P}(\mathcal{H}) \cong \mathbb{CP}^{d-1}$, pure quantum states are typically represented as the normalized vectors from the vector space $\mathcal{H}$, at the cost of introducing a global phase ambiguity. The most general quantum state of such a system is given by a density operator $\rho : \mathcal{H} \to \mathcal{H}$, which is a positive semi-definite (hence Hermitian) operator with trace 1.

The transformations of quantum states are described as mappings of density matrices, known as quantum channels. An important class of quantum processes on $\mathcal{H}$ consists of those represented by the unitary quantum channels. Such channels act as unitary operations when restricted to pure quantum states. Mathematically, the unitary quantum channel is the completely positive trace-preserving map $\mathbf{U}(\rho) = U\rho U^\dagger$, where $\rho : \mathcal{H} \to \mathcal{H}$ is any quantum state and $U \in \mathrm{U}(d)$ is some fixed unitary representative. Since two unitaries U, V which differ by a phase $U = e^{i\phi}V$ define the same unitary channel, the group of all unitary channels $\mathbf{U}(d)$ can be identified with the projective unitary group $\mathrm{PU}(d) = \mathrm{U}(d)/U(1)$, where the canonical projection $\mathrm{U}(d) \to \mathbf{U}(d)$ is mapping the unitary representatives to the corresponding unitary channels $U \mapsto \mathbf{U}$.

Unitary quantum channels can be used to describe the action of ideal (loss-less) quantum gates or communication channels. In this thesis, we consider only the unitary channels, so from now on, by a channel we will understand an element from $\mathrm{PU}(d)$.

[11] People with a mathematical background may protest to name such a finite-dimensional space a Hilbert space; however, in the physics community, such nomenclature is standard.

In practice, one is often interested in the closeness of different channels. Various norms (and induced metrics) can be used to quantify it. A prominent example is the diamond norm $||\cdot||_\diamond$ [62]. We denote the induced metric as $d_\diamond(\mathbf{U},\mathbf{V}) = ||\mathbf{U}-\mathbf{V}||_\diamond$.

The diamond metric has a clear operational meaning in terms of the statistical distinguishability of two channels (e.g. determines the maximal probability of success in a single-shot channel discrimination task).

Example 2.7. *Suppose an agent is given the unknown quantum channel - either* $\mathbf{U}$ *(with probability p) or* $\mathbf{V}$ *(with probability 1-p). The task is for the agent to determine which channel was given. The agent can prepare the quantum state* ρ*, pass it through the channel and measure the output. The maximal probability of agent's guess being correct is*

$$p_{\text{succ}} = \frac{1}{2} + \frac{1}{2} d_\diamond\left(p\,\mathbf{U}, (1-p)\,\mathbf{V}\right). \tag{2.120}$$

The relation between the diamond norm and the norm on $\mathrm{PU}(d)$ introduced in (2.24) is given by (see [41])

$$d_P(\mathbf{U},\mathbf{V}) \leq d_\diamond(\mathbf{U},\mathbf{V}) \leq 2\, d_P(\mathbf{U},\mathbf{V}). \tag{2.121}$$

We say that a finite subset of channels $\mathcal{A} \subset \mathbf{U}(d)$ is an ε-net if for every channel $\mathbf{U} \in \mathbf{U}(d)$, there exists a channel $\mathbf{V} \in \mathcal{A}$, such that $d_P(\mathbf{U},\mathbf{V}) \leq \varepsilon$. In other words, $\mathcal{A}$ represents all the possible channels, up to the error ε. Of course, the definition of ε-net can be applied in any metric space.

2.3.2 Quantum circuit model

A quantum circuit [4, 12] is a universal model for quantum computation, in which quantum information is processed through the application of a finite series of unitary operations, known as quantum logic gates (or simply gates), to a register of qubits, culminating in measurement. The model is used to describe the perfect unitary evolution of pure quantum states of the register, also referred to as state vectors.

The state vectors of an n-qubit quantum register belong to Hilbert space $\mathcal{H} \cong \left(\mathbb{C}^2\right)^{\otimes n}$ with computational orthonormal basis $\{|x\rangle| \quad x \in \{0,1\}^n\}$. The register is initialized in some fiducial quantum state $|\phi\rangle$, typically $|0\rangle^{\otimes n}$. The gates in the circuit can then be represented as a sequence $\mathcal{C} = (U_1, U_2, \ldots, U_\ell)$ of operations realizing the global unitary

operation $U_{\mathcal{C}} = U_\ell \dots U_2 U_1$ applied to the initial state. After the evolution, the projective measurement (possibly on a subset of qubits) is performed, resulting in the classical bitstring with probability governed by the Born rule.

Although we focus on the case of qubits, the quantum circuit model can also be applied to higher-dimensional building blocks, called qudits, for which $\mathcal{H} \cong \left(\mathbb{C}^d\right)^{\otimes n}$.

The state $|\psi\rangle$ is called separable if it can be written as a tensor product of single-qubit states

$$|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle, \quad |\psi_j\rangle \in \mathbb{C}^2. \tag{2.122}$$

The states that are not separable are called entangled. We say a gate is entangling if it can map separable states to entangled states. Prominent examples of such gates include CNOT, CZ and iSWAP.

Typically, each of the gates acts non-trivially only on a proper subset of an n-qubit register, say $k \leq n$ qubits, where usually $k = 1$ or 2. It is well known that an arbitrary global operation on an n-qubit register can be realized using single-qubit gates and two-qubit entangling gates (whose type can be fixed). The action of a k-local gate can be described as the matrix from $\mathrm{SU}(2^k)$, with the global action given by the appropriate tensor product with identity matrices on the remaining qubits (possibly with qubit swaps).

Similarly to a classical computer, whose computation can be described using the classical circuit model, every global quantum operation on a qubit-register can be realized using a universal set of elementary operations. A set of such quantum logic gates is called the universal gate-set $\mathcal{S}$ or, in the context of quantum hardware, the native gate-set. Typically, such a universal gate set consists of single-qubit gates and two-qubit entangling gates [12].

The action of the circuit $\mathcal{C}$ is represented pictorially using "wires" for each qubit and k-local gates as rectangles, circles or crosses through which the wires on which they act pass through. Vertical lines may be used to connect parts of the gate acting on different qubits. The wires represent the worldlines of an individual qubit, with the time passing from left to right as the computation progresses. This is an analogous description to the classical circuits describing the classical logical operations on classical bit registers. The current pictorial description of quantum circuits can be traced back to Penrose diagrams, which are used to describe tensor networks as a special case, adapted to represent the time-ordered unitary evolution. Indeed, quantum circuits can be equivalently defined as

[12] Certain architectures, such as ion-trap-based quantum computers, naturally admit physical gates with larger localities, including global entangling gates.

tensor networks on d-qubits, with k-local gates being the rank $2k$ tensors of type (k, k) and the wire connections between them, which correspond to tensor indices, determine tensor contractions. The topology of the circuit (i.e. the connectivity pattern of the constituent gates and wires) defines constraints on the temporal order of the execution of gates needed to obtain a target global operation. This leads to the concept of a circuit layer, which is a collection of gates, acting on disjoint subsets of qubits, that can be executed in parallel. The circuit can be represented as a sequence of layers, with the number of layers referred to as the circuit depth. The topology of the circuit can be captured, e.g., using the directed acyclic graphs (DAGs), frequently used in quantum compilers.

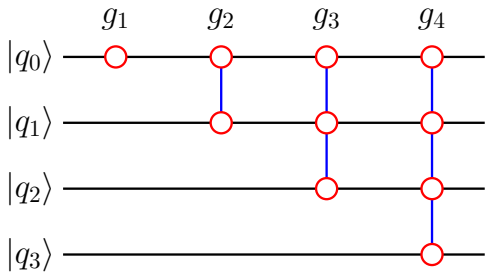


FIGURE 2.1: A 4-qubit quantum register with four gates g_i with increasing locality i, ranging from one-qubit to four-qubit gates. Red circles indicate the qubits on which a given gate acts.

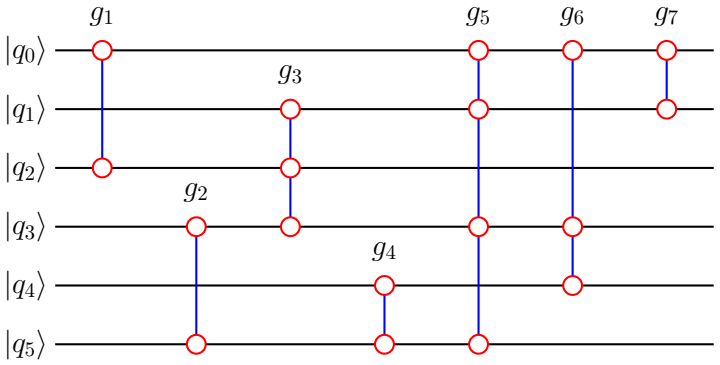


FIGURE 2.2: An example of a quantum circuit on 6 qubits of depth 5.

Importantly, the quantum circuits can be used not only to describe engineered quantum computing processes occurring inside quantum computers, but also to model the unitary dynamics of complex quantum systems [27, 28, 29]. In one direction, quantum circuits can be used to realize an approximate digitized version of continuous Hamiltonian dynamics through the product-formula decompositions (Trotterization). This is the subject of digital quantum simulation, in which a finite sequence of local gates approximates continuous-time

evolution [63, 64]. In another direction, quantum circuits can also be studied as discrete dynamical systems in their own right, without an underlying continuous Hamiltonian, serving as toy models for complex many-body dynamics. Such an approach enables the controlled analysis of key phenomena, including operator spreading, entanglement growth, and correlation functions. Specific settings, such as brickwork architecture circuits with generic (e.g. Haar-random) local gates and dual-unitary circuits, can capture many essential features of chaotic and thermalizing dynamics [29, 65]. On the other hand, random quantum circuits have found applications in describing entanglement growth and hydrodynamic universality classes [28, 66], as well as serving as toy models in high-energy physics. In that context, quantum circuits have been used to model information scrambling [30] and evaporation [27], as well as growth, saturation, and recurrence of quantum complexity in black hole interiors [67, 26, 68].

Finally, in the context of quantum computation and information, the group $\mathrm{SU}(d)$ is typically used as the group of all unitary quantum operations (noiseless quantum gates) acting on n qubits, where $d = 2^n$. However, the group elements are not in 1-1 correspondence with such operations due to the partial ambiguity of the global phase. To lift such ambiguity, the $\mathrm{PU}(d)$ can be used instead, which corresponds to the description via the quantum unitary channels.

2.3.3 Words, complexity and universality

In this subsection, we use the group $\mathbf{U}(d)$; however, all definitions can be adjusted to $\mathrm{SU}(d)$ and $\mathrm{U}(d)$.

In the setting of quantum computing, by $\mathcal{S}$ we denote a chosen set of elementary quantum gates from $\mathbf{U}(d)$, which are the basic unitary channels used to construct more complicated operations. It is convenient to think of circuits built out of $\mathcal{S}$ as the words over the alphabet $\mathcal{S}$. For $\mathcal{S} \subset \mathbf{U}(d)$, we denote the set of all words over an alphabet $\mathcal{S}$ as $\mathcal{S}^*$, the set of all words of length ℓ as

$$\mathcal{S}_\ell \coloneqq \{U_{i_1} U_{i_2} \cdots U_{i_\ell} | \quad U_{i_j} \in \mathcal{S},\, 1 \leq j \leq \ell\}, \tag{2.123}$$

and the set of all words with length at most ℓ by

$$\mathcal{S}_{\leq \ell} \coloneqq \bigcup_{1 \leq \ell' \leq \ell} \mathcal{S}_{\ell'}. \tag{2.124}$$

By circuit length ℓ, we understand the length of the corresponding word.

The smallest number of elements from $\mathcal{S}$ needed to implement a target unitary operation U with precision ε is called the quantum complexity $C_\varepsilon(U, \mathcal{S})$

$$C_\varepsilon(U, \mathcal{S}) \coloneqq \min_\ell\{\ell \,|\quad d_P(U, \mathcal{S}_\ell) \leq \varepsilon\} \tag{2.125}$$

Although the group of all unitary quantum channels on a smallest quantum register, i.e. a single-qubit one, corresponds to a group PU(2), it is informative to consider an even simpler example.

Example 2.8. *Consider a unitary group* U(1). *By* $R(\phi) = e^{i\phi}$ *we denote the rotation about the fixed angle* ϕ. *Then a set* $\mathcal{S} = \{R(\phi), 1\}$ *is universal for any irrational* ϕ/π. *Consider a target operation* $U = R(\phi)^\ell$, *which corresponds to the circuits* $\mathcal{S}_\ell$. *It is clear that for any accuracy* $\varepsilon > 0$, *the complexity of* U *can be upper bounded* $C_\varepsilon(U, \mathcal{S}) \leq \ell$, *however* $C_\varepsilon(R(\phi)^{\ell^*}, \mathcal{S}) = 0$ *for some large enough* ℓ^*.

This trivial example shows that the operations U realized by very long circuits (made of non-trivial operations) can have minimal complexity due to some complexity shortcuts. In fact, the computation of $C_\varepsilon(U, \mathcal{S})$ is known to be a very challenging problem, possibly residing at higher levels of the so-called polynomial hierarchy - specifically, above the NP complexity class. Checking whether a quantum circuit is almost equivalent to identity is known to be QMA-complete [69].

We introduce the uniform bound on the complexity $\ell(\mathcal{S}, \varepsilon)$

$$\ell(\mathcal{S}, \varepsilon) \coloneqq \sup_U C_\varepsilon(U, \mathcal{S}), \tag{2.126}$$

which is the minimal circuit length for which every operation U can be ε-approximated. The quantity $\ell(\mathcal{S}, \varepsilon)$ can be also reframed as the diameter of $\mathbf{U}(d)$ at scale ε [61].

Although the computation of $C_\varepsilon(U, \mathcal{S})$ is very hard, the pessimistic case $\ell(\mathcal{S}, \varepsilon)$ can be upper-bounded (see Section 2.4.3).

We say that $\mathcal{S}$ is universal (in $\mathbf{U}(d)$) if for any precision $\varepsilon > 0$, there exists a depth ℓ_ε such that every quantum operation from $\mathbf{U}(d)$ can be realized by a quantum circuit composed from the gates in $\mathcal{S}$ of depth at most ℓ_ε.

Using the language of ε-nets, the set $\mathcal{S}$ is universal if for any $\varepsilon > 0$ there exists ℓ_ε, such that $\mathcal{S}_{\leq \ell_\varepsilon}$ is an ε-net in $\mathbf{U}(d)$ [13]. Alternatively, using the language of topological groups, the set $\mathcal{S}$ is universal if it topologically generates $\mathbf{U}(d)$, i.e. $\overline{\langle \mathcal{S} \rangle} = \mathbf{U}(d)$.

For a universal gate-sets $\mathcal{S}$ and any $\varepsilon > 0$, the complexity $C_\varepsilon(U, \mathcal{S})$ is finite, for any target operation U on n-qubits. Hence, we can say $\mathcal{S}$ is universal if $\ell(\mathcal{S}, \varepsilon) < \infty$ for any $\varepsilon > 0$.

(a) The set $\mathcal{S}_\ell$ (red dots) does not realize the target operation (black dot) up to precision ε.

(b) The set $\mathcal{S}_\ell$ (red dots) realizes the target operation (black dot) up to precision ε.

FIGURE 2.3: The set $\mathcal{S}_\ell$ filling the group as ℓ increases.

In contrast to classical logical circuits, quantum circuits built from a finite set $\mathcal{S}$ can implement arbitrary n-qubit unitary operations only approximately, up to some precision ε (in a suitable metric).

We define a n-qubit Clifford $\mathcal{C}_n$ group as the normalizer of the n-qubit Pauli group

$$\mathcal{P}_n \coloneqq \{e^{i\phi\frac{\pi}{2}}\sigma_{j_1} \otimes \sigma_{j_2} \otimes \ldots \otimes \sigma_{j_n} | \quad \phi,\, j_k \in \{0,1,2,3\}\}, \tag{2.127}$$

up to the global phase

$$\mathcal{C}_n \coloneqq N_{\mathrm{U}(2^n)}(\mathcal{P}_n) \Big/ \mathrm{U}(1), \tag{2.128}$$

[13] If $e \in \mathcal{S}$, then $\mathcal{S}_{\ell_1} \subseteq \mathcal{S}_{\ell_2}$ for $\ell_1 \leq \ell_2$ so that $\mathcal{S}_{\leq \ell} = \mathcal{S}_\ell$.

where σ_j denotes the Pauli matrices (including identity σ_0). The Pauli matrices form a basis for density matrices as well as the unitary operations for a single-qubit. The group $\mathcal{C}_n$ is finite with $2^{n^2+2n}\prod_{j=1}^{n}(4^j-1)$ elements.

Introducing the gates

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}, \quad T = \begin{pmatrix}1 & 0\\ 0 & e^{i\pi/4}\end{pmatrix}, \tag{2.129}$$

the set $\{H, S\}$, where $S = T^2$, generates a single-qubit Clifford group $\mathcal{C}_1 = \langle H, S\rangle$. The $\mathcal{C}_n$ group can be generated by the $\mathcal{C}_1$ generators for each qubit and entangling gates, such as CNOT or CZ, acting on pairs of qubits (nearest neighbours are sufficient).

The Clifford circuits (i.e. circuits made of Clifford gates) are known to be efficiently simulable on classical computers. However, the Clifford gates are not universal. Surprisingly, expanding the Clifford-generating gate set with any non-Clifford gate yields a universal gate set. Gate sets that are constructed this way by adding a T gate are called Clifford+T.

Example 2.9. *The set $\mathcal{S} = \{H, T\}$ is universal for a single qubit. Notice that $T \notin \mathcal{C}_1$. It can be extended to a universal n-qubit gate set by applying it for each qubit and adding entangling operations (as for the Clifford group).*

It is known that the necessary condition for the universality of a finite $\mathcal{S}$ is the equality of the centralizers of the sets obtained by evaluating (2.85) on the whole group and on the set $\mathcal{S}$ for $t = 1$. The set $\mathcal{S}$ is then universal if $\langle\mathcal{S}\rangle$ is infinite, which can be checked by verifying it forms an $\frac{1}{2\sqrt{2}}$-net in Hilbert-Schmidt metric (for SU(d)) [70, 71]. In [52], the authors show that the infinite $\langle\mathcal{S}\rangle$ condition can be replaced with a statement about $\mathcal{S}$ forming δ-approximate t-designs (see 2.3.5) with $t = 6$ (for $d = 2$) and $t = 4$ (for $d \geq 3$). They also provide an alternative universality condition, which reduces to calculating the dimensions of centralizers for $t = 2$ or $t = 3$ and can be verified using a computationally efficient algorithm.

Finally, we note that universality is a generic property, i.e. for the fixed cardinality k of gate-sets $\mathcal{S}$, the set of universal gate-sets is a Zariski-open set in $\mathrm{SU}(d)^{\times k}$. Essentially, this means that almost all finite gate-sets are universal, and the ones that are not can be expressed as a system of polynomial equations in the matrix entries and their conjugates [72].

2.3.4 Efficiency and cost

The circuit length $\ell(\mathcal{S}, \varepsilon)$ can be understood as an absolute measure of the computational efficiency of the chosen set of operations $\mathcal{S}$ at the scale of ε-approximations. Since, in practice, the required ε can be lower-bounded, $\ell(\mathcal{S}, \varepsilon)$ fully characterizes the efficiency of $\mathcal{S}$. So far, we treated the efficiency rather abstractly, understanding it as the computational efficiency of some chosen generators, where each generator is counted with the same weight of 1, according to the definition of the quantum complexity (2.125). The quantum complexity may be a relevant measure in some models, e.g., random quantum circuit models of quantum scrambling.

However, to obtain measures that are of practical importance to quantum hardware, we need to take into account the implementation details. This leads us to the notion of a cost.

Quantum compilation [16, 73, 4] is a process that has two main objectives. The first objective is to approximate the target quantum circuit specified by a high-level, hardware-agnostic representation used by quantum programmers to the form expressible by the native physical gate set available for a specific quantum computer. The second objective of the compiler is circuit optimization, which, loosely speaking, boils down to the reduction of the cost of execution of quantum circuits, such as the depth of the circuit or the number of costly quantum gates used.

The cost of a quantum circuit is a rather general concept that can have both spatial and temporal components, depending on the specific context and architecture. The most basic distinction is between the current hardware, which is almost exclusively the noisy intermediate-scale quantum (NISQ) devices, and the fault-tolerant quantum computers, which are slowly emerging.

The NISQ devices are characterized by a modest number of noisy qubits. The number of qubits in current machines rarely exceeds 10^3, and the fidelity of the entangling operations, which are the noisiest, is usually below 99.9%. The NISQ machines do not support fault-tolerant quantum error correction schemes, so the depth of circuits that can be run on such computers is severely limited due to the accumulation of errors. Indeed, it is easy to calculate that the number of entangling gates is limited to hundreds or a few thousand. Hence, in the NISQ setting, the reduction of the lengths of circuits, their depth, and the number of noisy entangling gates is of utmost practical importance [13, 14, 15]. For example, if the CNOT gates are used as entangling gates, the number of CNOT gates (CNOT-count) is a widely used proxy for the overall cost of a circuit. The NISQ devices

often use parametrized quantum gates (at least for single qubits), so that the universal set $\mathcal{S}$ has infinitely many elements that can vary continuously. In such a case, an exact synthesis is possible through decomposition using Euler angles (for single qubits) and the Cartan/KAK decomposition (for 2 qubits) [4].

On the other hand, in the fault-tolerant regime, quantum error correction schemes enable reducing the errors of logical operations to arbitrarily small values, at the cost of certain spatial and temporal overheads in physical operations/qubits. In the fault-tolerant architectures, due to Eastin-Knill theorem, the cost mainly comes from the implementation of so-called non-transversal gates [74, 75, 76, 77, 78, 79]. For example, in the Clifford+T setting, the quantum error correction based on 2D color and surface codes, the T-count/T-depth is a commonly used proxy for the overall cost of the circuit. This is because the T-gate is then non-transversal and requires resource-heavy operations, such as magic state distillation and injection, to be implemented fault-tolerantly. On the other hand, the Clifford operations are relatively cheap. In such a scenario, reducing the T-count/T-depth leads to improvements in time and number of qubits required for fault-tolerant execution [18, 19, 20, 21, 22, 23, 24].

2.3.5 Approximate unitary t-designs

In this section, we introduce the notion of the (approximate) unitary t-design. Unitary t-designs are constructions ubiquitous in quantum information and computation. They found applications in quantum information protocols [80, 81], randomised benchmarking [82, 83], process tomography [84], shadow estimation [85], derandomisation of probabilistic constructions [86], decoupling [87], entanglement detection [88], quantum state discrimination [89], efficient quantum measurements [90], unitary codes [91] and the estimation of the properties of quantum systems [92]. Their connection with pseudo-random quantum circuits [93] makes them applicable to equilibration of quantum systems [94, 93], quantum metrology with random bosonic states [95], quantum complexity and information scrambling in black holes [31, 96, 97, 98, 99, 100]. Additionally, due to their anti-concentration property [101, 102], they have also been applied to the study of quantum speed-ups [103, 104, 105].

A (unitary) t-design is a probability measure ν on $\mathbf{U}(d)$ which mimics the Haar measure when applied to averaging the balanced polynomials $f_t(U) \in \mathcal{H}_t$,

$$\int_{\mathbf{U}(d)} d\nu(U) f_t(U) = \int_{\mathbf{U}(d)} d\mu(U) f_t(U). \tag{2.130}$$

Similarly, one can define the related notion of t-designs for quantum states, called the spherical t-designs. Unitary t-designs in $\mathrm{U}(d)$ are known to exist for any t and d, and many explicit constructions are known [106]. Finally, the construction of certain special cases, such as group designs and their generalizations, is an active area of research [107].

If the measure ν is supported on a finite number of points $\{\nu_i, U_i\}$, left-hand side integral of (2.130) becomes a sum

$$\sum_{U_i \in \mathcal{S}} \nu_i f_t(U_i) = \int_{\mathbf{U}(d)} d\mu(U) f_t(U), \tag{2.131}$$

where $\mathcal{S}$ denotes a finite set supporting the measure ν. Representing f_t as in (2.81), we can bound the difference between both sides of (2.130) using (2.22)

$$\left| \int d\nu_{\mathcal{S}}(U) f_t(U) - \int d\mu(U) f_t(U) \right| = |\mathrm{Tr}(A(T_{\nu_{\mathcal{S}},t} - T_{\mu,t}))| \leq ||A||_1 \cdot \delta(\nu_{\mathcal{S}}, t), \tag{2.132}$$

where $||A||_1$ is the trace norm (Schatten 1-norm).

Hence, using the notion of t-moment operators (2.84), we can relax the condition (2.130) and say that ν is a δ-approximate t-design if $\delta(\nu, t) < 1$ (2.86). If $\delta(\nu, t) = 0$, we call ν an (exact) t-design. We say $\mathcal{S}$ is a (δ-approximate) t-design, if it supports a probability measure which is a (δ-approximate) t-design.

Although intuitively, δ-approximate t-designs and ε-nets are related, the quantitative relations between them were rigorously studied only recently. In [41], the authors show for the group $\mathbf{U}(d)$ that $\mathcal{S}$ is an ε-net if $\mathcal{S}$ is a δ-approximate t-design with

$$t \simeq \frac{d^{5/2}}{\varepsilon}, \quad \delta \simeq \left(\frac{\varepsilon^{3/2}}{d} \right)^{d^2} \tag{2.133}$$

(see [41] for precise formulas with explicit constants). The authors of [41] additionally prove that t has to grow at least as fast as d^2/ε. Such relations can be used to obtain SKL theorems discussed in Section 2.4.3.

Finally, in the opposite way, (discrete) ε-nets in $\mathbf{U}(d)$ are δ-approximate t-designs with $\delta = 2\varepsilon t$. The authors of [41] prove this by constructing the discrete measure $\{\nu_i, U_i\}$ which is, in general, non-uniform. In general, it is interesting to ask what can be said solely about the uniform δ-approximate t-designs.

The proof from [41] relies on the construction of the polynomial approximate identity on $\mathbf{U}(d)$ via the periodisation and spectral truncation of Gaussians. However, the argument can be repeated for other polynomial or approximate identities. Below, we provide the sketch of the proof.

Let $\{\varphi_t\}_{1/t}$ be an approximate identity with net parameter $1/t$, so that $\varphi_t \in \mathcal{H}_t$, and fix any $\mathbf{V} \in \mathbf{U}(d)$. If we assume ν is an exact t-design, then T_ν acts as the projector onto constant functions, so

$$\int_{B_\varepsilon(\mathbf{V})} d\mu(\mathbf{U})(T_\nu\varphi_t)(\mathbf{U}) = \mu(B_\varepsilon(\mathbf{V}))). \tag{2.134}$$

Intuitively, $T_\nu\varphi_t = \nu * \varphi_t$ is a density obtained by smearing φ_t using ν. If the support of ν is not an ε-net, then we can pick a point $\mathbf{V}_0$, such that $d_P(\mathbf{V}_0, \mathrm{supp}(\nu)) \geq \varepsilon$. Then applying (2.134) to a ball $B_{\varepsilon/2}(\mathbf{V}_0)$ we have that the density mass in that ball is $\mu(B_{\varepsilon/2}(\mathbf{V}_0)))$. However, since φ_t is an approximate identity, as $t \to \infty$, the density will concentrate around the support of ν, e.g., for finitely-supported ν we have the Dirac delta-like picture. Thus,

$$\int_{B_{\varepsilon/2}(\mathbf{V}_0)} d\mu(\mathbf{U})(T_\nu\varphi_t)(\mathbf{U}) \to 0, \tag{2.135}$$

as $t \to \infty$. This contradiction proves that the support of ν forms an ε-net. To make this statement quantitative, we need to be able to bound the rate at which (2.135) vanishes and relate it to $\mu(B_{\varepsilon/2}(\mathbf{V}_0))$. This can be achieved using the bounds on φ_t for the vanishing outside ε-balls (see Property 3 from Definition 2.5).

Finally, if ν is a δ-approximate t-design, the difference between both sides of (2.134) can be bounded using the Cauchy-Schwartz inequality

$$|\langle 1 - T_\nu\varphi_t, \mathbb{1}_{B_\varepsilon(\mathbf{V})}\rangle| \leq ||1 - T_\nu\varphi_t||_2\sqrt{\mu(B_\varepsilon(\mathbf{V}))}, \tag{2.136}$$

and since

$$||1 - T_\nu\varphi_t||_2 = ||(T_\mu - T_\nu)\varphi_t||_2 \leq ||(T_\mu - T_\nu)\big|_{\mathcal{H}_t}||_\infty||\varphi_t||_2 \leq \delta||\varphi_t||_2, \tag{2.137}$$

we obtain

$$\left| \int_{B_\varepsilon(\mathbf{V})} d\mu(\mathbf{U})(T_\nu \varphi_t)(\mathbf{U}) - \mu(B_\varepsilon(\mathbf{V})) \right| \leq \delta \sqrt{\mu(B_\varepsilon(\mathbf{V}))} ||\varphi_t||_2. \tag{2.138}$$

Hence, assuming that $\delta \sim \frac{\sqrt{\mu(B_\varepsilon(\mathbf{V}))}}{||\varphi_t||_2}$, the situation is analogous to t-designs.

2.4 State of the art and research problems

2.4.1 Spectral gap, efficient and optimal gate sets

As already mentioned, the spectral gap problem has a historical connection to Kazhdan's Property (T) [38, 53].

The study of $\text{gap}(\nu_\mathcal{S}) = 1 - \delta(\nu_\mathcal{S})$ for finite universal $\mathcal{S}$ is a hard problem, as such a gap cannot be directly calculated. However, the existence of gap has been proven for specific types of gates; it is known that finite universal sets $\mathcal{S} \subset \text{SU}(d)$ with gates consisting of algebraic entries have the gap [108, 109]. This result was later generalized to any compact simple Lie group [110]. However, the question of whether the algebraic condition can be lifted is an open problem. Even the following question remains unanswered.

Example 2.10 (Open problem)**.** *Consider* $\mathcal{S} = \{U, V, U^{-1}, V^{-1}\}$ *in* $\text{SU}(d)$*,* $d \geq 2$*. Does* $\nu_\mathcal{S}$ *have a spectral gap for almost every pair* (U, V)*?*

There are many known gate sets in $\text{SU}(d)$ that exhibit a gap. Such gates are called (computationally) efficient as the existence of a gap implies the optimal asymptotic efficiency $\ell(\mathcal{S}, \varepsilon) = \Theta\left(\log(1/\varepsilon)\right)$ [111].

Moreover, there are known examples of universal gate sets that are not only efficient but also optimal in the sense that $\delta(\nu_\mathcal{S})$ is as small as possible. The optimal value is given by the Kesten bound

$$\delta(\nu_\mathcal{S}) \geq \frac{2\sqrt{|\mathcal{S}| - 1}}{|\mathcal{S}|}, \tag{2.139}$$

(c.f. with the weaker bound (2.80) and see Section (2.4.4) for explanation). A prominent example is the optimal family of single-qubit gates constructed using quaternion algebras/Hecke constructions with $|\mathcal{S}| = p + 1$ for $p \equiv 1 \bmod 4$ [112, 113]. For $p = 5$, they are known as V-gates [111]. Finally, some commonly used one-qubit gate sets are known to be

optimal [34, 114, 115, 116, 117], due to the number-theoretic arguments. More recently, so-called Golden and Super-Golden single-qubit gates have been proposed together with their multi-qubit generalizations [54, 118]. In principle, Super-golden gate sets can approximate generic two-qubit unitaries with asymptotically fewer costly T-type operations, compared to the standard Clifford+T gate sets.

Despite the lack of proof, nowadays it is widely believed that the generic n-qubit gate-sets $\mathcal{S}$ are efficient. This conjecture can be traced back to Sarnak's conjecture for the existence of a gap for universal discrete sets $\mathcal{S} \subset \mathrm{SU}(2)$ [119]. However, the quantitative methods to bound and compare the efficiency of various gate-sets $\mathcal{S}$ were not well-developed.

2.4.2 Decay of the spectral gap

An important property of a finite-scale spectral gap (2.77) is its poly-logarithmic decay as r increases. Theorem 2.6 provides such a state-of-the-art result (Theorem 6 from [40]).

Theorem 2.6. *For every semi-simple compact connected Lie group G, there are numbers c, r_0 and A such that the following holds. Let ν be an arbitrary probability measure on G. Then*

$$\mathrm{gap}_r(\nu) > c \cdot \mathrm{gap}_{r_0}(\nu) \cdot \log^{-A} r. \tag{2.140}$$

For simple groups, the value of A can be found in Table 2.2. For semi-simple groups A is the maximum of the corresponding values over all simple quotients of G. In particular, $A \leq 2$ for all groups.

TABLE 2.2: The value of $A(G)$ in terms of the Dynkin diagram.

A_n	B_n	C_n	D_n	E_6	E_7	E_8	F_4	G_2
$1+\frac{2}{n+1}$	$1+\frac{1}{n}$	$1+\frac{1}{n}$	$1+\frac{1}{n-1}$	$\frac{7}{6}$	$\frac{10}{9}$	$\frac{16}{15}$	$\frac{7}{6}$	$\frac{4}{3}$

However, note that Theorem 2.6 does not specify the constant c and (especially) r_0, which limits its practical applications.

2.4.3 Solovay-Kitaev-like theorems

In this section, we introduce the notion of the Solovay-Kitaev-like (SKL) theorem as a theorem that provides a poly-logarithmic upper bound on the efficiency $\ell(\mathcal{S}, \varepsilon)$. We start with recalling the definition of the seminal Solovay-Kitaev (SK) theorem.

Theorem 2.7 (Solovay-Kitaev)**.** *Assume $\mathcal{S}$ is a finite universal and symmetric set in* $\mathrm{SU}(d)$*. For every $U \in \mathrm{SU}(d)$, $\varepsilon > 0$ and*

$$\ell > A(\mathcal{S}) \cdot \log^c\left(\frac{1}{\varepsilon}\right)$$

there is $U_\ell \in \mathcal{S}_{\leq \ell}$ such that $||U - U_\ell||_\infty < \varepsilon$ (i.e. $\mathcal{S}_{\leq \ell}$ is an ε-net in $\mathrm{SU}(d)$*), where $c = 3+\alpha$ for any $\alpha > 0$.*

Using the notion of efficiency, the SK theorem implies $\ell(\mathcal{S}, \varepsilon) = \mathcal{O}(\log^{3+\alpha}(1/\varepsilon))$, for any $\alpha > 0$. In other words, all universal gate sets are rather efficient. Moreover, the proof of the SK theorem is constructive, so that an (efficient) [14] algorithm exists that can find U_ℓ. The SK algorithm became the cornerstone of modern quantum compilation. and many similar upper bounds on $\ell(\mathcal{S}, \varepsilon)$ were provided since its introduction. We refer to all such poly-logarithmic bounds $\ell(\mathcal{S}, \varepsilon) = \mathcal{O}(\mathrm{Poly}(\log(1/\varepsilon))$ (constructive and non-constructive) as Solovay-Kitaev-like (SKL) theorems.

In terms of constructive/algorithmic SKL theorems, the cubic scaling in the SK algorithm was broken recently in [33], with the exponent of $\log(1/\varepsilon)$ lowered to $\log_\phi(2) \approx 1.44042$, where $\phi = \frac{1+\sqrt{5}}{2}$ is the golden ratio. Similarly as in the SK theorem, the construction assumes that $\mathcal{S}$ is finite and inverse-closed. On the other hand, in [120] the authors provided the version of the SK algorithm without the symmetricity condition on $\mathcal{S}$, with $\ell(\mathcal{S}, \varepsilon) = \mathcal{O}(\log^{\gamma_d}(1/\varepsilon))$ and $\gamma_d = \frac{\log(8d^2+1)}{\log(3/2)} = \Theta(\log(d))$.

Ultimately, all poly-logarithmic upper bounds on $\ell(\mathcal{S}, \varepsilon)$ with $\log(1/\varepsilon)$ dependence on ε are asymptotically tight. Indeed, this is a consequence of a simple volume counting argument [111] which provides the lower bound $\ell(\mathcal{S}, \varepsilon) = \Omega(\log(1/\varepsilon))$.

In fact, the existence of a gap for $\mathcal{S}$ implies such optimal computational efficiency. For SU(d) we can formulate the following SKL theorem [111]

$$\ell(\mathcal{S}, \varepsilon) \leq \frac{(d^2-1)}{\log(1/\delta(\nu_{\mathcal{S}}))} \log\left(\frac{2}{a_v \varepsilon}\right), \tag{2.141}$$

where a_v is a constant (2.26).

Crucially, the existence of a gap is not necessary to obtain useful bounds on $\ell(\mathcal{S}, \varepsilon)$. Indeed, one can formulate the SKL theorems using the knowledge of a finite-scale spectral gap, which is sufficient in practice as it corresponds to studying the efficiency at a certain finite

[14] At least for efficiently calculable gates from $\mathcal{S}$.

precision ε. Moreover, for small values of t and d, the finite-scale spectral gap can be computed numerically. Such a finite-scale approach was studied in [61], where, aside from the spectral gap decay results from Section (2.4.2), the first part of Lemma 5 provides a SKL theorem for a compact connected semisimple Lie group G, of the following form

$$\ell(\mathcal{S}, \varepsilon) \leq \frac{C}{\mathrm{gap}_{r(\varepsilon)}(\nu_{\mathcal{S}})} \log\left(\frac{1}{\varepsilon}\right), \tag{2.142}$$

with $r(\varepsilon) = \frac{D}{\varepsilon^{2\dim(G)+2}}$ and C and D are positive group constants. Notice that the denominator of (2.142) is some function of ε with a priori unknown growth in $1/\varepsilon$.

Additionally, the second part of Lemma 5 from [61] provides the estimation of the finite-scale spectral gap from the diameter based on the argument of Jean Bourgain

$$\mathrm{gap}_r(\nu_{\mathcal{S}}) \leq \frac{1}{|\mathcal{S}|\ell^2(\mathcal{S}, \varepsilon(r))}, \tag{2.143}$$

with $\varepsilon(r) = 1/(Cr)^C$.

An improved SKL theorem akin to (2.142) was proved for $\mathbf{U}(d)$ in [41] and states that [15]

$$\ell(\mathcal{S}, \varepsilon) \lesssim \frac{d^2 - 1}{\log\left(1/\delta(\nu_{\mathcal{S}}, t(\varepsilon))\right)} \log\left(\frac{1}{\varepsilon}\right), \tag{2.144}$$

where $t(\varepsilon) \simeq \frac{d^{5/2}}{\varepsilon}$ is the bound stemming from the δ-approximate t-design and ε-net correspondence (see Section 2.3.5 and formula (2.133)). Note that for $\mathrm{SU}(d)$, $\dim(G) = d^2 - 1$, so that $r(\varepsilon)$ from (2.142) is $\frac{D}{\varepsilon^{2(d^2+1)+2}}$. In fact, (2.144) holds for an arbitrary probability measure ν, whose support generates $\mathbf{U}(d)$.

Thus, we can say that $\delta(\nu_{\mathcal{S}}, t(\varepsilon))$ determines the upper bounds on the efficiency of $\mathcal{S}$ in $\mathbf{U}(d)$ at the level of ε approximations, where the necessary scale $t(\varepsilon)$ is at least $\simeq d^2/\varepsilon$ and $\simeq d^{5/2}/\varepsilon$ is sufficient.

The proofs of both SLK theorems rely on the application of averaging operators to appropriate approximate identities with spectra bounded by the spectral gap scale [41, 61].

By combining the poly-logarithmic decay results with the SKL theorems based on finite-scale spectral gap (such as (2.142) and (2.144)), one can obtain an SKL theorem with

[15] Original Proposition 2 in [41] has $1 - \delta(\nu_{\mathcal{S}}, t) = \mathrm{gap}_t(\nu_{\mathcal{S}})$ in the denominator instead of $\log\left(1/\delta(\nu_{\mathcal{S}}, t)\right)$ due to unnecessary bounding.

clear ε-dependence. In particular, the decay (2.140) gives $\ell(\mathcal{S},\varepsilon) \propto \log^{A+1}(1/\varepsilon)$, which for $\mathrm{SU}(d)$ yields the exponent $2+\frac{2}{d}$, that is better than the one from the SK theorem for $d \geq 3$ [16].

Of course, one can apply the SKL theorems with explicit ε-dependence to obtain the bound on the finite-scale spectral gap of $\nu_{\mathcal{S}}$ via (2.143). However, it is clear that (2.140) cannot be improved this way.

2.4.4 Random walks, circuits and t-designs

We start by briefly explaining the connection between the averaging operators and random walks. Let G be a group and let $X_1, X_2, \ldots,$ be a sequence of i.i.d. random elements of G with law provided by the probability measure ν. For simplicity, we assume ν is symmetric. The random walk in G is the sequence of random elements

$$Y_1 = X_1, \quad Y_2 = X_2 X_1, \ldots, \quad Y_\ell = X_\ell X_{\ell-1} \ldots X_1, \ldots \tag{2.145}$$

We may study how fast the distribution of Y_ℓ, which has the law $\nu^{*\ell}$, converges to Haar measure by considering the averaging operator T_ν

$$(T_\nu f)(g) = (\nu * f)(g) = \mathbb{E}_{X_1 \sim \nu}(f(X_1 g)), \tag{2.146}$$

so that $(T_\nu^\ell f)(1) = \mathbb{E}_{Y_\ell \sim \nu^{*\ell}}(f(Y_\ell))$. Then, the existence of a gap for ν guarantees an exponential weak convergence of the law of Y_ℓ to the Haar measure.

We now move our attention to the properties of random quantum circuits. It is known that various architectures of random quantum circuits on n-qubits form δ-approximate t-designs in depth at most polynomial in n and t. Such results were obtained for random quantum circuits with fixed architecture (e.g., 1D brickwork) as well as nondeterministic architectures (see [121] for a good summary). Recently, a generic bound was obtained for all random quantum circuits with a fixed architecture of Haar-random two-site gates with tighter bounds for architectures satisfying certain conditions [121].

Finally, we briefly discuss the Haar-random gate sets. By a random gate set, we mean a gate set whose elements are Haar-randomly chosen elements.

[16] However worse that the striking recent improvement from [33].

It is known that the random gate sets $\mathcal{S} \subset \mathrm{U}(d)$ form decent t-designs with $t = \mathcal{O}(d^{1/6}/\log(d))$ [122, 123]. Recently, the distribution of $\delta(\nu_{\mathcal{S}}, t)$ for random gate sets was studied in [124], where the authors provide the bounds on the probabilities of $\mathcal{S}$ forming δ-approximate t-designs and the number of random gates needed to create such t-designs with given probability.

In [48], the authors propose a random matrix model based on Gaussian or Ginibre random matrix ensembles that aims to describe the probability distribution of $\delta(\nu_{\mathcal{S}}, t)$ for random $\mathcal{S}$. They prove that as $|\mathcal{S}| \to \infty$, the block-diagonal operator determining the gap at scale t (similar to (2.87) but without repetitions), converges in distribution to the corresponding block-diagonal operator T_t with i.i.d. random matrix blocks, after rescaling by $\sqrt{|\mathcal{S}|}$. They provide numerical evidence that their model is almost exact for all $|\mathcal{S}|$, with tail bounds of $\delta(\nu_{\mathcal{S}}, t)$ upper bounded by the corresponding tail bounds of their random matrix model. Since their random matrix model satisfies the spectral gap conjecture with probability 1, they conjecture the same is true for $\delta(\nu_{\mathcal{S}}, t)$ for all $|\mathcal{S}|$.

Following [48], we now comment on the limiting behaviour of the spectrum of the t-moment operator and the resulting optimal bound on the spectral gap for $\mathrm{U}(d)$. We recall that a spectral measure σ_{H_n} of a self-adjoint matrix $n \times n$, H_n, on a closed interval in $\mathbb{R}$ is the number of eigenvalues of H_n in this interval divided by n.

For a symmetric gate-set $\mathcal{S}$, the t-moment operator (2.84) is a bounded self-adjoint operator so that it has a well-defined spectrum. Its spectral measure $\sigma_{\mathcal{S},t}$ is compactly supported so it determined by its moments

$$\sigma_{\mathcal{S},t}^{(m)} = \frac{\mathrm{Tr}(T_{\nu_{\mathcal{S},t}}^m)}{d^{2t}}. \tag{2.147}$$

The asymptotic behavior of $\sigma_{\mathcal{S},t}^{(m)}$ moments, i.e., the limit $\lim_{t\to\infty} \sigma_{\mathcal{S},t}^{(m)}$ is determined by the number of length m spellings of the central elements

$$\lim_{t\to\infty} \sigma_{\mathcal{S},t}^{(m)} = \frac{1}{|\mathcal{S}|^m} \sum_{\substack{U_1,U_2,\ldots,U_m \in \mathcal{S} \\ U_1U_2\ldots U_m \propto I}} 1 \tag{2.148}$$

For $\mathcal{S}$ consisting of free generators, this boils down to the number of spellings of identity and was provided in [125]. Moreover in [125] it was shown that in this case there exists a measure $\sigma_{\mathcal{S}}$, such that $\sigma_{\mathcal{S}}^{(m)} = \lim_{t\to\infty} \sigma_{\mathcal{S},t}^{(m)}$. Such a measure is known as the Kesten-McKay measure [113, 126]

$$d\sigma_{\mathcal{S}}(x) = \frac{|\mathcal{S}|\sqrt{\delta_{\mathrm{opt}}^2(\mathcal{S}) - x^2}}{2\pi(1-x^2)} \mathbb{1}_{[-\delta_{\mathrm{opt}(\mathcal{S})}, \delta_{\mathrm{opt}(\mathcal{S})}]} dx, \tag{2.149}$$

where $\delta_{\mathrm{opt}(\mathcal{S})}$ is given by

$$\delta_{\mathrm{opt}}(|\mathcal{S}|) \coloneqq \frac{2\sqrt{|\mathcal{S}|-1}}{|\mathcal{S}|}. \tag{2.150}$$

It follows that that $\sigma_{\mathcal{S},t}$ converges weakly to $\sigma_{\mathcal{S}}$ in the limit $t \to \infty$.

Analogous results can be obtained for any (i.e. not necessarily inverse-closed) finite $\mathcal{S}$, for which $\mathcal{S} \cup \mathcal{S}^{-1}$ generates a free group [48]. Note that in the general case, the t-moment operator does not need to be self-adjoint, hence by the Kesten-McKay measure we understand the measure describing the singular values of $T_{\nu_{\mathcal{S}}}$, i.e. the spectral measure of $\sqrt{T_{\nu_{\mathcal{S}}} T_{\nu_{\mathcal{S}}}^*}$. Such a measure is given by

$$\frac{|\mathcal{S}|\sqrt{\delta_{\mathrm{opt}}^2(\mathcal{S}) - x^2}}{\pi(1-x^2)} \mathbb{1}_{[0, \delta_{\mathrm{opt}(\mathcal{S})}]} dx. \tag{2.151}$$

Crucially, although the moments converge to the Kesten-McKay measure only for the free generators, the introduction of the relations can only increase the number of spellings of the central elements, increasing the moments. Thus, the support remains contained in the support of the Kesten-McKay measure and we obtain a universal bound on the spectral gap for any universal discrete set $\mathcal{S}$

$$1 - \mathrm{gap}(\nu_{\mathcal{S}}) \geq \delta_{\mathrm{opt}}(|\mathcal{S}|). \tag{2.152}$$

Finally, the Kesten bound (2.152) is tighter than (2.80) starting from $|\mathcal{S}| \geq 2$.

2.4.5 Research problems

In this subsection, we formulate three main research problems related to the Thesis. Each problem is divided into subproblems, or rather stages, and we indicate which subproblems were addressed in the research part of the thesis (Chapters 3 to 5).

I. Decay of the spectral gap

a. Obtain poly-logarithmic bounds on $\mathrm{gap}_t(\mathcal{S})$ for $\mathbf{U}(d)$ with explicit or essentially calculable constants (addressed in Chapter 3).

b. Recreate Varju's result with exponent A or better for $\mathbf{U}(d)$ using more natural and simplified construction in the t-design language. Provide a way to calculate t_0 (not addressed in the thesis).

c. Show that t_0 can be as small as possible, e.g. at the level of universality detection $t_0 = 6$ (for $d = 2$) and $t_0 = 4$ (for $d \geq 3$) - see Section 2.3.3 (not addressed in the thesis).

II. The t-design $\equiv$ ε-net correspondence

a. Obtain the t-design $\implies$ ε-net correspondence for $\mathbf{U}(d)$ using more natural and simplified construction (addressed in Chapter 4).

b. Improve the scaling of t and δ in d and ε (partly addressed in Chapter 4).

c. Saturate the optimal scaling of $t \simeq d^2/\varepsilon$ or prove a tighter lower bound. Provide an upper bound on the scaling of δ and show it saturates (not addressed in the thesis).

III. Efficiency of quantum gates

a. Find a good and essentially calculable measure of the relative efficiency of quantum gate sets (addressed in Chapter 5).

b. Relate this measure to practically relevant proxies for physical efficiency/cost (addressed in Chapter 5).

c. Use this measure to analyse the physical efficiency of relevant gate sets (partly addressed in Chapter 5).

Chapter 3

Paper I: Calculable lower bounds on the efficiency of universal sets of quantum gates

3.1 Overview

In this first paper, we aimed to derive the poly-logarithmic lower bound on the finite-scale spectral gap, $\mathrm{gap}_t(\mathcal{S})$, which is given by an explicit formula with known or (in principle) calculable constants.

By $\mathcal{S} = \{U_1, \ldots, U_k, U_1^{-1}, \ldots U_k^{-1}\}$ we denote a universal symmetric set of d-dimensional quantum gates. The proof can be divided into three main steps:

1. We consider sets of gates, which can be derived from $\mathcal{S}$ by squaring its elements and removing a certain number of elements with their inverses $\{U_i^2, U_i^{-2}\}$. Using the reasoning based on Bourgain's argument (see [40]), we obtain a lower bound on $\mathrm{gap}_t(\mathcal{S})$ which depends on the diameters of such derived gate sets, at specific scales proportional to $1/t$ (c.f. (2.143)).

2. We then use the Solovay-Kitaev theorem (a version with calculable constants; c.f. Theorem 2.7), to bound such diameters. The constant A in the theorem we use depends on the circuit lengths needed to obtain the initial approximation for the Solovay-Kitaev theorem.

3. Finally, we use the SKL theorem based on the finite-scale spectral gap from [41] to bound such initial circuit lengths (see also (2.144)).

The final bound is given in Theorem 1 (c.f. Theorem 2.6):

$$\mathrm{gap}_t(\mathcal{S}) \geq \alpha \cdot g_{t_0}(\mathcal{S}) \cdot \log(\beta t)^{-2c} \tag{3.1}$$

where $t \geq t_0$ and the constants α and β are given by explicit formulas depending on d and additional construction parameter ε_0. The value of t_0 is given by the t-design ε-net correspondence and depends on d and ε_0. The parameter ε_0 can be understood as the parameter used to set the value of t_0, with the largest possible value of ε_0 being $1/(d+2)$. This corresponds to the smallest possible value of t_0 being 509 for $d = 2$. We provide numerical values for t_0, α and β with varying ε_0 for $d = 2$, 3, 4 in Tables 1 and 2. The value of c depends on the Solovay-Kitaev theorem used, namely, it is the exponent of $\log(1/\varepsilon)$. Finally, $g_{t_0}(\mathcal{S})$ is the quantity which can be found numerically by the calculations of the spectral gap of roughly 2^k gate sets, with a varying number of elements, derived from $\mathcal{S}$ at scale t_0. We supplement our results with the numerical simulations of such bounds for Haar-random single-qubit gate sets.

Although Theorem 1 requires a rather unusual condition, namely that for each $1 \leq i \neq j \leq k$, the set $\{U_i^2, U_j^2, U_i^{-2}, U_j^{-2}\}$ is universal, this condition is met for generic $\mathcal{S}$, e.g. Haar-random gates, with probability 1.

Another result provided in the paper is an alternative to [111] proof of the SKL theorem based on the (global) spectral gap, $\mathrm{gap}(\mathcal{S})$ (Theorem 3).

We note that the Solovay-Kitaev theorem in our construction can be replaced with other constructive SKL theorems with explicit ε-dependence and constants that are determined by some initial circuit lengths.

One should clarify that the results presented in this paper are not directly applicable to the efficiency bounds. Indeed, since our procedure involves Bourgain's argument to bound the spectral gap by the diameter, the resulting poly-logarithmic decay will have exponent $2c$, where c is the exponent of the Solovay-Kitaev theorem used. Thus, the resulting SKL theorem with explicit ε-dependence will have exponent $2c + 1$. Therefore, although the efficiency bounds are the motivation for studying the finite-scale spectral gap, the paper actually focuses on the bounds on the spectral gap itself. This is motivated by the lack of scale-defining constants in known poly-logarithmic bounds, such as (2.140).

3.2 Contribution statement

My contribution to this article was:

1. Co-formulation of the idea of the proof of Theorem 1 and preparation of the proof.
2. Preparation of the proof of Theorem 3.
3. Conduction of numerical experiments for Haar-random gates on the supercomputing cluster (based on the code to compute spectral gaps, shared with me by Piotr Dulian).
4. Writing the whole article.

3.3 Errata

1. Below (12), the regular representation restricted to space V_λ corresponds to $\pi_\lambda^{d_\lambda}$, not just π_λ.
2. Above (22), the Lie group $\mathfrak{su}(d)$ should be replaced with $\mathfrak{su}_{\mathbb{C}}(d) = \mathfrak{sl}(d, \mathbb{C})$.
3. In (71), the lower bound on volume should be used instead, which means C_v should be replaced with a_v.

OPEN ACCESS
IOP Publishing
Journal of Physics A: Mathematical and Theoretical
J. Phys. A: Math. Theor. **56** (2023) 115304 (20pp)
https://doi.org/10.1088/1751-8121/acbd24

Calculable lower bounds on the efficiency of universal sets of quantum gates

Oskar Słowik* **and Adam Sawicki**

Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland

E-mail: oslowik@cft.edu.pl

Received 22 February 2022; revised 20 January 2023
Accepted for publication 19 February 2023
Published 3 March 2023

Abstract

Currently available quantum computers, so called Noisy Intermediate-Scale Quantum devices, are characterized by relatively low number of qubits and moderate gate fidelities. In such scenario, the implementation of quantum error correction is impossible and the performance of those devices is quite modest. In particular, the depth of circuits implementable with reasonably high fidelity is limited, and the minimization of circuit depth is required. Such depths depend on the efficiency of the universal set of gates $\mathcal{S}$ used in computation, and can be bounded using the Solovay–Kitaev theorem. However, it is known that much better, asymptotically tight bounds of the form $\mathcal{O}(\log(\epsilon^{-1}))$, can be obtained for specific $\mathcal{S}$. Those bounds are controlled by so called spectral gap, denoted $\mathrm{gap}(\mathcal{S})$. Yet, the computation of $\mathrm{gap}(\mathcal{S})$ is not possible for general $\mathcal{S}$ and in practice one considers spectral gap at a certain scale $r(\epsilon)$, denoted $\mathrm{gap}_r(\mathcal{S})$. This turns out to be sufficient to bound the efficiency of $\mathcal{S}$ provided that one is interested in a physically feasible case, in which an error ε is bounded from below. In this paper we derive lower bounds on $\mathrm{gap}_r(\mathcal{S})$ and, as a consequence, on the efficiency of universal sets of d-dimensional quantum gates $\mathcal{S}$ satisfying an additional condition. The condition is naturally met for generic quantum gates, such as e.g. Haar random gates. Our bounds are explicit in the sense that all parameters can be determined by numerical calculations on existing computers, at least for small d. This is in contrast with known lower bounds on $\mathrm{gap}_r(\mathcal{S})$ which involve parameters with ambiguous values.

* Author to whom any correspondence should be addressed.

1751-8121/23/115304+20$33.00

Keywords: spectral gap, averaging operators, t-designs, quantum gates efficiency, quantum gates, quantum circuits, quantum computing

(Some figures may appear in colour only in the online journal)

1. Introduction and main results

Universal, scalable and fault-tolerant quantum computers are the holy grail of quantum computing. Such devices require quantum error correction that, due to quantum threshold theorem, can be implemented if the levels of gate errors are small enough [1–3]. However, recent quantum hardware, so called Noisy Intermediate-Scale Quantum (NISQ) devices, does not offer gate fidelities required for quantum error correction and their performance is heavily affected by gate imperfections [4–6]. Because of error accumulation effects, the depth of circuits feasible for NISQ devices is very modest. Hence it is imperative to find ways to minimize such depths. One of the ways to address this issue is to focus on the efficiency of universal sets [7, 8] of gates $\mathcal{S}$ used for the computations.

Spectral gap is a useful measure of efficiency of universal sets of quantum gates $\mathcal{S} \subset \mathrm{SU}(d)$. The value of gap for chosen $\mathcal{S}$, denoted $\mathrm{gap}(\mathcal{S})$, lies between 0 (no gap) and some optimal value $\mathrm{gap}_{\mathrm{opt}} < 1$, depending only on the number of gates $|\mathcal{S}|$ [9]. The higher the value of $\mathrm{gap}(\mathcal{S})$, the better is the upper bound on the minimal length (circuit depth) ℓ of a sequence of gates from $\mathcal{S}$ required to ε-approximate any unitary operation from $\mathrm{SU}(d)$. Recall that the Solovay–Kitaev theorem [10] provides such a bound for depth ℓ,

$$\ell = \tilde{A}(\mathcal{S}) \cdot \log^{3+\delta}(1/\epsilon)), \quad \delta > 0. \tag{1}$$

However, the existence of gap, i.e. $\mathrm{gap}(\mathcal{S}) > 0$, implies that

$$\ell = A(\mathcal{S}) \cdot \log(1/\epsilon)) + B(\mathcal{S}) \tag{2}$$

is enough, with the constants A and B proportional to $\log^{-1}(1/(1-\mathrm{gap}(\mathcal{S})))$ [11]. In fact, $\ell = \mathcal{O}(\log(1/\epsilon)))$ is optimal, which can be seen from a simple volumetric argument.

One should note that some properties of $\mathcal{S}$ with optimal spectral gap are known. For instance, if the gates from the universal set $\mathcal{S}$ have algebraic entries then the gap exists [12, 13]. Moreover, it has been conjectured that any universal $\mathcal{S}$ has the gap and there are explicit constructions of examples of $\mathcal{S}$ with the optimal spectral gap for $\mathrm{SU}(2)$ with $|\mathcal{S}| = p - 1$ for $p \equiv 1 \bmod 4$ [14, 15]. Finally, some commonly used one-qubit universal sets turned out to have the optimal spectral gap [16–19]. However, the construction of many-qubit gates with the optimal spectral gap remains an open problem.

The calculation of $\mathrm{gap}(\mathcal{S})$ is challenging and in practice one often considers the gap up to the certain scale r, denoted $\mathrm{gap}_r(\mathcal{S})$, such that $\mathrm{gap}(\mathcal{S})$ is the infimum of $\mathrm{gap}_r(\mathcal{S})$ over all scales r[1]. Since it is impossible to implement gates without any error, in practice ε can be bounded from below. In such a case, in order to bound ℓ it is sufficient to have the knowledge of $\mathrm{gap}_r(\mathcal{S})$ at some scale $r(\epsilon)$ instead of $\mathrm{gap}(\mathcal{S})$. This is due to the existence of the Solovay–Kitaev-like theorems involving $\mathrm{gap}_r(\mathcal{S})$. Specifically, it is known that for any universal $\mathcal{S}$ one can bound $\ell \propto \mathrm{gap}_r^{-1}(\mathcal{S}) \cdot \log(1/\epsilon)$ at some scale $r(\epsilon)$ (see the first part of lemma 5 in [20] and the improved version with $r(\epsilon) \simeq \mathcal{O}(1/\epsilon \cdot \log(1/\epsilon))$—proposition 2 in [21]). Thus, bounding $\mathrm{gap}_r(\mathcal{S})$ is imperative. From the seminal paper [20] it is known, in more general setting of

[1] The exact definition of a scale depends on the approach.

semisimple compact connected Lie groups, that there exist group constants c, A and r_0 such that

$$\mathrm{gap}_r(\mathcal{S}) \geqslant c \cdot \mathrm{gap}_{r_0}(\mathcal{S}) \cdot \log^{-A}(r), \tag{3}$$

for any $r \geqslant r_0$. Thus, the knowledge of gap at the certain scale r_0 enables to bound the rate at which $\mathrm{gap}_r(\mathcal{S})$ vanishes with growing $r \geqslant r_0$. However it is unclear what is the magnitude of the minimal scale r_0 from which the bound (3) holds, even for $\mathrm{SU}(2)$. Our preliminary analysis of this bound suggests that the value of r_0 for $\mathrm{SU}(2)$ resulting from the proof is enormous—orders of magnitude larger than the scale for which the numerical calculation of $\mathrm{gap}_{r_0}(\mathcal{S})$ is remotely possible.

In this paper we exploit Bourgain's argument for bounding $\mathrm{gap}_r(\mathcal{S})$ by the diameter of $\mathcal{S}$, which was communicated in the proof of the second part of lemma 5 from [20]. By introducing an additional assumption on $\mathcal{S}$ we obtain calculable bounds on $\mathrm{gap}_r(\mathcal{S})$ for universal sets of quantum gates. Our additional assumption on $\mathcal{S}$ is satisfied e.g. for generic quantum gates, such as Haar random gates (with probability 1). The main result of the paper is the following.

Theorem 1. *Let $\mathcal{S} = \{U_1, \ldots, U_k, U_1^{-1}, \ldots U_k^{-1}\}$ be a universal symmetric set of d-dimensional quantum gates, such that for any $U_i, U_j \in \mathcal{S}$, $i \neq j$, the set $\{U_i^2, U_j^2, U_i^{-2}, U_j^{-2}\}$ is universal. Then*

$$\mathrm{gap}_t(\mathcal{S}) \geqslant \alpha \cdot g_{t_0}(\mathcal{S}) \cdot \log^{-2c}(\beta t), \tag{4}$$

where $c = \log(5)/\log(3/2) \approx 4$, α and β are known constants and $g_{t_0}(\mathcal{S})$ can be determined by the numerical calculations of gaps at a known scale t_0 of certain universal sets that can be derived from $\mathcal{S}$.

The quantity $g_{t_0}(\mathcal{S})$ is defined in equation (137), see also equations (138) and (96). Crucially, we provide explicit formulas (136) and (139), (78) for $\alpha = \alpha(d, \epsilon_0)$, $\beta = \beta(d)$ and $t_0 = t_0(d, \epsilon_0)$, where ε_0 is the parameter in the construction leading to different bounds. The value of t_0 is small enough to enable numerical calculations of $g_{t_0}(\mathcal{S})$, at least for $d = 2, 3$ and 4. Hence, our bounds can be made explicit by numerical experiments for fixed $\mathcal{S}$. We provide examples of specific values of t_0, α and β, for $d = 2, 3$ and 4 in tables 1 and 2. The minimal possible values of t_0 are indicated by bold font and given by

$$t_{\min} := \lceil 5d^{5/2}/\epsilon_{0,\max} \cdot \tau(\epsilon_{0,\max}, d) \rceil, \quad \epsilon_{0,\max} := 1/(d+2), \tag{5}$$

where $\tau(\epsilon, d)$ is defined in equation (78). We present the values of t_0 up to the ones giving α around 1.

The value of α grows quickly with t_0 as can be seen in figure 1. Values of β and c do not depend on t_0. On the other hand, the value of $g_{t_0}(\mathcal{S})$ can decrease with increasing t_0.

In order to check the behavior of our bound (4) and demonstrate that it can be calculated on existing hardware, we performed a numerical simulation on a supercomputer. For the sake of this simulation, we chose 1000 Haar random sets $\mathcal{S}$ for $d = 2$, each consisting of three gates and their inverses. The computations took approximately two weeks and utilized 1008 CPU cores. We calculated the values of the lower bound for t_0 ranging from 550 to 900 (with increment 10) and plotted the bounds for t from t_0 to 1000. We also calculated the ratio of our bound and the true value of the gap at given t. We present those results averaged over all sets $\mathcal{S}$ in figure 2.

Table 1. Examples of values of t_0, α and β for $d=2$ (left) and $d=3$ (right). The parameter ε_0 is an element of the construction determining t_0 (along with d). Bold font indicates the choice of the smallest possible t_0.

ε_0	t_0	α	β	ε_0	t_0	α	β
0.04	4599	9.68×10^{-1}	0.393	0.02	29199	7.25×10^{-1}	0.251
0.05	3544	3.67×10^{-1}	0.393	0.03	18353	1.73×10^{-1}	0.251
0.06	2860	1.49×10^{-1}	0.393	0.04	13170	5.07×10^{-2}	0.251
0.07	2384	6.29×10^{-2}	0.393	0.05	10166	1.65×10^{-2}	0.251
0.08	2035	2.72×10^{-2}	0.393	0.06	8219	5.69×10^{-3}	0.251
0.09	1769	1.18×10^{-2}	0.393	0.07	6861	2.01×10^{-3}	0.251
0.1	1559	5.15×10^{-3}	0.393	0.08	5864	7.10×10^{-4}	0.251
0.11	1391	2.22×10^{-3}	0.393	0.09	5103	2.47×10^{-4}	0.251
0.12	1252	9.38×10^{-4}	0.393	0.1	4504	8.31×10^{-5}	0.251
0.13	1137	3.85×10^{-4}	0.393	0.11	4022	2.65×10^{-5}	0.251
0.14	1039	2.62×10^{-4}	0.393	0.12	3625	7.81×10^{-6}	0.251
0.15	955	5.68×10^{-5}	0.393	0.13	3295	2.07×10^{-6}	0.251
0.16	883	1.99×10^{-5}	0.393	0.14	3014	4.75×10^{-7}	0.251
0.17	820	6.36×10^{-6}	0.393	0.15	2775	8.82×10^{-8}	0.251
$\lessapprox$**0.25**	**509**	$\gtrapprox$**0**	**0.393**	$\lessapprox$**0.20**	**1958**	$\gtrapprox$**0**	**0.251**

Table 2. Examples of values of t_0, α and β for $d=4$. The parameter ε_0 is an element of the construction determining t_0 (along with d). Bold font indicates the choice of the smallest possible t_0.

ε_0	t_0	α	β
0.01	134232	8.46×10^{-1}	0.175
0.02	61313	1.06×10^{-1}	0.175
0.03	38602	2.18×10^{-2}	0.175
0.04	27738	5.49×10^{-3}	0.175
0.05	21435	1.52×10^{-3}	0.175
0.06	17347	4.34×10^{-4}	0.175
0.07	14494	1.24×10^{-4}	0.175
0.08	12398	3.43×10^{-5}	0.175
0.09	10797	8.86×10^{-6}	0.175
0.1	9537	2.07×10^{-6}	0.175
0.11	8522	4.14×10^{-7}	0.175
0.12	7687	6.59×10^{-8}	0.175
0.13	6990	7.37×10^{-9}	0.175
0.14	6400	4.54×10^{-10}	0.175
$\lessapprox$**0.1(6)**	**5195**	$\gtrapprox$**0**	**0.175**

The value of the lower bound looks qualitatively the same as the ratio in figure 2 rescaled by a constant. This is because the true value of the gap is practically constant for any chosen $\mathcal{S}$ in the inspected range of t. From figure 2 it is clear that our bound is far from being tight, at least in a tested range. However, obtained results are not far from our expectations taking into account the generality of our bounds. Moreover, evidently, the lower bounds improve quickly with t_0, due to the rising value of constant α that dominates possible deterioration of $g_{t_0}(\mathcal{S})$. In fact, the value of $g_{t_0}(\mathcal{S})$ is also constant for any $\mathcal{S}$ in the inspected range. Needless to say, such

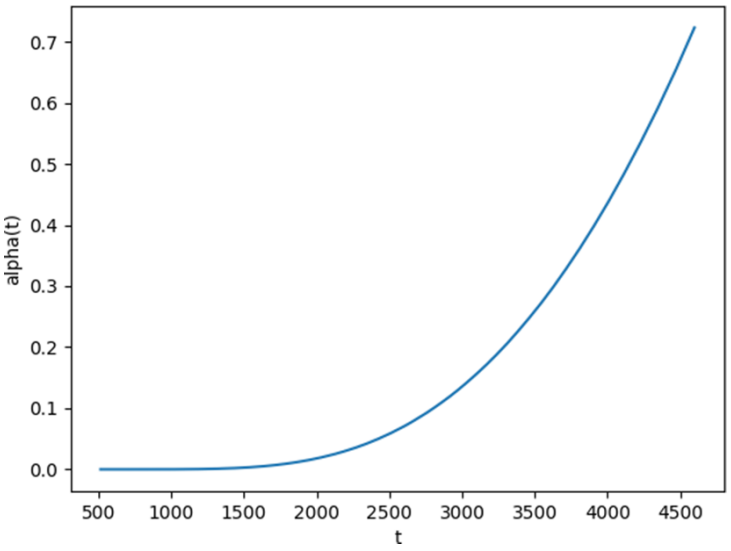


Figure 1. The value of α as a function of t_0 for $d = 2$.

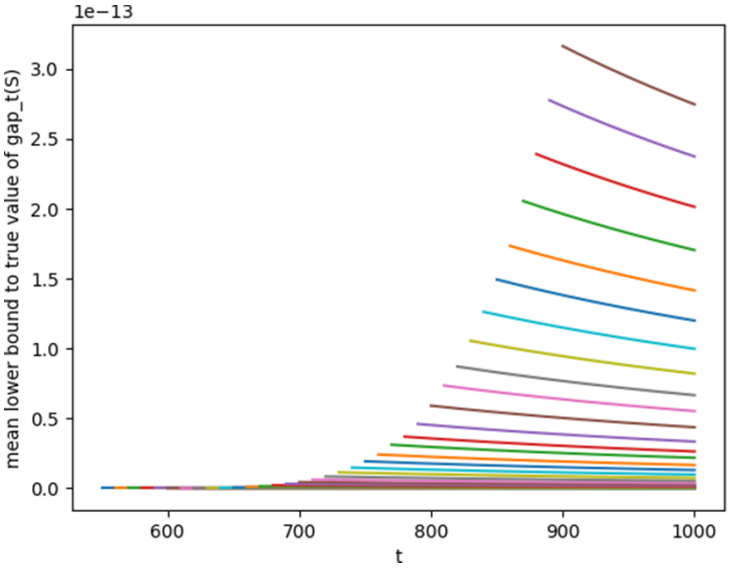


Figure 2. Ratio of lower bound (4) to true value of gap as a function of $t \in [t_0, 1000]$. The ratio was averaged over 1000 Haar random sets $\mathcal{S}$ with $d = 2$ and 3 gates on each set together with inverses. Each line corresponds to a lower bound calculated for different value of $t_0 \in [550, 900]$ with increment 10.

improvement cannot continue indefinitely, since the ratio must be at most 1. Unfortunately, we did not have enough resources to push our simulations further.

The structure of this paper is as follows. In section 2 we introduce the mathematics used in the paper, such as the averaging operators and their relevant spectral gaps. In section 3 we provide an alternative proof of the efficiency bound (2) from [11] with A and B proportional to $\mathrm{gap}^{-1}(\mathcal{S})$. In section 4 we present the proof of our main result, theorem 1.

2. Averaging operators and their spectral gaps

By G_d we denote the projective unitary group $\mathrm{PU}(d)$, which is the quotient of the unitary group $\mathrm{U}(d)$ by its center

$$\mathrm{U}(1) = \{e^{i\theta} | \theta \in [0, 2\pi)\}. \tag{6}$$

Consider the space $L^2(G_d)$ of square integrable complex functions on G_d with respect to μ, equipped with the standard scalar product $\langle \cdot, \cdot \rangle$ (linear on the second slot). Since G_d is compact we consider only unitary representations. A group G_d acts on $L^2(G_d)$ via (left) regular representation Reg. Given a function $f \in L^2(G_d)$ and element $g \in G_d$

$$(\mathrm{Reg}(g)f)(x) = f(g^{-1}x), \tag{7}$$

so the regular representation acts on functions by shifts. Regular representation is not irreducible. In fact, due to Peter–Weyl theorem, it decomposes into an orthogonal direct sum of all the irreducible unitary representations (irreps) with multiplicities equal to their dimensions

$$L^2(G_d) = \widehat{\bigoplus_{\lambda \in \Lambda}} V_\lambda^{\oplus d_\lambda}, \tag{8}$$

where Λ is the set of highest weights of G_d (enumerating all irreps up to isomorphism), V_λ is the representation space of irrep π_λ with highest weight λ and dimension d_λ and hat denotes the closure of an infinite direct sum. Moreover for each $\lambda \in \Lambda$

$$\left\{ \sqrt{d_\lambda} (\pi_\lambda)_{ij} | 1 \leqslant i, j \leqslant d_\lambda \right\}, \tag{9}$$

is an orthonormal basis of V_λ where matrix elements $(\pi_\lambda)_{ij}$ are functions in $L^2(G_d)$ given by

$$(\pi_\lambda)_{ij}(g) := \langle e_i, \pi_\lambda(g) e_j \rangle, \tag{10}$$

for some fixed orthonormal basis of V_λ, $\{e_k | 1 \leqslant k \leqslant d_\lambda\}$. Clearly, sum of all such basis form an orthonormal basis of $L^2(G_d)$

$$\left\{ \sqrt{d_\lambda} (\pi_\lambda)_{ij} | \lambda \in \Lambda, 1 \leqslant i, j \leqslant d_\lambda \right\}. \tag{11}$$

Hence any function $f \in V_\lambda$, as a linear combination of matrix elements, is given by

$$f(g) = \mathrm{Tr}[A \pi_\lambda(g)], \quad g \in G_d \tag{12}$$

for some complex $d_\lambda \times d_\lambda$ matrix A. The regular representation restricted to functions in V_λ is isomorphic to representation π_λ. If π is any (possibly reducible) representation of G_d, then π is isomorphic to the direct sum of irreps which can be identified with function spaces from Peter–Weyl decomposition $V_{\lambda_1}^{\otimes m_1} \oplus \ldots \oplus V_{\lambda_k}^{\otimes m_k}$, for some $k \geqslant 1$ and multiplicities $m_i \geqslant 1$. If $m_i \leqslant d_{\lambda_i}$ for all $1 \leqslant i \leqslant k$, then representation π will appear as a subrepresentation of $L^2(G)$. The corresponding space of functions consists of all functions obtained via

$$f(g) = \mathrm{Tr}[A \pi(g)], \quad g \in G_d \tag{13}$$

for all matrices A.

We now comment on how one may naturally choose a scale up to which one would like to consider irreps of G_d.

The Lie algebra of G_d is isomorphic to $\mathfrak{su}(d)$ since

$$G_d = \mathrm{PU}(d) \cong \mathrm{PSU(d)} = \mathrm{SU(d)}/Z(\mathrm{SU}(d)), \tag{14}$$

where $Z(\mathrm{SU}(d)) \simeq \mathbb{Z}_d$, the center of $\mathrm{SU}(d)$, is discrete.

The adjoint representation Ad of $\mathrm{U}(d)$ descents into the quotient group G_d forming the adjoint representation Ad of a group G_d acting on its representation space $\mathfrak{su}(d)$ via

$$\mathrm{Ad}_U(X) = \hat{U}X\hat{U}^{-1}, \quad U \in G_d, X \in \mathfrak{su}(d), \tag{15}$$

where $\hat{U} \in \mathrm{U}(d)$ is any representative of $U \in G_d$. Importantly Ad is faithful hence every representation of G_d is realized inside $\mathrm{Ad}^{\otimes n}$ for n large enough. The defining representation U of $\mathrm{U}(d)$ does not descend into a well-defined representation U of G_d but $U \otimes \overline{U}$ does, where $\overline{U}$ is the adjoint of U. In fact,

$$U \otimes \overline{U} \cong \mathrm{Ad} \oplus I, \tag{16}$$

where by I we denote the one-dimensional trivial representation. Thus each irrep of G_d appears in rep $(U \otimes \overline{U})^{\otimes t}$ for some t. Moreover this rep contains only projective irreps of $\mathrm{U}(d)$ hence reps of G_d.

Consider $t \geqslant 2$. Then

$$(U \otimes \overline{U})^{\otimes t} \cong (U \otimes \overline{U})^{\otimes t-1} \otimes (\mathrm{Ad} \oplus I) \cong [(U \otimes \overline{U})^{\otimes t-1} \otimes \mathrm{Ad}] \oplus (U \otimes \overline{U})^{\otimes t-1} \tag{17}$$

and applying this reasoning inductively we see that all irreps of $(U \otimes \overline{U})^{\otimes s}$ appear in $(U \otimes \overline{U})^{\otimes t}$ for $s \leqslant t$. Thus we see that with t increasing the rep $(U \otimes \overline{U})^{\otimes t}$ contains more and more irreps of G_d and each irrep of G_d is contained in this rep for t large enough. In the language of Peter–Weyl theorem the corresponding functions in $L^2(G)$ are

$$f(U) = \mathrm{Tr}[A(U \otimes \overline{U})^{\otimes t}], g \in G \tag{18}$$

so they are balanced polynomials in U and $\overline{U}$ of degree t. Thus increasing t corresponds to considering polynomials with higher degrees. This motivates us to consider the following function spaces in $L^2(G_d)$,

$$L^2_t(G_d) = \bigoplus_{\lambda \in \Lambda_t} V_\lambda, \tag{19}$$

where Λ_t is the set of unique (i.e. without repetitions and up to isomorphism) highest weights of irreps of G_d appearing in $(U \otimes \overline{U})^{\otimes t}$. In the case $t = 0$, we set

$$L^2_0(G_d) = I. \tag{20}$$

Additionally we define the following related symbols. The set $\tilde{\Lambda}_t$ which equals Λ_t without the weight of the trivial representation and the set of all unique highest weights

$$\Lambda_\infty := \bigcup_{t=0}^{\infty} \Lambda_t. \tag{21}$$

Fortunately the weights Λ_t have a nice description in terms of the sequences of integers.

Lemma 2. *The set Λ_t consists precisely of weights indexed by nonincreasing length d integer sequences λ such that $|\lambda| = 0$ and $|\lambda_+| \leqslant t$, where $|\lambda|$ denotes the sum of entries and λ_+ is the subsequence of positive entries.*

Each sequence $\lambda = (\lambda_1, \ldots, \lambda_d) \in \Lambda_t$ corresponds to a weight (the linear functional on the Cartan subalgebra $\mathfrak{h} \subset \mathfrak{su}(d)$)

$$\lambda = \lambda_1 L_1 + \ldots \lambda_d L_d, \tag{22}$$

where L_i are the standard basis elements[2]. Since $L_1 + \ldots + L_d = 0$ in $\mathfrak{h}^*$, adding a constant sequence, $(c, \ldots, c)$ for some $c \in \mathbb{Z}$, to λ does not change the weight.

Example 1. Consider the system of two qutrits $\mathbb{C}^3 \otimes \mathbb{C}^3$ and $t = 2$. Then, from lemma 2, we have

$$\Lambda_t = \{(2,0,-2),(2,-1,-1),(1,1,-2),(1,0,-1),(0,0,0)\} \tag{23}$$

which is equivalent to

$$\Lambda_t = \{(4,2,0),(3,0,0),(3,3,0),(2,1,0),(0,0,0)\}, \tag{24}$$

and for example $\lambda = (2,1,0)$ corresponds to the highest weight $2L_1 + L_2$ i.e. to the adjoint representation. Similarly we can represent $\tilde{\Lambda}_t$ as

$$\tilde{\Lambda}_t = \{(4,2,0),(3,0,0),(3,3,0),(2,1,0)\}. \tag{25}$$

We introduce the following norm on the space of weights of G_d

$$||\lambda||_1 := \sum_{i=1}^{d} |\lambda_i|. \tag{26}$$

It is clear that for each $\lambda \in \Lambda_t$,

$$||\lambda||_1 \leqslant 2t. \tag{27}$$

From now on we represent each irrep λ by the sequence with smallest $||\lambda||_1$. In particular, the trivial representation is given by $\lambda = (0,0,\ldots,0)$.

By choosing the orthonormal basis of function spaces (11) we have the isomorphisms

$$V_\lambda \simeq \mathcal{H}_\lambda, \tag{28}$$

where $\mathcal{H}_\lambda := \mathbb{C}^{d_\lambda}$. We define

$$\mathcal{H}_t := \bigoplus_{\lambda \in \Lambda_t} \mathcal{H}_\lambda \simeq L^2_t(G_d), \tag{29}$$

and analogously we define $\mathcal{H}_\infty$[3]. By $\mathcal{H}$ we denote the vector space isomorphic to $L^2(G_d)$. Clearly,

$$L^2(G_d) \simeq \mathcal{H} := \bigoplus_{\lambda \in \Lambda_\infty} \mathcal{H}_\lambda^{\oplus d_\lambda}. \tag{30}$$

For any representation of G_d and any finite Borel measure ν on G_d we define the operator

$$\pi(\nu) := \int_{G_d} \pi(g) d\nu(g), \tag{31}$$

acting on the representation space of π. We use can use equation (31) to define various averaging operators. By $\mathcal{S}$ we denote a finite set of generators of G_d and $\nu_\mathcal{S}$ is the counting measure of $\mathcal{S}$ on G_d.

The t-averaging operator wrt to $\mathcal{S}$, $T_{\nu_\mathcal{S},t} : \mathcal{H}_t \to \mathcal{H}_t$ is

$$T_{\nu_\mathcal{S},t} := \bigoplus_{\lambda \in \Lambda_t} \pi_\lambda(\nu_S), \tag{32}$$

[2] The linear functional L_i returns the ith diagonal entry of a matrix in $\mathfrak{h}$.

[3] Here we use the closure of the direct sum.

and can be represented as a block-diagonal matrix. Analogously we define the ∞-averaging operator wrt to $\mathcal{S}$, $T_{\nu_\mathcal{S},\infty} : \mathcal{H}_\infty \to \mathcal{H}_\infty$,

$$T_{\nu_\mathcal{S},\infty} := \bigoplus_{\lambda\in\Lambda_\infty} \pi_\lambda(\nu_S). \tag{33}$$

Finally, the (global) averaging operator wrt to $\mathcal{S}$, $T_{\nu_\mathcal{S}} : \mathcal{H} \to \mathcal{H}$ is

$$T_{\nu_\mathcal{S}} := \bigoplus_{\lambda\in\Lambda_\infty} \pi_\lambda(\nu_S)^{\oplus d_\lambda}. \tag{34}$$

In the language of functions, introduced averaging operators correspond to restrictions of $\mathrm{Reg}(\nu_\mathcal{S})$ to corresponding function subspaces. We denote such isomorphic averaging operators using the same symbols. For example, the global averaging operator is

$$T_{\nu_\mathcal{S}} = \mathrm{Reg}(\nu_\mathcal{S}), \tag{35}$$

so the action on $f \in L^2(G_d)$ is

$$(T_{\nu_\mathcal{S}} f)(h) = \int_{G_d} (\mathrm{Reg}(g)f)(h)\mathrm{d}\nu_\mathcal{S}(g) = \frac{1}{|\mathcal{S}|}\sum_{i=1}^{|\mathcal{S}|} f(g_i^{-1}h). \tag{36}$$

The justification for the name averaging operator is clear from equation (36). Indeed, $T_{\nu_\mathcal{S}}$ replaces the function f with the averaged function, whose value at h is the average of the values of f over all translates of h by the elements of $\mathcal{S}$.

Similarly, the t-averaging operator is

$$T_{\nu_\mathcal{S},t} = \mathrm{Reg}(\nu_\mathcal{S})\big|_{L^2_t(G_d)}, \tag{37}$$

so it acts just like $T_{\nu_\mathcal{S}}$ but on a restricted domain of functions.

Since $T_{\nu_\mathcal{S},t}$ is a sum of $|\mathcal{S}|$ left shift operators, normalized by $1/|\mathcal{S}|$, and due to left-invariance of Haar measure, each such operator is unitary on $L^2_t(G_d)$, we see that $||T_{\nu_\mathcal{S},t}||_{\mathrm{op}} \leqslant 1$, where by $||\cdot||_{\mathrm{op}}$ we denote the operator norm. On the other hand $T_{\nu_\mathcal{S},t}$ acts trivially on $\mathcal{H}_{\lambda_0}$, where $\lambda_0 = (0,0,\ldots,0)$ so $||T_{\nu_\mathcal{S},t}||_{\mathrm{op}} \geqslant 1$ and hence $||T_{\nu_\mathcal{S},t}||_{\mathrm{op}} = 1$.

The subspace $\mathcal{H}_{\lambda_0}$ corresponds to the subspace of constant functions $L^2_0(G_d) = V_{\lambda_0}$, with orthogonal compliment being the space of functions with Haar-average zero. Let T_μ denote the projector onto $\mathcal{H}_{\lambda_0}$. At the level of function spaces, T_μ is the projector onto $L^2_0(G_d)$ which assigns to each function f the constant function with value being the Haar average[4] of f,

$$(T_\mu f)(h) = \int_{G_d} f(g)\mathrm{d}\mu(g). \tag{38}$$

In order to assess how quick the words in $\mathcal{S}$ fill the group G_d, we compare the averaging operator $T_{\nu_\mathcal{S}}$ with T_μ by checking the operator of their difference. Since $T_{\nu_\mathcal{S}}\big|_{\mathcal{H}_{\lambda_0}} = T_\mu\big|_{\mathcal{H}_{\lambda_0}}$, the norm $||T_{\nu_\mathcal{S}} - T_\mu||_{\mathrm{op}}$ equals the norm of the operator

$$\tilde{T}_{\nu_\mathcal{S}} := \bigoplus_{\lambda\in\tilde{\Lambda}_\infty} \pi_\lambda(\nu_S)^{\oplus d_\lambda}. \tag{39}$$

Similarly, we define $\tilde{T}_{\nu_\mathcal{S},t}$ and $\tilde{T}_{\nu_\mathcal{S},\infty}$. Clearly,

$$\sigma(T_{\nu_\mathcal{S}}) = \sigma(T_{\nu_\mathcal{S},\infty}), \tag{40}$$

[4] From now on the symbol μ denotes the Haar measure on G_d.

so we have

$$||T_{\nu_S} - T_\mu||_{\text{op}} = ||\tilde{T}_{\nu_S}||_{\text{op}} = ||\tilde{T}_{\nu_S,\infty}||_{\text{op}}. \tag{41}$$

This motivates us to define the spectral gap of $\mathcal{S}$ as

$$\text{gap}(\mathcal{S}) := 1 - ||\tilde{T}_{\nu_S}||_{\text{op}}. \tag{42}$$

The spectral gap is an useful numerical value describing the set $\mathcal{S}$ via the properties of the corresponding averaging operator.

Similarly, we define spectral gap of $\mathcal{S}$ at scale t as

$$\text{gap}_\text{t}(\mathcal{S}) := 1 - ||\tilde{T}_{\nu_S,t}||_{\text{op}}. \tag{43}$$

In general we can define analogous gaps for any finite Borel measure ν on G_d. For example,

$$\text{gap}_\text{t}(\nu) := 1 - ||\tilde{T}_{\nu,t}||_{\text{op}}, \tag{44}$$

where $\tilde{T}_{\nu,t}$ is defined as in equation (32) with $\nu_\mathcal{S}$ substituted by ν. It is clear that

$$\text{gap}(\mathcal{S}) = \inf_t \text{gap}_\text{t}(\mathcal{S}), \tag{45}$$

and the gaps (42)–(44) belong to $[0,1]$.

We argue that we can assume that $\mathcal{S}$ is symmetric without the loss of generality. For a measure ν on G_d we define its conjugate $\tilde{\nu}$ via the property

$$\int_{G_d} f(g)\text{d}\tilde{\nu}(g) = \int_{G_d} f(g^{-1})\text{d}\nu(g) \tag{46}$$

for all continuous functions f on G_d. We say a measure ν is symmetric if $\nu = \tilde{\nu}$. For two measures ν_1 and ν_2 on G_d, their convolution $\nu_1 * \nu_2$ is a measure on G_d defined via

$$\nu_1 * \nu_2(\Omega) = \int_{G_d} \mathbb{1}_\Omega(gh)\nu_1(g)\nu_2(h). \tag{47}$$

Going back to the definition (31) we have

$$\pi(\nu_1 * \nu_2) = \pi(\nu_1)\pi(\nu_2). \tag{48}$$

It is easy to see that $\pi(\tilde{\nu}) = \pi(\nu)^*$. In particular if ν is symmetric then $\pi(\nu)$ is self-adjoint and hence $\sigma(\pi(\nu))$ is real. Note also that $\nu * \tilde{\nu}$ is automatically symmetric. We can write

$$\pi(\tilde{\nu} * \nu) = \pi(\tilde{\nu}) \cdot \pi(\nu) = \pi(\nu)^* \cdot \pi(\nu), \tag{49}$$

which means that

$$||\pi(\tilde{\nu} * \nu)||_{\text{op}} = ||\pi(\nu)||^2_{\text{op}}. \tag{50}$$

Finally, because $\sqrt{1-x} \leqslant 1 - \frac{x}{2}$ for any $0 \leqslant x \leqslant 1$,

$$\text{gap}_t(\tilde{\nu}_\mathcal{S} * \nu_\mathcal{S}) \geqslant \text{gap}_t(\mathcal{S}) \geqslant \frac{1}{2}\text{gap}_t(\tilde{\nu}_\mathcal{S} * \nu_\mathcal{S}). \tag{51}$$

Since $\mathcal{S}$ is symmetric, $T_{\nu_S,t}$ is Hermitian and so its spectrum $\sigma(T_{\nu_S,t})$ is contained in $[-1,1]$. The same is true for $T_{\nu_\text{S},\infty}$ and T_{ν_S}. Note that since the subspace $\mathcal{H}_{\lambda_0}$ is excluded, the question if $\text{gap}(\mathcal{S}) > 0$ is non-trivial. The gap exists, i.e. $\text{gap}(\mathcal{S}) > 0$, if and only if 1 belongs to the spectrum $\sigma(T_{\nu_S})$ i.e. it is the accumulation point of $\sigma(T_{\nu_S})$. In such a case we say that T_{ν_S} has a spectral gap.

Let's denote by $\mathcal{S}_\ell$ a set of words in G_d of length ℓ built from elements of $\mathcal{S}$

$$\mathcal{S}_\ell := \{g_1 g_2 \dots g_\ell | \, g_1, g_2, \dots g_\ell \in \mathcal{S}\}. \tag{52}$$

The corresponding averaging operator is $T^{\ell}_{\nu_{\mathcal{S}}}$. Indeed,

$$T^{\ell}_{\nu_{\mathcal{S}}} = T_{\nu^{*(\ell)}_{\mathcal{S}}}, \tag{53}$$

and since $\nu^{*(\ell)}_{\mathcal{S}}$ is the law for $\mathcal{S}_\ell$, $T^{\ell}_{\nu_{\mathcal{S}}}$ is the averaging operator with respect to $\mathcal{S}_\ell$. At the level of functions we have

$$(T^{\ell}_{\nu_{\mathcal{S}}} f)(h) = \frac{1}{|\mathcal{S}|^{\ell}} \sum_{w \in S_\ell} f(w^{-1} g). \tag{54}$$

Importantly, $\mathrm{gap}(\mathcal{S})$ can be interpreted as the exponential rate of convergence of the (global) averaging operator $T_{\nu_{\mathcal{S}}}$ to T_μ in the operator norm with ℓ increasing. Indeed, due to left-invariance of Haar measure

$$T_{\nu_{\mathcal{S}}} T_\mu = T_\mu = T_\mu T_{\nu_{\mathcal{S}}}, \tag{55}$$

so we have

$$||T^{\ell}_{\nu_{\mathcal{S}}} - T_\mu||_{\mathrm{op}} = ||(T_{\nu_{\mathcal{S}}} - T_\mu)^{\ell}||_{\mathrm{op}} \leqslant ||(T_{\nu_{\mathcal{S}}} - T_\mu)||^{\ell}_{\mathrm{op}} = ||\tilde{T}_{\nu_{\mathcal{S}}}||^{\ell}_{\mathrm{op}}, \tag{56}$$

and using the notion of a spectral gap (42) we have

$$||T^{\ell}_{\nu_{\mathcal{S}}} - T_\mu||_{\mathrm{op}} \leqslant (1 - \mathrm{gap}(\mathcal{S}))^{\ell} \leqslant e^{-\ell\, \mathrm{gap}(\mathcal{S})}. \tag{57}$$

Thus, if the gap exists then $T^{\ell}_{\nu_{\mathcal{S}}}$ converges to T_μ as $\ell \to \infty$ exponentially fast in the operator norm. Moreover, the rate of convergence improves exponentially with $\mathrm{gap}(\mathcal{S})$ increasing. This motivates us to study $\mathrm{gap}(\mathcal{S})$.

3. Bound on gates efficiency from spectral gap

In [11] it has been shown in case of group $\mathrm{SU}(d)$ that if $\mathcal{S}$ is universal and $T_{\nu_{\mathcal{S}}}$ has a spectral gap then words of length $\ell = \mathcal{O}\left(\log\left(\frac{1}{\epsilon}\right)\right)$ form an ε-net.

In this section we present an alternative proof of this fact for G_d. By $D(\cdot,\cdot)$ we denote a G_d-invariant metric on G_d defined as follows. For $g, h \in G_d$ let $\hat{g}, \hat{h} \in \mathrm{U}(d)$ be the corresponding representatives. Then

$$D(g,h) := \inf_{\theta \in [0,2\pi)} ||\mathrm{e}^{\mathrm{i}\theta} \hat{g} - \hat{h}||_{\mathrm{op}}. \tag{58}$$

Equivalently, we can take the infimum over representatives

$$D(g,h) = \inf_{\hat{g},\hat{h}} ||\hat{g} - \hat{h}||_{\mathrm{op}}. \tag{59}$$

We introduce $B(x,r)$ as the closed ball in G_d with radius r centered at x with respect to D and $B(r)$ is such ball centered at $\mathbb{I}$.

By $\mathrm{Vol}(\Omega)$ we mean the Haar volume of a subset $\Omega \subset G_d$,

$$\mathrm{Vol}(\Omega) = \int_{G_d} \mathbb{1}_\Omega(g) \mathrm{d}\mu(g), \tag{60}$$

where $\mathbb{1}_\Omega$ denotes the indicator function of Ω.

We start with the following simple observation. Let $f \in L^2_t(G_d)$, $\int_{G_d} f(g) \mathrm{d}\mu(g) = 1$, and pick some region $\Omega \subseteq G_d$. Since

$$\mathrm{Vol}(\Omega) = \int_\Omega 1 \mathrm{d}\mu(g) \tag{61}$$

we have that

$$\int_\Omega 1 \mathrm{d}\mu(g) - \int_\Omega (T^\ell_{\nu_S,t} f)(g) \mathrm{d}\mu(g) = \langle 1 - T^\ell_{\nu_S,t} f | \mathbb{1}_\Omega \rangle \leqslant ||1 - T^\ell_{\nu_S,t} f||_2 \cdot \sqrt{\mathrm{Vol}(\Omega)} \tag{62}$$

and

$$||1 - T^\ell_{\nu_S,t} f||_2 = ||(T^\ell_{\nu_S,t} - T_\mu) f||_2 \leqslant ||(T^\ell_{\nu_S,t} - T_\mu)||_{\mathrm{op}} \cdot ||f||_2 \leqslant e^{-\ell \,\mathrm{gap}_t(\mathcal{S})} \cdot ||f||_2, \tag{63}$$

where T_μ is a projector onto $L^2_0(G_d)$ on $L^2_t(G_d)$. Thus,

$$\int_\Omega \left(T^\ell_{\nu_S,t} f\right)(g) \geqslant \mathrm{Vol}(\Omega) - e^{-\ell \,\mathrm{gap}_t(\mathcal{S})} ||f||_2 \sqrt{\mathrm{Vol}(\Omega)}. \tag{64}$$

Clearly, analogous results are true for other averaging operators, in particular for T_{ν_S}.

Theorem 3. *Assume $\mathcal{S}$ is such that $T_\mathcal{S}$ has a spectral gap. Then S_ℓ is an ε-net for every ℓ*

$$\ell \geqslant \frac{\dim G_d}{\mathrm{gap}(\mathcal{S})} \log\left(\frac{1}{\epsilon}\right) + B.$$

Proof. Pick an element $U_0 \in G_d$ and a ball $\Omega = B(U_0, \epsilon/2)$ centered at it. Pick ℓ such that there is no $w_\ell \in \mathcal{S}_\ell$ which ε-approximates U_0, i.e. such that $D(w_\ell, U_0) \leqslant \epsilon$. Let f be a normalized indicator function of $B(\mathbb{I}, \epsilon/2)$, i.e.

$$f(x) = \frac{1}{\mathrm{Vol}(B(\mathbb{I}, \epsilon/2))} \mathbb{1}_{B(\mathbb{I}, \epsilon/2)}(x). \tag{65}$$

We have

$$\int_\Omega \left(T^\ell_{\nu_S} f\right)(g) \mathrm{d}\mu(g) = \frac{1}{|\mathcal{S}|} \sum_{w_\ell \in \mathcal{S}^\ell} \int_{G_d} f(w_\ell^{-1} g) \mathrm{d}\mu(g) \tag{66}$$

but for each g in Ω

$$D(w_\ell^{-1} g, \mathbb{I}) = D(g, w_\ell) > \epsilon/2, \tag{67}$$

hence

$$\int_\Omega \left(T^\ell_{\nu_S} f\right)(g) \mathrm{d}\mu(g) = 0. \tag{68}$$

Using equation (64) we get

$$e^{-\ell \,\mathrm{gap}(\mathcal{S})} \geqslant \mathrm{Vol}(\Omega), \tag{69}$$

since $||f||_2 = 1/\sqrt{\mathrm{Vol}(\Omega)}$. Hence, if

$$e^{-\ell \,\mathrm{gap}(\mathcal{S})} < \mathrm{Vol}(\Omega), \tag{70}$$

we get a contradiction, which means that $\mathcal{S}_\ell$ is an ε-net. On the other hand

$$\mathrm{Vol}(\Omega) \leqslant C_V (\epsilon/2)^{\dim G_d}, \tag{71}$$

where C_V is some group constant. Thus,

$$\ell \geqslant \frac{\dim G_d}{\mathrm{gap}(\mathcal{S})} \log\left(\frac{1}{\epsilon}\right) + B, \tag{72}$$

with

$$B = -\frac{\log(C_V) - \dim G_d \cdot \log(2)}{\mathrm{gap}(\mathcal{S})}. \tag{73}$$

□

We have $\dim G_d = d^2 - 1$ and in the case of G_d can put $C_V = (9.5)^{d^2-1}$, so

$$B = -\frac{d^2-1}{\mathrm{gap}(\mathcal{S})}\log(4.75). \tag{74}$$

The values of constant C_V bounding the volume of a ball in various groups can be obtained by techniques from [22].

Note that theorem 3 cannot be stated in analogous form for the t-averaging operators $T_{\nu_S,t}$, since the normalized indicator function (65) does not belong to $L^2_t(G_d)$ for any t so we cannot write equation (68) for $T_{\nu_S,t}$ instead of T_{ν_S}. However, by considering appropriate approximations of Dirac delta by polynomials from $L^2_t(G_d)$, we can show that

$$\int_\Omega \left(T^\ell_{\nu_S,t} f\right)(g)\mathrm{d}\mu(g) \tag{75}$$

is sufficiently small and hence obtain analogous results. In particular, it is known that

$$\ell \geqslant \frac{C \cdot \log\left(\frac{1}{\epsilon}\right)}{\mathrm{gap}_r(\mathcal{S})}, \tag{76}$$

where $r = D/\epsilon^{2(d^2-1)+2}$ and C, D are some constants [20]. This result has been improved in case of $\mathrm{U}(d)$ in [21], where

$$\ell \geqslant \frac{(d^2-1)(2\log\left(\frac{1}{\epsilon}\right) + \log(4C_b^{3/2}d)) + \log(32)}{\mathrm{gap}_t(\mathcal{S})}, \tag{77}$$

for some absolute constant C_b and $t \geqslant 5d^{5/2}/\epsilon \cdot \tau(\epsilon, d)$, where $\tau(\epsilon, d)$ is

$$\tau(\epsilon,d) = \log^{\frac{1}{2}}(6C_b/\epsilon) \cdot \sqrt{\frac{1}{32}\log^{\frac{1}{2}}(6C_b/\epsilon) + \log\left(\frac{d}{\epsilon} \cdot \log^{\frac{1}{2}}(6C_b/\epsilon)\right)}. \tag{78}$$

4. Calculable lower bound on spectral gap

In this section, we derive lower bounds on the spectral gap at scale t for $\mathcal{S} \subset G_d$, such that any two pairs in $\mathcal{S}$ (of gate with its inverse) form an universal set themselves. This condition can be verified numerically by known universality criteria, see e.g. [23].

Our bound for any t can be calculated from the knowledge of certain gaps up to some fixed $t_0 = t_0(d)$ and is of the form

$$\mathrm{gap}_t(\mathcal{S}) \geqslant \alpha \cdot g_{t_0}(\mathcal{S}) \cdot \log(\beta t)^{-2c}, \tag{79}$$

where $\alpha, \beta, c > 0$ are some specific calculable constants and $g_{t_0}(\mathcal{S})$ can be determined numerically by calculating gaps of certain sets derived from $\mathcal{S}$ up to some calculable scale t_0.

We study the action of the t-averaging operator wrt to $\mathcal{S}$,

$$T_{\nu_S,t} := \bigoplus_{\lambda \in \Lambda_t} \pi_\lambda(\nu_S), \tag{80}$$

acting on the Hilbert space

$$\mathcal{H}_t = \bigoplus_{\lambda\in\Lambda_t} \mathcal{H}_\lambda. \tag{81}$$

By $S(\mathcal{H}_\lambda)$ we denote the unit sphere in $\mathcal{H}_\lambda$,

$$S(\mathcal{H}_\lambda) = \{w \in \mathcal{H}_\lambda \,|\, ||w|| = 1\}. \tag{82}$$

We choose the orthonormal basis

$$\{w_{ij}^\lambda \,|\, 1 \leqslant i,j \leqslant d_\lambda,\, \lambda \in \Lambda_t\} \tag{83}$$

of $\mathcal{H}_t$, induced by the basis (11). Clearly, $||T_{\nu_{\mathcal{S},t}}||_{\mathrm{op}} \leqslant 1$ and our goal is to improve this bound. The irreps Λ_t of G_d can be divided into three disjoint sets, based on the type of the representation of $\mathrm{U}(d)$ they come from:

$$\Lambda_t = \Lambda_{t,\mathbb{H}} \cap \Lambda_{t,\mathbb{R}} \cap \Lambda_{t,\mathbb{C}}, \tag{84}$$

where $\mathbb{H}$, $\mathbb{R}$ and $\mathbb{C}$ stands for quaternionic, real and complex representations. In fact, $\Lambda_{t,\mathbb{H}} = \emptyset$ since quaternionic representations of $\mathrm{U}(d)$ do not contribute to projective representations.

Since

$$||\tilde{T}_{\nu_{\mathcal{S}},t}||_{\mathrm{op}} = \max_{\lambda\in\tilde{\Lambda}_t} ||\pi_\lambda(\nu_{\mathcal{S}})||_{\mathrm{op}}, \tag{85}$$

we fix any $\lambda \in \tilde{\Lambda}_t$ and consider $||\pi_\lambda(\nu_{\mathcal{S}})||_{\mathrm{op}}$.

Additionally we assume $\mathcal{S} = \{U_1, \ldots, U_k, U_1^{-1}, \ldots U_k^{-1}\}$ is generic so that for each $1 \leqslant i \neq j \leqslant k$, the set $\{U_i^2, U_j^2, U_i^{-2}, U_j^{-2}\}$ is a universal symmetric set.

4.1. Strategy of the proof

Our strategy is to show that for any $\lambda \in \tilde{\Lambda}_t$, any $w \in S(\mathcal{H}_\lambda)$ and any generator $U_m \in \mathcal{S}$, except for at most one, say U_k,

$$||(\pi_\lambda(U_i) + \pi_\lambda(U_i^{-1}))w|| \leqslant 2 - (b_i/2)^2 \tag{86}$$

for some coefficients $b_i^2 = b_i^2(\lambda) > 0$ which can be bounded by gaps of certain subsets of the set $\mathcal{S}^2 = \{U_1^2, \ldots, U_k^2, U_1^{-2}, \ldots U_k^{-2}\}$ at some known scale t_0. Hence,

$$||\pi_\lambda(\nu_{\mathcal{S}})w|| \leqslant \frac{1}{|\mathcal{S}|}\left[\sum_{1\leqslant i<k} ||(\pi_\lambda(U_i) + \pi_\lambda(U_i^{-1}))w|| + 2\right] \leqslant 1 - \frac{1}{4|\mathcal{S}|}\sum_{i=1}^{|\mathcal{S}|/2-1} b_i^2(\lambda), \tag{87}$$

which implies

$$\mathrm{gap}_t(\mathcal{S}) \geqslant \min_{\lambda\in\tilde{\Lambda}_t} \frac{1}{4|\mathcal{S}|}\sum_{i=1}^{|\mathcal{S}|/2-1} b_i^2(\lambda) > 0. \tag{88}$$

This means that we can obtain a non-trivial lower bound on $\mathrm{gap}_t(S)$ for any $t \geqslant t_0$. Crucially, the value of t_0 can be easily determined and is not large, so the numerical calculations of the bound are feasible.

4.2. The main reasoning

Since $\pi_\lambda(g_i)$ is unitary we have

$$||(\pi_\lambda(g_i) + \pi_\lambda(g_i^{-1}))w||^2 = 4 - ||(\pi_\lambda(g_i) - \pi_\lambda(g_i^{-1}))w||^2 \tag{89}$$

for any $w \in S(\mathcal{H}_\lambda)$. Let $\imath_\lambda$ denote the Frobenius–Schur indicator of π_λ,

$$\imath_\lambda = \int_{G_d} \chi_\lambda(g^2)\mathrm{d}\mu(g) = \begin{cases} -1, & \text{if } \lambda \in \Lambda_{t,\mathbb{H}}, \\ 0, & \text{if } \lambda \in \Lambda_{t,\mathbb{C}}, \\ 1, & \text{if } \lambda \in \Lambda_{t,\mathbb{R}}. \end{cases} \quad . \tag{90}$$

Note that

$$\int_{G_d} \pi_\lambda(g^2)\mathrm{d}\mu(g) = \frac{\imath_\lambda}{d_\lambda}\mathbb{I}_\lambda \tag{91}$$

since the LHS is a self-intertwiner.

Observe that since $||\lambda||_1 > 0$, for any i,j and p,q we have

$$\int_G \langle \pi_\lambda(g^2) w^\lambda_{ij}, w^\lambda_{pq}\rangle \mathrm{d}\mu(g) = \frac{\imath_\lambda}{d_\lambda}||w^\lambda_{ij}||^2 \delta_{ip}\delta_{jq}. \tag{92}$$

Hence, for any $w \in S(\mathcal{H}_\lambda)$

$$\int_G \langle \pi_\lambda(g^2) w, w\rangle \mathrm{d}\mu(g) = \frac{\imath_\lambda}{d_\lambda}, \tag{93}$$

so for any $\lambda \in \Lambda_t$, there exists $h = h(w) \in G_d$ such that $\mathfrak{Re}\langle \pi_\lambda(h^2)w, w\rangle < \frac{\imath_\lambda}{d_\lambda}$ and

$$||(\pi_\lambda(h) - \pi_\lambda(h^{-1}))w|| > \sqrt{2\left(1 - \frac{\imath_\lambda}{d_\lambda}\right)}. \tag{94}$$

Note that if λ is quaternionic the bound is even better.

We want to connect h^2 with a square of some generator g_i^2, so that the large value of the norm (94) will propagate to the large value of $||(\pi_\lambda(g_i^2) - \pi_\lambda(e))w||$. Let

$$\mathcal{S}^2 := \{U_1^2, \ldots, U_k^2, U_1^{-2}, \ldots U_k^{-2}\} \tag{95}$$

and by $\mathcal{S}^2_{j_1\ldots j_m}$ we denote the set $\mathcal{S}^2$ without elements $U_{j_1}^2, \ldots, U_{j_m}^2$ (and their inverses),

$$\mathcal{S}^2_{j_1\ldots j_m} := \mathcal{S}^2 \setminus \{U_{j_1}^2, U_{j_1}^{-2} \ldots, U_{j_m}^2, U_{j_m}^{-2}\}. \tag{96}$$

By the assumption, each set $\mathcal{S}^2_{j_1\ldots j_m}$ is universal for $1 \leqslant m \leqslant k-2$. Consider any $\mathcal{S}^2_{j_1\ldots j_m}$ (we allow $\mathcal{S}^2$ as the special case with $m=0$). We find an $\epsilon_{j_1\ldots j_m}$-approximation of h^2 in terms of squares generators, namely we write $\tilde{h} = g_1^2 g_2^2 \ldots g^2_{\ell_{j_1\ldots j_m}}$, where each $g_i^2 \in \mathcal{S}^2_{j_1\ldots j_m}$, so that $D(h^2, \tilde{h}) < \epsilon$ and we specify $1 > \epsilon_{j_1\ldots j_m} > 0$ later. We have

$$\begin{aligned} \sqrt{2\left(1 - \frac{\imath_\lambda}{d_\lambda}\right)} < ||(\pi_\lambda(h^2) - \pi_\lambda(e))w|| &\leqslant ||(\pi_\lambda(e) - \pi_\lambda(\tilde{h}))w|| \\ &+ ||(\pi_\lambda(\tilde{h}) - \pi_\lambda(h^2))w|| \end{aligned} \tag{97}$$

so

$$\begin{aligned} ||(\pi_\lambda(e) - \pi_\lambda(\tilde{h}))w|| &\geqslant \sqrt{2\left(1 - \frac{\imath_\lambda}{d_\lambda}\right)} - ||(\pi_\lambda(\tilde{h}) - \pi_\lambda(h^2))w|| \\ &\geqslant \sqrt{2\left(1 - \frac{\imath_\lambda}{d_\lambda}\right)} - ||\pi_\lambda(\tilde{h}) - \pi_\lambda(h^2)||_{\mathrm{op}}. \end{aligned} \tag{98}$$

From the unitary invariance of operator norm

$$||\pi_\lambda(\tilde{h}) - \pi_\lambda(h^2)||_{\mathrm{op}} = ||\pi_\lambda(e) - \pi_\lambda(\bar{h})||_{\mathrm{op}}, \tag{99}$$

where $\bar{h} = \tilde{h}^{-1}h^2$ and $D(e,\bar{h}) < \epsilon_{j_1\ldots j_m}$. Let us fix a maximal torus $T \subset G_d$ with Lie algebra $\mathfrak{t} \subset \mathfrak{su}(d)$. We can write $\bar{h} = gtg^{-1}$, where $t \in T$ and $g \in G$. Clearly,

$$\mathrm{Spec}(\pi_\lambda(\bar{h})) = \mathrm{Spec}(\pi_\lambda(t)). \tag{100}$$

Let $\{e^{i\gamma_1},\ldots,e^{i\gamma_{d_\lambda}}\}$ be the spectrum of $\pi_\lambda(t)$ and $\{w_1,\ldots,w_{d_\lambda}\}$ be an orthonormal basis of $\mathcal{H}_\lambda$ in which

$$\pi_\lambda(t)w_j = e^{i\gamma_j}w_j \tag{101}$$

for $1 \leqslant j \leqslant d_\lambda$. By the definition of a real weight we have

$$|\gamma_j| = |\langle \mu_j, H\rangle| = |(H_{\mu_j}, H)| \leqslant ||\lambda||_1 \cdot \max_i|\theta_i| \tag{102}$$

for some weight μ_j of irrep π_λ and $H = \log(t) = \mathrm{diag}(i\theta_1,\ldots,i\theta_d) \in \mathfrak{t}$[5]. We assume $\theta_i \in (-\pi,\pi]$ for each i. Since $D(e,t) < \epsilon_{j_1\ldots j_m}$ we have

$$|\theta_i|/\pi \leqslant\ |\sin(\theta_i/2)| < \epsilon_{j_1\ldots j_m}/2 \tag{103}$$

for each i, so

$$|\gamma_j| \leqslant \pi||\lambda||_1\epsilon_{j_1\ldots j_m}/2. \tag{104}$$

Finally,

$$||\pi_\lambda(e) - \pi_\lambda(\bar{h})||_{\mathrm{op}} = ||\pi_\lambda(e) - \pi_\lambda(t)||_{\mathrm{op}} \leqslant 2\max_i|\sin(\gamma_i/2)| \leqslant \max_i|\gamma_i| \tag{105}$$

hence

$$||\pi_\lambda(e) - \pi_\lambda(\bar{h})||_{\mathrm{op}} \leqslant C||\lambda||_1 \cdot \epsilon_{j_1\ldots j_m}, \tag{106}$$

where $C = \pi/2$. Thus,

$$||(\pi_\lambda(e) - \pi_\lambda(\tilde{h}))w|| \geqslant \sqrt{2\left(1 - \frac{\imath_\lambda}{d_\lambda}\right)} - C||\lambda||_1 \cdot \epsilon_{j_1\ldots j_m}. \tag{107}$$

We use triangle inequality to propagate the result into some generator

$$\begin{aligned}\sqrt{2\left(1 - \frac{\imath_\lambda}{d_\lambda}\right)} - C||\lambda||_1\epsilon_{j_1\ldots j_m} &\leqslant ||(\pi_\lambda(e) - \pi_\lambda(\tilde{h}))w|| \\ &= ||(\pi_\lambda(e) - \pi_\lambda(g_1^2g_2^2\ldots g_\ell^2))w|| \leqslant ||(\pi_\lambda(e) - \pi_\lambda(g_1^2)))w|| \\ &\quad + ||(\pi_\lambda(g_1^2) - \pi_\lambda(g_1^2g_2^2)))w|| + \ldots \\ &\quad + ||(\pi_\lambda(g_1^2g_2^2\ldots g_{i-1}^2) - \pi_\lambda(g_1^2g_2^2\ldots g_i^2)))w|| + \ldots + ||(\pi_\lambda(g_1^2g_2^2\ldots g_{\ell-1}^2) \\ &\quad - \pi_\lambda(g_1^2g_2^2\ldots g_\ell^2)))w||\end{aligned} \tag{108}$$

so there exists i such that

$$||(\pi_\lambda(g_1^2g_2^2\ldots g_{i-1}^2) - \pi_\lambda(g_1^2g_2^2\ldots g_i^2)))w|| \geqslant b_{j_1\ldots j_m}(\lambda), \tag{109}$$

[5] Note that $D(e,\bar{h}) = D(e,t) < \epsilon_{j_1\ldots j_m} < 1$ so $\log(t)$ exists.

where

$$b_{j_1\ldots j_m}(\lambda) = \frac{\sqrt{2\left(1-\frac{\imath_\lambda}{d_\lambda}\right)} - C||\lambda||_1\epsilon_{j_1\ldots j_m}}{\ell_{j_1\ldots j_m}} \tag{110}$$

and from the unitary invariance of operator norm

$$||(\pi_\lambda(g_1^2 g_2^2\ldots g_{i-1}^2) - \pi_\lambda(g_1^2 g_2^2\ldots g_i^2)))w|| = ||(\pi_\lambda(e) - \pi_\lambda(g_i^2)))w||. \tag{111}$$

Since g_i^2 is an element from $\mathcal{S}^2_{j_1\ldots j_m}$ and

$$||(\pi_\lambda(e) - \pi_\lambda(g_i^2))w|| = ||(\pi_\lambda(e) - \pi_\lambda(g_i^{-2}))w||, \tag{112}$$

there exists i_q, where $1 \leqslant q \leqslant m$, such that

$$||(\pi_\lambda(e) - \pi_\lambda(U_{i_q}^2))w|| \geqslant b_{j_1\ldots j_m}(\lambda) \geqslant b_m(\lambda), \tag{113}$$

where

$$b_m(\lambda) := \min_{i_1,\ldots,i_m} b_{i_1\ldots i_m}(\lambda) \tag{114}$$

is the bound for the worst choice of $i_1,\ldots,i_m$, which we denote $\mathcal{S}_m^2$[6]. The set $\mathcal{S}_m^2$ has the corresponding ε_m and ℓ_m via equation (110).

We proceed as follows. First, we consider above procedure for $\mathcal{S}^2$ and obtain

$$||(\pi_\lambda(e) - \pi_\lambda(U_{i_1}^2))w|| \geqslant b_0(\lambda), \tag{115}$$

for some $U_{i_1}^2 \in \mathcal{S}^2$. Next, we repeat the argument for $\mathcal{S}_{i_1}^2$ and get

$$||(\pi_\lambda(e) - \pi_\lambda(U_{i_2}^2))w|| \geqslant b_{i_1}(\lambda) \geqslant b_1(\lambda), \tag{116}$$

for some $U_{i_2}^2 \in \mathcal{S}_{i_1}^2$. We proceed in this manner until $m = k-2$, which gives

$$||(\pi_\lambda(e) - \pi_\lambda(U_{i_{k-1}}^2))w|| \geqslant b_{i_1 i_2\ldots i_{k-2}}(\lambda) \geqslant b_{k-2}(\lambda), \tag{117}$$

for some $U_{i_{k-1}}^2 \in \mathcal{S}_{i_1\ldots i_{k-2}}^2$.

This way we obtain bounds for each pair generators except for one pair $\{U_{i_k}^2, U_{i_k}^{-2}\}$, where $1 \leqslant i_k \leqslant k$ is the remaining index. Thus, using equation (89), for all i_m with $m \in \{1,\ldots,k-1\}$ we have

$$||(\pi_\lambda(U_{i_m}) + \pi_\lambda(U_{i_m}^{-1})))w|| \leqslant \sqrt{4 - b_{m-1}^2(\lambda)} \tag{118}$$

provided that

$$\epsilon_m \leqslant \frac{\sqrt{2\left(1-\frac{\imath_\lambda}{d_\lambda}\right)}}{C||\lambda||_1}, \epsilon_m < 1. \tag{119}$$

For $\ell_{m-1} \gg 1$, the good approximation is

$$\sqrt{4 - b_{m-1}^2(\lambda)} \leqslant 2 - \left(\frac{b_{m-1}(\lambda)}{2}\right)^2. \tag{120}$$

Hence,

$$\mathrm{gap}_t(\mathcal{S}) \geqslant \min_{\lambda\in\tilde{\Lambda}_t} \frac{1}{8k}\sum_{m=0}^{k-2} b_m^2(\lambda) > 0. \tag{121}$$

[6] Note that $b_m(\lambda) \geqslant b_n(\lambda)$ for $m < n$.

Using similar argument by considering only $\mathcal{S}^2_{i_1\ldots i_{k-2}}$ we have the following, weaker bound

$$\mathrm{gap}_t(\mathcal{S}) \geqslant \min_{\lambda\in\tilde{\Lambda}_t} \frac{k-1}{8k} b^2_{k-2}(\lambda) > 0. \tag{122}$$

Indeed, comparing equations (121) and (122) we have the inequality

$$\frac{1}{8k}\sum_{m=0}^{k-2} b^2_m(\lambda) \geqslant \frac{k-1}{8k} b^2_{k-2}(\lambda). \tag{123}$$

Moreover, the ratio between LHS and RHS of equation (123) is

$$\frac{\frac{1}{k-1}\sum_{m=0}^{k-2} b^2_m(\lambda)}{b^2_{k-2}(\lambda)}, \tag{124}$$

i.e. it is ratio between the average of a nonincreasing sequence $b^2_0, b^2_1, \ldots, b^2_{k-2}$ and its smallest element b^2_{k-2}. Since we expect this sequence to (generically) quickly decrease, we suppose that the bound (121) is (relatively) much better than (122), at least generically.

It remains to somehow simultaneously bound the coefficients $b_m(\lambda)$ for all $\lambda \in \tilde{\Lambda}_t$. Since $\ell_m \leqslant \mathrm{diam}_\epsilon(G, \mathcal{S}^2_m)$, from (121), (110) we obtain the bound for the gap from the diameter

$$\mathrm{gap}_t(\mathcal{S}) \geqslant \min_{\lambda\in\tilde{\Lambda}_t} \frac{1}{8k}\sum_{m=0}^{k-2}\left(\sqrt{2\left(1-\frac{\imath_\lambda}{d_\lambda}\right)} - C||\lambda||_1\epsilon_m(\lambda)\right)^2 \frac{1}{\mathrm{diam}_{\epsilon_m(\lambda)}(G,\mathcal{S}^2_m)^2} \tag{125}$$

valid for

$$0 < \epsilon_m(\lambda) \leqslant \frac{\sqrt{2\left(1-\frac{\imath_\lambda}{d_\lambda}\right)}}{C||\lambda||_1}, \epsilon_m(\lambda) < 1, \tag{126}$$

which can be weakened to the following simplified bound

$$\mathrm{gap}_t(\mathcal{S}) \geqslant \frac{1}{8k}\sum_{m=0}^{k-2}(1-2Ct\epsilon_m)^2 \frac{1}{\mathrm{diam}_{\epsilon_m}(G,\mathcal{S}^2_m)^2} \tag{127}$$

valid for

$$0 < \epsilon_m \leqslant \frac{1}{2Ct}, \epsilon_m < 1. \tag{128}$$

We have a trade-off between the contribution of ε_m to the numerator of multiplicative term (the smaller the ε_m the better) and to the diameter (the larger the ε_m the better).

Because we do not know how $\mathrm{diam}_{\epsilon_m}(G,\mathcal{S}^2_m)$ depends on ε_m, in order to proceed we can use Solovay–Kitaev theorem for $\mathcal{S}^2_m$ to bound

$$\mathrm{diam}_{\epsilon_m}(G,\mathcal{S}^2_m) \leqslant A_m \cdot \log^c\left(\frac{1}{c_s^2\epsilon_m}\right), \quad A_m = \frac{1}{\left[2\log\left(\frac{1}{c_s\epsilon_{0,m}}\right)\right]^c}\ell_{0,m} \tag{129}$$

where $c = \log(5)/\log(3/2) \approx 4$, c_s is some constant ($c_s = d + 2 + \mathcal{O}(\epsilon)$), $\epsilon_{0,m}$ is the ε of initial approximation in Solovay–Kitaev algorithm and $\ell_{0,m}$ is the word length of this approximation. Thus,

$$\mathrm{gap}_t(\mathcal{S}) \geqslant \frac{1}{8k}\sum_{m=0}^{k-2}\left(\frac{1-2Ct\epsilon_m}{A_m}\right)^2 \log^{-2c}(c_s^{-2}\epsilon_m^{-1}) \tag{130}$$

for any

$$0 < \epsilon_m \leqslant \frac{1}{2Ct}, \epsilon_m < 1. \tag{131}$$

We can bound $\ell_{0,m}$ by equation (77),

$$\ell_{0,m} \geqslant \frac{(d^2-1)(2\log\left(\frac{1}{\epsilon_{0,m}}\right) + \log(4C_b^{3/2}d)) + \log(32)}{\mathrm{gap}_{t_{0,m}}(\mathcal{S}_m^2)}, \tag{132}$$

for $C_b = 9\pi$ and $t_{0,m} \geqslant 5d^{5/2}/\epsilon_{0,m} \cdot \tau(\epsilon_{0,m}, d)$.

For simplicity, we set the common $\epsilon_{0,m} = \epsilon_0$ and put $\epsilon_m = 1/(4Ct)$, which yields

$$\begin{aligned}\mathrm{gap}_t(\mathcal{S}) &\geqslant \frac{1}{32k}\sum_{m=0}^{k-2}\frac{1}{A_m^2}\log^{-2c}(4c_s^{-2}Ct) \\ &= \frac{1}{32k}\sum_{m=0}^{k-2}\frac{1}{\ell_{0,m}^2}\left[2\log(c_s^{-1}\epsilon_0^{-1})\right]^{2c}\log^{-2c}(4c_s^{-2}Ct),\end{aligned} \tag{133}$$

and by setting the common scale $t_{0,m} = t_0 := 5d^{5/2}/\epsilon_0 \cdot \tau(\epsilon_0, d)$ and using equation (132) to set the value of $\ell_{0,m}$ we obtain

$$\begin{aligned}\mathrm{gap}_t(\mathcal{S}) \geqslant \frac{1}{32k}\sum_{m=0}^{k-2}&\frac{\mathrm{gap}_{t_0}^2(\mathcal{S}_m^2)}{\left((d^2-1)(2\log(\epsilon_0^{-1}) + \log(4C_b^{3/2}d)) + \log(32)\right)^2} \\ &\times\left[\frac{2\log(c_s^{-1}\epsilon_0^{-1})}{\log(4c_s^{-2}Ct)}\right]^{2c},\end{aligned} \tag{134}$$

which can be rewritten as

$$\mathrm{gap}_t(\mathcal{S}) \geqslant \alpha \cdot g_{t_0}(\mathcal{S}) \cdot \log^{-2c}(\beta t), \tag{135}$$

where α and β are

$$\alpha := \frac{\left[2\log(c_s^{-1}\epsilon_0^{-1})\right]^{2c}}{16 \cdot \left((d^2-1)(2\log(\epsilon_0^{-1}) + \log(4C_b^{3/2}d)) + \log(32)\right)^2}, \quad \beta := \frac{4C}{c_s^2}, \tag{136}$$

and

$$g_{t_0}(\mathcal{S}) := \frac{1}{|\mathcal{S}|}\sum_{m=0}^{|\mathcal{S}|/2-2}\mathrm{gap}_{t_0}^2(\mathcal{S}_m^2). \tag{137}$$

Finally, we can redefine $\mathrm{gap}_{t_0}(\mathcal{S}_m^2)$ to be the smallest value of a gap at scale t_0 over all sets $\mathcal{S}_{i_1\ldots i_m}^2$,

$$\mathrm{gap}_{t_0}(\mathcal{S}_m^2) := \mathrm{argmin}_{i_1,\ldots,i_m}\mathrm{gap}_{t_0}(\mathcal{S}_{i_1\ldots i_m}^2) \tag{138}$$

and this way $g_{t_0}(\mathcal{S})$ can be determined numerically by the calculations at scale

$$t_0 := 5d^{5/2}/\epsilon_0 \cdot \tau(\epsilon_0, d). \tag{139}$$

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

This research was funded by the National Science Centre, Poland under the grant OPUS: UMO-2020/37/B/ST2/02478.

ORCID iD

Oskar Słowik https://orcid.org/0000-0003-4138-3063

References

[1] Knill E, Laflamme R and Zurek W H 1998 Resilient quantum computation *Science* **279** 342–5
[2] Kitaev A Y 2003 Fault-tolerant quantum computation by anyons *Science* **303** 2–30
[3] Aharonov D and Ben-Or M 2008 Fault-tolerant quantum computation with constant error rate *SIAM J. Comput.* **38** 1207–82
[4] Harrow A W and Montanaro A 2017 Quantum computational supremacy *Nature* **549** 203–9
[5] Preskill J 2018 Quantum computing in the NISQ era and beyond *Quantum* **2** 79
[6] Boixo S, Isakov S V, Smelyanskiy V N, Babbush R, Ding N, Jiang Z, Bremner M J, Martinis J M and Neven H 2018 Characterizing quantum supremacy in near-term devices *Nat. Phys.* **14** 595–600
[7] Sawicki A and Karnas K 2017 Universality of single-qudit gates *Ann. Henri Poincaré* **18** 3515–52
[8] Sawicki A and Karnas K 2017 Criteria for universality of quantum gates *Phys. Rev.* A **95** 062303
[9] Kesten H 1959 Symmetric random walks on groups *Trans. Am. Math. Soc.* **92** 336–54
[10] Kitaev A Y, Shen A and Vyalyi M N 2002 *Classical and Quantum Computation* (Providence, RI: American Mathematical Society)
[11] Harrow A W, Recht B and Chuang I L 2002 Efficient discrete approximations of quantum gates *J. Math. Phys.* **43** 4445–51
[12] Bourgain J and Gamburd A 2008 On the spectral gap for finitely-generated subgroups of SU(2) *Invent. Math.* **171** 83–121
[13] Bourgain J and Gamburd A 2012 A spectral gap theorem in SU(d) *J. Eur. Math. Soc.* **14** 1455–511
[14] Lubotzky A, Phillips R and Sarnak P 1986 Hecke operators and distributing points on the sphere I *Communications on Pure and Applied Mathematics. Supplement: Proc. Symp. on Frontiers of the Mathematical Sciences: 1985* vol 39 pp S149–86
[15] Lubotzky A, Phillips R and Sarnak P 1987 Hecke operators and distributing points on S2. II *Communications on Pure and Applied Mathematics* **40** 401–20
[16] Bocharov A, Gurevich Y and Svore K M 2013 Efficient decomposition of single-qubit gates into V basis circuits *Phys. Rev.* A **88** 012313
[17] Selinger P 2015 Efficient Clifford+T approximation of single-qubit operators *Quantum Inf. Comput.* **15** 159–80
[18] Sarnak P 2015 *Letter to Scott Aaronson and Andy Pollington on the Solovay-Kitaev theorem*
[19] Kliuchnikov V, Maslov D and Mosca M 2016 Practical approximation of single-qubit unitaries by single-qubit quantum Clifford and T circuits *IEEE Trans. Comput.* **65** 161–72
[20] Varjú P P 2013 Random walks in compact groups *Doc. Math.* **18** 1137–75
[21] Oszmaniec M, Sawicki A and Horodecki M 2022 Epsilon-nets, unitary designs and random quantum circuits *IEEE Trans. Inf. Theory* **68** 989–1015
[22] Szarek S J 1998 Metric Entropy of Homogeneous Spaces *Banach Center Publications, Institute of Mathematics, Polish Academy of Sciences* **43**
[23] Sawicki A, Mattioli L and Zimborás Z 2021 How to check universality of quantum gates? (arXiv:2111.03862)

Chapter 4

Paper II: Fundamental solutions of the heat equation on unitary groups establish an improved relation between ϵ-nets and approximate unitary t-designs

4.1 Overview

In this second paper, our goal was to establish the quantitative correspondence between unitary (δ-approximate) t-designs and ϵ-nets using heat kernels on the space of unitary channels $\mathbf{U}(d)$.

Our point of reference was the work [41], in which authors prove that δ-approximate t-designs form ϵ-nets for $t \simeq d^{5/2}/\epsilon$ and $\delta \simeq \left(\epsilon^{3/2}/d\right)^{d^2}$ (see [41] for precise statement and formulas). The authors construct polynomial approximations to the Dirac delta by periodising Gaussians on the torus. Our approach is, in our opinion, more natural as it is based on trimming the heat kernels.

The main idea of our proof is the same as in [41], and is based on the following observation (see also Section 2.3.5). Let ν be a discrete probability measure on $\mathbf{U}(d)$ which does not form an ϵ-net. Then, we can pick an element $\mathbf{V_0}$, so that every element from the

support of ν is at least ϵ-separated from $\mathbf{V_0}$. We choose some polynomial approximate identity $\varphi_t \in \mathcal{H}_t$ on $\mathbf{U}(d)$, apply the averaging operator (2.71) and integrate the resulting smeared density, over $\epsilon/2$-ball around $\mathbf{V_0}$. Then, from (2.134), the value of the integral is $\mu(B_{\epsilon/2}(\mathbf{V_0}))$, but since φ_t is an approximate identity, the value of this integral must vanish as φ_t goes towards the Dirac-delta distribution, i.e. $t \to \infty$. A similar argument can be applied to the δ-approximate case, using (2.138) instead. Such a proof by contradiction can be used to extract the scaling of t and δ in d and ϵ. Clearly, the resulting scaling depends on the polynomial approximate identity used.

We obtain the heat kernel on $\mathrm{PU}(d)$ by applying the averaging map to the heat kernel on $\mathrm{SU}(d)$ written in two forms: the character decomposition (c.f. 2.110) and the Poisson form [58]. Each form is applied to a different part of the proof. Intuitively, the Poisson summation formula moves the parameter σ to the denominator of the exponent, which helps bound the behaviour of the heat kernel for small σ and obtain the bounds on the vanishing of integrals outside ϵ-balls. On the other hand, the character form is used to determine the L^2 trimming error.

The proof can be divided into 5 main steps.

1. We "trim" the heat kernel on $\mathrm{PU}(d)$ to obtain a balanced polynomial approximation of order t. Moreover, we prove an L^2-norm error bound for this approximation and argue that the trimming procedure is optimal.

2. We show that the heat kernel on $\mathrm{PU}(d)$ is an approximation to the Dirac delta. In particular, its integral vanishes outside any ϵ-ball as $\sigma \to 0$. We provide bounds on the rate at which such an integral vanishes.

3. We combine bounds from 1. and 2. to obtain a bound for the integral of the absolute value of the trimmed heat kernel outside an ϵ-ball. We use such bounds to show that the trimmed heat kernel is a polynomial approximation to the Dirac delta.

4. We provide the upper bounds on the L^2-norm of the heat kernel on $\mathrm{PU}(d)$.

5. Using the bounds from 3. and 4., we obtain the scaling of t and δ sufficient for a unitary δ-approximate t-design to form an ϵ-net. This step essentially uses the main idea of the proof outlined above.

The resulting bounds on t and δ are provided as explicit formulas in the main theorem of the paper - Theorem 2, and enjoy the scaling of $t \simeq d^{5/2}/\epsilon$ and $\delta \simeq \left(\epsilon^{3/2}/d\right)^{d^2}$. Essentially,

compared to [41], we were able to significantly improve the scaling of δ while retaining the same scaling of t. We also provide the proof for the (ideal) t-design case, with the same scaling of t. Additionally, we summarise the properties of the polynomial approximations to the Dirac delta on $\mathrm{PU}(d)$ we constructed, as approximate identities. Such characterisation is provided in Theorem 3 and includes the bounds on the vanishing of integrals outside ϵ-balls, blow-up of the L^2-norm, and approximate non-negativity (via the L^1-norm estimates). All the formulas are provided with explicit constants for $\mathrm{PU}(d)$.

In Section III, we comment on the possible applications of our results in areas such as the efficiency of quantum gates (in particular, Quantum Circuit Overhead, which is a subject of Paper III), inverse-free SKL theorems, and quantum complexity and black hole physics. Finally, we suspect that our construction of polynomial approximations of the Dirac delta based on heat kernels can find applications in other domains of science. In particular, due to tunable filtering properties (see Section (2.2.6)) and good analytic control, we suspect that trimmed heat kernels can be used to obtain an alternative and appealing proof to the poly-logarithmic spectral gap decay results for $\mathrm{PU}(d)$, analogous to (2.140).

4.2 Contribution statement

My contribution to this article was:

1. Writing the bulk of Sections I-III and VII and a significant portion of Section IV in the main text. In particular: Remark 1, Applications 1-3, Examples 1 and 2.
2. Co-writing of Sections V, VI, and VIII. In particular: Remarks 2-5, proof of Lemma 1, joint proof of Theorem 1 and Theorem 2, proof of Theorem 3.
3. Preparation of Appendix C, D, and E. In particular: the proofs of Lemmas 9-16 and Corollary 1.

IOP Publishing Journal of Physics A: Mathematical and Theoretical
J. Phys. A: Math. Theor. **58** (2025) 445301 (45pp) https://doi.org/10.1088/1751-8121/ae0edc

Fundamental solutions of the heat equation on unitary groups establish an improved relation between ϵ-nets and approximate unitary t-designs

Oskar Słowik[1,*], **Oliver Reardon-Smith**[1] **and Adam Sawicki**[1,2]

[1] Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warszawa, Poland
[2] Guangdong Technion—Israel Institute of Technology, 241 Daxue Road, Jinping District, Shantou, Guangdong Province, People's Republic of China

E-mail: oslowik@cft.edu.pl

Received 8 July 2025; revised 18 September 2025
Accepted for publication 1 October 2025
Published 28 October 2025

Abstract

The concepts of ϵ-nets and unitary (δ-approximate) t-designs are important and ubiquitous across quantum computation and information. Both notions are closely related and the quantitative relations between t, δ and ϵ find applications in areas such as (non-constructive) inverse-free Solovay–Kitaev like theorems and random quantum circuits. In recent work, quantitative relations have revealed the close connection between the two constructions, with ϵ-nets functioning as unitary δ-approximate t-designs and vice-versa, for appropriate choice of parameters. In this work we improve these results, significantly increasing the bound on the δ required for a δ-approximate t-design to form an ϵ-net from $\delta \simeq \left(\epsilon^{3/2}/d\right)^{d^2}$ to $\delta \simeq \left(\epsilon/d^{1/2}\right)^{d^2}$. We achieve this by constructing polynomial approximations to the Dirac delta using heat kernels on the projective unitary group $\mathrm{PU}(d) \cong \mathbf{U}(d)$, whose properties we studied and which may be applicable more broadly. We also outline the possible applications of our results in quantum circuit overheads, quantum complexity and black hole physics.

Keywords: unitary t-designs, quantum complexity, Solovay–Kitaev theorem, heat kernel

* Author to whom any correspondence should be addressed.

1. Introduction

Unitary t-designs are a fundamental construction, finding widespread applications across quantum information and computation. They have been employed in areas such as randomised benchmarking [1, 2], process tomography [3], quantum information protocols [4, 5], unitary codes [6], derandomisation of probabilistic constructions [7], decoupling [8], entanglement detection [9], quantum state discrimination [10], shadow estimation [11], efficient quantum measurements [12] and estimation of the properties of quantum systems [13]. Moreover, their link to pseudo-random quantum circuits [14] makes them applicable in areas such as the equilibration of quantum systems [14, 15], quantum metrology with random bosonic states [16], quantum complexity and information scrambling in black holes [17–20]. They have also been applied to the study of quantum speed-ups [21–23], due to their anti-concentration property [24, 25].

Epsilon-nets are of similar importance, finding broad application and, in particular, serving as the natural language for quantum compilation. Solovay–Kitaev like (SKL) theorems [26, 27] provide joint bounds on the complexity of quantum operations $\mathbf{U}$, for a given error ϵ and gateset $\mathcal{S}$. In other words, they bound how many operations are required for circuits of gates from a given gateset to form an ϵ-net. Moreover, constructive SKL theorems say how to find the approximating circuits, which makes them the cornerstone of quantum compilation. The original SK theorem, which is constructive, bounds the length of the sequence of gates as $\ell = \mathcal{O}(\log^c\left(\frac{1}{\epsilon}\right))$, where $c \approx 3.97$. In fact, it is well-known that any $c > 3$ works and recently a constructive SKL theorem with $c \approx 1.44$ was provided in [26], which is significantly closer to the optimal value $c = 1$.

The parameter δ of the (unitary) δ-approximate t-design generated by $\mathcal{S}$ can be studied on finite scales, say given by the highest considered degree t, denoted $\delta(\nu_{\mathcal{S}}, t)$. Such a finite-scale approach was explored in [28, 29]. For fixed ϵ and $\mathcal{S}$, the knowledge of $\delta(\nu_{\mathcal{S}}, t)$ at a suitably chosen scale $t(\epsilon)$, is sufficient to bound ℓ via a non-constructive SKL theorem $\ell = \mathcal{O}(\frac{1}{\log\left(1/\delta\left(\nu_{\mathcal{S},t(\epsilon)}\right)\right)}\log\left(\frac{1}{\epsilon}\right))$ with explicit form (see e.g. [28]). Such SKL theorems can be used to bound the efficiency of various gate sets $\mathcal{S}$, e.g. their (T-)Quantum Circuit Overhead [27]. In particular, if the supremum of $\delta(\nu_{\mathcal{S}}, t)$ over all t is smaller than 1, then we obtain an asymptotically optimal scaling $\ell = \Theta(\log(\frac{1}{\epsilon}))$ [30, 31]. However, the analysis of such a supremum is a hard problem and is computationally intractable. Hence, the SKL theorems based on a finite-scale $\delta(\nu_{\mathcal{S}}, t(\epsilon))$ are of significant practical interest. The tightness of such theorems depends on the tightness of the $t(\epsilon)$ scaling, which can be understood as the t sufficient for a (δ-approximate) t-design to form an ϵ-net.

In light of the importance of both ϵ-nets and t-designs, it is interesting that there is a strong link between the two constructions. Indeed a (possibly approximate) t-design of sufficiently large t forms an ϵ-net, while an ϵ-net of sufficiently small ϵ forms an approximate t-design. To our knowledge, the first systematic study of the quantitative relations between them was surprisingly recent, in [28], where the authors show that an ϵ-net is formed by δ-approximate t-designs on the space of unitary channels $\mathbf{U}(d)$ for $t \simeq \frac{d^{5/2}}{\epsilon}$ and $\delta \simeq \left(\frac{\epsilon^{3/2}}{d}\right)^{d^2}$, where $\simeq$ can be understood as 'ignoring logarithmic factors' and 'infinitesimal corrections to the exponents'. The authors of [28] were able to prove that t has to grow at least as $1/\epsilon$ (for fixed d) and as d^2 (for fixed ϵ). Thus they were able to show that this scaling of t with ϵ is essentially optimal, while the scaling of t with d is (at worst) not very far from optimal, with a 'gap' of $\sqrt{d}$ between the known lower and upper bounds. They conjectured that a scaling of $t \simeq d^2$ was possible but were not able to prove this.

In this work we build on these results, obtaining (up to logarithmic factors) the same scaling of t as the authors of [28], but dramatically improving the scaling of δ with ϵ and d in the δ-approximate case. We are able to show that a δ-approximate t-design forms an ϵ-net if delta obeys an inequality which scales like $\delta \simeq \left(\frac{\epsilon}{d^{\frac{1}{2}}}\right)^{d^2}$.

Our method involves the construction of a polynomial approximation to a Dirac delta on the space of quantum unitary channels $\mathrm{PU}(d)$. Our construction of the approximate Dirac delta is a natural one, based on the properties of the heat kernel on $\mathrm{SU}(d)$. As running the evolution of the heat equation 'forwards' leads to 'heat' spreading out over time, naturally running it backward and considering very small times leads to a sharp delta-like peak at times close to 0. As has been known since the work of Fourier himself, the heat equation has a close link to Fourier analysis on the appropriate space. Indeed, our key bounds are based on the results from [32], which may be viewed as a generalization of the well-known Poisson formula [33] to compact semi-simple simply-connected Lie groups. The heat kernel is at the heart of many important methods across mathematical physics and beyond. As a tool to study the eigenvalues and eigenfunctions of the Laplacian, it has been used as far back as Kac's seminal 1966 paper 'Can One Hear the Shape of a Drum' [34] and has been of prime importance in the study of the Laplacian on Riemannian manifolds throughout the subject's history [35]. It has been applied extensively in the context of quantum field theory [36], mathematical finance [37] and quantum gravity [38] and used to prove a diverse range of important theorems, including the Atiyah–Singer index theorem [39, 40] and the Poincaré conjecture [41, 42]. For a thorough review, we invite the reader to the textbooks [37, 43].

Our core results—the bounds on t and δ are the subjects of theorem 1 (for t-designs) and theorem 2 (for δ-approximate t-designs). We also provide a technical result about the properties of our approximate Dirac delta (theorem 3), which may be useful for other applications.

Outline of the proof—the proof of the main theorem (theorem 2) can be divided into five steps:

1. We 'trim' the full heat kernel on $\mathrm{PU}(d)$ to obtain an approximation of it by a balanced polynomial of order t, and prove an error bound for this approximation.
2. We prove that the heat kernel on $\mathrm{PU}(d)$ is an approximation to the Dirac delta. In particular, its integral vanishes outside any ϵ-ball as $\sigma \to 0$ at a rate we can bound.
3. By combining the above two bounds, we obtain a bound for the integral of the absolute value of the trimmed heat kernel outside an ϵ-ball, thereby showing the trimmed heat kernel is also an approximation to the Dirac delta.
4. We derive the bounds on the L^2-norm of the heat kernel on $\mathrm{PU}(d)$.
5. We combine the bounds to obtain a bound for the t and δ sufficient for a projective unitary δ-approximate t-design to be an ϵ-net. Essentially, this argument follows from applying the t-design property to the order t balanced polynomial we obtained in step 1.

Structure of the paper—the paper is organised as follows:

- In section 2, we briefly explain the main ideas behind the paper, such as ϵ-nets, t-designs and heat kernels.
- In section 3 we summarise the **main results** and their **applications**.
- In section 4 we address step 1 of the proof.
- In section 5, we address steps 2 and 3 of the proof and combine them to prove a bound for the t sufficient for a projective unitary t-design to be an ϵ-net (theorem 1).

- In section 6 address step 4 of the proof and combine the bounds from steps 2–4 to derive the bounds on t and δ sufficient for a projective unitary δ-approximate t-design to be an ϵ-net, realising step 5 of the proof (theorem 2).
- In section 7, we summarize the technical properties of trimmed heat kernels as approximations to the Dirac delta (theorem 3).
- Finally, in section 8, we provide a summary and outline the future research directions.
- The appendix contains the proofs of various technical lemmas.

2. Main ideas

Central in quantum information theory is the concept of unitary channels. Such channels act via unitary operations (lossless quantum gates) when restricted to pure quantum states.

The unitary channel $\mathbf{U}$ acting on a Hilbert space $\mathcal{H} \cong \mathbb{C}^d$ is the CPTP map defined via $\mathbf{U}(\rho) = U\rho U^\dagger$, for any quantum state $\rho : \mathcal{H} \to \mathcal{H}$ and some fixed unitary $U \in \mathrm{U}(d)$. Since two unitaries U and V which differ by a phase $U = V\mathrm{e}^{\mathrm{i}\phi}$ define the same unitary channel, the set of all unitary channels can be identified with the projective unitary group $\mathrm{PU}(d) = \mathrm{U}(d)/\mathcal{Z}(\mathrm{PU}(d))$, where $\mathcal{Z}(\mathrm{PU}(d)) = \{\mathrm{e}^{\mathrm{i}\phi} I, \phi \in (-\pi, \pi]\} \cong \mathrm{U}(1)$ is the centre of $\mathrm{U}(d)$.

Since we prefer to work with the $\mathrm{SU}(d)$ group, in our considerations we assume $U \in \mathrm{SU}(d)$ and use $\mathrm{PU}(d) = \mathrm{SU}(d)/\mathcal{Z}(\mathrm{SU}(d))$, where $\mathcal{Z}(\mathrm{SU}(d)) = \{\mathrm{e}^{\mathrm{i}\frac{2\pi}{d}k} I, k \in \mathbb{Z}\} \cong \mathbb{Z}_d$ is the centre of $\mathrm{SU}(d)$ (group of dth roots of unity). From now on we denote $\Gamma := \mathcal{Z}(\mathrm{SU}(d))$ and use square brackets to denote the elements of the projective group as equivalence classes of elements of $\mathrm{SU}(d)$, i.e. U is mapped to the unitary channel $\mathbf{U}$ under the quotient map $\pi : \mathrm{SU}(d) \to \mathrm{PU}(d)$.

In practice, one is often interested in the closeness of different unitary channels. Various norms (and induced metrics) can be used to quantify this. A prominent example is the diamond norm $||\cdot||_\diamond$, which has a clear operational meaning in terms of the statistical distinguishability of two channels (e.g. determines the maximal probability of success in a single-shot channel discrimination task). We denote the induced metric as $d_\diamond(\mathbf{U}, \mathbf{V}) = ||\mathbf{U} - \mathbf{V}||_\diamond$.

We define $d(\cdot,\cdot)$ to be a metric on $\mathrm{SU}(d)$ induced by the operator norm

$$d(U, V) := \|U - V\|_\infty. \tag{1}$$

Since we want to work with the group $\mathrm{SU}(d)$, we define the metric on $\mathrm{PU}(d)$ in terms of the former

$$d_P(\mathbf{U}, \mathbf{V}) := \min_{\gamma \in \Gamma} d(U, \gamma V). \tag{2}$$

Clearly, due to the unitary invariance of the operator norm, the metrics $d(\cdot,\cdot)$ and $d_P(\cdot,\cdot)$ are translation-invariant.

One may show [28] that $d_\diamond(\cdot,\cdot)$ and $d_P(\cdot,\cdot)$ are related as

$$d_P(\mathbf{U}, \mathbf{V}) \leqslant d_\diamond(\mathbf{U}, \mathbf{V}) \leqslant 2 d_P(\mathbf{U}, \mathbf{V}). \tag{3}$$

We say that a finite subset of channels $\mathcal{A} \subset \mathrm{PU}(d)$ is an ϵ-net if for every channel $\mathbf{U} \in \mathrm{PU}(d)$, there exists a channel $\mathbf{V} \in \mathcal{A}$, such that $d_P(\mathbf{U}, \mathbf{V}) \leqslant \epsilon$. In other words, $\mathcal{A}$ represents all the possible channels, up to the error ϵ.

To consider unitary designs, we need to define integration of functions on $\mathrm{PU}(d)$. The Haar measure μ_P on $\mathrm{PU}(d)$ is the pushforward of the Haar measure μ_S on $\mathrm{SU}(d)$, i.e. $\mu_P(\mathcal{A}) = \mu_S(\pi^{-1}(\mathcal{A}))$, whenever $\pi^{-1}(\mathcal{A})$ is μ_S-measurable.

Every function f on $\mathrm{PU}(d)$ can be lifted to a unique function $\tilde{f}$ on $\mathrm{SU}(d)$, so that $\tilde{f}(U) = f(\mathbf{U})$. Clearly, such a function is constant on the equivalence classes (fibres of π), i.e. all the elements U that define the same unitary channel. Conversely, every function $\tilde{f}$ on $\mathrm{SU}(d)$ which is constant on the equivalence classes, descends to a unique function f on $\mathrm{PU}(d)$, so that $\tilde{f}(U) = f(\mathbf{U})$. Hence, we can write

$$\int_{\mathrm{PU}(d)} f \mathrm{d}\mu_P = \int_{\mathrm{SU}(d)} \tilde{f} \mathrm{d}\mu_S. \tag{4}$$

If $X \subseteq \mathrm{PU}(d)$ is some Haar-measurable set then inserting indicator functions into (4) we obtain

$$\int_X f \mathrm{d}\mu_P = \int_{\tilde{X}} \tilde{f} \mathrm{d}\mu_S, \tag{5}$$

where $\tilde{X} = \pi^{-1}(X)$. This allows us to move freely between the $\mathrm{PU}(d)$ and $\mathrm{SU}(d)$ settings.

The (unitary) t-design on $\mathrm{PU}(d)$ is the probability measure ν on $\mathrm{PU}(d)$ which mimics the averaging properties of the Haar-measure with respect to the polynomials of degree at most t. Specifically, let $\mathcal{H}_t$ denote the space of homogeneous polynomials of degree t in the matrix elements of U and in $\overline{U}$.

A probability measure ν on G is a unitary t-design if for any $f \in \mathcal{H}_t$ we have

$$\int_G \mathrm{d}\nu(U) f(U) = \int_G \mathrm{d}\mu(U) f(U). \tag{6}$$

From the practical point of view, one is often interested in the case of ν being a discrete finitely supported measure, so that the averaging takes place over a finite set of elements $\{\nu_i, U_i\}$

$$\sum_i \nu_i f(U_i) = \int_G \mathrm{d}\mu(U) f(U). \tag{7}$$

For example, ν can be the probability measure supported on a finite universal set of quantum gates $\mathcal{S} = \{U_1, U_2, \ldots, U_k\}$, which we denote as $\nu_{\mathcal{S}}$. In this work, we assume all t-designs have finite support since we directly apply lemma 2 of [28]. However, this lemma can be generalised to infinitely supported or even continuous measures, so that our results hold for such cases as well, if the definition of the ϵ-net is relaxed by removing the finiteness condition.

Moreover, it is useful to consider the cases in which (6) is satisfied only approximately. To do so, it is useful to define so-called t-moment operators

$$T_{\mu,t} := \int_G \mathrm{d}\mu(U)\, U^{t,t}, \quad T_{\nu,t} := \int_G \mathrm{d}\nu(U)\, U^{t,t}. \tag{8}$$

One may check that the space $\mathcal{H}_t$ is spanned by the entries of $U^{t,t} := U^{\otimes t} \otimes \bar{U}^{\otimes t}$. Indeed for every $f \in \mathcal{H}_t$ there exists a matrix A such that $f(U) = \mathrm{Tr}(AU^{t,t})$.

This way, the deviation from ν being a t-design (6) can be measured as (see [28])

$$\delta(\nu, t) := \|T_{\nu,t} - T_{\mu,t}\|_\infty \in [0, 1], \tag{9}$$

where $\|\cdot\|_\infty$ is an operator norm, leading to the notion of δ-approximate t-designs, for which $\delta(\nu, t) < 1$ and exact t-designs, for which $\delta(\nu, t) = 0$.

Finally, we recall that for $s < t$ we have $\mathcal{H}_s \subset \mathcal{H}_t$, hence t-designs are also s-designs.

The techniques used in this paper are similar to the ones from [28] and include the usage of the approximations to the Dirac delta on compact groups. However, in this paper, we employ approximations based on the heat kernel—a natural and well-known object, contrary to the

periodised Gaussian construction from [28]. We will first introduce the heat kernel with an elementary classical example.

Example 1 (heat equation on a circle and the Poisson summation formula). Consider a circle $S^1 \cong \mathbb{R}/\mathbb{Z}$ as an example of a 1-dimensional flat torus. Denoting the coordinate as ϕ, the metric tensor induced from the Euclidean metric on $\mathbb{R}$ is $g = d\phi^2$. Hence $\Delta = \frac{\partial^2}{\partial \phi^2}$ and the heat equation [3] on such a manifold reads $u_t(t,\phi) = u_{\phi\phi}(t,\phi)$ with the initial condition $u(0,\phi) = f(\phi)$. We assume that $f(\phi)$ is square integrable, i.e. $f(\phi) \in L^2(S^1)$. Such a problem is typically solved by considering the corresponding equation on $\mathbb{R}$ with the periodic boundary conditions, separation of variables and the expansion of the initial data $f(\phi)$ to the Fourier series. Here, we take a different approach—we find the fundamental solution to the corresponding problem on $\mathbb{R}$ and periodise it. We use the definition of the Fourier transform of a (complex) function in $L^2(\mathbb{R})$ as the unique unitary extension of the map $g \mapsto \hat{g}$, $g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, where

$$\hat{g}(\xi) = \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}2\pi\xi x} g(x)\,\mathrm{d}x, \quad \forall \xi \in \mathbb{R}. \tag{10}$$

From now on, we fix t and consider the corresponding single-variable functions on $\mathbb{R}$. We unwrap the initial datum f into an interval $[0,1) \subset \mathbb{R}$. We denote the corresponding functions on $\mathbb{R}$ using the same symbols as for S^1. Applying the Fourier transform, we obtain

$$\hat{u}_t(\xi,t) + 4\pi^2\xi^2\hat{u}(\xi,t) = 0 \tag{11}$$

with the initial condition $\hat{u}(\xi,0) = \hat{f}(\xi)$, where $4\pi^2\xi^2$ is the eigenvalue of $-\Delta$. Multiplying (11) by $\mathrm{e}^{4\pi^2\xi^2 t}$ we obtain $\frac{\partial}{\partial t}\left(\mathrm{e}^{4\pi^2\xi^2 t}\hat{u}(\xi,t)\right) = 0$. Hence $\mathrm{e}^{4\pi^2\xi^2 t}\hat{u}(\xi,t)$ is some function of ξ and from the initial condition we see that $\hat{u}(\xi,t) = \mathrm{e}^{-4\pi^2\xi^2 t}\hat{f}(\xi)$. Denoting the inverse Fourier transform of $\mathrm{e}^{-4\pi^2\xi^2 t}$ as $H_{\mathbb{R}}(\phi,t)$ we obtain

$$H_{\mathbb{R}}(\phi,t) = \frac{1}{\sqrt{4\pi t}}\mathrm{e}^{-\frac{\phi^2}{4t}}. \tag{12}$$

Thus, the solution on $\mathbb{R}$ is the convolution (with respect to the ϕ variable)

$$u(\phi,t) = (H_{\mathbb{R}}(\cdot,t) * f)(\phi) = \int_{-\infty}^{\infty} K_{\mathbb{R}}(\phi,\phi',t) f(\phi')\,\mathrm{d}\phi', \tag{13}$$

where

$$K_{\mathbb{R}}(\phi,\phi',t) = H_{\mathbb{R}}(\phi-\phi',t) = \frac{1}{\sqrt{4\pi t}}\mathrm{e}^{-\frac{(\phi-\phi')^2}{4t}}, \tag{14}$$

and $u(\phi,t)$ is smooth for all $t > 0$. To find the fundamental solution on S^1 we periodise $H_{\mathbb{R}}(\phi,t)$ obtaining a 1-periodic function on $\mathbb{R}$ and an equivalent function on $\mathbb{R}/\mathbb{Z}$

$$H_{S^1}(\phi,t) = \frac{1}{\sqrt{4\pi t}}\sum_{n\in\mathbb{Z}} \mathrm{e}^{-\frac{(\phi+n)^2}{4t}}, \quad \phi \in \mathbb{R}/\mathbb{Z}. \tag{15}$$

To rewrite (15) we can use the Poisson summation formula, which states that for a complex-valued function $s(x)$ on $\mathbb{R}$ whose all derivatives decay at infinity (i.e. a Schwartz function)

$$\sum_{n=-\infty}^{\infty} s(n) = \sum_{k=-\infty}^{\infty} \hat{s}(k). \tag{16}$$

[3] Physically we consider heat equations with unit conductivity.

Treating (15) as a one-periodic function on $\mathbb{R}$ we apply the Poisson summation formula with $s(x) = \mathrm{e}^{-\frac{(\phi+x)^2}{4t}}$ and obtain

$$H_{S^1}(\phi,t) = \sum_{k\in\mathbb{Z}} \mathrm{e}^{-i2\pi k\phi}\mathrm{e}^{-(2\pi k)^2 t}, \quad \phi \in \mathbb{R}/\mathbb{Z}, \tag{17}$$

which is the complex Fourier series expansion. Rewriting it into the sine-cosine form yields

$$H_{S^1}(\phi,t) = 1 + 2\sum_{k=1}^{\infty} \cos(2\pi k\phi)\,\mathrm{e}^{-(2\pi k)^2 t} \quad \phi \in \mathbb{R}/\mathbb{Z}, \tag{18}$$

which is a linear combination of eigenfunctions $2\cos(2\pi k\phi)$ with eigenvalues $-4\pi^2k^2$ and is of the same form as the fundamental solution obtained via the typical Fourier series expansion approach.

Generalizing the heat kernel on $\mathbb{R}$ shown in equation (12), the heat kernel on $\mathbb{R}^d$ has the form

$$K(t,x,y) = \frac{1}{(4\pi t)^{d/2}}\mathrm{e}^{-||x-y||^2/4t}, \tag{19}$$

where $||\cdot||$ is the Euclidean norm, defined for any $x, y \in \mathbb{R}^d$ and $t > 0$. This is the fundamental solution to the heat equation

$$u_t(t,x) = \Delta u(t,x), \tag{20}$$

where Δ is the Laplacian on $\mathbb{R}^d$. One can consider the generalisation of the heat equation (20) to other spaces, e.g. Riemannian manifolds (M, g), by replacing Δ with the Laplace-Beltrami operator (in local coordinates)

$$\Delta f = \frac{1}{\sqrt{|g|}}\partial_i\left(\sqrt{|g|}g^{ij}\partial_j f\right), \tag{21}$$

acting on differentiable functions f on M.

Following the method demonstrated in example 1, we can derive the heat kernel for the flat d-dimensional torus $\mathbb{R}^d/\Lambda$ by taking the heat kernel on $\mathbb{R}^d$ shown in equation (19) and periodising the solution over the lattice $\Lambda \cong \mathbb{Z}^d$. The Poisson formula also generalises to higher dimensions. A thorough introduction to this topic may be found in [44].

However, in our work, we are interested in heat kernels on Lie groups. To make sense of the heat equation on a Lie group, the proper Riemannian structure needs to be chosen. For compact semi-simple Lie group G, the Riemannian structure (G, g) can be defined via Ad-invariant positive definite inner product $(\cdot,\cdot)$ stemming from the Killing form[4].

Notice that although the group $\mathrm{U}(d)$ is compact, it is not semi-simple. Hence, the metric tensor stemming from the Killing form is only positive semi-definite. Indeed, one can check that such a metric tensor for $\mathrm{U}(1) \cong S^1$ is identically zero, so it does not equip S^1 with the Riemannian structure. This is in contrast with the construction from example 1.

Of course, general Lie groups are not commutative. Hence, in order to study the heat equation on a compact Lie group G, non-commutative Fourier/harmonic analysis is needed. Fourier coefficients on a compact Lie group are calculated with respect to the irreducible representations (irreps) of the group. Generally, the object being transformed is the regular Borel

[4] Taking the negative of the negative-definite Killing form leads to the positive-definite scalar product.

measure on G. However, we focus on the related case of integrable functions f. In this case (see e.g. [45]), the Fourier coefficient $\hat{f}(\lambda)$ is the operator in $\mathrm{End}(V_{\pi_\lambda})$ defined via

$$\hat{f}(\lambda) = \int_G \pi_\lambda\left(g^{-1}\right) f(g)\,\mathrm{d}\mu(g), \tag{22}$$

where by V_{π_λ} we denote the representation space of irrep π_λ with highest weight λ. Equipping the space $\mathrm{End}(V_{\pi_\lambda})$ with the norm $\sqrt{d_\lambda}||\cdot||_{HS}$, where $d_\lambda := \dim(V_{\pi_\lambda})$ and the Hilbert–Schmidt norm $||u||_{HS}^2 = \mathrm{Tr}(uu^*)$, one can show that such the Fourier transform is an isomorphism of Hilbert spaces

$$L^2(G) \cong \bigoplus_{\pi\in\hat{G}} \mathrm{End}(V_{\pi_\lambda}), \tag{23}$$

where $\hat{G}$ is the set of equivalence classes of irreps of G. Namely, we obtain a generalisation of the Plancherel's theorem

$$||f||_2^2 = \int_G |f(g)|^2 \mathrm{d}\mu(g) = \sum_{\lambda\in\hat{G}} d_\lambda ||\hat{f}(\lambda)||_{HS}^2. \tag{24}$$

This is a consequence of the Peter–Weyl theorem.

Remark 1. The transform (22) is a generalization of the Fourier series. Indeed, suppose a compact group G is additionally abelian and connected (so is a torus). Take one-dimensional torus $\mathrm{U}(1) \cong S^1$ for example. The unitary irreps π_λ of $\mathrm{U}(1)$ are the homomorphisms $\mathrm{U}(1) \to \mathrm{U}(1)$ so they are of the form $\mathrm{e}^{\mathrm{i}\phi} \mapsto \mathrm{e}^{\mathrm{i}\lambda\phi}$ for some integer λ. All irreps are one-dimensional and $\hat{S^1} \cong \mathbb{Z}$. The Fourier coefficients of a function $f\colon \mathrm{U}(1) \to \mathbb{C}$ are

$$\hat{f}(\lambda) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}\lambda\phi} f\left(\mathrm{e}^{\mathrm{i}\phi}\right)\mathrm{d}\phi, \tag{25}$$

which coincides with the Fourier coefficients of the corresponding 2π-periodic complex-valued function $\tilde{f}\colon \mathbb{R} \to \mathbb{C}$, $\tilde{f}(x) = f(\mathrm{e}^{\mathrm{i}x})$. Similarly, other results such as the completeness, orthogonality relations and Plancherel's theorem generalise to the non-abelian case via the Peter–Weyl theorems and representation theory.

Heat kernels on simply-connected compact semi-simple Lie groups were studied in [32], together with a useful Poisson form. In section 4 we show how to apply those results for $\mathrm{PU}(d)$, which is not simply-connected.

3. Main results and applications

Below we summarise the main results of this paper and outline some of their applications.

Result 1. The main technical result of the paper is the construction of the polynomial approximation to the Dirac delta function $H_P^{(t)}(\cdot,\sigma)$ on $\mathrm{PU}(d)$, together with some of its properties. This allows us to summarise the key properties of the family of polynomial approximations of Dirac delta based on the trimmed heat kernels (see theorem 3 for a precise statement).

Result 2. The supports of exact t-designs in $\mathrm{PU}(d)$ with $d \geqslant 2$ are ϵ-nets for $t \simeq \frac{d^{\frac{5}{2}}}{\epsilon}$ (see theorem 1 for a precise statement).

Result 3. The supports of approximate t-designs in $\mathrm{PU}(d)$ with $d \geqslant 2$ are ϵ-nets for $t \simeq \frac{d^{\frac{5}{2}}}{\epsilon}$ and $\delta \simeq \left(\frac{\epsilon}{d^{1/2}}\right)^{d^2}$. (see theorem 2 and its proof for a precise statement). This provides an 'essentially yes' answer to the conjecture about the optimal scaling of $t(\epsilon, d)$ from section IV in [28].

Application 1 (efficiency of quantum gates). This result is analogous to proposition 2 from [28] and is a simple consequence of result 3. For example, using the bound (129) from the proof of theorem 2, one may prove that if ν is a discrete probability measure on $\mathrm{PU}(d)$ with $d \geqslant 2$, which is a δ-approximate t-design with $\delta = \delta(\nu, t)$ for

$$t \geqslant 32 \frac{d^{\frac{5}{2}}}{\epsilon} \log(d) \log\left(\frac{4}{a_\nu \epsilon}\right), \tag{26}$$

where $C = 9\pi$, then the support of $\nu^{*\ell}$ forms an ϵ-net in $\mathrm{PU}(d)$ for

$$\ell \geqslant \frac{\log(1/\kappa(d)) + \left(d^2-1\right)\left(\frac{5}{4}\log\left(\frac{1}{\epsilon}\right) + \frac{3}{4}\log(Dd)\right)}{\log(1/\delta(\nu, t))}, \tag{27}$$

where

$$D = 8C^{2/3}\log^{1/3}(2C), \tag{28}$$

and $\log(1/\kappa(d)) < 5$. Moreover $\log(1/\kappa(d)) < 0$ for $d \geqslant 9$. Hence, in the case of the measure $\nu_{\mathcal{S}}$, the support of $\nu_{\mathcal{S}}^{*\ell}$ are the length ℓ words built out of the elements of $\mathcal{S}$ and this result is the SKL theorem with $\log(\frac{1}{\epsilon})$ term but also the multiplicative factor $\log^{-1}(1/\delta(\nu, t))$, which depends on t (or ϵ e.g. by taking (26) as equality). Such SKL theorems can be used to bound the overhead of quantum circuits [27].

Application 2 (inverse-free SK theorem). Similarly as in [28], application 1 can be turned into the inverse-free non-constructive SKL theorem without the ϵ-dependent multiplicative factor $\log^{-1}(1/\delta(\nu, t))$, by bounding the decay of $1 - \delta(\nu, t)$ with growing t, using the results from [29]. Namely, let $\nu_{\mathcal{S}}$ be a uniform probability measure on $\mathcal{S} \subset \mathrm{PU}(d)$. Then the support of $\nu_{\mathcal{S}}^{*\ell}$ is an ϵ-net in $\mathrm{PU}(d)$ for

$$\ell \geqslant A \frac{\log^3\left(\frac{1}{\epsilon}\right) + B}{\log(1/\delta(\nu, t_0))}, \tag{29}$$

where A, B and t_0 are some positive group constants. However, the constants are unknown due to the ambiguity of constants presented in [29].

Application 3 (quantum complexity and black hole physics). This application comes from the [20] in which the authors use the approximation of Dirac delta construction from [28] to prove the results about the approximate equidistribution of δ-approximate t-designs in the space $\mathbf{U}(d)$. This is then used to obtain results about the saturation and recurrence of the complexity of random local quantum circuits with gate set $\mathcal{S}$ without the assumptions on the spectral gap or inverse-closeness of $\mathcal{S}$. Such circuits can be used to model the chaotic dynamics of quantum many-body systems, which may be applicable in areas such as the physics of black hole interiors.

We believe that after some work, using our construction, one may obtain the approximate equidistribution of δ-approximate t-designs (theorem 16 from [20]) with better scaling in ϵ and d, which translates to the saturation and recurrence results.

4. The heat kernel on the projective unitary group

In the sequel, we employ formulae which are known for the heat kernel on $\mathrm{SU}(d)$, but which do not appear to be readily available for that on $\mathrm{PU}(d)$. Using standard techniques, we are able to write the latter in terms of the former in order to generalise the formulae we need.

Before we do so, we recall some facts from the representation theory of Lie groups (see e.g. [46–48]) and fix some notation and relevant conventions.

We work over the field of complex numbers. Let K be a (real) compact simply-connected Lie group (e.g. $\mathrm{SU}(d)$). Due to compactness, we can restrict ourselves to unitary complex representations. The complex representation theory of K is equivalent to the complex representation theory of its Lie algebra $\mathfrak{k}$, which is equivalent to the complex representation theory of its complexification $\mathfrak{g} = \mathfrak{k} + i\mathfrak{k}$.

The Cartan subalgebra of Lie algebra $\mathfrak{g}$ is an abelian and diagonalisable subalgebra of $\mathfrak{g}$, which is maximal under set inclusion. In general, there are many ways to choose the Cartan subalgebra. In our case, we can fix it by choosing the maximal torus in the Lie group. Let T be the maximal torus in K with Lie algebra $\mathfrak{t}$. Then the corresponding Cartan subalgebra $\mathfrak{h}$ of $\mathfrak{g}$ is $\mathfrak{h} = \mathfrak{t} + i\mathfrak{t}$.

For $K = \mathrm{SU}(d)$, we have $\mathfrak{g} = \mathfrak{sl}(d,\mathbb{C})$, which is a finite-dimensional complex semi-simple Lie algebra. The theory of finite-dimensional complex representations of such algebras is well-known and particularly nice. For example, such algebras are classified by their root systems/Dynkin diagrams and such representations are characterised by the theorem of the highest weight. Here, to match the notation of [32], we take a slightly different approach than usual, which is more suitable for compact groups K. In particular, we consider real weights and roots.

Let (Π, V) be a (finite-dimensional) representation of K and π be the associated representation of $\mathfrak{g}$. The (real) weight of V with respect to $\mathfrak{t}$ is an element λ from the dual space $\mathfrak{t}^*$, such that the corresponding weight space

$$V_\lambda := \{v \in V| \quad \pi(H)\, v = i\,\lambda(H)\, v, \forall H \in \mathfrak{t}\} \tag{30}$$

is not zero. Hence, the (real) root of $\mathfrak{g}$ with respect to $\mathfrak{t}$ is the non-zero element α from the dual space $\mathfrak{t}^*$, such that the corresponding root space

$$\mathfrak{g}_\alpha := \{E \in \mathfrak{g}| \quad [H,E] = i\,\alpha(H)\, E, \forall H \in \mathfrak{t}\} \tag{31}$$

is not zero. We denote the root system of $\mathfrak{g}$ as Φ, the set of all positive roots as Φ^+ and the set of simple roots as Δ.

Additionally, we assume that K is simply-connected. The algebra $\mathfrak{k}$ is equipped with $Ad(K)$-invariant positive-definite inner product $(\cdot,\cdot)$ defined as the negative of its Killing form (which is non-degenerate and negative-definite)

$$(X,Y) := -\mathrm{Tr}\left(\mathrm{ad}(X) \circ \mathrm{ad}(Y)\right). \tag{32}$$

The restriction of $(\cdot,\cdot)$ to $\mathfrak{t}$ is non-degenerate (hence, it defines the inner product on $\mathfrak{t}$). Thus, can use $(\cdot,\cdot)$ to identify $\mathfrak{t} \cong \mathfrak{t}^*$ via $X \mapsto \lambda_X$ for $X \in \mathfrak{t}$, where $\lambda_X(Y) = (X,Y)$ for any $Y \in \mathfrak{t}$ and $\lambda \mapsto X_\lambda$ for $\lambda \in \mathfrak{t}^*$, where $\lambda(Y) = (X_\lambda, Y)$ for any $Y \in \mathfrak{t}$. This way we also define $(\cdot,\cdot)$ on $\mathfrak{t}^*$ as $(\lambda,\kappa) = (X_\lambda, X_\kappa)$ for $\lambda,\kappa \in \mathfrak{t}^*$ and the induced norm $||\cdot||$. The inner product (32) defines the Riemannian metric on K, hence also the Laplace-Beltrami operator Δ. Thus, we can study the corresponding heat kernels.

Additionally for $\lambda \in \mathfrak{t}^*, \lambda \neq 0$ we define

$$\lambda^* := \frac{2}{(\lambda,\lambda)}\lambda \tag{33}$$

and the Weyl vector

$$\delta := \frac{1}{2}\sum_{\alpha\in\Phi^+}\alpha. \tag{34}$$

We aim to describe the heat kernel on $\mathrm{PU}(d)$ in terms of the heat kernel on $\mathrm{SU}(d)$. Specialising to the case $K=\mathrm{SU}(d)$, we introduce

$$\Gamma = \left\{\exp\left(\frac{2k\pi}{d}\right) I \,|\, k\in\mathbb{Z}\right\} \cong \mathbb{Z}_d, \tag{35}$$

so that $K/\Gamma \cong \mathrm{PU}(d)$. Our approach is based on the averaging map

$$f(x) \mapsto \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} f(\gamma x)\,. \tag{36}$$

Every irrep of $\mathrm{PU}(d)$ extends to an irrep of $\mathrm{SU}(d)$ by making it constant on Γ-cosets. It follows from lemmas 3 in appendix A that every irrep of $\mathrm{PU}(d)$ is obtained by applying the averaging map to a corresponding irrep of $\mathrm{SU}(d)$. Let's consider an elementary example.

Example 2 (irreps of $\mathrm{SU}(2)$ and $\mathrm{SO}(3)$). In this example we consider the irreps of $\mathrm{SU}(2)$ and aim to find the corresponding representations of $\mathrm{PU}(2)\cong\mathrm{PSU}(2)\cong\mathrm{SO}(3)$.

The irreps of $\mathrm{SU}(2)$ can be enumerated by the corresponding particle spin $j=0,\frac{1}{2},1,\ldots$ and have dimensions $2j+1$ (i.e. single irrep in each dimension). The centre $\Gamma\cong\mathbb{Z}_2$ acts by π-shifts. We want to find the irrep obtained via the averaging map applied to the irrep with spin j. We start with the spin j character of $\mathrm{SU}(2)$

$$\chi_j(\theta) = \frac{\sin((2j+1)\theta)}{\sin(\theta)}. \tag{37}$$

Averaging (37) yields

$$\frac{1}{2}\left(\frac{\sin((2j+1)\theta)-\sin((2j+1)\theta+(2j+1)\pi)}{\sin(\theta)}\right) = \begin{cases}\chi_j(\theta), & \text{for j being full-integer,}\\ 0, & \text{for j being half-integer.}\end{cases} \tag{38}$$

Hence, we obtain a well-known fact that the full-integer spin irreps of $\mathrm{SU}(2)$ are projective.

Thus, we can focus on the description of the heat kernel on $\mathrm{SU}(d)$. The first formula we employ is the standard expression for the heat kernel as the combination of characters, valid for compact semi-simple simply-connected Lie groups

$$H_S(g,\sigma) = \sum_\lambda d_\lambda \exp(-k_\lambda\sigma)\,\chi_\lambda(g)\,, \tag{39}$$

where λ is the highest weight vector and the sum is over complex irreps, d_λ is the dimension of the irrep, χ_λ is the character and $k_\lambda := (\lambda+2\delta,\lambda)$—see [32, 45, 49]. The parameter $\sigma>0$ plays the role of time and the subscript S indicates that this is the heat kernel on $\mathrm{SU}(d)$; later we will use H_P to denote the equivalent object on $\mathrm{PU}(d)$. The formula (39) is, in fact, the decomposition in terms of the eigenfunctions of the Laplace–Beltrami operator Δ, which are the characters χ_λ, where

$$\Delta\chi_\lambda = -k_\lambda\chi_\lambda. \tag{40}$$

In order to describe the highest weights λ for $\mathrm{SU}(d)$ using vectors, we introduce the linear functionals on $\mathfrak{t}$ (see (57)) acting as

$$L_j : \begin{pmatrix} i\phi_1 & & & \\ & i\phi_2 & & \\ & & \ddots & \\ & & & i\phi_d \end{pmatrix} \mapsto \phi_j, \tag{41}$$

so that $\lambda = \sum_{i=1}^{d} \lambda_i L_i$. Then the highest weights of $\mathrm{U}(d)$ can be labelled by integer-valued vectors $(\lambda_1, \lambda_2, \ldots, \lambda_d)$ with non-increasing entries, i.e. $\lambda_i \geqslant \lambda_{i+1}$ for $1 \leqslant i \leqslant d-1$. One can show that any irreducible representation of $\mathrm{U}(d)$ restricts to an irreducible representation of $\mathrm{SU}(d)$, while any irreducible representation of $\mathrm{SU}(d)$ extends to one of $\mathrm{U}(d)$. However, this mapping is not one-to-one. Since $\sum_{i=1}^{d} L_i(x) = 0$ for any $x \in \mathfrak{sl}(d,\mathbb{C})$ any irreducible representations of $\mathrm{U}(d)$ labelled by vectors which differ by a constant vector $(n, n, \ldots, n)$ for some $n \in \mathbb{Z}$ correspond to the same irreducible representation of $\mathrm{SU}(d)$.

We will also need to consider the irreducible representations of $\mathrm{PU}(d)$, which consists of equivalence classes of members of $\mathrm{U}(d)$ under the equivalence relation $U \sim \mathrm{e}^{\mathrm{i}\phi} U$. Any irrep of $\mathrm{PU}(d)$ extends to an irrep of $\mathrm{U}(d)$ by choosing it to be constant on equivalence classes so we can again label irreps of $\mathrm{PU}(d)$ with the same labels as those of $\mathrm{U}(d)$. An irrep of $\mathrm{U}(d)$ corresponds to an irrep of $\mathrm{PU}(d)$ exactly when it is constant on equivalence classes, which happens when the highest weight vector satisfies $\sum_j \lambda_j = 0$. We denote[5]

$$||\lambda||_1 := \sum_{i=1}^{d} |\lambda_i|. \tag{42}$$

By restricting to $||\lambda||_1 \leqslant 2t$, we obtain the set of vectors labelling the projective irreps corresponding to the t-design, i.e. appearing in the decomposition of the representation $U^{t,t}$ [50]. For $\mathrm{SU}(d)$ the dimension of the representation d_λ and the eigenvalue k_λ may be expressed as (see [45])

$$d_\lambda = \chi_\lambda(e) = \frac{\prod_{j<l}(\lambda_j - \lambda_l + l - j)}{\prod_{j<l}(l-j)} \leqslant (1 + \|\lambda\|_1)^{d(d-1)/2}, \tag{43}$$

$$k_\lambda = \frac{1}{2d} \sum_j \left(\lambda_j^2 + (d - 2j + 1)\lambda_j\right) - \frac{1}{2d^2} \left(\sum_j \lambda_j\right)^2. \tag{44}$$

If we have $\sum_j \lambda_j = 0$, so the $\mathrm{SU}(d)$ irrep is also a $\mathrm{PU}(d)$ irrep then we have the bound

$$k_\lambda \geqslant \frac{\|\lambda\|_1^2}{2d^2} + \frac{1}{4}\|\lambda\|_1. \tag{45}$$

Since $\mathrm{SU}(d)$ and $\mathrm{PU}(d)$ share a Lie algebra and d_λ with k_λ can be computed in terms of properties of the Lie algebra, these are identical for special and projective unitary representations. Therefore, the averaging map may be applied term-wise to the formula for the heat kernel on

[5] Not to be confused with the norm $||\cdot||$ stemming from the Killing form.

SU(d) to obtain the corresponding one for PU(d)

$$\frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} H_{\mathrm{S}}(\gamma g,\sigma) = \frac{1}{d}\sum_{\gamma\in\Gamma}\sum_{\lambda} d_\lambda \exp(-k_\lambda\sigma)\,\chi_\lambda(\gamma g) \tag{46}$$

$$= \sum_{\lambda} d_\lambda \exp(-k_\lambda\sigma)\,\frac{1}{d}\sum_{\gamma\in\Gamma}\chi_\lambda(\gamma g) \tag{47}$$

$$= \sum_{\lambda} d_\lambda \exp(-k_\lambda\sigma)\,\delta_{\mathrm{P}}(\lambda)\,\chi_\lambda(g) \tag{48}$$

$$= H_{\mathrm{P}}(g,\sigma), \tag{49}$$

where H_{S} and H_{P} are the special and projective unitary heat kernels, respectively, and from equation (35), we have $|\Gamma| = d$. Here δ_{P} is a Kronecker-delta like function, taking value 1 for irreps of SU(d) which are also irreps of PU(d) (i.e. projective representations) and value 0 otherwise. In fact,

$$\delta_P(\lambda) = \begin{cases} 1, & \sum_j \lambda_j = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{50}$$

Notice that the heat kernel (39) is a class function, hence it can be defined instead on the maximal torus of SU(d). With a mild abuse of notation, we will not distinguish between the two descriptions.

Let us formulate a second formula for the heat kernel on compact, semi-simple, simply-connected Lie groups from [32]

$$j(\exp(X)) = (2i)^m \prod_{\alpha\in\Phi^+} \sin\left(\frac{\alpha(X)}{2}\right), \tag{51}$$

$$K(X,\sigma) = \sum_{\gamma\in\Gamma} \pi(\lambda_X+\gamma)\exp\left(-\frac{1}{4\sigma}\|\lambda_X+\gamma\|^2\right), \tag{52}$$

$$H(\exp(X),\sigma) = \frac{c}{\pi(\delta)}(2\pi)^{l+m} i^m j(\exp(X))^{-1}\exp\left(\|\delta\|^2\sigma\right)(4\pi\sigma)^{-N/2}K(X,\sigma), \tag{53}$$

where

$$\pi(\lambda) := \prod_{\alpha\in\Phi^+}(\lambda,\alpha), \quad \lambda\in\mathfrak{t}^*, \tag{54}$$

$m = |\Phi^+|$, N is the group dimension, c is a (known) dimension-dependent group constant and Γ is a lattice generated by $l = \dim(\mathfrak{t})$ elements α_j^* (see (33)) corresponding to the simple roots $\Delta = \{\alpha_1,\ldots,\alpha_l\}$

$$\Gamma := 2\pi\sum_{j=1}^{l}\mathbb{Z}\alpha_j^*. \tag{55}$$

Formally, the heat kernel given by (53) is only defined for the regular elements X from $\mathfrak{t}$, i.e. the ones with distinct eigenvalues. However, the corresponding set of group elements for which formula (53) is not well defined is of Haar-measure zero. The function defined by the formula (53) extends to a unique continuous function, that defined by (39) on the whole group. Hence, with a slight abuse of notation, we treat (53) as defined on the whole group,

e.g. when integrating. More explicitly, one can see that the non-definedness of (53) at non-regular elements arises exactly from the factor of $j(\exp(X))^{-1}$ giving factors of $\sin(\alpha(X)/2)^{-1}$ as $\alpha(X) \to 0$. However, these apparently singular terms are balanced by terms linear in $\alpha(X)$ arising from the term $\pi(\lambda_X + \gamma)$. We will sketch this in more detail in appendix F.

The formula (53) is equivalent to (39) via the Poisson summation formula and we refer to it as the Poisson form (of the heat kernel). The Poisson form is relevant to us exactly because of the factor of σ^{-1} appearing in the exponent in (52). Roughly speaking, this formula is useful for bounding the behaviour of the heat kernel when the σ is small, while equation (39) is useful when σ is large.

The maximal torus T of $\mathrm{SU}(d)$ may be identified with the group of determinant 1 diagonal matrices parametrised by a vector $\phi \in \mathbb{R}^{d-1}$ as

$$T(\phi) := \begin{pmatrix} \mathrm{e}^{\mathrm{i}\phi_1} & & & & \\ & \mathrm{e}^{\mathrm{i}\phi_2} & & & \\ & & \ddots & & \\ & & & \mathrm{e}^{\mathrm{i}\phi_{d-1}} & \\ & & & & \mathrm{e}^{-\mathrm{i}\sum_{j=1}^{d-1}\phi_j} \end{pmatrix}. \tag{56}$$

The Lie algebra $\mathfrak{t}$ of T consists of traceless diagonal purely imaginary matrices parametrised by $\phi \in \mathbb{R}^{d-1}$ as

$$X(\phi) := i \begin{pmatrix} \phi_1 & & & & \\ & \phi_2 & & & \\ & & \ddots & & \\ & & & \phi_{d-1} & \\ & & & & -\sum_{j=1}^{d-1}\phi_j \end{pmatrix}. \tag{57}$$

Clearly, one can restrict the parameters e.g. $\phi_i \in (-\pi, \pi]$ for $1 \leqslant i \leqslant d-1$.

The complexified Lie algebra of K is $\mathfrak{g} = \mathfrak{sl}(d, \mathbb{C})$ and consists of traceless complex matrices. Let $E_{ij} \in \mathfrak{sl}(d, \mathbb{C})$ where $i \neq j$ denote the matrix with 1 in the (i, j) position and 0 elsewhere. The root system of $\mathfrak{g}$ with respect to $\mathfrak{t}$ is $\Phi = \{\alpha_{ij} | \quad 1 \leqslant i \neq j \leqslant d\}$ where the linear functionals α_{ij} act as

$$\alpha_{ij} : X(\phi) \mapsto \phi_i - \phi_j \tag{58}$$

and the corresponding one-dimensional root spaces are $\mathfrak{g}_{\alpha_{ij}} = \mathbb{C}E_{ij}$. Noting that $\alpha_{ji} = -\alpha_{ij}$ we choose positive roots $\Phi^+ = \{\alpha_{ij} | \quad 1 \leqslant i < j \leqslant d\}$ and a set of simple roots to be $\Delta = \{\alpha_{i,i+1} | \quad 1 \leqslant i \leqslant d-1\}$. We identify the Lie algebra $\mathfrak{t}$ with its dual $\mathfrak{t}^*$ under the inner product obtained from the Killing form

$$(X, Y) = -2d \operatorname{tr}(XY). \tag{59}$$

Under this identification α_{ij} is mapped to a diagonal matrix $X_{\alpha_{ij}}$ from $\mathfrak{t}$ with $\pm i/2d$ appearing as the only two non-zero entries of the ith and jth positions on the diagonal, respectively. Let X_δ be the element of $\mathfrak{t}$ which is identified with the Weyl vector δ, defined in equation (34). Then

$$X_\delta = \frac{1}{2} \sum_{i<j} X_{\alpha_{ij}}, \tag{60}$$

$$(X_\delta)_{kk} = i\left(\frac{d+1}{4d} - \frac{k}{2d}\right), \tag{61}$$

$$\|\delta\|^2 = \|X_\delta\|^2 \tag{62}$$

$$= 2d\sum_{k=1}^{d}\left(\frac{d+1}{4d} - \frac{k}{2d}\right)^2 \tag{63}$$

$$= \frac{d^2-1}{24}. \tag{64}$$

The duals of elements of Γ may be indexed by length $d-1$ integer vectors k

$$X(k) := 2\pi i \begin{pmatrix} k_1 & & & & \\ & k_2 & & & \\ & & \ddots & & \\ & & & k_{d-1} & \\ & & & & -\sum_{j=1}^{d-1} k_j \end{pmatrix}. \tag{65}$$

To simplify the notation we rename $X(\phi)$ and $X(k)$ as X_ϕ and X_k. Specialising the Poisson form of the heat kernel (53) to this parametrisation of the maximal torus of $\mathrm{SU}(d)$, one obtains

$$H_S(\exp(X_\phi),\sigma) := C(d,\sigma)j(\exp(X_\phi))^{-1}\sum_{k\in\mathbb{Z}^{d-1}}\pi(X_\phi + X_k)\exp\left(-\frac{1}{4\sigma}\|X_\phi + X_k\|^2\right), \tag{66}$$

where

$$C(d,\sigma) := \frac{c}{\pi(\delta)}(2\pi)^{l+m} i^m \exp\left(\|\delta\|^2\sigma\right)(4\pi\sigma)^{-N/2} \tag{67}$$

and for convenience we have written everything in terms of elements of the Lie algebra, converting elements of the dual where necessary, e.g. $\pi(X) = \prod_{\alpha\in\Phi^+}(\alpha,\lambda_X) = \prod_{\alpha\in\Phi^+}\alpha(X)$.

In order to obtain the corresponding heat kernel on $\mathrm{PU}(d)$ we proceed as above, and again we average this expression over the normal subgroup Γ given by dth roots of unity. We obtain an expression for the heat kernel on $\mathrm{PU}(d)$ in the Poisson form

$$H_P(\exp(X_\phi),\sigma) := \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} H_S(\gamma\exp(X_\phi),\sigma) \tag{68}$$

$$= \frac{C(d,\sigma)}{|\Gamma|}\sum_{\gamma\in\Gamma} j(\exp(X_\phi))^{-1}\sum_{k\in\mathbb{Z}^{d-1}}\pi(X_\phi + X_k + \log(\gamma))$$
$$\times\exp\left(-\frac{1}{4\sigma}\|X_\phi + X_k + \log(\gamma)\|^2\right), \tag{69}$$

where we have used that $j(\gamma e^X) = j(e^X)$ and to match how we parametrised the torus in (57) we choose the logarithm to be

$$\log\left(e^{i\frac{2\pi r}{d}I}\right) = i\frac{2\pi}{d}\begin{pmatrix} r & & & \\ & \ddots & & \\ & & r & \\ & & & -r(d-1) \end{pmatrix}. \tag{70}$$

15

We stress that formally $H_P(\exp(X_\phi),\sigma)$, is a function on $\mathrm{SU}(d)$ that is a lift of the heat kernel on $\mathrm{PU}(d)$.

5. Bounds for exact *t*-designs

In this section, we prove an error bound for a polynomial approximation to the heat kernel. In order to connect to projective t-designs, it is necessary for this approximation to be in terms of *balanced polynomials*. We call a function on $\mathrm{PU}(d)$ a balanced polynomial of order t if

$$f(\mathbf{U}) = \mathrm{tr}\left((U\otimes U^*)^{\otimes t}A\right), \tag{71}$$

holds for all $U\in\pi^{-1}(\mathbf{U})$, where A is some fixed matrix and recalling that in our present notation each $\mathbf{U}\in\mathrm{PU}(d)$ is an equivalence class of elements $\pi^{-1}(\mathbf{U})\subset\mathrm{SU}(d)$.

The approximation we seek follows directly from the formula

$$H_P(g,\sigma) = \sum_\lambda d_\lambda \exp(-\sigma k_\lambda)\chi_\lambda(g), \tag{72}$$

where $\sum_i \lambda_i = 0$, given by (49), upon noticing that each character χ_λ, of the projective unitary group is a balanced polynomial of order $\frac{\|\lambda\|_1}{2}$. This follows since they may be written in terms of Schur functions, see e.g. [45] for details. Let us denote by $H_P^{(t)}(g,\sigma)$ the restriction of the sum in (72) to balanced polynomials of order at most t, that is

$$H_P^{(t)}(g,\sigma) = \sum_{\lambda,\,\|\lambda\|_1\leqslant 2t} d_\lambda \exp(-\sigma k_\lambda)\chi_\lambda(g), \tag{73}$$

where $\sum_i \lambda_i = 0$. We refer to the polynomial approximations $H_P^{(t)}$ of the heat kernel H_P as trimmed heat kernels.

We seek to bound the 2-norm of the difference between the trimmed heat kernel $H_P^{(t)}$ and the full heat kernel H_P, where the 2-norm here is the one induced by the Haar measure

$$\|f\|_2^2 = \int_{\mathrm{SU}(d)} |f|^2 \,\mathrm{d}\mu. \tag{74}$$

Using expressions (72) and (73) one may bound the trimming error $\left\|H_P(\cdot,\sigma) - H_P^{(t)}(\cdot,\sigma)\right\|_2$ for t large enough (for fixed σ and d). For a precise statement and proof, see lemma 4 from appendix B. This allows us to focus on the properties of the full heat kernel H_P.

Remark 2. (optimality of the trimming procedure). The trimming procedure given by (73) is optimal in the following sense. The trimmed heat kernel $H_P^{(t)}(\cdot,\sigma)$ is the unique function in $\mathcal{H}_t$ closest to the heat kernel $H_P(\cdot,\sigma)$ in L^2-norm. Indeed, $H_P^{(t)}(\cdot,\sigma)$ is the orthogonal projection of $H_P(\cdot,\sigma)$ onto a finite-dimensional subspace $\mathcal{H}_t$ of the Hilbert space $L^2(\mathrm{PU}(d))$. Hence, the result follows from Hilbert's projection theorem.

The next step is to bound the complement of the integral of an absolute value of a heat kernel on $\mathrm{PU}(d)$ over the complement of a small ball $B_{P,\epsilon}$. As explained in section 4, we reduce this problem to considerations on $\mathrm{SU}(d)$. Recall that an element of $\mathrm{PU}(d)$ consists of

an equivalence class of elements of $\mathrm{SU}(d)$ where two matrices are equivalent if they differ by an element of Γ. By $B_\epsilon(V)$ we denote the closed operator-norm ϵ-ball in $\mathrm{SU}(d)$ centred at V

$$B_\epsilon(V) := \{U \in \mathrm{SU}(d) \,|\, d(U,V) \leqslant \epsilon\}. \tag{75}$$

By $B_{P,\epsilon}(\mathbf{V})$ be denote a closed ϵ-ball centred at $\mathbf{V}$ in metric $d_P(\cdot,\cdot)$

$$B_{P,\epsilon}(\mathbf{V}) := \{\mathbf{U} \in \mathrm{PU}(d) \,|\, d_P(\mathbf{U},\mathbf{V}) \leqslant \epsilon\}. \tag{76}$$

By $\tilde{B}_{P,\epsilon}(V) \subseteq \mathrm{SU}(d)$, where $V \in \pi^{-1}(\mathbf{V})$, we denote the inverse image of $B_{P,\epsilon}(\mathbf{V})$ under the quotient map $\pi : \mathrm{SU}(d) \to \mathrm{PU}(d)$

$$\tilde{B}_{P,\epsilon}(V) := \pi^{-1}(B_{P,\epsilon}(\mathbf{V})) = \bigcup_{\gamma\in\Gamma} \gamma B_\epsilon(V). \tag{77}$$

If the centre $\mathbf{V}$ is not specified, the ball is centred at the group identity. By $\mathcal{H}_r^d$ we denote an ∞-norm closed ball/hypercube in $\mathbb{R}^d$ of radius r

$$\mathcal{H}_r^d := \left\{v \in \mathbb{R}^d \,|\, ||v||_\infty \leqslant r\right\} \tag{78}$$

and by $\mathcal{Z}^{d-1}$ we denote the hyperplane in $\mathbb{R}^d$ consisting of vectors y with $\sum_{j=1}^d y_j = 0$.

Every element $U \in \mathrm{SU}(d)$ can be written as $U = VDV^{-1}$ for some $V \in \mathrm{SU}(d)$ and $D \in T$. Since the operator norm is unitary invariant, a ball B_ϵ corresponds to a unique ball $B_{T,\epsilon} \subset T$, via $B_\epsilon = \bigcup_{V\in\mathrm{SU}(d)} VB_{T,\epsilon}V^{-1}$, where $B_{T,\epsilon} = \{D \in T | d(D,I) \leqslant \epsilon\}$.

Hence, a ball $B_\epsilon \subset \mathrm{SU}(d)$ corresponds to a ball in $T \subset \mathrm{SU}(d)$ which is an image of $\mathcal{H}_{\tilde{\epsilon}}^{d-1}$ (identified with a subset of $\mathfrak{t}$), under the exponential map $\exp : \mathfrak{t} \to T$, where

$$\tilde{\epsilon} = 2 \cdot \arcsin(\epsilon/2) \in [0,\pi], \tag{79}$$

so $\epsilon \leqslant \tilde{\epsilon}$.

We first prove a lemma allowing us to remove the summation over Γ obtained when we express the $\mathrm{PU}(d)$ heat kernel in terms of that of $\mathrm{SU}(d)$.

Lemma 1. *1 Let φ be a non-negative function on $\mathrm{SU}(d)$ Haar-normalised to 1. Fix $\epsilon > 0$ and consider a set $\tilde{B}_{P,\epsilon}$ defined by (77), let its complement be $\tilde{B}^c_{P,\epsilon}$. Then*

$$\int_{\tilde{B}^c_{P,\epsilon}} \frac{1}{|\Gamma|} \sum_{\gamma\in\Gamma} \varphi(\gamma g)\, \mathrm{d}\mu(g) \leqslant \int_{B^c_\epsilon} \varphi(g)\, \mathrm{d}\mu(g). \tag{80}$$

Proof.

$$\int_{\tilde{B}_{P,\epsilon}} \frac{1}{|\Gamma|} \sum_{\gamma\in\Gamma} \varphi(\gamma g)\, \mathrm{d}\mu(g) = \frac{1}{|\Gamma|} \sum_{\gamma\in\Gamma} \int_{\cup_{\kappa\in\Gamma} \kappa B_\epsilon} \varphi(\gamma g)\, \mathrm{d}\mu(g) \tag{81}$$

$$= \frac{1}{|\Gamma|} \sum_{\gamma\in\Gamma} \int_{\cup_{\kappa\in\gamma\Gamma} \kappa B_\epsilon} \varphi(g)\, \mathrm{d}\mu\left(\gamma^{-1} g\right) \tag{82}$$

$$= \frac{1}{|\Gamma|} \sum_{\gamma\in\Gamma} \int_{\cup_{\kappa\in\Gamma} \kappa B_\epsilon} \varphi(g)\, \mathrm{d}\mu(g) \tag{83}$$

$$= \int_{\cup_{\kappa\in\Gamma} \kappa B_\epsilon} \varphi(g)\, \mathrm{d}\mu(g) \tag{84}$$

$$\geqslant \int_{B_\epsilon} \varphi(g)\,\mathrm{d}\mu(g) \tag{85}$$

hence

$$1-\int_{\tilde{B}_{P,\epsilon}} \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\varphi(\gamma g)\,\mathrm{d}\mu(g) \leqslant 1-\int_{B_\epsilon}\varphi(g)\,\mathrm{d}\mu(g). \tag{86}$$

The bound in lemma 1 may seem crude, however, the more of a mass of φ is concentrated in a ball B_ϵ the tighter it becomes. This corresponds e.g. to the heat kernel $H_S(g,\sigma)$ with decreasing σ (see also figure 1). □

Applying lemma 1 with $\varphi = H_S$ and the Weyl integration formula (see e.g. [45]) we can bound the integral of $H_P(g,\sigma)$ over $\tilde{B}^c_{P,\epsilon}$ as follows

$$\int_{\tilde{B}^c_{P,\epsilon}} H_P(g,\sigma)\,\mathrm{d}\mu(g) \leqslant \int_{B^c_\epsilon} H_S(g,\sigma)\,\mathrm{d}\mu(g) \tag{87}$$

$$= \frac{C(d,\sigma)}{|W|}\sum_{k\in\mathbb{Z}^{d-1}}\int_{\mathcal{H}^{d-1}_\pi\setminus\mathcal{H}^{d-1}_{\tilde{\epsilon}}} j(\exp(X_\phi))^*\,\pi(X_\phi+X_k)\exp\left(-\frac{1}{4\sigma}\|X_\phi+X_k\|^2\right)\mathrm{d}\mu(\phi), \tag{88}$$

where $\mathrm{d}\mu(\phi)=\frac{\mathrm{d}\phi_1\mathrm{d}\phi_2\ldots\mathrm{d}\phi_{d-1}}{(2\pi)^{d-1}}$ stems from the Haar measure on T, W is the Weyl group and we have cancelled the j^{-1} with part of the $|j|^2$ term in the Weyl integration formula.

Using the triangle inequality, we obtain

$$\int_{\tilde{B}^c_{P,\epsilon}} |H_P(g,\sigma)|\mathrm{d}\mu(g) \leqslant \frac{C(d,\sigma)}{|W|}\sum_{k\in\mathbb{Z}^{d-1}}\int_{\mathcal{H}^{d-1}_\pi\setminus\mathcal{H}^{d-1}_{\tilde{\epsilon}}} |j(\exp(X_\phi))\,\pi(X_\phi+X_k)| \times\exp\left(-\frac{1}{4\sigma}\|X_\phi+X_k\|^2\right)\mathrm{d}\mu(\phi). \tag{89}$$

We seek to express the right-hand side of (89) in terms of the dominant term $\mathcal{I}_0$ ($k=0$) and some smaller correction $\mathcal{R}$ which we will bound in terms of $\mathcal{I}_0$

$$\int_{\tilde{B}^c_{P,\epsilon}} |H_P(g,\sigma)|\mathrm{d}\mu(g) \leqslant \mathcal{I}_0+\mathcal{R}, \tag{90}$$

where

$$\mathcal{I}_0 := \frac{C(d,\sigma)}{|W|}\int_{\mathcal{H}^{d-1}_\pi\setminus\mathcal{H}^{d-1}_{\tilde{\epsilon}}} |j(\exp(X_\phi))\,\pi(X_\phi)|\exp\left(-\frac{1}{4\sigma}\|X_\phi\|^2\right)\mathrm{d}\mu(\phi), \tag{91}$$

and

$$\mathcal{R} := \frac{C(d,\sigma)}{|W|}\sum_{k\in\mathbb{Z}^{d-1}\setminus\{0\}}\int_{\mathcal{H}^{d-1}_\pi\setminus\mathcal{H}^{d-1}_{\tilde{\epsilon}}} |j(\exp(X_\phi))\,\pi(X_\phi+X_k)|\exp\left(-\frac{1}{4\sigma}\|X_\phi+X_k\|^2\right)\mathrm{d}\mu(\phi). \tag{92}$$

We provide the bounds on $\mathcal{I}_0$ and $\mathcal{R}$ via lemma 8 proved in appendix C and lemma 13 in appendix D respectively. Our bounds apply for σ small enough (for fixed ϵ and d).

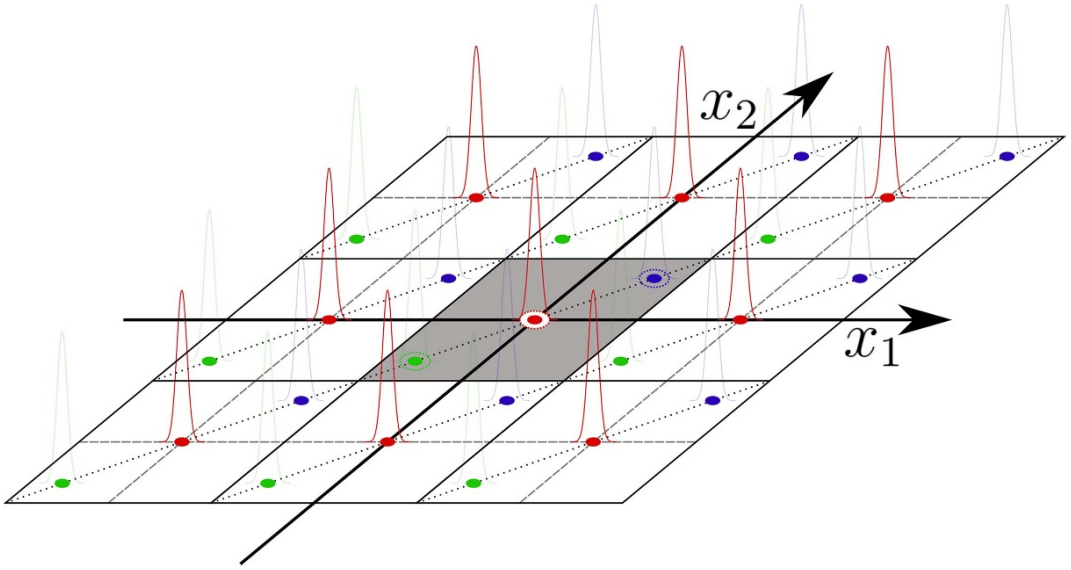


Figure 1. Illustration of the distribution of the components of a heat kernel H_P on PU(3), obtained via the averaging map applied to a heat kernel H_S in Poisson form. Actual shapes and relative sizes are not depicted. The averaging takes place over Γ, which consists of three roots of unity, denoted by red, green and blue points in the central square region. The elements of Γ act by shifting by the roots of unity along the dotted grey lines, corresponding to a torus. The heat kernel H_S corresponds to the red peaks. Each repeated square region corresponds to the contribution from a different k-vector in the Poisson form, which lies on the grey dashed grid. Notice that only the central square region ($k=0$) corresponds to points in a group. However, the tails of the peaks from non-central square regions ($k\neq 0$) overlap with the central square region, contributing to the heat kernel. A ball $\tilde{B}_{P,\epsilon}$ corresponds to a sum of three balls in a central region, denoted by dotted lines. A ball B_ϵ corresponds to the red ball at the origin, and the grey region corresponds to its complement. lemma 1 states that the integral of H_P over $\tilde{B}_{P,\epsilon}$ can be upper bounded by the integral of H_S (proportional to the red component) over the grey region. This is outlined by the opacity of the blue and green peaks. Lemma 13 shows that this integral can be bounded by bounding the contribution from the central ($k=0$) red peak, which is obtained in lemma 8.

By combining the trimming error bound with the vanishing properties of the full heat kernel H_P (lemmas 4, 8 and 13) we obtain a bound for the vanishing of the absolute value of the trimmed heat kernel $H_P^{(t)}$, stated in lemma 2.

Lemma 2. *Provided that*

$$2t \geqslant \frac{d^2}{\sqrt{\sigma}}\sqrt{2\log\left(\frac{d^4}{\sigma}\right)} \tag{93}$$

and

$$\sigma \leqslant \frac{\epsilon^2}{32d\log(d)} \tag{94}$$

for any

$$\eta \geqslant \frac{1}{\prod_{k=1}^{d} k!} \tag{95}$$

we have

$$\int_{\tilde{B}^c_{P,\epsilon}} \left| H_P^{(t)}(g,\sigma) \right| \mathrm{d}\mu(g) \leqslant 2^{\frac{d}{2}} \exp\left(-\sigma \frac{t^2}{d^2} - \frac{1}{2}\sigma t \right) + \frac{1+\eta}{2} \exp\left(-\frac{d}{16\sigma}\epsilon^2 + \frac{d^2-1}{24}\sigma \right). \tag{96}$$

Proof.

$$\int_{\tilde{B}^c_{P,\epsilon}} \left| H_P^{(t)}(g,\sigma) \right| \mathrm{d}\mu(g) \leqslant \int_{\tilde{B}^c_{P,\epsilon}} |H_P^{(t)}(g,\sigma) - H_P(g,\sigma)| \mathrm{d}\mu(g) + \int_{\tilde{B}^c_{P,\epsilon}} |H_P(g,\sigma)| \mathrm{d}\mu(g) \tag{97}$$

$$= \int_{\mathrm{SU}(d)} |H_P^{(t)}(g,\sigma) - H_P(g,\sigma)| \chi_{\tilde{B}^c_{P,\epsilon}}(g) \mathrm{d}\mu(g) + \int_{\tilde{B}^c_{P,\epsilon}} |H_P(g,\sigma)| \mathrm{d}\mu(g), \tag{98}$$

where χ_X denotes the indicator function of the set X. Applying Hölder's inequality to the first term of (98) gives

$$\int_{\tilde{B}^c_{P,\epsilon}} \left| H_P^{(t)}(g,\sigma) \right| \mathrm{d}\mu(g) \leqslant \left\| H_P^{(t)}(g,\sigma) - H_P(g,\sigma) \right\|_2 + \int_{\tilde{B}^c_{P,\epsilon}} |H_P(g,\sigma)| \mathrm{d}\mu(g). \tag{99}$$

Finally, substituting the bounds from lemmas 4, 8 and 13 with $\gamma = 1/2$ and applying $\epsilon \leqslant \tilde{\epsilon}$ gives the result. The condition (94) is the result of multiplying the bounds on σ we require for each lemma; one could obtain a slightly improved, but more complicated, bound by taking the minimum rather than the product. □

We are now able to prove our first theorem

Theorem 1. *Let ν be an exact t-design in* $\mathrm{PU}(d)$, $d \geqslant 2$, *then* $\mathrm{supp}(\nu)$ *is an ϵ-net provided*

$$t \geqslant 32 \frac{d^{\frac{5}{2}}}{\epsilon} \log(d) \log\left(\frac{4}{a_\nu \epsilon} \right), \tag{100}$$

where $C = 9\pi$.

Proof. We proceed via a proof by contradiction. Assume $\mathrm{supp}(\nu)$ is not an ϵ-net, then according to [28], lemmas 1 and 2, we know there exists a $\mathbf{V_0} \in \mathrm{PU}(d)$ such that for any $\kappa \leqslant \epsilon$

$$\mathrm{Vol}\left(B_{P,\kappa}(\mathbf{V_0}) \right) \leqslant \max_{V \in \tilde{B}^c_{P,\epsilon}} \int_{\tilde{B}_{P,\kappa}(V)} H_P^{(t)}(g,\sigma) \mathrm{d}\mu(g). \tag{101}$$

Note that $B_{P,\kappa}(\mathbf{V}) \subset B^c_{P,\epsilon-\kappa}$, hence also $\tilde{B}_{P,\kappa}(V) \subset \tilde{B}^c_{P,\epsilon-\kappa}$, so

$$\max_{V \in \tilde{B}^c_{P,\epsilon}} \int_{B_\kappa(V)} H_P^{(t)}(g,\sigma) \mathrm{d}\mu(g) \leqslant \max_{V \in \tilde{B}^c_{P,\epsilon}} \int_{\tilde{B}_{P,\kappa}(V)} \left| H_P^{(t)}(g,\sigma) \right| \mathrm{d}\mu(g) \tag{102}$$

$$\leqslant \int_{\tilde{B}^c_{P,\epsilon-\kappa}} \left| H_P^{(t)}(g,\sigma) \right| \mathrm{d}\mu(g). \tag{103}$$

The Haar (μ_P) volume of κ-ball (described in metric $d_P(\cdot,\cdot)$) in PU(d) can be bounded from below as follows:

$$\mathrm{Vol}\left(B_{P,\kappa}\right) \geqslant \left(a_v \kappa\right)^{d^2-1}, \tag{104}$$

where $a_v = \frac{1}{9\pi}$ (see [28]). Such a volume does not depend on the centre of the ball, due to the translation-invariance of the Haar measure and the metric $d_P(\cdot,\cdot)$. We take $\kappa = \frac{\epsilon}{2}$ and therefore have a contradiction if

$$\int_{\tilde{B}^c_{P,\epsilon/2}} \left| H^{(t)}\left(g,\sigma\right)\right| \mathrm{d}\mu\left(g\right) < \left(\frac{1}{2}a_v\epsilon\right)^{d^2-1}. \tag{105}$$

Hence, under the assumptions of lemmas 2 (with $\epsilon/2$ instead of ϵ), with the choice of $\eta = 1$, in order to get a contradiction, we can demand for example

$$\exp\left(-\frac{d}{64\sigma}\epsilon^2 + \frac{d^2-1}{24}\sigma\right) < \frac{1}{2}\left(\frac{1}{2}a_v\epsilon\right)^{d^2-1}, \tag{106}$$

$$2^{\frac{d}{2}}\exp\left(-\sigma\frac{t^2}{d^2} - \frac{1}{2}\sigma t\right) < \frac{1}{2}\left(\frac{1}{2}a_v\epsilon\right)^{d^2-1}. \tag{107}$$

The inequality (106) constrains σ as a function of ϵ and d and is satisfied whenever

$$\sigma < \sigma_* = \frac{\epsilon^2}{128 d \log\left(d\right)\log\left(\frac{2}{a_v\epsilon}\right)}, \tag{108}$$

which may be seen by taking logarithms of both sides of (106) and bounding the term containing $\left(d^2-1\right)\frac{\sigma}{24}$ using assumption (94). We can now bound the sufficient t, assuming $\sigma = \sigma^*$, using (107). Simply taking logarithms of (107) and substituting in $\sigma = \sigma^*$ we obtain

$$t^2 \geqslant 128\frac{d^5}{\epsilon^2}\log\left(d\right)\log^2\left(\frac{4}{a_v\epsilon}\right), \tag{109}$$

however we additionally need t to satisfy the assumption of lemma 2, so we obtain a final scaling

$$t^2 \geqslant 1024\frac{d^5}{\epsilon^2}\log^2\left(d\right)\log^2\left(\frac{4}{a_v\epsilon}\right), \tag{110}$$

□

6. Bounds for δ-approximate t-designs

In order to derive the version of theorem 1 for δ-approximate t-designs, we bound the L^2-norm of the heat kernel.

Since we want to apply the results from the previous sections, we use the Poisson form of the heat kernel, which allows us to group the terms as follows

$$\|H_S(\cdot,\sigma)\|_2^2 = \frac{C(d,\sigma)^2}{|W|}\int \sum_{k\in\mathbb{Z}^{d-1}}\sum_{l\in\mathbb{Z}^{d-1}} \pi(X_\phi + X_k)\pi^*(X_\phi + X_l) \times e^{-\frac{1}{4\sigma}\left(\|X_\phi + X_k)\|^2 + \|X_\phi + X_l)\|^2\right)} \mathrm{d}\mu(\phi) \tag{111}$$

$$\leqslant \frac{C(d,\sigma)^2}{|W|}\int \sum_{k\in\mathbb{Z}^{d-1}}\sum_{l\in\mathbb{Z}^{d-1}} |\pi(X_\phi + X_k)\pi^*(X_\phi + X_l)| \times \mathrm{e}^{-\frac{1}{4\sigma}\left(\|X_\phi + X_k)\|^2 + \|X_\phi + X_l)\|^2\right)} \mathrm{d}\mu(\phi) \tag{112}$$

$$= \mathcal{I}_{0,0}^2 + \mathcal{R}_{*,0}^2 + \mathcal{R}_{0,*}^2 + \mathcal{R}_{*,*}^2, \tag{113}$$

where $\mathcal{I}_{0,0}^2$ is the $k=0$ and $l=0$ term, $\mathcal{R}_{*,0}^2$ is the sum of the terms with $k\neq 0$ and $l=0$, $\mathcal{R}_{0,*}^2$ is the sum of the terms with $k=0$ and $l\neq 0$ and $\mathcal{R}_{*,*}^2$ is the sum of the terms with $k\neq 0$ and $l\neq 0$. We bound the contributions from $\mathcal{I}_{0,0}^2$, $\mathcal{R}_{*,0}^2$ and $\mathcal{R}_{*,*}^2$ separately in appendix E. The joint bound for $\|H_P^{(t)}(\cdot,\sigma)\|_2$ is provided in lemma 16 from appendix E.

We now have all the prerequisites to prove our main theorem, which is a generalisation of theorem 1 to δ approximate t-designs.

Theorem 2. *Let ν be a δ-approximate t-design in* $\mathrm{PU}(d)$*, $d\geqslant 2$, with*

$$\delta \leqslant \left(\frac{1}{4C\log^{1/4}\left(\frac{2C}{\epsilon}\right)\log^{1/4}(d)}\frac{\epsilon}{d^{1/2}}\right)^{d^2-1} \tag{114}$$

where

$$C = 9\pi\,, \tag{115}$$

then $\mathrm{supp}(\nu)$ *is an ϵ-net provided*

$$t \geqslant 32\frac{d^{\frac{5}{2}}}{\epsilon}\log(d)\log\left(\frac{4}{a_\nu\epsilon}\right). \tag{116}$$

Proof. We proceed as in the proof of theorem 1. Assume $\mathrm{supp}(\nu)$ is not an ϵ-net, then according to [28], lemma 2 and 3, we know there exists a $\mathbf{V_0}\in\mathrm{PU}(d)$ such that for any $\kappa\leqslant\epsilon$

$$\mathrm{Vol}\left(B_{P,\kappa}(\mathbf{V_0})\right) \leqslant \int_{\tilde{B}^c_{P,\epsilon-\kappa}} \left|H_P^{(t)}(g,\sigma)\right|\mathrm{d}\mu(g) + \delta\sqrt{\mathrm{Vol}\left(B_{P,\kappa}(\mathbf{V_0})\right)}||H_P^{(t)}(\cdot,\sigma)||_2 \tag{117}$$

$$\leqslant 2^{\frac{d}{2}}\exp\left(-\sigma\frac{t^2}{d^2} - \frac{1}{2}\sigma t\right) + \frac{1 + \left(1 + \delta\sqrt{\mathrm{Vol}\left(B_{P,\epsilon/2}\right)}\frac{d\sqrt{d!}}{2^{m-1}}\right)\frac{1}{\prod_k k!}}{2} \times \exp\left(-\frac{d}{64\sigma}\epsilon^2 + \frac{d^2-1}{24}\sigma\right) \tag{118}$$

$$+ \delta\sqrt{\mathrm{Vol}\left(B_{P,\epsilon/2}\right)}d\mathcal{I}_{0,0} \tag{119}$$

$$\leqslant 2^{\frac{d}{2}}\exp\left(-\sigma\frac{t^2}{d^2} - \frac{1}{2}\sigma t\right) + \exp\left(-\frac{d}{64\sigma}\epsilon^2 + \frac{d^2-1}{24}\sigma\right) + \delta\sqrt{\mathrm{Vol}\left(B_{P,\epsilon/2}\right)}d\mathcal{I}_{0,0} \tag{120}$$

where we put $\kappa = \epsilon/2$ and used lemmas 2 and 16 with $\eta = \frac{1}{\prod_{k=1}^{d} k!}$. Moreover, we assumed δ is not too large, so that

$$\left(1 + \delta\sqrt{\mathrm{Vol}\left(B_{P,\epsilon/2}\right)}\frac{d\sqrt{d!}}{2^{m-1}}\right)\frac{1}{\prod_{k=1}^{d} k!} \leqslant 1, \tag{121}$$

e.g.

$$\delta \leqslant \frac{1}{2\sqrt{2}} \leqslant \frac{2^{m-1}}{d\sqrt{d!}\sqrt{\mathrm{Vol}\left(B_{P,\epsilon/2}\right)}}. \tag{122}$$

We take $\kappa = \frac{\epsilon}{2}$ and therefore have a contradiction if, under the assumptions of lemmas 2 we have three inequalities

$$\exp\left(-\frac{d}{64\sigma}\epsilon^2 + \frac{d^2-1}{24}\sigma\right) < \frac{1}{2}\left(\frac{1}{2}a_v\epsilon\right)^{d^2-1} \tag{123}$$

$$2^{\frac{d}{2}}\exp\left(-\sigma\frac{t^2}{d^2} - \frac{1}{2}\sigma t\right) < \frac{1}{4}\left(\frac{1}{2}a_v\epsilon\right)^{d^2-1} \tag{124}$$

$$\delta d\mathcal{I}_{0,0} < \frac{1}{4}\left(\frac{1}{2}a_v\epsilon\right)^{\frac{d^2-1}{2}}. \tag{125}$$

Inequality (123) is the same as (106) from the proof of theorem 1, hence it is satisfied for the same $\sigma = \sigma^*$. Inequality (124) differs from (107) by the factor of $1/4$ instead of $1/2$, one may check that this inequality is still satisfied as long as t satisfies the bound in equation (100).

It remains to ensure that (125) is satisfied. Using lemmas 14, 7 and 11, then taking the logarithms of both sides of (125) and bounding the terms that do not depend on ϵ or σ by the $\Theta(d^2\log(d))$ term, we obtain

$$\log\left(\frac{1}{\delta}\right) \geqslant \left(\frac{d^2-1}{4}\right)\log\left(\frac{1}{\sigma}\right) + \frac{d^2-1}{24}\sigma + \left(\frac{3d^2}{16} + 4\right)\log(d) + \frac{d^2-1}{2}\log\left(\frac{2}{a_v\epsilon}\right). \tag{126}$$

Plugging $\sigma = \sigma^*$ leads to

$$\delta \leqslant \left(\frac{a_v}{2^{9/2}}\right)^{\frac{d^2-1}{2}}\left(\frac{\epsilon}{\log^{1/4}\left(\frac{2}{a_v\epsilon}\right)\log^{1/4}(d)}\right)^{d^2-1}\frac{\exp\left(-\frac{\left(d^2-1\right)\epsilon^2}{3072d\log(d)\log\left(\frac{2}{a_v\epsilon}\right)}\right)}{d^{\frac{7}{16}d^2+\frac{15}{4}}}. \tag{127}$$

One may check that (127) is stronger than (122). Moreover, since $\epsilon \leqslant 2$, we can lower bound the exponential term as

$$\exp\left(-\frac{\left(d^2-1\right)\epsilon^2}{3072d\log(d)\log\left(\frac{2}{a_v\epsilon}\right)}\right) \geqslant \exp\left(-\frac{d}{768\log(d)\log\left(\frac{1}{a_v}\right)}\right) = d^{-\frac{d}{768\log^2(d)\log\left(\frac{1}{a_v}\right)}}. \tag{128}$$

Hence,

$$\delta \leqslant \left(\frac{a_v}{2^{9/2}}\right)^{\frac{d^2-1}{2}} \left(\frac{\epsilon}{\log^{1/4}\left(\frac{2}{a_v\epsilon}\right)\log^{1/4}(d)\, d^{1/2}}\right)^{d^2-1} \kappa(d), \tag{129}$$

where

$$\kappa(d) := d^{\frac{d^2}{16}-\frac{17}{4}-\frac{d}{768\log^2(d)\log\left(\frac{1}{a_v}\right)}}. \tag{130}$$

□

We now observe that the function $\kappa(d)^{\frac{1}{d^2-1}}$ is increasing for $d \geqslant 2$, which may be demonstrated by computing the derivative. We may, therefore, lower bound it by bounding the value at $d = 2$. For example,

$$\kappa(d)^{\frac{1}{d^2-1}} \geqslant 2^{-\frac{3}{2}}. \tag{131}$$

Combining this bound with the above reasoning, we obtain

$$\delta \leqslant \left(\frac{1}{16\sqrt{9\pi}\log^{1/4}\left(\frac{2}{a_v\epsilon}\right)\log^{1/4}(d)}\frac{\epsilon}{d^{1/2}}\right)^{d^2-1}, \tag{132}$$

which may easily be seen to imply the bound shown in the Theorem statement.

Remark 3. The bound on δ provided in theorem 2 is significantly looser than e.g. (127) or (129). In the provided form theorem 2 is appropriate for comparison with the results of [28], however in applications we expect one of the more precise bounds to be more appropriate.

7. Trimmed heat kernel as a polynomial approximation of Dirac delta

In this section, we summarise various properties of our construction of the polynomial approximation of the Dirac delta. These properties are either obvious or were addressed in previous sections. The only missing property was the behaviour of the L^1-norm of the trimmed heat kernel, which also bounds its negativity.

Theorem 3. *The trimmed heat kernel $H_P^{(t)}(\cdot,\sigma)$ for $\mathbf{U}(d)$ with $d \geqslant 2$ has the following properties:*

1. $H_P^{(t)}(\cdot,\sigma) \in \mathcal{H}_t$.
2. $H_P^{(t)}(\cdot,\sigma)$ *is Haar-normalised to 1 and also approximately non-negative for t large enough (see point 5).*
3. *Controllable vanishing outside the ball of radius ϵ as $\sigma \to 0$*

$$\int_{\tilde{B}^c_{P,\epsilon}} \left|H_P^{(t)}(g,\sigma)\right| \mathrm{d}\mu(g) \leqslant 2^{\frac{d}{2}} \exp\left(-\sigma\frac{t^2}{d^2} - \frac{1}{2}\sigma t\right) + \frac{1+\eta}{2}\exp\left(-\frac{d}{16\sigma}\epsilon^2 + \frac{d^2-1}{24}\sigma\right) \tag{133}$$

for

$$t \geqslant t_* := \frac{d^2}{2\sqrt{\sigma}}\sqrt{2\log\left(\frac{d^4}{\sigma}\right)} \tag{134}$$

and

$$\sigma \leqslant \frac{\epsilon^2}{32d\log(d)} \tag{135}$$

with $\eta \geqslant \frac{1}{\prod_{k=1}^{d} k!}$. *Hence, for any* $\epsilon > 0$,

$$\lim_{\sigma\to 0}\int_{\tilde{B}^c_{P,\epsilon}} |H_P^{(t_*)}(g,\sigma)|\mathrm{d}\mu(g) = 0. \tag{136}$$

4. *Controllable blow-up of the* L^2*-norm*

$$||H_P^{(t)}(\cdot,\sigma)||_2 \leqslant c\left(\frac{d}{\sigma}\right)^{\frac{d^2-1}{4}} \tag{137}$$

for $\sigma \leqslant \frac{1}{d\log(d)}$ *and some positive group constant c. One can take* $c = 8$ *for* $d \geqslant 2$ *and* $c = 1$ *for* $d \geqslant 12$.

5. *Bounded* L^1*-norm*

$$||H_P^{(t)}(\cdot,\sigma)||_1 \leqslant 1 + 2^{\frac{d}{2}}\exp\left(-\sigma\frac{t^2}{d^2} - \frac{1}{2}\sigma t\right) \tag{138}$$

for $t \geqslant t_*$. *Hence, for fixed d*

$$\lim_{\sigma\to 0}||H_P^{(t_*)}(\cdot,\sigma)||_1 = 1. \tag{139}$$

Proof. Point 1 follows from the construction detailed above. Point 2 is a consequence of the orthogonality of characters and point 5. Point 3 is the lemma 2. Point 4 is corollary 1 from appendix E. Point 5 can be proved using the triangle inequality and Hölder's inequality. Indeed, since $H_P(\cdot,\sigma)$ is normalised to 1 and non-negative, using lemma 4 we can write

$$||H_P^{(t)}(\cdot,\sigma)||_1 \leqslant ||H_P(\cdot,\sigma)||_1 + ||\left(H_P^{(t)}(\cdot,\sigma) - H_P(\cdot,\sigma)\right)\chi_{\mathrm{SU}(d)}||_1 \tag{140}$$

$$\leqslant 1 + ||H_P^{(t)}(\cdot,\sigma) - H_P(\cdot,\sigma)||_2 \tag{141}$$

$$\leqslant 1 + 2^{\frac{d}{2}}\exp\left(-\sigma\frac{t^2}{d^2} - \frac{1}{2}\sigma t\right) \tag{142}$$

for

$$t \geqslant \frac{d^2}{2\sqrt{\sigma}}\sqrt{2\log\left(\frac{d^4}{\sigma}\right)}. \tag{143}$$

□

Remark 4. One can also write down similar properties of the (full) heat kernel $H_P(\cdot,\sigma)$, which follow from the proofs of the same Lemmas as theorem 3. In this case, point 1 is not true for any t. The normalisation from point 2 is true, but $H_P(\cdot,\sigma)$ is non-negative so $||H_P(\cdot,\sigma)||_1 = 1$. The bound from point 4 is valid. The bound from point 3 simplifies to

$$\int_{\tilde{B}^c_{P,\epsilon}} |H_P(g,\sigma)|\mathrm{d}\mu(g) \leqslant \frac{1+\eta}{2}\exp\left(-\frac{d}{16\sigma}\epsilon^2 + \frac{d^2-1}{24}\sigma\right) \tag{144}$$

for

$$\sigma \leqslant \frac{\epsilon^2}{32d\log(d)} \tag{145}$$

and $\eta \geqslant \frac{1}{\prod_{k=1}^{d} k!}$. Hence, for any $\epsilon > 0$

$$\lim_{\sigma\to 0}\int_{\tilde{B}^c_{P,\epsilon}} |H_P(g,\sigma)|\mathrm{d}\mu(g) = 0. \tag{146}$$

Remark 5. Formally, theorem 3 shows in particular that for fixed $d \geqslant 2$, the family of L^1-integrable functions $\{k_\lambda\}_{\lambda>0}$, where $k_\lambda(g) := H_P^{t_*(\lambda)}(g, 1/\lambda)$ and $t_*(\lambda) := \frac{d^2}{2}\sqrt{2\lambda\log(d^4\lambda)}$, is an approximate identity on $\mathrm{PU}(d)$.

8. Summary and future work

We have improved on the state-of-the-art for constructing ϵ-nets in the group of unitary channels from unitary t-designs, that of [28]. Our method involves the construction of a polynomial approximation to a Dirac delta on the space of quantum unitary channels $\mathrm{PU}(d)$ stemming from the natural object—a heat kernel on $\mathrm{SU}(d)$.

In the case of exact t-designs we obtain results very close to those of [28], but in the more practically relevant approximate case, our results significantly improve on the state-of-the-art, showing that δ-approximate t-designs form ϵ nets for much larger values of δ than was previously known.

While our scaling of δ with d and ϵ for approximate t designs substantially improves on prior work, it is not obvious that it is optimal. We leave for future work the task of either improving this scaling even further or proving that no improvements are possible.

The construction of [28] is used in [20] to prove saturation and recurrence results for the complexity of random quantum circuits without the assumptions on the gap of the universal set $\mathcal{S}$. In future work, it may be possible to apply our construction to improve the known results in this setting. It would also be interesting to understand the unknown constants A and B appearing in (29), obtaining an inverse-free non-constructive SKL theorem from our results. This amounts to deriving explicit constants of the polylogarithmic spectral gap decay (see Theorem 6 from [29]), especially r_0. We expect that the trimmed heat kernel construction may be applied to obtain explicit bounds on such decay.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

This research was funded by the National Science Centre, Poland under the Grant OPUS: UMO2020/37/B/ST2/02478. AS would like to thank Luz Roncal for inspiring discussions regarding heat kernels on compact Lie groups. OS would like to thank Michał Oszmaniec for the discussion about our results.

Appendix A. Connecting the heat kernels on the projective and special unitary groups

In this appendix, we prove lemma 3, which can be used to show how one can obtain a heat kernel on $\mathrm{PU}(d)$ using the heat kernel on $\mathrm{SU}(d)$ via averaging.

Lemma 3. *Let K be a simply-connected compact Lie group and Γ a finite normal subgroup so that K/Γ is a compact Lie group. Let ρ be a finite-dimensional unitary irrep of K. Then the function*

$$\tilde{\rho} : g \mapsto \frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} \rho(\gamma g), \tag{A1}$$

is either

1. *identical to ρ if ρ is constant on Γ-cosets in K or*
2. *identically zero.*

Proof. We observe

$$\tilde{\rho}(g)\,\tilde{\rho}(g') = \tilde{\rho}(gg'), \tag{A2}$$

so $\tilde{\rho}$ is a group homomorphism exactly if it maps the identity in K to the identity operator. Since $\tilde{\rho}(e)$ is self-adjoint and is easily seen to be idempotent, it is an orthogonal projector. Since $\tilde{\rho}(e)$ commutes with every operator in the image of K under ρ and ρ is irreducible by Schur's lemma, $\rho(e) \propto I$ and is therefore equal to either I or 0.

In the case that $\tilde{\rho}(e)$ is the identity operator compute

$$\begin{aligned}
\rho\left(g^{-1}\right)\tilde{\rho}(g) &= \frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} \rho\left(g^{-1}\right)\rho(\gamma g) && \text{(A3)}\\
&= \frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} \rho\left(g^{-1}\gamma g\right) && \text{(A4)}\\
&= \frac{1}{|\Gamma|} \sum_{\gamma' \in \Gamma} \rho\left(g^{-1} g \gamma' g^{-1} g\right) && \text{(A5)}\\
&= \tilde{\rho}(e) = I, && \text{(A6)}
\end{aligned}$$

so that for all $g \in K$ we have $\rho(g) = \tilde{\rho}(g)$ implying that the ρ is constant on Γ-cosets. □

Proof of (A2).

$$\tilde{\rho}(g)\,\tilde{\rho}(g') = \frac{1}{|H|^2}\sum_{h,h'\in H}\rho(hg)\,\rho(h'g') \tag{A7}$$

$$= \frac{1}{|H|^2}\sum_{h,h'\in H}\rho(hgh'g') \tag{A8}$$

$$= \frac{1}{|H|^2}\sum_{h\in H}\sum_{h''\in H}\rho\left(hgg^{-1}h''gg'\right) \tag{A9}$$

$$= \frac{1}{|H|^2}\sum_{h\in H}\sum_{h''\in H}\rho(hh''gg') \tag{A10}$$

$$= \frac{1}{|H|}\sum_{h\in H}\rho(hgg') = \tilde{\rho}(gg') \tag{A11}$$

□

Appendix B. Bounding the polynomial approximation of the heat kernel

In this appendix, we prove lemma 4, which quantifies the L^2-norm difference between the heat kernel H_P and the trimmed heat kernel $H_P^{(t)}$.

Lemma 4 ('Trimming' the heat kernel). *The trimmed heat kernel $H_P^{(t)}$ satisfies*

$$\left\|H_P(\cdot,\sigma) - H_P^{(t)}(\cdot,\sigma)\right\|_2 \leqslant 2^{\frac{d}{2}}\exp\left(-2\sigma(1-\gamma)\frac{t^2}{d^2} - \frac{1}{2}\sigma t\right), \tag{B1}$$

for any $0<\gamma<1$, provided that

$$2t \geqslant \frac{d^2}{\sqrt{\gamma\sigma}}\sqrt{\log\left(\frac{d^4}{2\gamma\sigma}\right)}. \tag{B2}$$

Proof. In terms of L^2-norm, as a function of x the approximation error may be computed

$$\left\|H_P(\cdot,\sigma) - H_P^{(t)}(\cdot,\sigma)\right\|_2^2 = \left\|\sum_{\lambda,\|\lambda\|_1>2t} d_\lambda \exp(-\sigma k_\lambda)\,\chi_\lambda(x)\right\|_2^2 \tag{B3}$$

$$= \int_G \mathrm{d}\mu(x)\left|\sum_{\lambda,\|\lambda\|_1>2t} d_\lambda \exp(-\sigma k_\lambda)\,\chi_\lambda(x)\right|^2 \tag{B4}$$

$$= \sum_{\nu,\|\nu\|_1>2t}\sum_{\lambda,\|\lambda\|_1>2t} d_\lambda d_\nu \exp(-\sigma(k_\lambda+k_\nu))\int_G \mathrm{d}\mu(x)\,\chi_\nu^*(x)\,\chi_\lambda(x) \tag{B5}$$

$$= \sum_{\lambda,\|\lambda\|_1>2t} d_\lambda^2 \exp(-2\sigma k_\lambda), \tag{B6}$$

using the orthonormality of the characters. The weights λ are integer-valued d dimensional vectors with non-increasing entries, which satisfy the condition $\sum_j \lambda_j = 0$. We now need a bound for $\alpha(j)$, the number of highest weights λ satisfying $\|\lambda\|_1 = j$. Each highest weight is uniquely determined by an integer-valued $d-1$ dimensional vector with non-increasing entries, since e.g. the last element of the vector is fixed by the constraint $\sum_j \lambda_j = 0$. In order to simplify the reasoning we will ignore the non-increasing property and obtain a slightly looser bound than we would by including it. Such $d-1$ dimensional vector clearly has 1-norm less than $||\lambda||_1$. Since the infinity-norm lower bounds the one-norm it follows that the number of highest-weight vectors with 1-norm equal to j is upper bounded by the number of integer vectors with infinity norm less than j, which is exactly the number of integer points in an $d-1$ dimensional hypercube of side length $2j$. We therefore obtain the very crude upper bound

$$\alpha(j) \leqslant (1+2j)^{d-1}. \tag{B7}$$

Substituting this, along with the bounds from (43) and (45) into (B6) we obtain

$$\left\| H_P(\cdot,\sigma) - H_P^{(t)}(\cdot,\sigma) \right\|_2^2 = \sum_{j>2t} \alpha(j)\, d_\lambda^2 \exp(-2\sigma k_\lambda) \tag{B8}$$

$$\leqslant \sum_{j>2t} (1+2j)^{(d-1)} (1+j)^{d(d-1)} \exp\left(-2\sigma\left(\frac{j^2}{2d^2} + \frac{j}{4}\right)\right) \tag{B9}$$

$$\leqslant \left(2+\frac{1}{2t}\right)^{d-1} \left(1+\frac{1}{2t}\right)^{d(d-1)} \sum_{j>2t} j^{(d+1)(d-1)} \exp\left(-2\sigma\left(\frac{j^2}{2d^2} + \frac{j}{4}\right)\right). \tag{B10}$$

Our approach is to bound the expression by a Gaussian integral, so we first bound the polynomial term in the sum by a Gaussian. An easy bound follows from the bound on the Lambert W function

$$W_{-1}\left(-\mathrm{e}^{u-1}\right) > -1 - \sqrt{2u} - u, \tag{B11}$$

obtained in [51], namely that

$$j^2 \geqslant \frac{d^4}{\gamma\sigma} \log\left(\frac{d^4}{2\gamma\sigma}\right) \implies (1+2j)^{(d+1)(d-1)} \leqslant \exp\left(\gamma\sigma\frac{j^2}{d^2}\right). \tag{B12}$$

Assuming that

$$j > 2t \geqslant \frac{d^2}{\sqrt{\gamma\sigma}} \sqrt{\log\left(\frac{d^4}{2\gamma\sigma}\right)} \implies (1+2j)^{(d+1)(d-1)} \leqslant \exp\left(\gamma\sigma\frac{j^2}{d^2}\right), \tag{B13}$$

where $0 < \gamma < 1$ is a constant we have introduced. We substitute this bound to obtain

$$\left\|H_P(\cdot,\sigma)-H_P^{(t)}(\cdot,\sigma)\right\|_2^2 \leqslant \left(2+\frac{1}{2t}\right)^{d-1}\left(1+\frac{1}{2t}\right)^{d(d-1)}\sum_{j>2t}\exp\left(-\sigma\left((1-\gamma)\frac{j^2}{d^2}+\frac{j}{2}\right)\right). \tag{B14}$$

The summand is decreasing in j, so we may bound the sum by an integral to obtain

$$\left\|H_P(\cdot,\sigma)-H_P^{(t)}(\cdot,\sigma)\right\|_2^2 \leqslant \left(2+\frac{1}{2t}\right)^{d-1}\left(1+\frac{1}{2t}\right)^{d(d-1)} \times\int_{2t}^{\infty}\exp\left(-\sigma\left((1-\gamma)\frac{x^2}{d^2}+\frac{x}{2}\right)\right)\mathrm{d}x. \tag{B15}$$

We compute the integral and employ the standard bound

$$\operatorname{erfc}(x)\leqslant\frac{1}{x\sqrt{\pi}}\exp\left(-x^2\right), \tag{B16}$$

to obtain

$$\left\|H_P(\cdot,\sigma)-H_P^{(t)}(\cdot,\sigma)\right\|_2^2 \leqslant \left(2+\frac{1}{2t}\right)^{d-1}\left(1+\frac{1}{2t}\right)^{d(d-1)}\frac{2d^2}{\sigma}\frac{1}{d^2+8t(1-\gamma)} \times\exp\left(-4\sigma(1-\gamma)\frac{t^2}{d^2}-\sigma t\right). \tag{B17}$$

Recalling that we are still assuming that $2t\geqslant\frac{d^2}{\sqrt{\gamma\sigma}}\sqrt{\log\left(\frac{d^4}{2\gamma\sigma}\right)}$ we can obtain the very simple bound

$$\left\|H_P(\cdot,\sigma)-H_P^{(t)}(\cdot,\sigma)\right\|_2^2 \leqslant 2^d\frac{e}{1+d^2}\frac{1}{5-4\gamma}\exp\left(-4\sigma(1-\gamma)\frac{t^2}{d^2}-\sigma t\right) \tag{B18}$$

$$\leqslant 2^d\exp\left(-4\sigma(1-\gamma)\frac{t^2}{d^2}-\sigma t\right). \tag{B19}$$

□

Appendix C. Bounding the dominant term $\mathcal{I}_0$

In this appendix, we provide a proof of lemma 8, which bounds the $\mathcal{I}_0$ term (91). To do that, we use the following lemmas 5–7.

Lemma 5. *Let* GUE_d^0 *denote a GUE ensemble of traceless* $d\times d$ *matrices. Then for any* $r\geqslant 0$ *(see e.g. [52])*

$$\Pr_{A\sim\mathrm{GUE}_d^0}\left(\|A\|_\infty\leqslant r\right)=\left(\prod_{j=1}^{d}\frac{1}{j!}\right)(2\pi)^{-\frac{d-1}{2}}2^{\frac{d^2-1}{2}}\int_{\mathcal{Z}^{d-1}\cap\mathcal{H}_r^d}\mathrm{d}y\exp\left(-\sum_{j=1}^{d}y_j^2\right)\prod_{1\leqslant i<j\leqslant d}\left(y_i-y_j\right)^2. \tag{C1}$$

Lemma 6. *For any $r \geqslant 2\sqrt{d}$ (see [53])*

$$\Pr_{A\sim \mathrm{GUE}_d^0}\left(\|A\|_\infty \geqslant r\right) \leqslant \frac{1}{2}\exp\left(-\frac{d}{2}\left(\frac{r}{\sqrt{d}}-2\right)^2\right). \tag{C2}$$

Proof. From [53] we have

$$\Pr_{A\sim \mathrm{GUE}_d^0}\left(\frac{1}{\sqrt{d}}\|A\|_\infty \geqslant 2+x\right) \leqslant \frac{1}{2}\exp\left(-\frac{dx^2}{2}\right) \tag{C3}$$

valid for $x \geqslant 0$. The result follows via $x = \frac{r}{\sqrt{d}} - 2$. □

Lemma 7. *For* SU(d) *we can evaluate (see (67))*

$$\frac{C(d,\sigma)}{|W|} = \frac{\sqrt{d}\,(2d)^{(d-1)/2+m}}{\prod_{k=1}^{d} k!}(2\pi)^{d-1+m}\,\mathrm{e}^{\frac{d^2-1}{24}\sigma}\,(4\pi\sigma)^{-\left(d^2-1\right)/2}, \tag{C4}$$

where $m = d(d-1)/2$.

Proof. From [32] we know that $c = 2^{l/2}\frac{\sqrt{D}}{\prod_{i=1}^{l}|\alpha_i|}$, where D is the determinant of the Cartan matrix. It is known that for A_{d-1} root system, $D = d$ and $|W| = d!$. Moreover, we have that $N = d^2-1$, $l = d-1$, $m = d(d-1)/2$, $|\alpha_i| = 1/\sqrt{d}$. The expression $\pi(\delta)$ can be calculated from equation (61) as

$$\pi(\delta) = (2d)^{-m} i^m \prod_{1\leqslant p<q\leqslant d}(q-p) = (2d)^{-m} i^m \prod_{k=1}^{d-1} k!. \tag{C5}$$

□

Lemma 8. *Assume $\sigma \leqslant \frac{\tilde{\epsilon}^2}{32}$. Then*

$$\mathcal{I}_0 \leqslant \overline{\mathcal{I}_0} := \frac{1}{2}\exp\left(-\frac{d}{16\sigma}\tilde{\epsilon}^2 + \frac{d^2-1}{24}\sigma\right). \tag{C6}$$

Proof. The $\mathcal{I}_0$ can be bounded as

$$\mathcal{I}_0 \leqslant \frac{C(d,\sigma)\,2^m}{|W|}\int_{\mathcal{H}_\pi^{d-1}\setminus\mathcal{H}_{\tilde{\epsilon}}^{d-1}} \mathrm{d}\mu(\phi)\left|\prod_{\alpha>0}\alpha(X_\phi)\sin\left(\frac{\alpha(X_\phi)}{2}\right)\right|\exp\left(\frac{-1}{4\sigma}\|X_\phi\|^2\right) \tag{C7}$$

$$\leqslant \frac{C(d,\sigma)}{|W|}\int_{\left(\mathcal{H}_{\tilde{\epsilon}}^{d-1}\right)^c} \mathrm{d}\mu(\phi)\left(\prod_{\alpha>0}\alpha(X_\phi)^2\right)\exp\left(\frac{-1}{4\sigma}\|X_\phi\|^2\right). \tag{C8}$$

Recalling (59) and (57) we can write

$$\|X_\phi\|^2 = 2d\left(\sum_{j=1}^{d-1}\phi_j^2 + \left(\sum_{j=1}^{d-1}\phi_j\right)^2\right), \tag{C9}$$

and it is convenient to introduce the variables $y_j := \phi_j\sqrt{\frac{d}{2\sigma}}$ and $y_d := -\sum_{j=1}^{d-1} y_j$. The expression in (C8) is then equal to

$$\frac{C(d,\sigma)}{|W|}\left(\frac{2\sigma}{d}\right)^{m+\frac{l}{2}}(2\pi)^{-(d-1)}\frac{1}{\sqrt{d}}\int_{\mathcal{Z}^{d-1}\cap\left(\mathcal{H}^d_{\tilde{\epsilon}\sqrt{\frac{d}{2\sigma}}}\right)^c}\left(\prod_{1\leqslant i<j\leqslant d}(y_i-y_j)^2\right)$$
$$\times\exp\left(-\sum_{j=1}^{d}y_j^2\right)\mathrm{d}\mu_{\mathcal{Z}}(y) \tag{C10}$$
$$=\mathrm{e}^{\|\delta\|^2\sigma}\left(\prod_{k=1}^{d}\frac{1}{k!}\right)2^{\frac{1}{2}(d-1)d}\pi^{\frac{1}{2}-\frac{d}{2}}\int_{\mathcal{Z}^{d-1}\cap\left(\mathcal{H}^d_{\tilde{\epsilon}\sqrt{\frac{d}{2\sigma}}}\right)^c}\left(\prod_{1\leqslant i<j\leqslant d}(y_i-y_j)^2\right)$$
$$\times\exp\left(-\sum_{j=1}^{d}y_j^2\right)\mathrm{d}\mu_{\mathcal{Z}}(y), \tag{C11}$$

where $\mu_{\mathcal{Z}}$ is the Euclidean measure on the hyperplane $\mathcal{Z}^{d-1}$ and a factor of $d^{-\frac{1}{2}}$ appears in (C10) from changing the measure from the Euclidean one on the $d-1$ variables $y_1\ldots y_{d-1}$ to the Euclidean measure intrinsic to the hyperplane. The transition from (C10) to (C11) comes from the application of lemma 7.

Using the normalisation of the probability to 1, we can apply lemma 5 with $r=\sqrt{\frac{d}{2\sigma}}\tilde{\epsilon}$ to (C11) to write

$$\mathcal{I}_0\leqslant\mathrm{e}^{\|\delta\|^2\sigma}\Pr_{A\sim\mathrm{GUE}_d^0}\left(\|A\|_\infty\geqslant\tilde{\epsilon}\sqrt{\frac{d}{2\sigma}}\right). \tag{C12}$$

Applying lemma 6 with $r=\sqrt{\frac{d}{2\sigma}}\tilde{\epsilon}$ to (C12) and assuming

$$\tilde{\epsilon}\geqslant\frac{2\sqrt{2\sigma}}{1-\beta}, \tag{C13}$$

for some $0<\beta<1$ (which is stronger than the assumption in lemma 6) we obtain the following bound

$$\mathcal{I}_0\leqslant\frac{1}{2}\exp\left(-\frac{d}{4\sigma}\tilde{\epsilon}^2+\frac{4d}{2\sqrt{2\sigma}}\tilde{\epsilon}-2d+\|\delta\|^2\sigma\right) \tag{C14}$$
$$\leqslant\frac{1}{2}\exp\left(-\frac{d\beta^2}{4\sigma}\tilde{\epsilon}^2+\|\delta\|^2\sigma\right) \tag{C15}$$
$$=\frac{1}{2}\exp\left(-\frac{d\beta^2}{4\sigma}\tilde{\epsilon}^2+\frac{d^2-1}{24}\sigma\right). \tag{C16}$$

We set $\beta=1/2$ and for $\sigma\leqslant\frac{\tilde{\epsilon}^2}{32}$ (C13) we get

$$\mathcal{I}_0\leqslant\frac{1}{2}\exp\left(-\frac{d}{16\sigma}\tilde{\epsilon}^2+\frac{d^2-1}{24}\sigma\right). \tag{C17}$$

□

Appendix D. Bounding the correction term $\mathcal{R}$

In this appendix, we bound the remaining terms $\mathcal{R}$, defined in (92). Our strategy is to bound $\mathcal{R}$ by the volume of the complement of ϵ-ball times the upper bound on the integrand outside of the ball (lemma 12). We then show that $\mathcal{R}$ is indeed a correction to $\mathcal{I}_0$ (see (91)), which relatively decays very fast with growing d (lemma 13). To do so, we employ the following lemmas 9–11.

Lemma 9. *The number of k-vectors in each* $||\cdot||_\infty$ *norm shell of radius* $r>0$

$$S_{r,d} := \left\{(n_1,\ldots,n_d)\in\mathbb{Z}^d |\ \max_i |n_i| = r\right\} \tag{D1}$$

can be upper bounded as

$$|S_{r,d}| \leqslant 2^d (2r)^{d-1}. \tag{D2}$$

Proof. Consider the corresponding balls

$$B_{r,d} := \left\{(n_1,\ldots,n_d)\in\mathbb{Z}^d |\ \max_i |n_i| \leqslant r\right\}. \tag{D3}$$

It is easy to see that $|B_{r,d}| = (2r+1)^d$. Thus,

$$|S_{r,d}| = (2r+1)^d - (2r-1)^d. \tag{D4}$$

Using the binomial expansion, we can write

$$|S_{r,d}| = 2\sum_{k=0}^{\lfloor\frac{d-1}{2}\rfloor}\binom{d}{2k+1}(2r)^{d-2k-1} \leqslant 2(2r)^{d-1}\sum_{k=0}^{\lfloor\frac{d-1}{2}\rfloor}\binom{d}{2k+1} = 2^d(2r)^{d-1}. \tag{D5}$$

□

Lemma 10. *Let* $\Gamma(s,x)$ *denote the upper incomplete Gamma function*

$$\Gamma(s,x) := \int_x^\infty t^{s-1}\mathrm{e}^{-t}\mathrm{d}t. \tag{D6}$$

Then, assuming $s\geqslant 1$ *and* $x>s-1$

$$\Gamma(s,x) \leqslant \frac{\mathrm{e}^{-x}x^s}{x-s+1}. \tag{D7}$$

Proof.

$$\Gamma(s,x) = \mathrm{e}^{-x}\int_0^\infty (t+x)^{s-1}\mathrm{e}^{-t}\mathrm{d}t \leqslant \mathrm{e}^{-x}x^{s-1}\int_0^\infty \mathrm{e}^{\frac{t}{x}(s-1)-t}\mathrm{d}t \tag{D8}$$

This bound can be improved using continued fraction representation. □

Lemma 11.

$$\left(\frac{d}{4}\right)^{-d^2/8} \geqslant \frac{1}{\prod_{k=1}^d k!} \tag{D9}$$

Proof. We first lower bound the value of $\log(k!)$ from below. It is clear that a sum

$$\log(k!) = \log(1)+\log(2)+\ldots+\log(k) \tag{D10}$$

can be lower bounded by $(k-j+1)\log(j)$ for any $1 \leqslant j \leqslant k$. Picking $j = \lfloor \frac{k}{2} \rfloor$ and using the monotonicity of $x\log(x)$ we obtain

$$\frac{k}{2}\log\left(\frac{k}{2}\right) \leqslant \log(k!)\,. \tag{D11}$$

Using this bound and repeating the argument for

$$\log\left(\prod_{k=1}^{d} k!\right) = \log(1!) + \log(2!) + \ldots + \log(d!) \tag{D12}$$

we obtain

$$\frac{d^2}{8}\log\left(\frac{d}{4}\right) \leqslant \log\left(\prod_{k=1}^{d} k!\right). \tag{D13}$$

The result follows via exponentiation. □

Lemma 12.

$$\mathcal{R} \leqslant \overline{\mathcal{R}} := \frac{C(d,\sigma)\,2^m}{|W|} 2^{d-1} \cdot (2\pi)^m \left(1+\frac{d}{2}\right)^m 2^{m+d-1} \mathrm{e}^{-\frac{d\pi^2}{2\sigma}} \tag{D14}$$

for $\sigma \leqslant \frac{2\pi^2 d}{d^2+d-2}$.

Proof. Consider a summand for some fixed $k \neq 0$

$$\frac{C(d,\sigma)\,2^m}{|W|} \int_{\mathcal{H}_\pi^{d-1}\setminus\mathcal{H}_{\tilde{\epsilon}}^{d-1}} \mathrm{d}\mu(\phi) \left|\prod_{\alpha>0} \alpha(X_\phi + X_k) \sin\left(\frac{\alpha(X_\phi)}{2}\right)\right| \exp\left(-\frac{1}{4\sigma}\|X_\phi + X_k\|^2\right) \tag{D15}$$

$$\leqslant \frac{C(d,\sigma)\,2^m}{|W|} \int_{\mathcal{H}_\pi^{d-1}\setminus\mathcal{H}_{\tilde{\epsilon}}^{d-1}} \mathrm{d}\mu(\phi) \left|\prod_{\alpha>0} \alpha(X_\phi + X_k)\right| \exp\left(\frac{-1}{4\sigma}\|X_\phi + X_k\|^2\right). \tag{D16}$$

We have

$$\prod_{\alpha>0} \alpha(X_\phi + X_k) = \prod_{1\leqslant i<j<d} (\phi_i - \phi_j + 2\pi(k_i - k_j)) \prod_{1\leqslant i\leqslant d-1} (\phi_i - \phi_d + 2\pi(k_i - k_d)) \tag{D17}$$

and on the domain of integration we can bound $|\phi_i - \phi_j| \leqslant 2\pi$ and $|k_i - k_j| \leqslant d||k||_\infty$ for $i < j \leqslant d$. Thus,

$$\left|\prod_{\alpha>0} \alpha(X_\phi + X_k)\right| \leqslant (2\pi + 2\pi d||k||_\infty)^m \tag{D18}$$

$$= (2\pi)^m (1 + d||k||_\infty)^m. \tag{D19}$$

Let us find lower bounds on the exponents. We have

$$||X_\phi + X_k||^2 = 2d\left(\sum_{i=1}^{d-1} (\phi_i + 2\pi k_i)^2 + (\phi_d + 2\pi k_d)^2\right). \tag{D20}$$

The first term of (D20) is just the square of the Euclidean distance from the origin in coordinate space. This way, figure 1 can be used to understand the summation and bounding process

better. The second term is non-negative and is a square of the sum of all coordinates. As such, it has a minimum of zero on the hyperplane crossing the origin. To simplify the reasoning, we discard the second term altogether and obtain the isotropic bound

$$||X_\phi + X_k||^2 \geqslant 2d\left(2\pi\,||k||_\infty - \pi\right)^2 \tag{D21}$$

$$= 2d\pi^2\left(2||k||_\infty - 1\right)^2. \tag{D22}$$

Denoting

$$\Psi(\phi,k) := \int_{\mathcal{H}_\pi^{d-1}\setminus\mathcal{H}_{\tilde{\epsilon}}^{d-1}} \mathrm{d}\mu(\phi)\left|\prod_{\alpha>0}\alpha\left(X_\phi + X_k\right)\sin\left(\frac{\alpha\left(X_\phi\right)}{2}\right)\right|\exp\left(-\frac{1}{4\sigma}\left\|X_\phi + X_k\right\|^2\right) \tag{D23}$$

and assuming σ is small enough, namely

$$\sigma \leqslant \frac{2\pi^2 d}{d^2+d-2}, \tag{D24}$$

we can use lemmas 9 and 10 to bound the correction term $\mathcal{R}$ as follows[6]

$$\mathcal{R} \leqslant \frac{C(d,\sigma)\,2^m}{|W|}\sum_{k\neq 0}\Psi(\phi,k) \tag{D25}$$

$$\leqslant \frac{C(d,\sigma)\,2^m}{|W|}\sum_{k\neq 0}\cdot\left(1-\mathrm{Vol}\left(B_\epsilon\right)\right)\sum_{k=1}^{\infty}\left|S_{k,d-1}\right|(2\pi)^m(1+dk)^m\exp\left(-\frac{d\pi^2}{2\sigma}(2k-1)^2\right) \tag{D26}$$

$$\leqslant \frac{C(d,\sigma)\,2^m}{|W|}\cdot 2^{2(d-2)+1}(2\pi)^m\sum_{k=1}^{\infty}k^{d-2}(1+dk)^m\exp\left(-\frac{d\pi^2}{2\sigma}(2k-1)^2\right) \tag{D27}$$

$$= \frac{C(d,\sigma)\,2^m}{|W|}2^{d-1}(2\pi)^m\sum_{u=1,u\,\mathrm{odd}}^{\infty}(1+u)^{d-2}\left(1+\frac{d}{2}(u+1)\right)^m\exp\left(-\frac{d\pi^2}{2\sigma}u^2\right) \tag{D28}$$

$$= \frac{C(d,\sigma)\,2^m}{|W|}2^{d-1}(2\pi)^m\left(1+\frac{d}{2}\right)^m\sum_{u=1,u\,\mathrm{odd}}^{\infty}(1+u)^{d-2}\left(1+\frac{d}{d+2}u\right)^m\exp\left(-\frac{d\pi^2}{2\sigma}u^2\right) \tag{D29}$$

$$\leqslant \frac{C(d,\sigma)\,2^m}{|W|}2^{d-1}\cdot(2\pi)^m\left(1+\frac{d}{2}\right)^m 2^{m+d-2}\sum_{u=1,u\,\mathrm{odd}}^{\infty}u^{m+d-2}\exp\left(-\frac{d\pi^2}{2\sigma}u^2\right) \tag{D30}$$

[6] Below we slightly abused the notation by denoting the $||k||_\infty$ as k.

$$\leqslant \frac{C(d,\sigma)\,2^m}{|W|} 2^{d-1}\cdot(2\pi)^m \left(1+\frac{d}{2}\right)^m 2^{m+d-2}\left(\mathrm{e}^{-\frac{d\pi^2}{2\sigma}}+\frac{1}{2}\int_1^\infty u^{m+d-2}\,\exp\left(-\frac{d\pi^2}{2\sigma}u^2\right)\mathrm{d}u\right) \tag{D31}$$

$$= \frac{C(d,\sigma)\,2^m}{|W|} 2^{d-1}\cdot(2\pi)^m \left(1+\frac{d}{2}\right)^m 2^{m+d-2}\left(\mathrm{e}^{-\frac{d\pi^2}{2\sigma}}+\frac{1}{2^2\left(\frac{d\pi^2}{2\sigma}\right)^{\frac{m+d-1}{2}}}\right.$$
$$\left.\times\,\Gamma\left(\frac{m+d-1}{2},\frac{d\pi^2}{2\sigma}\right)\right) \tag{D32}$$

$$= \frac{C(d,\sigma)\,2^m}{|W|} 2^{d-1}\cdot(2\pi)^m \left(1+\frac{d}{2}\right)^m 2^{m+d-2}\left(1+\frac{1}{4\,(A\,(d,\sigma)+1)}\right)\mathrm{e}^{-\frac{d\pi^2}{2\sigma}} \tag{D33}$$

$$\leqslant \frac{C(d,\sigma)\,2^m}{|W|} 2^{d-1}\cdot(2\pi)^m \left(1+\frac{d}{2}\right)^m 2^{m+d-1}\mathrm{e}^{-\frac{d\pi^2}{2\sigma}}, \tag{D34}$$

where

$$A\,(d,\sigma) := \frac{d\pi^2}{2\sigma}-\frac{m+d-1}{2}. \tag{D35}$$

Let us explain the bounding process in more detail. We applied the bound $1-\mathrm{Vol}(B_\epsilon)\leqslant 1$ and lemma 9 to bound (D26) by (D27). We substituted $u=2k-1$ to (D27). We bounded $d/(d+2)\leqslant 1$ [7] and applied a very crude bound

$$(1+u)^{m+d-2}\leqslant 2^{m+d-2}\cdot u^{m+d-2} \tag{D36}$$

to bound (D29) by (D30). The function

$$f(u) := u^{m+d-1}\mathrm{e}^{-\frac{d\pi^2}{2\sigma}u^2} \tag{D37}$$

is increasing from 0 at $u=0$ to its local maximum and then is decreasing. The condition (D24) guarantees that the local maximum of $f(u)$ for $u>0$ is smaller or equal to 1, since this requires a weaker condition

$$\sigma\leqslant\frac{d\pi^2}{m+d-2}. \tag{D38}$$

This requirement allows us to bound the sum over odd u (D30) in terms of its first term plus an appropriate integral (D31). We applied a well-known formula

$$\int_a^\infty x^d\mathrm{e}^{-\alpha x^2}\mathrm{d}x = \frac{\Gamma\left(\frac{d+1}{2},a^2\alpha\right)}{2\alpha^{\frac{d+1}{2}}} \tag{D39}$$

to (D31). Finally, due to (D24), we have that $A(d,\sigma)\geqslant 0$ which allows us to apply lemma 10 [8] to bound (D32) by (D33) and bound (D33) by (D34).

□

[7] Otherwise, $u=1$ needs to be considered separately when bounding binomial by the power of two, since $du/(d+2)<1$. for $u=1$.

[8] The application of lemma 10 requires $A(d,\sigma)>-1$.

We want to compare the upper bound on $\mathcal{R}$ from lemma 12 with an upper bound on $\mathcal{I}_0$ from lemma 8.

Lemma 13. *Let* $\sigma \leqslant \frac{1}{d\log(d)}$. *Then,*

$$\overline{\mathcal{R}} \leqslant \cdot\eta\overline{\mathcal{I}_0} \tag{D40}$$

for $\eta \geqslant \frac{1}{\prod_{k=1}^{d} k!}$ *and* $d \geqslant 2$. *In particular, we can take* $\eta \geqslant 1/2$ *to obtain a uniform bound for all* $d \geqslant 2$.

Proof. Clearly, from lemma 8

$$\overline{\mathcal{I}_0} \geqslant \frac{1}{2}\exp\left(-\frac{d\pi^2}{16\sigma} + \frac{d^2-1}{24}\sigma\right), \tag{D41}$$

so that using lemmas 7 and 12

$$\frac{\overline{\mathcal{R}}}{\overline{\mathcal{I}_0}} \leqslant \frac{R(d)}{\prod_{k=1}^{d} k!}\mathrm{e}^{-\frac{7}{16}\frac{d\pi^2}{\sigma}}, \tag{D42}$$

where

$$R(d) := 2^{\frac{7d}{2}-d^2+4m-\frac{3}{2}} \cdot \pi^{d-\frac{d^2}{2}+2m-\frac{1}{2}} d^{\frac{d}{2}+m} (d+2)^m. \tag{D43}$$

Demanding the ratio bound (D42) to be smaller than η yields

$$\sigma \leqslant \frac{7d\pi^2}{16}\frac{1}{\log(R) - \log\left(\prod_{k=1}^{d} k!\right) - \log(\eta)}. \tag{D44}$$

Bounding $\log(R) \leqslant 3d^2\log(d)$ (for $d \geqslant 2$) we obtain

$$\frac{7\pi^2}{48d\log(d)} \leqslant \frac{7d\pi^2}{16}\frac{1}{\log(R)} \leqslant \frac{7d\pi^2}{16}\frac{1}{\log(R) - \log\left(\prod_{k=1}^{d} k!\right) - \log(\eta)}, \tag{D45}$$

so (D44) is satisfied for

$$\sigma \leqslant \frac{1}{d\log(d)} \tag{D46}$$

and $\eta \geqslant \frac{1}{\prod_{k=1}^{d} k!}$, with the right-hand side decaying very fast with d and upper bounded using lemma 11.

Note that (D46) is stronger than (D24).

□

Appendix E. Bounding the L^2-norm

In this appendix, we prove lemma 16 and corollary 1, which bound the L^2-norm of the trimmed heat kernel $H_P^{(t)}$. The L^2-norm is divided into two contributions (see (111)) which are bounded separately in lemmas 14 and 15.

Lemma 14.

$$\mathcal{I}_{0,0}^2 = \frac{C(d,\sigma)}{2^{m+\frac{l}{2}}}\mathrm{e}^{\|\delta\|^2\sigma} \tag{E1}$$

Proof. We have

$$\mathcal{I}_{0,0}^2 = \frac{C(d,\sigma)^2}{|W|}\int\left(\prod_{\alpha>0}\alpha(X_\phi)^2\right)\exp\left(-\frac{1}{2\sigma}\|X_\phi\|^2\right)\mathrm{d}\mu(\phi), \tag{E2}$$

which is very similar to the integral analysed previously in lemma 8. We introduce the variables $y_j = \phi_j\sqrt{\frac{d}{\sigma}}$ and $y_d = -\sum_{j=1}^{d-1} y_j$ and obtain

$$\mathcal{I}_{0,0}^2 = \frac{C(d,\sigma)^2}{|W|}\left(\frac{\sigma}{d}\right)^{m+\frac{l}{2}}(2\pi)^{-(d-1)}\frac{1}{\sqrt{d}}\int_{\mathcal{Z}}\mathrm{d}\mu_{\mathcal{Z}}(y)\left(\prod_{1\leqslant i<j\leqslant d}(y_i-y_j)^2\right)\exp\left(-\sum_j y_j^2\right) \tag{E3}$$

$$= \frac{C(d,\sigma)}{2^{m+\frac{l}{2}}}\frac{C(d,\sigma)}{|W|}\left(\frac{2\sigma}{d}\right)^{m+\frac{l}{2}}(2\pi)^{-(d-1)}\frac{1}{\sqrt{d}}\int_{\mathcal{Z}}\left(\prod_{1\leqslant i<j\leqslant d}(y_i-y_j)^2\right)$$
$$\times\exp\left(-\sum_j y_j^2\right)\mathrm{d}\mu_{\mathcal{Z}}(y) \tag{E4}$$

$$= \frac{C(d,\sigma)}{2^{m+\frac{l}{2}}}\mathrm{e}^{\|\delta\|^2\sigma}\left(\prod_{k=1}^{d}\frac{1}{k!}\right)2^{\frac{1}{2}(d-1)d}\pi^{\frac{1}{2}-\frac{d}{2}}\int_{\mathcal{Z}}\left(\prod_{1\leqslant i<j\leqslant d}(y_i-y_j)^2\right)$$
$$\times\exp\left(-\sum_j y_j^2\right)\mathrm{d}\mu_{\mathcal{Z}}(y), \tag{E5}$$

$$= \frac{C(d,\sigma)}{2^{m+\frac{l}{2}}}\mathrm{e}^{\|\delta\|^2\sigma}\Pr_{A\sim\mathrm{GUE}_k^0}(\|A\|_\infty\leqslant\infty) \tag{E6}$$

$$= \frac{C(d,\sigma)}{2^{m+\frac{l}{2}}}\mathrm{e}^{\|\delta\|^2\sigma}. \tag{E7}$$

□

Lemma 15.

$$\mathcal{R}_{*,0}^2+\mathcal{R}_{0,*}^2+\mathcal{R}_{*,*}^2 \leqslant \frac{9}{8}\frac{d!}{2^{2m}}\overline{\mathcal{R}}^2 \tag{E8}$$

for $\sigma\leqslant\frac{2\pi^2 d}{d^2+d-2}$ *and* $d\geqslant 2$.

Proof. The proof goes along the same lines as in lemma 12, so we direct the reader there for an explanation.

Denoting

$$\Psi(\phi,k,l) := \int_{\mathcal{H}_\pi^{d-1}}\left|\prod_{\alpha>0}\alpha(X_\phi+X_k)\prod_{\alpha>0}\alpha^*(X_\phi+X_l)\right|$$
$$\times\exp\left(-\frac{1}{4\sigma}\left(\|X_\phi+X_k\|^2+\|X_\phi+X_l\|^2\right)\right)\mathrm{d}\mu(\phi) \tag{E9}$$

we have

$$\mathcal{R}^2_{*,*} = \frac{C(d,\sigma)^2}{|W|}\sum_{k\neq 0}\sum_{l\neq 0}\Psi(\phi,k,l) \tag{E10}$$

$$\leqslant \frac{C(d,\sigma)^2}{|W|}\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}|S_{k,d-1}||S_{l,d-1}|(2\pi)^{2m}(1+dk)^m(1+dl)^m \times \exp\left(-\frac{d\pi^2}{2\sigma}\left((2k-1)^2+(2l-1)^2\right)\right) \tag{E11}$$

$$\leqslant \frac{C(d,\sigma)^2}{|W|}\cdot 2^{4(d-2)+2}(2\pi)^{2m}\left(\sum_{k=1}^{\infty}k^{d-2}(1+dk)^m\exp\left(-\frac{d\pi^2}{2\sigma}(2k-1)^2\right)\right)^2 \tag{E12}$$

$$= \frac{C(d,\sigma)^2}{|W|}2^{2(d-1)}(2\pi)^{2m}\left(\sum_{u=1,u\,\mathrm{odd}}^{\infty}(1+u)^{d-2}\left(1+\frac{d}{2}(u+1)\right)^m\exp\left(-\frac{d\pi^2}{2\sigma}u^2\right)\right)^2 \tag{E13}$$

$$= \frac{C(d,\sigma)^2}{|W|}2^{2(d-1)}(2\pi)^{2m}\left(1+\frac{d}{2}\right)^{2m}\left(\sum_{u=1,u\,\mathrm{odd}}^{\infty}(1+u)^{d-2}\left(1+\frac{d}{d+2}u\right)^m \times \exp\left(-\frac{d\pi^2}{2\sigma}u^2\right)\right)^2 \tag{E14}$$

$$\leqslant \frac{C(d,\sigma)^2}{|W|}2^{2(d-1)}\cdot(2\pi)^{2m}\left(1+\frac{d}{2}\right)^{2m}2^{2(m+d-2)}\left(\sum_{u=1,u\,\mathrm{odd}}^{\infty}u^{m+d-2}\exp\left(-\frac{d\pi^2}{2\sigma}u^2\right)\right)^2 \tag{E15}$$

$$\leqslant \frac{C(d,\sigma)^2}{|W|}2^{2(d-1)}\cdot(2\pi)^{2m}\left(1+\frac{d}{2}\right)^{2m}2^{2(m+d-2)}\left(\mathrm{e}^{-\frac{d\pi^2}{2\sigma}}+\frac{1}{2}\int_1^{\infty}u^{m+d-2} \times \exp\left(-\frac{d\pi^2}{2\sigma}u^2\right)\mathrm{d}u\right)^2 \tag{E16}$$

$$= \frac{C(d,\sigma)^2}{|W|}2^{2(d-1)}\cdot(2\pi)^{2m}\left(1+\frac{d}{2}\right)^{2m}2^{2(m+d-2)}\left(\mathrm{e}^{-\frac{d\pi^2}{2\sigma}}+\frac{1}{2^2\left(\frac{d\pi^2}{2\sigma}\right)^{\frac{m+d-1}{2}}} \times \Gamma\left(\frac{m+d-1}{2},\frac{d\pi^2}{2\sigma}\right)\right)^2 \tag{E17}$$

$$= \frac{C(d,\sigma)^2}{|W|}2^{2(d-1)}\cdot(2\pi)^{2m}\left(1+\frac{d}{2}\right)^{2m}2^{2(m+d-2)}\left(1+\frac{1}{4(A(d,\sigma)+1)}\right)^2\mathrm{e}^{-\frac{d\pi^2}{\sigma}} \tag{E18}$$

$$\leqslant \frac{C(d,\sigma)^2}{|W|} 2^{2(d-1)} \cdot (2\pi)^{2m} \left(1+\frac{d}{2}\right)^{2m} 2^{2(m+d-1)} \mathrm{e}^{-\frac{d\pi^2}{\sigma}}, \tag{E19}$$

where $A(d,\sigma)$ is defined by (D35). Similarly,

$$\mathcal{R}^2_{*,0} = \frac{C(d,\sigma)^2}{|W|} \sum_{k\neq 0} \Psi(\phi,k,0) \tag{E20}$$

$$\leqslant \frac{C(d,\sigma)^2}{|W|} \sum_{k=1}^{\infty} |S_{k,d-1}| (2\pi)^{2m} (1+dk)^m \exp\left(-\frac{d\pi^2}{2\sigma}\left((2k-1)^2+1\right)\right) \tag{E21}$$

$$\leqslant \frac{C(d,\sigma)^2}{|W|} \cdot 2^{2(d-2)+1} (2\pi)^{2m} \mathrm{e}^{-\frac{d\pi^2}{2\sigma}} \left(\sum_{k=1}^{\infty} k^{d-2} (1+dk)^m \exp\left(-\frac{d\pi^2}{2\sigma}(2k-1)^2\right)\right) \tag{E22}$$

$$= \frac{C(d,\sigma)^2}{|W|} 2^{d-1} (2\pi)^{2m} \mathrm{e}^{-\frac{d\pi^2}{2\sigma}} \left(\sum_{u=1,u\,\mathrm{odd}}^{\infty} (1+u)^{d-2} \left(1+\frac{d}{2}(u+1)\right)^m \exp\left(-\frac{d\pi^2}{2\sigma}u^2\right)\right) \tag{E23}$$

$$= \frac{C(d,\sigma)^2}{|W|} 2^{d-1} (2\pi)^{2m} \mathrm{e}^{-\frac{d\pi^2}{2\sigma}} \left(1+\frac{d}{2}\right)^m \left(\sum_{u=1,u\,\mathrm{odd}}^{\infty} (1+u)^{d-2} \left(1+\frac{d}{d+2}u\right)^m \times \exp\left(-\frac{d\pi^2}{2\sigma}u^2\right)\right) \tag{E24}$$

$$\leqslant \frac{C(d,\sigma)^2}{|W|} 2^{d-1} \cdot (2\pi)^{2m} \mathrm{e}^{-\frac{d\pi^2}{2\sigma}} \left(1+\frac{d}{2}\right)^m 2^{m+d-2} \left(\sum_{u=1,u\,\mathrm{odd}}^{\infty} u^{m+d-2} \exp\left(-\frac{d\pi^2}{2\sigma}u^2\right)\right) \tag{E25}$$

$$\leqslant \frac{C(d,\sigma)^2}{|W|} 2^{d-1} \cdot (2\pi)^{2m} \mathrm{e}^{-\frac{d\pi^2}{2\sigma}} \left(1+\frac{d}{2}\right)^m 2^{m+d-2} \left(\mathrm{e}^{-\frac{d\pi^2}{2\sigma}} + \frac{1}{2}\int_1^{\infty} u^{m+d-2} \times \exp\left(-\frac{d\pi^2}{2\sigma}u^2\right) \mathrm{d}u\right) \tag{E26}$$

$$= \frac{C(d,\sigma)^2}{|W|} 2^{d-1} \cdot (2\pi)^{2m} \mathrm{e}^{-\frac{d\pi^2}{2\sigma}} \left(1+\frac{d}{2}\right)^m 2^{m+d-2} \left(\mathrm{e}^{-\frac{d\pi^2}{2\sigma}} + \frac{1}{2^2\left(\frac{d\pi^2}{2\sigma}\right)^{\frac{m+d-1}{2}}} \times \Gamma\left(\frac{m+d-1}{2}, \frac{d\pi^2}{2\sigma}\right)\right) \tag{E27}$$

$$= \frac{C(d,\sigma)^2}{|W|} 2^{d-1} \cdot (2\pi)^{2m} \left(1+\frac{d}{2}\right)^m 2^{m+d-2} \left(1+\frac{1}{4(A(d,\sigma)+1)}\right) \mathrm{e}^{-\frac{d\pi^2}{\sigma}} \tag{E28}$$

$$\leqslant \frac{C(d,\sigma)^2}{|W|} 2^{d-1} \cdot (2\pi)^{2m} \left(1+\frac{d}{2}\right)^m 2^{m+d-1} \mathrm{e}^{-\frac{d\pi^2}{\sigma}} \tag{E29}$$

and $\mathcal{R}^2_{0,*} = \mathcal{R}^2_{0,*}$. Thus, for $d \geqslant 2$ we have

$$\mathcal{R}^2_{*,0} + \mathcal{R}^2_{0,*} + \mathcal{R}^2_{*,*} \leqslant \frac{C(d,\sigma)^2}{|W|} 2^{2(d-1)} \cdot (2\pi)^{2m} \left(1 + \frac{d}{2}\right)^{2m} 2^{2(m+d-1)} \times \mathrm{e}^{-\frac{d\pi^2}{\sigma}} \left(1 + \frac{2}{2^{d-1}\left(1+\frac{d}{2}\right)^m 2^{m+d-1}}\right) \tag{E30}$$

$$= \frac{d!}{2^{2m}} \overline{\mathcal{R}}^2 \left(1 + \frac{2}{2^{d-1}\left(1+\frac{d}{2}\right)^m 2^{m+d-1}}\right) \tag{E31}$$

$$\leqslant \frac{9}{8} \frac{d!}{2^{2m}} \overline{\mathcal{R}}^2. \tag{E32}$$

□

Lemma 16 (bound on the L^2-norm of the (trimmed) heat kernel).

$$||H_P^{(t)}(\cdot,\sigma)||_2 \leqslant d\mathcal{I}_{0,0} + d\frac{\sqrt{d!}}{2^{m-1}} \eta \overline{\mathcal{I}_0} \tag{E33}$$

for $\sigma \leqslant \frac{1}{d\log(d)}$, $d \geqslant 2$ *and any* $\eta \geqslant \frac{1}{\prod_{k=1}^d k!}$.

Proof. It is easy to see that $\|H_P(\cdot,\sigma)\|_2 \leqslant |\Gamma| \cdot ||H_S(\cdot,\sigma)||_2$. Thus, from lemmas 14 and 15

$$\|H_P(\cdot,\sigma)\|_2 \leqslant d\sqrt{\mathcal{I}^2_{0,0} + \frac{9}{8}\frac{d!}{2^{2m}}\overline{\mathcal{R}}^2} \leqslant d\mathcal{I}_{0,0} + \frac{3\sqrt{2}}{4} d\frac{\sqrt{d!}}{2^m}\overline{\mathcal{R}}. \tag{E34}$$

Since the terms in (111) are non-negative, it is clear that (E34) can be applied to trimmed heat kernels. The result follows from bounding $\frac{3\sqrt{2}}{4} \leqslant 2$ and the application of lemma 13. □

Corollary 1.

$$||H_P^{(t)}(\cdot,\sigma)||_2 \leqslant c\left(\frac{d}{\sigma}\right)^{\frac{d^2-1}{4}} \tag{E35}$$

for $\sigma \leqslant \frac{1}{d\log(d)}$ *and* $d \geqslant 2$, *where* c *is some positive group constant. For example, one can take* $c = 8$ *for* $d \geqslant 2$ *and* $c = 1$ *for* $d \geqslant 12$.

Proof. Using the proof of lemma 16

$$\|H_P(\cdot,\sigma)\|_2 \leqslant d^{\frac{3}{16}d^2+1}\sqrt{d!}2^{-\frac{d^2}{8}+\frac{d}{4}}\pi^{\frac{d-1}{4}}\mathrm{e}^{\frac{d^2-1}{24}\sigma}\sigma^{-\frac{d^2-1}{4}} + \frac{3\sqrt{2}}{4} d\frac{\sqrt{d!}}{2^m}\frac{1}{\prod_{k=1}^d k!}\mathrm{e}^{\frac{d^2-1}{24}\sigma} \tag{E36}$$

$$\leqslant d^{\frac{3}{16}d^2+1}\sqrt{d!}2^{-\frac{d^2}{8}+\frac{d}{4}}\pi^{\frac{d-1}{4}}\mathrm{e}^{\frac{d^2-1}{24}\sigma}\sigma^{-\frac{d^2-1}{4}} + \frac{3\sqrt{2}}{4} d\frac{\sqrt{d!}}{2^m}\left(\frac{d}{4}\right)^{-d^2/8}\mathrm{e}^{\frac{d^2-1}{24}\sigma} \tag{E37}$$

$$= d\frac{\sqrt{d!}}{2^m}\left(\frac{d}{4}\right)^{-d^2/8}\mathrm{e}^{\frac{d^2-1}{24}\sigma}\left(\frac{3\sqrt{2}}{4} + d^{\frac{5}{16}d^2}2^{d^2/8-d/4}\pi^{\frac{d-1}{4}}\sigma^{-\frac{d^2-1}{4}}\right) \tag{E38}$$

$$\leqslant d^{\frac{3}{16}d^2+1}\sqrt{d!}2^{-\frac{d^2}{8}+\frac{d}{4}+1}\pi^{\frac{d-1}{4}}\mathrm{e}^{\frac{d^2-1}{24}\frac{1}{d\log(d)}}\sigma^{-\frac{d^2-1}{4}} \tag{E39}$$

for $\sigma \leqslant \frac{1}{d\log(d)}$. The logarithm of the sigma-independent terms can be upper bounded by $\frac{1}{4}(d^2-1)\log(d)$ for $d \geqslant 8$. For $d \geqslant 2$, it can be upper bounded by $\frac{1}{4}(d^2-1)\log(d) + \log(19)$. □

Appendix F. Well-definedness of the Poisson form of the heat kernel at non-regular points

A complete proof of this is beyond the scope of this manuscript, and perhaps the easiest such proof is that of Urakawa [32], showing that the two forms of the heat kernel are equivalent. In this example, we will sketch the idea by demonstrating the phenomenon in the limit as two elements of the vector ϕ become equal to each other.

Fix all elements of ϕ in equation (66) other than ϕ_1 and ϕ_2, set these to be $a+b$ and $a-b$, respectively. Assume all of the ϕ_j are not equal to each other, and none are in the interval $(a-b, a+b)$. We will consider the limit as $b \to 0$. The relevant part of (66) becomes

$$j\left(\exp\left(X_\phi\right)\right)^{-1} \sum_{k\in\mathbb{Z}^{d-1}} \pi\left(X_\phi + X_k\right) \exp\left(-\frac{1}{4\sigma}\left\|X_\phi + X_k\right\|^2\right) = \sum_{k\in\mathbb{Z}^{d-1}} \prod_{1\leqslant i<j<d} \frac{\alpha_{ij}\left(X_\phi + X_k\right)}{\sin\left(\frac{\alpha_{ij}(X_\phi)}{2}\right)}$$
$$\times \exp\left(-\frac{1}{4\sigma}\left\|X_\phi + X_k\right\|^2\right) \tag{F1}$$
$$= \sum_{k\in\mathbb{Z}^{d-1}} \frac{2b + 2\pi\left(k_1 - k_2\right)}{\sin\left(b\right)} \prod_{i,j\,\text{remaining}} \frac{\alpha_{ij}\left(X_\phi + X_k\right)}{\sin\left(\frac{\alpha_{ij}(X_\phi)}{2}\right)} \exp\left(-\frac{1}{4\sigma}\left\|X_\phi + X_k\right\|^2\right), \tag{F2}$$

where $\prod_{i,j\,\text{remaining}}$ denotes all of the terms in the product $\prod_{1\leqslant i<j\leqslant d}$ except for the $i=1, j=2$ term which has now been written explicitly. We now split the sum over k into parts for the cases $k_1 \neq k_2$ and $k_1 = k_2$ To streamline the notation the vector k is now parametrised by the two elements k_1 and k_2 and the remaining $d-3$ dimensional vector k'

$$\sum_{k'\in\mathbb{Z}^{d-3}} \sum_{k_1=k_2} \frac{2b}{\sin\left(b\right)} \prod_{i,j\,\text{remaining}} \frac{\alpha_{ij}\left(X_\phi + X_k\right)}{\sin\left(\frac{\alpha_{ij}(X_\phi)}{2}\right)} \exp\left(-\frac{1}{4\sigma}\left\|X_\phi + X_k\right\|^2\right)$$
$$+ \sum_{k'\in\mathbb{Z}^{d-3}} \sum_{k_1\neq k_2} \frac{2b + 2\pi\left(k_1 - k_2\right)}{\sin\left(b\right)} \prod_{i,j\,\text{remaining}} \frac{\alpha_{ij}\left(X_\phi + X_k\right)}{\sin\left(\frac{\alpha_{ij}(X_\phi)}{2}\right)} \exp\left(-\frac{1}{4\sigma}\left\|X_\phi + X_k\right\|^2\right). \tag{F3}$$

As can be seen, the singularity in the first term now has the expected form $b/\sin(b)$ which converges to 1 in the limit $b \to 0$ and can easily be extended to a continuous function. For $k_1 \neq k_2$ we need to match the term with $k_1 = c$, $k_2 = d$ with the term with $k_1 = d$, $k_2 = c$ in order to obtain the cancellation that we need. Let $\hat{k}$ be the vector with elements c, d, k', and $\tilde{k}$ be the vector with elements d, c, k'. With this notation, the second term in equation (F3) becomes

$$\sum_{k'\in\mathbb{Z}^{d-3}} \sum_{c<d} \left(\frac{2b + 2\pi\left(c-d\right)}{\sin\left(b\right)} \prod_{i,j\,\text{remaining}} \frac{\alpha_{ij}\left(X_\phi + X_{\hat{k}}\right)}{\sin\left(\frac{\alpha_{ij}(X_\phi)}{2}\right)} \exp\left(-\frac{1}{4\sigma}\left\|X_\phi + X_{\hat{k}}\right\|^2\right) \right.$$
$$\left. + \frac{2b + 2\pi\left(d-c\right)}{\sin\left(b\right)} \prod_{i,j\,\text{remaining}} \frac{\alpha_{ij}\left(X_\phi + X_{\tilde{k}}\right)}{\sin\left(\frac{\alpha_{ij}(X_\phi)}{2}\right)} \exp\left(-\frac{1}{4\sigma}\left\|X_\phi + X_{\tilde{k}}\right\|^2\right) \right), \tag{F4}$$

and we can now take the limit $b \to 0$ to observe that

$$\lim_{b\to 0} \prod_{i,j\,\text{remaining}} \frac{\alpha_{ij}\left(X_\phi + X_{\hat{k}}\right)}{\sin\left(\frac{\alpha_{ij}(X_\phi)}{2}\right)} \exp\left(-\frac{1}{4\sigma}\left\|X_\phi + X_{\hat{k}}\right\|^2\right)$$
$$= \lim_{b\to 0} \prod_{i,j\,\text{remaining}} \frac{\alpha_{ij}\left(X_\phi + X_{\tilde{k}}\right)}{\sin\left(\frac{\alpha_{ij}(X_\phi)}{2}\right)} \exp\left(-\frac{1}{4\sigma}\left\|X_\phi + X_{\tilde{k}}\right\|^2\right). \tag{F5}$$

Denoting this limit $L(k)$, we obtain

$$\sum_{k'\in\mathbb{Z}^{d-3}} \sum_{c<d} L(k) \lim_{b\to 0}\left(\frac{2b + 2\pi(c-d)}{\sin(b)} + \frac{2b + 2\pi(d-c)}{\sin(b)}\right) = 4 \sum_{k'\in\mathbb{Z}^{d-3}} \sum_{c<d} L(k), \tag{F6}$$

where we have cancelled the $(c-d)$ term with the $(d-c)$ term and used the well-known fact that $\lim_{b\to 0} \frac{b}{\sin(b)}$ converges. An identical phenomenon appears when more than 2 eigenvalues become equal to each other, but proving this directly is considerably more cumbersome.

ORCID iDs

Oskar Słowik 0000-0003-4138-3063
Oliver Reardon-Smith 0000-0002-0124-1389
Adam Sawicki 0000-0003-4906-2459

References

[1] Wallman J J and Flammia S T 2014 Randomized benchmarking with confidence *New J. Phys.* **16** 103032
[2] Epstein J M, Cross A W, Magesan E and Gambetta J M 2014 Investigating the limits of randomized benchmarking protocols *Phys. Rev. A* **89** 062321
[3] Scott A J 2008 Optimizing quantum process tomography with unitary 2-designs *J. Phys. A: Math. Theor.* **41** 055308
[4] Dankert C, Cleve R, Emerson J and Livine E 2009 Exact and approximate unitary 2-designs and their application to fidelity estimation *Phys. Rev. A* **80** 012304
[5] Abeyesinghe A, Devetak I, Hayden P and Winter A 2009 The mother of all protocols: restructuring quantum information's family tree *Proc. R. Soc. A* **465** 2537–63
[6] Roy A and Scott A J 2009 Unitary designs and codes, Designs, Codes and Cryptography *Des. Codes Cryptogr.* **53** 13–31
[7] Gross D, Krahmer F and Kueng R 2014 A partial derandomization of phaselift using spherical designs *J. Fourier Anal. Appl.* **21** 229–66
[8] Szehr O, Dupuis F, Tomamichel M and Renner R 2013 Decoupling with unitary approximate two-designs *New J. Phys.* **15** 053022
[9] Bae J, Hiesmayr B C and McNulty D 2019 Linking entanglement detection and state tomography via quantum 2-designs *New J. Phys.* **21** 013012
[10] Sen P 2005 Random measurement bases, quantum state distinction and applications to the hidden subgroup problem (arXiv:quant-ph/0512085)
[11] Helsen J and Walter M 2023 Thrifty shadow estimation: reusing quantum circuits and bounding tails *Phys. Rev. Lett.* **131** 240602
[12] Czartowski J, Goyeneche D, Grassl M and Życzkowski K 2020 Isoentangled mutually unbiased bases, symmetric quantum measurements and mixed-state designs *Phys. Rev. Lett.* **124** 090503
[13] Huang H-Y, Kueng R and Preskill J 2020 Predicting many properties of a quantum system from very few measurements *Nat. Phys.* **16** 1050–7
[14] Brandão F G S L, Harrow A W and Horodecki M 2016 Local random quantum circuits are approximate polynomial-designs *Commun. Math. Phys.* **346** 397–434

[15] Masanes L, Roncaglia A J and Acín A 2013 Complexity of energy eigenstates as a mechanism for equilibration *Phys. Rev. E* **87** 032137
[16] Oszmaniec M, Augusiak R, Gogolin C, Koński J, Acín A and Lewenstein M 2016 Random bosonic states for robust quantum metrology *Phys. Rev. X* **6** 041044
[17] Roberts D A and Yoshida B 2017 Chaos and complexity by design *J. High Energy Phys.* JHEP04(2017)121
[18] Nakata Y, Hirche C, Koashi M and Winter A 2017 Efficient quantum pseudorandomness with nearly time-independent Hamiltonian dynamics *Phys. Rev. X* **7** 021006
[19] Brandão F G, Chemissany W, Hunter-Jones N, Kueng R and Preskill J 2021 Models of quantum complexity growth *PRX Quantum* **2** 030316
[20] Oszmaniec M, Kotowski M, Horodecki M and Hunter-Jones N 2024 Saturation and recurrence of quantum complexity in random local quantum dynamics *Phys. Rev. X* **14** 041068
[21] Harrow A W and Montanaro A 2017 Quantum computational supremacy *Nature* **549** 203–9
[22] Boixo S, Isakov S V, Smelyanskiy V N, Babbush R, Ding N, Jiang Z, Bremner M J, Martinis J M and Neven H 2018 Characterizing quantum supremacy in near-term devices *Nat. Phys.* **14** 595–600
[23] Arute F *et al* 2019 Quantum supremacy using a programmable superconducting processor *Nature* **574** 505–10
[24] Hangleiter D, Bermejo-Vega J, Schwarz M and Eisert J 2018 Anticoncentration theorems for schemes showing a quantum speedup *Quantum* **2** 65
[25] Yoganathan M, Jozsa R and Strelchuk S 2019 Quantum advantage of unitary Clifford circuits with magic state inputs *Proc. R. Soc. A* **475** 20180427
[26] Kuperberg G 2023 Breaking the cubic barrier in the Solovay-Kitaev algorithm (arXiv:2306.13158)
[27] Słowik O, Dulian P, and Sawicki A 2025 Quantum circuit overhead (arXiv:2505.00683)
[28] Oszmaniec M, Sawicki A and Horodecki M 2022 Epsilon-nets, unitary designs and random quantum circuits *IEEE Trans. Inf. Theory* **68** 989
[29] Varjú P P 2015 Random walks in compact groups (arXiv:1209.1745)
[30] Harrow A W, Recht B and Chuang I L 2002 Efficient discrete approximations of quantum gates *J. Math. Phys.* **43** 4445–51
[31] Słowik O and Sawicki A 2023 Calculable lower bounds on the efficiency of universal sets of quantum gates *J. Phys. A: Math. Theor.* **56** 115304
[32] Urakawa H 1974 The heat equation on a compact Lie group *Osaka J. Math.* **2** 285
[33] Zygmund A 2003 *Trigonometric Series* 3rd edn (Cambridge University Press)
[34] Kac M 1966 Can one hear the shape of a drum? *Am. Math. Month.* **73**
[35] Chavel I 1984 *Eigenvalues in Riemannian Geometry* (Academic)
[36] Vassilevich D 2003 Heat kernel expansion: user's manual *Phys. Rep.* **388** 279
[37] Avramidi I G 2015 *Heat Kernel Method and its Applications* (Springer)
[38] Avramidi I G 2000 *Heat Kernel and Quantum Gravity* (Springer)
[39] Atiyah M, Bott R and Patodi V K 1973 On the heat equation and the index theorem *Invent. math.*
[40] Berline N, Getzler E and Vergne M 2004 *Heat Kernels and Dirac Operators* (Springer)
[41] Perelman G 2002 The entropy formula for the Ricci flow and its geometric applications (arXiv:math/0211159)
[42] Morgan J and Tian G 2007 *Ricci Flow and the Poincaré Conjecture* (American Mathematical Society)
[43] Grigor'yan A 2009 *Heat Kernel and Analysis on Manifolds (AMS/IP Studies in Advanced Mathematics* vol 47) (American Mathematical Society)
[44] Maher D G 2006 Brownian motion and heat kernels on compact Lie groups and symmetric spaces *PhD Thesis* The University of New South Wales
[45] Faraut J 2008 *Analysis on Lie Groups: An Introduction, Cambridge Studies in Advanced Mathematics* (Cambridge University Press)
[46] Hall B 2015 Lie Groups, Lie Algebras and Representations: Elementary Introduction *Graduate Texts in Mathematics* vol 222 (Springer)
[47] Fulton W and Harris J 1991 Representation theory: a first course *Graduate Texts in Mathematics* vol 129 (Springer)
[48] Kirillov A 2008 An introduction to Lie groups and Lie algebras *Cambridge Studies in Advanced Mathematics* vol 113 (Cambridge University Press)
[49] Sugiura M 1971 Fourier series of smooth functions on compact Lie groups *Osaka J. Math.* **8** 33–47

[50] Benkart G, Chakrabarti M, Halverson T, Leduc R, Lee C and Stroomer J 1994 Tensor product representations of general linear groups and their connections with Brauer algebras *J. Algebr.* **166** 529
[51] Chatzigeorgiou I 2013 Bounds on the Lambert function and their application to the outage analysis of user cooperation *IEEE Commun. Lett.* **17** 1505
[52] Tracy C A and Widom H 2001 On the distributions of the lengths of the longest monotone subsequences in random words *Probab Theor. Relat Fields*
[53] Aubrun G and Szarek S J 2017 *Alice and Bob meet Banach* (American Mathematical Society)

Chapter 5

Paper III: Quantum Circuit Overhead

5.1 Overview

In the third paper, we aimed to introduce an informative and computable quantity to evaluate the efficiency of universal gate sets $\mathcal{S}$ and use it to gain insight into the efficiency of some commonly-used single-qubit gate sets.

Although the SKL theorems (e.g. based on the finite-scale spectral gap) can be used to derive upper bounds on the efficiency of discrete $\mathcal{S}$, such bounds depend on the number of gates in $\mathcal{S}$. This dependence is also reflected in the optimal value of $\delta(\nu_{\mathcal{S}})$, which scales as $\Theta(1/\sqrt{\mathcal{S}})$ (see Section 2.4.1). On the other hand, the volumetric bound on the efficiency also depends on $|\mathcal{S}|$. As it is intuitive that allowing more gates in $\mathcal{S}$ should result in better efficiency, in practice, each gate requires specific experimental procedures to be reliably executed. Thus, one can argue that the sets $\mathcal{S}$ shouldn't contain too many elements, and it is informative to compare the sets $\mathcal{S}$ within the gates of the same number of elements $|\mathcal{S}|$. This is the primary rationale behind the measure of efficiency introduced in this paper, called the Quantum Circuit Overhead (QCO).

In this paper, we introduce the notion of the QCO, which is the ratio of the lengths of the shortest circuits built out of gates from $\mathcal{S}$ producing an ε-net, i.e. the computational efficiency $\ell(\mathcal{S}, \varepsilon)$, compared to the efficiency of the optimal gate set with $|\mathcal{S}|$ elements. Although the QCO cannot be computed directly, we explain how it can be bounded from above by the quantity Q given by a simple formula

$$Q(\mathcal{S}, \varepsilon) \coloneqq \frac{\log(|\mathcal{S}|)}{\log\left(1/\delta(\nu_{\mathcal{S}}, t(\varepsilon))\right)}, \tag{5.1}$$

where $t(\varepsilon)$ is the t stemming from the t-design and ε-net correspondence, so that $Q(\mathcal{S}, \varepsilon)$ can be calculated by the numerical spectral gap computations at scale $t(\varepsilon) \simeq d^{5/2}/\varepsilon$. This formula arises from upper-bounding the efficiency of $\mathcal{S}$ using the SKL theorem, based on the finite-scale spectral gap and lower-bounding the corresponding efficiency for the best gate set using a simple volumetric argument. The formula (5.1) can also be understood as an upper bound on the so-called covering exponent in the case of uniform weights [114].

Moreover, we introduce the related notion of the T-Quantum Circuit Overhead (T-QCO), where we focus on the occurrence of specific gates, denoted T_i, which are assumed to be considerably more costly than the remaining gates. This is the case for many fault-tolerant implementations based on the Clifford+T gate set, where the fault-tolerant implementation of the T gate (also known as $P(\pi/4)$ gate) typically dominates the overall cost. We conveniently choose the remaining "free" gates to form a non-universal group C, which does not intersect with the set $\{T_i\}$. Similarly to QCO, the T-QCO can be upper bounded by the quantity Q_T, obtained by applying formula (5.1) to the gate set $\mathcal{S}_T$ derived from $\mathcal{S}$ by conjugating the gates T_i with the elements from C. This simple conjugation trick directly relates the computational complexity of the circuits over $\mathcal{S}_T$ with the T-complexity, i.e., the complexity in which we count only the occurrences of operations in $\{T_i\}$.

We note that the notion of QCO and T-QCO can be applied in the NISQ setting too. In particular, the gate set $\mathcal{S}$ does not need to be discrete; however, in this case, the notion of overhead reduces to the numerator $\ell(\mathcal{S}, \epsilon)$. For example, one can use the T-QCO to analyse the T-complexity of some fixed entangling operations while allowing the "free" group C to contain all the single-qubit unitary channels. Finally, the T-QCO is a good proxy for the overall cost-effectiveness of gate sets for the architectures, with a clear separation into the set of costly gates $\{T_i\}$ and the gates whose cost can be neglected, C. We provide examples of architectures for which such an assumption is reasonable, including the fault-tolerant architectures based on 2D surface and color codes.

We perform extensive numerical simulations of the upper bounds Q/Q_T for the random ensembles [1] of single-qubit gates, focusing on ensembles obtained by complementing the Clifford and Hurwitz groups with a Haar-random gate and purely Haar-random gates with the corresponding number of gates $|\mathcal{S}|$. We distinguish the case of generic Haar-random gates (of infinite order) and Haar-random gates with fixed finite order (8 for the Clifford

[1] We use approx. 10^4 gates sets per ensemble.

and 2 for the Hurwitz group). We compare the resulting histograms of Q/Q_T with each other as well as with certain specific values, corresponding to "special" completions of both finite groups. Namely, the Super-Golden gates and the $P(\pi/4)$ gate (in the case of the Clifford group). We argue that in each case, the value of Q/Q_T (which depends on ε) can be lower-bounded in the $\varepsilon \to 0$ limit by the optimal value stemming from considerations involving the Kesten-McKay measure.

Our numerical results show that the histograms of Q/Q_T enjoy fast stabilization with growing t (i.e. diminishing ε), allowing the computations to be terminated at the scales t below the theoretical value $t(\varepsilon)$. Moreover, the Super-Golden gate for the Hurwitz group saturates the optimal value within the inspected range of t. Interestingly, this is not the case for the Clifford group. Our numerical Monte-Carlo search for the optimal completions of those groups confirms that the optimal completion for the Hurwitz group is close to the Super-Golden gate. Similar analysis for the Clifford group found the optimal completion with $Q_T \approx 3.7$ by the gates of the form $UP(3\pi/4)U^\dagger$, where U is a Bloch sphere rotation around any axis $(x, y, 0)$ with $|x| \neq |y|$ by an angle in $[\pi/8, \pi/2]$. This is close to the optimal value $Q_T \approx 3.4$. Somewhat intuitively, the worst completions were obtained for U belonging to the Clifford group. In particular, the completion by the famous $P(\pi/4)$ resulted in a very poor value of $Q_T \approx 52$, lying far away from the center of mass of the Q_T histogram. In this sense, the famous $P(\pi/4)$ is a very non-optimal completion of the single-qubit Clifford group.

Finally, one should acknowledge that we are comparing what we can calculate, namely the upper bounds Q/Q_T on the (T)-QCO. This does not guarantee that the actual value of the T-QCO for the $P(\pi/4)$ completion is poor. However, we believe that the actual value of the overhead is sufficiently correlated with the upper bound to make some conclusions. The second delicate matter is whether a representative of the family of optimal completions for the Clifford group we found can be implemented fault-tolerantly with a sufficiently low cost compared to the $P(\pi/4)$ gate.

5.2 Contribution statement

My contribution to this article was:

1. Writing the paper, except figure generation, Appendix E, and the description of the optimal completions for the Clifford and Hurwitz group in Section V.

2. Refinement of the notion of QCO, including the idea of T-QCO.

3. Co-planning of the numerical experiments to be performed.

Quantum Circuit Overhead

Oskar Słowik*
Center for Theoretical Physics, Polish Academy of Sciences, Aleja Lotników 32/46, 02-668 Warszawa, Poland

Piotr Dulian
Center for Theoretical Physics, Polish Academy of Sciences, Aleja Lotników 32/46, 02-668 Warszawa, Poland and Centre for Quantum Optical Technologies, Centre of New Technologies, University of Warsaw, Banacha 2c, 02-097 Warsaw, Poland

Adam Sawicki†
Center for Theoretical Physics, Polish Academy of Sciences, Aleja Lotników 32/46, 02-668 Warszawa, Poland and Guangdong Technion - Israel Institute of Technology, 241 Daxue Road, Jinping District, Shantou, Guangdong Province, China

(Dated: October 3, 2025)

We introduce a measure for evaluating the efficiency of finite universal quantum gate sets $\mathcal{S}$, called the Quantum Circuit Overhead (QCO), and the related notion of T-Quantum Circuit Overhead (T-QCO). The overhead is based on the comparison between the efficiency of $\mathcal{S}$ versus the optimal efficiency among all gate sets with the same number of gates. We demonstrate the usefulness of the (T-)QCO by extensive numerical calculations of its upper bounds, providing insight into the efficiency of various choices of single-qubit $\mathcal{S}$, including Haar-random gate sets and the gate sets derived from finite subgroups, such as Clifford and Hurwitz groups. In particular, our results suggest that, in terms of the upper bounds on the T-QCO, the famous T gate is a highly non-optimal choice for the completion of the Clifford gate set, even among the gates of order 8. We identify the optimal choices of such completions for both finite subgroups.

I. INTRODUCTION

Quantum circuit [1, 2] is a universal model for quantum computation in which quantum information is processed via the application of a series of unitary operations called quantum logic gates. Similarly to a classical computer, whose computation can be described using the classical circuit model, every global quantum operation on a qubit register can be realized using a universal finite set of elementary operations. A set of such quantum logic gates is referred to as the universal gate set or, in the context of quantum hardware, the native gate set.

Contrary to the classical case, the finite length quantum circuits built out of a finite discrete set $\mathcal{S}$ of quantum gates can be used to implement arbitrary multi-qubit (global) unitary operations only approximately, up to some error ϵ (in a suitable metric). The number of elementary gates needed to implement a target unitary operation U with precision ϵ using gates from $\mathcal{S}$ is a measure of the complexity of U with respect to $\mathcal{S}$ [1–3]. For a universal gate set $\mathcal{S}$ and any finite ϵ, the complexity of any U is finite and thus can be upper bounded by the shortest circuit length, $\ell(\mathcal{S}, \epsilon)$, so that any U can be ϵ-approximated by a quantum circuit built out of $\mathcal{S}$ of length at most $\ell(\mathcal{S}, \epsilon)$. This number can be understood as an absolute measure of the efficiency of $\mathcal{S}$ at the scale of ϵ-approximations. Since the implementation of quantum gates is always flawed, for reasonably small nonzero ϵ, this number fully characterizes the efficiency of $\mathcal{S}$.

Quantum compilation [1, 4, 5] is a process whose main objective is to approximate the target quantum circuit from the high-level hardware-agnostic representation used by quantum programmers to the form expressible by the native gate set executable on a specific quantum computer. Another task handled by the compiler is circuit optimization, which, loosely speaking, involves reducing the resources of quantum circuits, such as the depth of the circuit or the number of specific gates used. In the case of the current noisy intermediate-scale quantum (NISQ) machines, which do not enjoy quantum error correction, the reduction of the circuit depth and the number of costly gates (such as the noisy entangling gates) is of utmost practical importance [6–8]. On the other hand, in the fault-tolerant regime, due to the Eastin-Knill theorem [9–11], the number of resource-costly non-traversal gates often determines the bottleneck [12–14]. For example, in the case of Clifford+T gate sets realized using many topological codes, such as 2D surface or color codes, the focus is usually on the reduction of the T-count i.e. the number of non-transversal T gates (also known as the $P(\pi/4)$ or $\pi/8$ gates [1]), which

* oslowik@cft.edu.pl
† a.sawicki@cft.edu.pl

[1] To avoid confusion with the T symbol occurring in T-QCO, we

leads to an improvement in error rates, runtime and the number of qubits needed to perform the computations [15–22]. However, the compilation process is fundamentally limited by the efficiency of the used gate set $\mathcal{S}$.

Aside from the applications in the description of information processing occurring in quantum computers, quantum circuits can be used to describe the discrete unitary dynamics of general discrete quantum systems [23–25]. Such an approach has been recently proposed to gain insight into the physics of black hole interiors, and interesting results regarding the saturation and recurrence of the complexity of such systems have been obtained [26, 27]. Such behaviour also depends on the efficiency of gate sets $\mathcal{S}$ used to model the system.

Although it is conjectured that the generic universal gate sets $\mathcal{S}$ have, so called spectral gap, which implies the optimal asymptotic efficiency $\ell(\mathcal{S},\epsilon) = \Theta\left(\log(1/\epsilon)\right)$, the quantitative methods to bound and compare the efficiency of various gate sets $\mathcal{S}$ are not well-developed.

In this work, we introduce and study the relative measure of the efficiency of universal gate sets $\mathcal{S}$ that we call the quantum circuit overhead (QCO) and the related notion of T-Quantum Circuit Overhead (T-QCO). The notion of overhead is based on the comparison of the efficiency $\ell(\mathcal{S},\epsilon)$ among the gate sets $\mathcal{S}$ having the same number of elements, where the optimal efficiency is denoted $\ell_{\text{opt}}(|\mathcal{S}|,\epsilon)$. Crucially, both overheads can be upper-bounded by essentially calculable quantities, namely Q and Q_T, respectively, which can be obtained from numerical simulations.

To demonstrate the feasibility of our method and its applications, we provide extensive numerical examples in which we calculate Q/Q_T, focusing on the comparison between the two scenarios for single-qubit gate sets:

1. A Haar-random set $\mathcal{S}$ with a fixed number of elements (of infinite or fixed finite order r),

2. A set $\mathcal{S}$ composed of a finite group (such as Clifford or Hurwitz group) completed with a single Haar-random gate (of infinite or fixed finite order r), making the set universal.

In the second scenario, we compare such random ensembles with some "special" choices, e.g. the $P(\pi/4)$ gate in the case of the Clifford group, gaining insight into their efficiency. The inclusion of the finite order cases is motivated by the fault-tolerance considerations and the analysis of the so-called Super-Golden Gates [28]. Surprisingly, our results suggest that the $P(\pi/4)$ gate is a highly non-optimal choice among all gates of order $r = 8$ in terms of Q_T. We also identified the best possible gates of orders $r = 8$ and $r = 2$ in the Clifford and Hurwitz group cases, respectively.

refer to the T gate as $P(\pi/4)$ gate.

Although our numerical experiments focus on a single-qubit case, our framework can be applied in any dimension, in particular to the multiqubit gates. Moreover, it can be (in principle) applied to the setting in which the universal gate set is not discrete, e.g. consists of parametrized gates. We refrained from performing such experiments due to their computational costs.

In order to upper bound the overhead, we need to be able to upper bound $\ell(\mathcal{S},\epsilon)$ and lower bound $\ell_{\text{opt}}(|\mathcal{S}|,\epsilon)$.

II. SOLOVAY-KITAEV LIKE THEOREMS

Lossless unitary quantum operations on n-qubit register are described via the unitary channels $\mathbf{U}(\rho) = U\rho U^\dagger$, which form a group $\mathbf{U}(d)$, where $d = 2^n$. This group can be naturally identified with the projective unitary group $\mathrm{PU}(d)$. We use the following metric on $\mathbf{U}(d)$

$$d(\mathbf{U},\mathbf{V}) \coloneqq \min_{\varphi} ||U - e^{i\varphi}V||_\infty, \tag{1}$$

where by $||\cdot||_\infty$ we denote the operator norm and U, V are the unitary representatives of the channels $\mathbf{U}$ and $\mathbf{V}$ respectively (see Appendix A for more details).

The famous Solovay-Kitaev (SK) theorem states that if $\mathcal{S} \subset \mathbf{U}(d)$ is a finite universal symmetric (i.e. inverse-closed) set of quantum gates, then $\ell(\mathcal{S},\epsilon) = \mathcal{O}(\log^c(1/\epsilon))$, where the constant c depends on the proof and typically $c \approx 3.97$ or $c = 3 + \alpha$, for any $\alpha > 0$ [1, 2, 29]. The proofs are constructive, so that an (efficient) algorithm exists that can find the desired decompositions. As a result, the SK algorithm serves as the foundation of modern quantum compilation. Since its introduction, many similar (constructive and non-constructive) poly-logarithmic upper bounds $\ell(\mathcal{S},\epsilon) = \mathcal{O}(\mathrm{Poly}(\log(1/\epsilon))$ have been provided [30–37]. Such theorems often work for groups other than $\mathbf{U}(d)$, e.g., semi-simple compact Lie groups, and use different assumptions on the gates in $\mathcal{S}$; we refer to them as Solovay-Kitaev-like (SKL) theorems.

For example, in terms of constructive/algorithmic SKL theorems, the cubic $\ell(\mathcal{S},\epsilon)$ scaling in the SK algorithm was recently improved in [30] to $\log_\phi(2) \approx 1.44$, where ϕ is the golden ratio. The construction assumes that $\mathcal{S}$ is finite and inverse-closed. On the other hand, in [31], the authors provided the generalization of the SK algorithm working for any finite universal (i.e., not necessarily inverse-closed) sets $\mathcal{S}$, with $\ell(\mathcal{S},\epsilon) = \mathcal{O}(\log^{\gamma_d}(1/\epsilon))$ and $\gamma_d = \Theta(\log(d))$.

However, it is known that for finite $\mathcal{S}$, all poly-logarithmic bounds with exponent 1 are asymptotically tight. The Haar volume[2] of an ϵ-ball $B_\epsilon \subset \mathbf{U}(d)$ can be bounded as

$$(a_v\epsilon)^{d^2-1} \leq \mathrm{Vol}(B_\epsilon) \leq (A_v\epsilon)^{d^2-1}, \tag{2}$$

[2] Due to translational invariance of Haar measure and the metric, the volume of a ball does not depend on its origin.

with known constants $a_v = \frac{1}{9\pi}$ and $A_v = 87$. These constants were provided in [36] using methods from [38]. Then, using the simple volume counting argument [1, 32], one may express the lower bound on $\ell(\mathcal{S}, \epsilon)$ as

$$\ell_{\mathrm{vol}}(|\mathcal{S}|, \epsilon) \approx \frac{d^2 - 1}{\log(|\mathcal{S}|)} \log\left(\frac{1}{A_v \epsilon}\right), \tag{3}$$

where $A_v = 87$, so $\ell(\mathcal{S}, \epsilon) = \Omega(\log(1/\epsilon))$.

This lower bound depends only on the number of elements in $\mathcal{S}$. Hence, it can be used to lower bound $\ell_{\mathrm{opt}}(|\mathcal{S}|, \epsilon)$, which yields

$$\ell(\mathcal{S}, \epsilon) \geq \ell_{\mathrm{opt}}(|\mathcal{S}|, \epsilon) \geq l_{\mathrm{vol}}(|\mathcal{S}|, \epsilon) \tag{4}$$

It is known that such an optimal scaling $\Theta(\log(1/\epsilon))$ can be obtained for $\mathcal{S}$, having a so-called spectral gap. It is useful to reformulate this property to the language of unitary δ-approximate t-designs.

A unitary δ-approximate t-design is a probability measure ν on $\mathbf{U}(d)$ which mimics the averaging properties of Haar measure μ when applied to balanced polynomials with degree bounded by t, up to some discreptancy $\delta(\nu, t) := \|T_{\nu,t} - T_{\mu,t}\|_\infty$, where

$$T_{\mu,t} := \int_{\mathbf{U}(d)} d\mu(U) U^{t,t}, \quad T_{\nu,t} := \int_{\mathbf{U}(d)} d\nu(U) U^{t,t}, \tag{5}$$

are so called t-moment (averaging) operators, $U^{t,t} := U^{\otimes t} \otimes \bar{U}^{\otimes t}$, and we require $\delta(\nu, t) < 1$ (see Appendix B for more information). For any gate set $\mathcal{S}$, by $\nu_{\mathcal{S}}$ we denote the uniform probability measure supported on its elements.

Note that for symmetric $\mathcal{S}$,

$$||T^{\ell}_{\nu_{\mathcal{S}},t} - T_{\mu,t}||_\infty = \delta^{\ell}(\nu_{\mathcal{S}}, t) \tag{6}$$

quantifies the difference between the averaging over the circuits of length ℓ and over the Haar measure [3]. Therefore, the smaller $\delta(\nu_{\mathcal{S}}, t)$ is, the shorter the circuits needed to mimic the Haar averaging.

The spectral gap of $\mathcal{S}$ is then $1 - \delta(\nu_{\mathcal{S}})$, where $\delta(\nu_{\mathcal{S}})$ is the supremum of $\delta(\nu_{\mathcal{S}}, t)$ over all scales t, so that the spectral gap property reads $\delta(\nu_{\mathcal{S}}) < 1$. The quantitative version of the statement about the efficiency of gate sets $\mathcal{S}$ with a spectral gap is a non-constructive SKL theorem [32, 33] and it states that if $\delta(\nu_{\mathcal{S}}) > 0$, then for any precision ϵ every operation U from $\mathbf{U}(d)$ can be approximated by a sequence of gates from $\mathcal{S}$ of the length

$$\frac{d^2 - 1}{\log(1/\delta(\nu_{\mathcal{S}}))} \log\left(\frac{2}{A_v \epsilon}\right). \tag{7}$$

Notice that although the scaling is optimal, the pre-factor may be arbitrarily large. Moreover, in our examples, the pre-factor is bounded from below via $\delta(\nu_{\mathcal{S}}) \geq \delta_{\mathrm{opt}}(\mathcal{S})$, where

$$\delta_{\mathrm{opt}}(\mathcal{S}) := \frac{2\sqrt{|\mathcal{S}| - 1}}{|\mathcal{S}|}. \tag{8}$$

[39] (see Appendix C for more detailed explaination). We say a finite gate set $\mathcal{S}$ is efficient if $\delta(\nu_{\mathcal{S}}) = \delta_{\mathrm{opt}}(\mathcal{S})$ and refer to $\delta_{\mathrm{opt}}(\mathcal{S})$ as the optimal value. Note that the optimal value depends only on the number of gates $|\mathcal{S}|$.

The study of $\delta(\nu_{\mathcal{S}})$ for generic $\mathcal{S}$ is a hard problem as $\delta(\nu_{\mathcal{S}})$ can not be directly calculated. However, some properties of $\delta(\nu_{\mathcal{S}})$ are known. For example, it is known that $\delta(\nu_{\mathcal{S}}) < 1$ for the finite universal sets $\mathcal{S}$ consisting of algebraic elements [40, 41]. This result was later generalized to any compact, simple Lie group [42]. Moreover, it has been conjectured (and is now commonly believed) that $\delta(\nu_{\mathcal{S}}) < 1$ for any finite universal $\mathcal{S}$ and there are known examples of efficient finite single-qubit gate sets $\mathcal{S}$ with $|\mathcal{S}| = p - 1$ for $p \equiv 1 \bmod 4$ [43, 44]. Finally, some commonly used one-qubit gate sets are known to be efficient [28, 45–47]. To the best of our knowledge, the construction of efficient many-qubit gates remains an open problem.

Fortunately, one can still obtain useful non-constructive SKL theorems using the knowledge of $\delta(\nu_{\mathcal{S}}, t)$. Such a finite-scale approach was studied in [34–37] and is sufficient in practice, as it corresponds to studying efficiency at a certain finite precision ϵ. The approach from [36, 37] utilizes the relation between ϵ-nets and δ-approximate t-designs.

A subset of channels $\mathcal{E}$ from $\mathbf{U}(d)$ is an ϵ-net if for every channel $\mathbf{U}$ from $\mathbf{U}(d)$, there exists a channel $\mathbf{V}$ from $\mathcal{E}$, such that $d(\mathbf{U}, \mathbf{V}) \leq \epsilon$. In other words, $\mathcal{E}$ contains all the possible channels up to the error ϵ. It is intuitively clear that ϵ-nets formed by quantum circuits built from $\mathcal{S}$ and δ-approximate t-designs supported on them are related. However, the quantitative relations between them were not known until recently. Such bounds for the group $\mathbf{U}(d)$ were first rigorously studied in [36], where the authors show [4] that a set is an ϵ-net if it is a support of a δ-approximate t-design with the parameters obeying the following scalings

$$t(\epsilon) \gtrsim \frac{d^{5/2}}{\epsilon}, \quad \delta(\epsilon) \lesssim \left(\frac{\epsilon^{3/2}}{d}\right)^{d^2} \tag{9}$$

(see [36] for precise formulas). A more recent study improves the second scaling to $\delta(\epsilon) \lesssim (\epsilon/d^{1/2})^{d^2}$ [37].

From the point of view of nonabelian Fourier analysis on groups, such reciprocal relation between t and ϵ can be intuitively understood as the relation between distances on the group and its corresponding "frequency" space,

[3] For non-symmetric $\mathcal{S}$ we have an inequality.

[4] The result is more general as it does not assume that the measure is uniform.

so that smaller ϵ corresponds to faster varying functions. The quantitative version of such SKL theorem was proved in [36] and states [5] that for a fixed precision ϵ, every operation U from $\mathbf{U}(d)$ can be ϵ-approximated by sequences of gates from $\mathcal{S}$ of the length $\ell_\delta(\mathcal{S},\epsilon)$

$$\ell(\mathcal{S},\epsilon) \leq \ell_\delta(\mathcal{S},\epsilon) \sim \frac{d^2-1}{\log\left(1/\delta(\nu_{\mathcal{S}},t(\epsilon))\right)} \log\left(\frac{1}{\epsilon}\right), \quad (10)$$

where $t(\epsilon)$ is the bound of type (9) stemming from the ϵ-net t-design correspondence. Thus, we can say that $\delta(\nu_{\mathcal{S}},t(\epsilon))$ upper bounds the efficiency of $\mathcal{S}$ on the level of ϵ-approximations. Moreover, for not too large values of t and d, the value of $\delta(\nu_{\mathcal{S}},t)$ can be calculated using supercomputing clusters. Conveniently, contrary to the Solovay-Kitaev theorem, such SKL theorem can be applied to arbitrary $\mathcal{S}$, in particular to continuous $\mathcal{S}$.

The distribution of $\delta(\nu_{\mathcal{S}},t)$ for (fully) Haar-random ensembles of finite $\mathcal{S}$ was studied in [48], with the extensive numerical analysis suggesting fast stabilization of the distribution with growing t. Our numerical experiments further validate this observation and extend it to all types of ensembles of gate sets studied in this paper. Hence, although the bounds (9) provide some theoretical guarantees on the scales t needed to gain insight into the ϵ-scale efficiency (via (10)), our results suggest that in practice, it suffices to compute $\delta(\nu_{\mathcal{S}},t)$ for t much smaller than the bounds $t(\epsilon)$.

Although from (10) it seems like $\delta(\nu_{\mathcal{S}},t)$ is a good measure of the efficiency of finite $\mathcal{S}$, the value of $\delta(\nu_{\mathcal{S}},t)$ is sensitive to the number of gates $|\mathcal{S}|$. In particular, as the number of gates $|\mathcal{S}|$ goes to infinity, the optimal value (8), which lower bounds the supremum of $\delta(\nu_{\mathcal{S}},t)$ over t, goes to 0. Since the implementation of gate sets $\mathcal{S}$ with large $|\mathcal{S}|$ is costly in practice, e.g. due to the necessary calibrations of quantum hardware, it makes sense to compare the gate sets $\mathcal{S}$ of fixed $|\mathcal{S}|$. This motivates us to introduce the notion of the overhead of quantum circuits.

III. QUANTUM CIRCUIT OVERHEAD

We define the Quantum Circuit Overhead (QCO) of a finite universal gate set $\mathcal{S}$ for ϵ-approximations as the ratio between the smallest length of circuits over $\mathcal{S}$ which form an ϵ-net, $\ell(\mathcal{S},\epsilon)$, and the optimal length $\ell_{\text{opt}}(|\mathcal{S}|,\epsilon)$ achievable using gate sets with the same number of gates $|\mathcal{S}|$. Such a quantity is very hard to calculate in general, however we can bound it from above by bounding $\ell(\mathcal{S},\epsilon)$ from above and $\ell_{\text{opt}}(|\mathcal{S}|,\epsilon)$ from below using (3), (4) and (10) as follows

$$\frac{\ell(\mathcal{S},\epsilon)}{\ell_{\text{opt}}(|\mathcal{S}|,\epsilon)} \leq \frac{\ell_\delta(\mathcal{S},\epsilon)}{\ell_{\text{vol}}(|\mathcal{S}|,\epsilon)} \lesssim Q(\mathcal{S},\epsilon), \quad (11)$$

where we define the computable upper bound on QCO as

$$Q(\mathcal{S},\epsilon) \coloneqq \frac{\log(|\mathcal{S}|)}{\log\left(1/\delta(\nu_{\mathcal{S}},t(\epsilon))\right)}, \quad (12)$$

and $t(\epsilon)$ is the bound stemming from the ϵ-net t-design correspondence of type (9). Note that $Q(\mathcal{S},\epsilon)$ is a non-increasing function of ϵ. It is interesting to study the asymptotic behavior of (12) in the limit of $\epsilon \to 0$ (i.e. $t \to \infty$), namely we define

$$\overline{Q}(\mathcal{S}) \coloneqq \limsup_{\epsilon\to 0} Q(\mathcal{S},\epsilon). \quad (13)$$

For efficient gates, we can use (8) to obtain

$$\overline{Q}_{\text{opt}}(\mathcal{S}) \coloneqq \frac{\log(|\mathcal{S}|)}{\log\left(\frac{|\mathcal{S}|}{2\sqrt{|\mathcal{S}|-1}}\right)} \geq 2, \quad (14)$$

where $\overline{Q}_{\text{opt}}(\mathcal{S}) \gtrapprox 2$ for large $|\mathcal{S}|$. We refer to $\overline{Q}_{\text{opt}}(\mathcal{S})$ as the optimal value, since it is a lower bound on $\overline{Q}(\mathcal{S})$ attainable on the efficient gate sets $\mathcal{S}$.

Notably, our definition of QCO still makes sense for the infinite $\mathcal{S}$, however then it simplifies to the efficiency $\ell(\mathcal{S},\epsilon)$ due to $\ell_{\text{opt}}(|\mathcal{S}|,\epsilon)) = 1$ being realized trivially.

The notion of QCO is suitable for scenarios in which one is interested in the pure computational efficiency of the gate sets or, in the context of quantum computers, the total gate count of the circuits (see Example 1). Practical architectures in which such a scenario may be relevant include the homogeneous-cost models based on anyons (see Table I).

Example 1 (single-qubit gate count) *Consider a single-qubit NISQ architecture with a gate set $\mathcal{S}$, consisting of gates with similar fidelities. Then the QCO of $\mathcal{S}$, which boils down to the analysis of the gate count/circuit depth, is a sensible measure of the efficiency of $\mathcal{S}$.*

IV. T-QUANTUM CIRCUIT OVERHEAD

In many architectures, it is reasonable to count the occurrence of the specific gates, which are considered to be particularly costly, while discarding the occurrences of remaining operations, regarded as relatively "free" (see Examples 2 and 3). This motivates us to introduce the following definition of the T-Quantum Circuit Overhead (T-QCO).

Let C be a group of quantum operations in $\mathbf{U}(d)$ and suppose our chosen set of gates is of the form

$$\mathcal{S} = C \cup \{T_1, \ldots T_n\}, \quad (15)$$

where $T_i \notin C$ are additional operations which make $\mathcal{S}$ universal. We consider the operations in C as free resources and want to focus on the occurrences of the costly operations, denoted as T_i. Thus, we are interested in the

[5] Original Proposition 2 in [36] has $1-\delta(\nu_{\mathcal{S}},t)$ instead of $\log\left(1/\delta(\nu_{\mathcal{S}},t)\right)$ due to unnecessary bounding.

T-complexities of operations U in $\mathbf{U}(d)$, i.e. the smallest number of T_i gates needed to ϵ-approximate U using operations from $\mathcal{S}$. Hence, in analogy to the definition of QCO, we define the T-Quantum Circuit Overhead (T-QCO) of a finite gate set $\mathcal{S}$ for ϵ-approximations as the ratio between the smallest T-count of the circuits over $\mathcal{S}$ which form an ϵ-net and the optimal T-count over all gate sets of the form (15), with the same number of gates.

To bound the T-QCO of the set $\mathcal{S}$, we consider the following derived set of operations

$$\mathcal{S}_T \coloneqq \bigcup_{i\in[n]} \left\{cT_ic^\dagger, c\in C\right\}, \tag{16}$$

which allows us to upper bound the T-QCO by

$$Q_T(\mathcal{S},\epsilon) \coloneqq Q(\mathcal{S}_T,\epsilon) \tag{17}$$

(see Appendix D for a detailed explanation).

Of course, in practice, the physical gate set does not need to include the entire group of free operations C, but rather some chosen generators. In such a case, the group C should be understood as the group generated by the "free" gates. Such a procedure is justified as long as the elements of C can be considered sufficiently cheap.

Similarly to QCO, the definition of T-QCO is also applicable to infinite gate sets.

Finally, the T-QCO is well-defined for reasonably small ϵ, so that the denominator is non-zero.

Example 2 (CNOT-count flavour) *Consider a NISQ n-qubit architecture with the parametrized 2-qubit entangling gates $\mathrm{Ent}_{i,i+1}(\overline{\phi})$ with similar fidelities, acting on qubits i and $i+1$ for $1\le i\le n-1$. We pick $\mathcal{S}$ as in (15) where $C = \boldsymbol{U}(2)^{\otimes n}$ [6] and $T_i = \mathrm{Ent}_{i,i+1}(\overline{\phi})$, for $1\le i\le n-1$. Then the T-QCO of $\mathcal{S}$ is the sensible measure of efficiency wrt to the choice of $\overline{\phi}$.*

Example 3 (T-count flavour) *Consider a fault-tolerant architecture with n (logical) qubits, such that the Clifford gates are low-cost compared to the parametrized family of non-Clifford phase gates $P(\phi)$, which can be implemented with similar cost. We pick $\mathcal{S}$ as in (15), where $C = \mathcal{C}_n$ is the n-qubit Clifford group and T_i, for $1\le i\le n$, is the non-Clifford $P(\phi)$ gate acting on the i-th qubit. Then the T-QCO of $\mathcal{S}$ is the sensible measure of efficiency wrt to the choice of ϕ.*

Contrary to the QCO, the notion of T-QCO is most suitable for scenarios in which the gate set can be strongly separated into a group of gates with negligible cost and a group with (similar) high cost. For example, in NISQ architectures, the T-QCO can be applied with T_i being the chosen entangling gates (see Example 2). For fault-tolerant architectures, see Table I and Example 3.

[6] In this example we used $\mathbf{U}(2)$ as the set of single qubit operations to integrate them out and focus on the impact of the entangling gates. However, any single qubit gate set can be used.

V. NUMERICAL EXAMPLES

We provide the numerical examples focusing on the calculation of the upper bounds on QCO and T-QCO, given by Q (12) and Q_T (17), respectively (see Appendix E for more details about the methods used in numerical experiments). The calculations were performed on a supercomputing cluster.

We consider two types of one-qubit finite universal gate sets:

1. Haar-random gate sets with n elements of (finite or infinite) order r, denoted $\mathcal{S}_{\mu,n,r}$,
2. gate sets derived from a finite subgroup $C\subset\mathbf{U}(2)$:
   (a) completed with a fixed gate T, denoted C_T,
   (b) completed with a single Haar-random gate of (infinite or fixed finite) order r, denoted $C_{\mu,r}$,

   following the setting (15).

We analyze two choices of one-qubit C - the Clifford group $\mathcal{C}$ and the Hurwitz group $\mathcal{H}$. For each C, we construct a random ensemble of $\approx 10^4$ derived universal gate sets of type $C_{\mu,r}$, where r is ∞ or equal to either 8 or 2 for $\mathcal{C}$ and $\mathcal{H}$, respectively. This way, we obtain histograms representing the probability density of Q_T for a fixed t. We increase the value of t until the histograms stabilize and mark the corresponding optimal values of Q_T (see Fig. 1 and Fig. 2 for $\mathcal{C}_{\mu,r}$ ensembles and Fig. 4 and Fig. 5 for $\mathcal{H}_{\mu,r}$ ensembles). The optimal value does not depend on the scale t and lower bounds the histograms in $t\to\infty$ limit.

Moreover, we compare such histograms with analogous histograms of Q for the same-size ensembles of type $\mathcal{S}_{\mu,n,r}$ containing the corresponding number of gates $n=|C|$ (see Fig. 3 for Clifford group and Fig. 6 for Hurwitz group) and with the values of Q_T for gate sets of type C_T with "special" choices of T.

The comparison with the purely random ensembles $\mathcal{S}_{\mu,n,r}$ is relevant from the theoretical point of view, as such gate sets are generic and the distribution of $\delta(\nu_{\mathcal{S}},t)$ can be studied using Random Matrix models [58].

Finally, we identify the choices of T giving the best values of Q_T, among all gates of order $r=8$ (for the Clifford group) and $r=2$ (for the Hurwitz group). We achieve this by the Monte Carlo search over the relevant random completions $C_{\mu,r}$.

Additionally, we check the tightness of the bound (8) in the case of ensembles of type $C_{\mu,r}$ with finite r by calculating the distributions of singular values of the corresponding t-moment operator (see Appendix C and Fig. 7 and Fig. 8 for more details).

A. Clifford group

The one-qubit Clifford subgroup $\mathcal{C}\subset\mathbf{U}(2)$ has 24 elements and is generated by

TABLE I. Examples of fault-tolerant architectures to which (T-)QCO can be applied as a reasonable proxy for an overall efficiency.

Category	**Architecture**	**Cheap operations (C)**	**Costly operations ($\{T_i\}$)**	**Cost split**	**Metric**
NISQ	NISQ devices (general) [7]	Local single-qubit group on each qubit (Euler/ZXZ primitives treated as free)	Entangling gates (e.g. CNOT/CZ/MS/Rydberg)	**High**; 2-qubit gates dominate time/error	**T-QCO**
Fault-tolerant (Code-based)	2D surface code [49, 50]	Clifford subgroup via lattice surgery / Pauli-based computation	T via magic state distillation (factory limited)	**Very high**; T-state throughput bottleneck	**T-QCO**
	2D color code [51, 52]	Transversal Clifford subgroup (baseline codes)	T via magic state distillation or gauge-fixing / code-switching	**Very high**; T preparation dominates	
	3D surface code [53]	Subgroup generated by Cliffords *and* transversal CCZ (treat CCZ as cheap)	T via magic state distillation	**High**; cheap CCZ leaves T as the main costly gate	
	3D color code (baseline) [52, 54]	Transversal Clifford + CCZ subgroup (cheap); some boundary/gauge variants enable transversal T	T via injection/gauge-fixing (when not made transversal)	**High**; clear cheap/costly split in baseline setting	
	Triorthogonal codes / CSS-T (factories) [55]	Clifford subgroup inside distillation circuits	Production/consumption of high-fidelity T/CCZ resource states	**Very high** within factory; resource states dominate	
Fault-tolerant (Anyonic)	Ising / Majorana anyons [56, 57]	Clifford subgroup by braiding	T via magic state injection (or equivalent)	**High**; injections dominate	**T-QCO**
	Fibonacci anyons (braiding-universal) [56]	No robust cheap subgroup; all gates from braids	All gates via braiding (cost by compiled braid length)	**Homogeneous cost**; no cheap/costly split	**QCO**

$$\mathcal{C} = \left\langle \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \right\rangle, \tag{18}$$

up to normalization. The special choices of T gates include the $P(\pi/4)$ gate (of order $r = 8$) and the so-called Super-Golden gate [59] (of order $r = 2$), denoted T_{24}

$$P(\pi/4) = \begin{pmatrix} 1 & 0 \\ 0 & 1+i \end{pmatrix}, \quad T_{24} = \begin{pmatrix} -1-\sqrt{2} & 2-\sqrt{2}+i \\ 2-\sqrt{2}-i & 1+\sqrt{2} \end{pmatrix}, \tag{19}$$

up to normalization.

The value for the gate set $\mathcal{C}_{P(\pi/4)}$ is way outside the range of Fig. 1 and Fig. 2, with $Q_T \approx 52$ for $t = 500$.

For the $\mathcal{C}_{\mu,8}$ ensemble, the additional Haar-random gate of order $r = 8$ has two possible forms

$$U^\dagger P(\pi/4) U \quad \text{and} \quad U^\dagger P(3\pi/4) U, \tag{20}$$

where U is a Haar-random gate. These two cases correspond to the rotation on the Bloch sphere by $\pi/4$ and $3\pi/4$ around a random axis. The best T-QCO upper bound found in our numerical computations is $Q_T \approx 3.7$ for $t = 500$, which is close to the optimal value $\overline{Q}_{\text{opt}} \approx 3.4$. It can be attained for the second form from (20) with U being a Bloch sphere rotation around any axis $(x, y, 0)$ with $|x| \neq |y|$ by an angle in $[\pi/8, \pi/2]$. Interestingly, the worst T-QCO upper bound with $Q_T \approx 52$ for $t = 500$ was achieved when U was an element of the Clifford group.

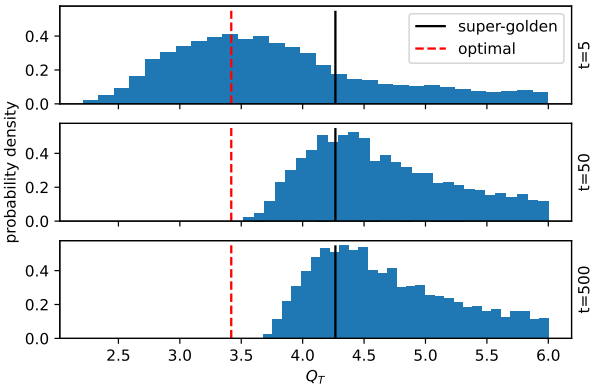


FIG. 1. The histograms of Q_T probability density for an ensemble of type $\mathcal{C}_{\mu,\infty}$ with increasing t. The dashed line denotes the corresponding optimal value. The solid line corresponds to a Super-Golden gate set $\mathcal{C}_{T_{24}}$.

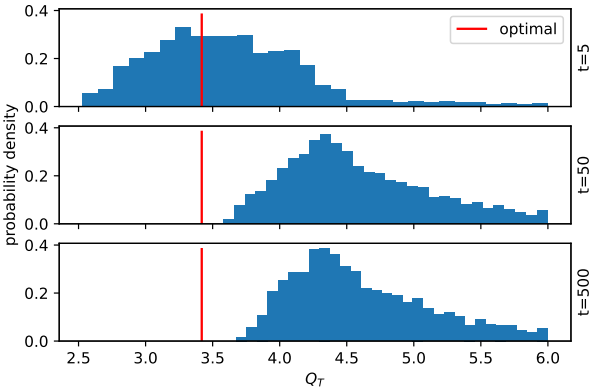


FIG. 2. The histograms of Q_T probability density for an ensemble of type $\mathcal{C}_{\mu,8}$. The solid line denotes the corresponding optimal value.

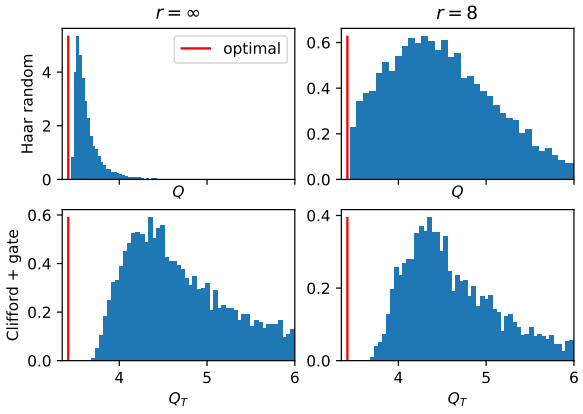


FIG. 3. The histograms of Q_T probability density for ensembles of type $\mathcal{C}_{\mu,r}$ (bottom) vs the histogram of Q for the corresponding ensembles of type $\mathcal{S}_{\mu,24,r}$ (top) for $t = 500$. The solid line denotes the corresponding optimal value. Note that the scales on the Y-axis differ.

B. Hurwitz group

The one-qubit Hurwitz subgroup $\mathcal{H} \subset \mathbf{U}(2)$ has 12 elements and is generated by

$$\mathcal{H} = \left\langle \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \right\rangle, \tag{21}$$

up to normalization. The special choice of T gate is the Super-Golden gate (of order $r = 2$), denoted T_{12}

$$T_{12} = \begin{pmatrix} 3 & 1-i \\ 1+i & -3 \end{pmatrix}, \tag{22}$$

up to normalization.
For the $\mathcal{H}_{\mu,2}$ ensemble, the additional Haar-random gate

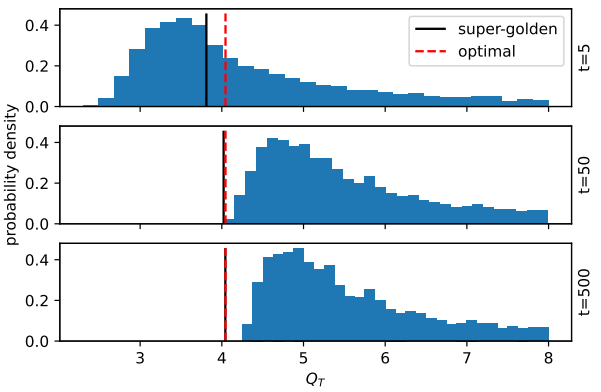


FIG. 4. The histograms of Q_T probability density for an ensemble of type $\mathcal{H}_{\mu,\infty}$ with increasing t. The dashed line denotes the corresponding optimal value. The solid line corresponds to a Super-Golden gate set $\mathcal{H}_{T_{12}}$.

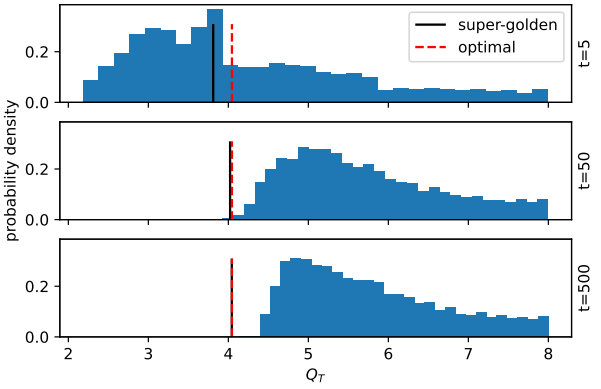


FIG. 5. The histograms of Q_T probability density for an ensemble of type $\mathcal{H}_{\mu,2}$ with increasing t. The dashed line denotes the corresponding optimal value. The solid line corresponds to a Super-Golden gate set $\mathcal{H}_{T_{12}}$.

of order $r = 2$ is a Bloch sphere rotation by π around a random axis. According to our numerical results, the optimal T-QCO bound $\overline{Q}_{\text{opt}} \approx 4$ is attained for a Super-Golden gate set $\mathcal{H}_{T_{12}}$, where T_{12} is a rotation around $(1, 1, \sqrt{9})/\sqrt{11}$. Computations for random gates also showed that the best $Q_T \approx 4.1$ for $t = 500$ is obtained for gates close to T_{12}.

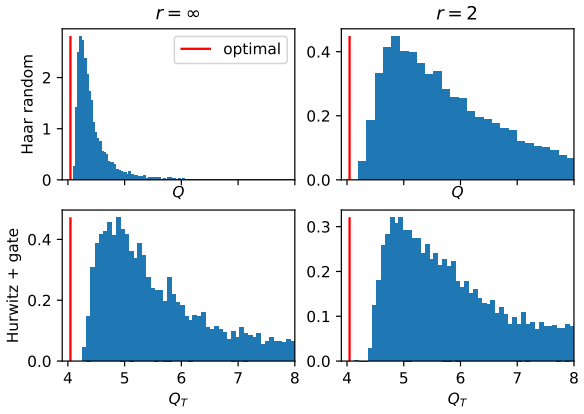


FIG. 6. The histograms of Q_T probability density for ensembles of type $\mathcal{H}_{\mu,r}$ (bottom) vs the histogram of Q for the corresponding ensembles of type $\mathcal{S}_{\mu,12,r}$ (top) for $t = 500$. The solid line denotes the corresponding optimal value. Note that the scales on the Y-axis differ.

VI. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we introduce the new measure of efficiency of universal sets of quantum gates, called the Quantum Circuit Overhead (QCO) and the related notion of T-Quantum Circuit Overhead (T-QCO). Our measure quantifies the overhead of a fixed gate set's efficiency compared to the optimal gate set with the same number of gates, at a given approximation scale. The concept of overhead can be applied to various NISQ and fault-tolerant architectures as a reasonable first approximation of the real cost-effectiveness of gate sets. We provide formulas for Q and Q_T, which are the upper bounds on QCO and T-QCO, respectively, as well as their asymptotically optimal values (lower bounds) for all settings considered in the numerical examples. We performed extensive numerical calculations on a supercomputing cluster to study various random ensembles of universal single-qubit gate sets, particularly those derived as completions of a Clifford and Hurwitz group with a Haar-random gate of infinite or finite order r. In our experiments, we compare various gate sets using the Q/Q_T quantity.

Our numerical examples demonstrate that computing upper bounds on (T-)QCO is tractable on existing supercomputing infrastructure, at least for single-qubit gate sets, with the Q/Q_T distributions stabilizing rapidly. Generic gate sets $\mathcal{S}_{\mu,n,r}$ consistently scored better in Q/Q_T than the structured ones. Interestingly, in the case of the Clifford group, the gate sets completed with the $P(\pi/4)$ gate turned out to perform significantly worse than the generic completions in terms of Q_T. Moreover, our analysis shows that the $P(\pi/4)$ gate is a highly non-optimal choice among the gates of order $r = 8$ in this metric. In this case, we identified the best-performing gates of the same order as the family of the conjugates of $P(3\pi/4)$ by the Bloch sphere rotation around any axis $(x, y, 0)$ with $|x| \neq |y|$ by an angle in $[\pi/8, \pi/2]$. Finally, our results suggest that so-called single-qubit Super-Golden-Gates based on the Hurwitz group enjoy the optimal asymptotic value of Q_T. Interestingly, it does not seem to be the case for the Clifford group construction.

Clearly, one should be cautious about drawing conclusions about the overhead from the comparison of the upper bounds Q/Q_T. Our preliminary numerical analysis of $\ell(\mathcal{S}, \epsilon)$ for Haar-random gate sets with three gates indicates that a small Q is related to small overhead. Although we have not observed the opposite, i.e. it seems like large Q does not imply significant overhead, we suspect that such behaviour should be apparent as $\epsilon \to 0$. Indeed, we have observed the separation of the values of $\delta(\nu_{\mathcal{S}}, \epsilon)$ from 1 for the gate sets with lowest $\ell(\mathcal{S}, \epsilon)$ and the smallest value of ϵ we were able to use, $\epsilon = 0.1$.

Moreover, the optimisation of gates based on T-QCO is relevant in the quantum computing context only if compared gate sets can be implemented with similar cost.

In terms of future directions, it would be interesting to perform numerical experiments for gate sets with larger locality, particularly those containing entangling gates. Such an approach may help identify good entangling gates within some parametrized families. Additionally, one would like to find and study fault-tolerant architectures that admit efficient implementations of the conjugate of $P(3\pi/4)$, as found in the paper, to enhance the practical importance of this result. Finally, although the explicit calculation of (T-)QCO is, in general, intractable, it may be worthwhile to extend our preliminary analysis further to study smaller values of ϵ.

ACKNOWLEDGMENTS

This research was funded by the National Science Centre, Poland under the grant OPUS: UMO2020/37/B/ST2/02478. We gratefully acknowledge Polish high-performance computing infrastructure PLGrid (HPC Center: ACK Cyfronet AGH) for providing computer facilities and support within computational grant no. PLG/2024/017436.

DATA AVAILABILITY

The code used in the numerical experiments is publicly available [60].

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).

[2] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi, *Classical and Quantum Computation* (American Mathematical Society, USA, 2002).

[3] S. Aaronson, The complexity of quantum states and transformations: From quantum money to black holes (2016), arXiv:1607.05256 [quant-ph].

[4] Y. Ge, W. Wenjie, C. Yuheng, P. Kaisen, L. Xudong, Z. Zixiang, W. Yuhan, W. Ruocheng, and Y. Junchi, Quantum circuit synthesis and compilation optimization: Overview and prospects (2024), arXiv:2407.00736 [quant-ph].

[5] T. Häner, D. S. Steiger, K. Svore, and M. Troyer, A software methodology for compiling quantum programs, Quantum Science and Technology **3**, 020501 (2018).

[6] V. Gheorghiu, J. Huang, S. M. Li, M. Mosca, and P. Mukhopadhyay, Reducing the CNOT count for Clifford+T circuits on NISQ architectures, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **42**, 1873–1884 (2023).

[7] J. Preskill, Quantum Computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[8] K. Noh, L. Jiang, and B. Fefferman, Efficient classical simulation of noisy random quantum circuits in one dimension, Quantum **4**, 318 (2020).

[9] B. Eastin and E. Knill, Restrictions on transversal encoded quantum gate sets, Phys. Rev. Lett. **102**, 110502 (2009).

[10] M. P. Woods and Á. M. Alhambra, Continuous groups of transversal gates for quantum error correcting codes from finite clock reference frames, Quantum **4**, 245 (2020).

[11] P. Faist, S. Nezami, V. V. Albert, G. Salton, F. Pastawski, P. Hayden, and J. Preskill, Continuous symmetries and approximate quantum error correction, Physical Review X **10**, 10.1103/physrevx.10.041018 (2020).

[12] D. Gottesman, Quantum error correction and fault-tolerance (2005), arXiv:quant-ph/0507174 [quant-ph].

[13] Y.-H. Luo, M.-C. Chen, M. Erhard, H.-S. Zhong, D. Wu, H.-Y. Tang, Q. Zhao, X.-L. Wang, K. Fujii, L. Li, N.-L. Liu, K. Nemoto, W. J. Munro, C.-Y. Lu, A. Zeilinger, and J.-W. Pan, Quantum teleportation of physical qubits into logical code spaces, Proceedings of the National Academy of Sciences **118**, e2026250118 (2021), https://www.pnas.org/doi/pdf/10.1073/pnas.2026250118.

[14] B. Eastin and E. Knill, Restrictions on transversal encoded quantum gate sets, Physical Review Letters **102**, 10.1103/physrevlett.102.110502 (2009).

[15] M. E. Beverland, A. Kubica, and K. M. Svore, Cost of universality: A comparative study of the overhead of state distillation and code switching with color codes, PRX Quantum **2**, 020341 (2021).

[16] V. Gheorghiu, M. Mosca, and P. Mukhopadhyay, T-count and T-depth of any multi-qubit unitary, npj Quantum Information **8**, 10.1038/s41534-022-00651-y (2022).

[17] F. J. R. Ruiz, T. Laakkonen, J. Bausch, M. Balog, M. Barekatain, F. J. H. Heras, A. Novikov, N. Fitzpatrick, B. Romera-Paredes, J. van de Wetering, A. Fawzi, K. Meichanetzidis, and P. Kohli, Quantum circuit optimization with AlphaTensor, Nature Machine Intelligence **7**, 374 (2025).

[18] D. Gosset, V. Kliuchnikov, M. Mosca, and V. Russo, An algorithm for the T-count, Quantum Info. Comput. **14**, 1261–1276 (2014).

[19] V. Vandaele, Lower T-count with faster algorithms (2024), arXiv:2407.08695 [quant-ph].

[20] H. Zhou, C. Zhao, M. Cain, D. Bluvstein, C. Duckering, H.-Y. Hu, S.-T. Wang, A. Kubica, and M. D. Lukin, Algorithmic fault tolerance for fast quantum computing (2024), arXiv:2406.17653 [quant-ph].

[21] L. Heyfron and E. T. Campbell, An efficient quantum compiler that reduces T count (2018), arXiv:1712.01557 [quant-ph].

[22] A. G. Fowler, A. M. Stephens, and P. Groszkowski, High-threshold universal quantum computation on the surface code, Physical Review A **80**, 10.1103/physreva.80.052312 (2009).

[23] T. Tokusumi, A. Matsumura, and Y. Nambu, Quantum circuit model of black hole evaporation, Classical and Quantum Gravity **35**, 235013 (2018).

[24] M. P. Fisher, V. Khemani, A. Nahum, and S. Vijay, Random quantum circuits, Annual Review of Condensed Matter Physics **14**, 335–379 (2023).

[25] P. W. Claeys, M. Henry, J. Vicary, and A. Lamacraft, Exact dynamics in dual-unitary quantum circuits with projective measurements, Physical Review Research **4**, 10.1103/physrevresearch.4.043212 (2022).

[26] P. Hayden and J. Preskill, Black holes as mirrors: quantum information in random subsystems, Journal of High Energy Physics **2007**, 120–120 (2007).

[27] M. Oszmaniec, M. Kotowski, M. Horodecki, and N. Hunter-Jones, Saturation and recurrence of quantum complexity in random local quantum dynamics, Phys. Rev. X **14**, 041068 (2024).

[28] P. Sarnak, Letter to Scott Aaronson and Andy Pollington on the Solovay-Kitaev theorem (2015).

[29] C. M. Dawson and M. A. Nielsen, The Solovay-Kitaev algorithm (2005), arXiv:quant-ph/0505030 [quant-ph].

[30] G. Kuperberg, Breaking the cubic barrier in the Solovay-Kitaev algorithm (2023), arXiv:2306.13158 [quant-ph].

[31] A.Bouland and T. Giurgica-Tiron, Efficient universal quantum compilation: An inverse-free Solovay-Kitaev algorithm (2021), arXiv:2112.02040 [quant-ph].

[32] A. W. Harrow, B. Recht, and I. L. Chuang, Efficient discrete approximations of quantum gates, Journal of Mathematical Physics **43**, 4445–4451 (2002).

[33] O. Słowik and A. Sawicki, Calculable lower bounds on the efficiency of universal sets of quantum gates, Journal of Physics A: Mathematical and Theoretical **56**, 115304 (2023).

[34] D. Dolgopyat, On mixing properties of compact group extensions of hyperbolic systems, Israel Journal of Mathematics **130**, 157 (2002).

[35] P. P. Varjú, Random walks in compact groups, Documenta Mathematica **18**, 1137 (2013).

[36] M. Oszmaniec, A. Sawicki, and M. Horodecki, Epsilon-nets, unitary designs, and random quantum circuits, IEEE Transactions on Information Theory **68**, 989 (2022).

[37] O. Słowik, O. Reardon-Smith, and A. Sawicki, Fundamental solutions of heat equation on unitary groups establish an improved relation between ϵ-nets and approximate unitary t-designs (2025), arXiv:2503.08577 [quant-ph].
[38] S. J. Szarek, Metric entropy of homogeneous spaces, Banach Center Publications **43** (1998).
[39] H. Kesten, Symmetric random walks on groups, Transactions of the American Mathematical Society **92**, 336 (1959).
[40] J. Bourgain and A. Gamburd, On the spectral gap for finitely-generated subgroups of SU(2), Inventiones mathematicae **171**, 83 (2007).
[41] J. Bourgain and A. Gamburd, A spectral gap theorem in SU(d) (2011), arXiv:1108.6264 [math.GR].
[42] Y. Benoist and N. de Saxcé, A spectral gap theorem in simple Lie groups (2014), arXiv:1405.1808 [math.RT].
[43] A. Lubotzky, R. Phillips, and P. Sarnak, Hecke operators and distributing points on the sphere I, Communications on Pure and Applied Mathematics. Supplement: Proceedings of the Symposium on Frontiers of the Mathematical Sciences: 1985. **39**, S149 (1986).
[44] A. Lubotzky, R. Phillips, and P. Sarnak, Hecke operators and distributing points on S^2. II, Communications on Pure and Applied Mathematics **40**, 401 (1987).
[45] A. Bocharov, Y. Gurevich, and K. M. Svore, Efficient decomposition of single-qubit gates into V basis circuits, Physical Review A **88** (2013).
[46] P. Selinger, Efficient Clifford+T approximation of single-qubit operators, Quantum Information and Computation **15**, 159 (2015).
[47] V. Kliuchnikov, D. Maslov, and M. Mosca, Practical approximation of single-qubit unitaries by single-qubit quantum Clifford and T circuits, IEEE Transactions on Computers **65**, 161 (2016).
[48] P. Dulian and A. Sawicki, Matrix concentration inequalities and efficiency of random universal sets of quantum gates, Quantum **7**, 983 (2023).
[49] D. Horsman, A. G. Fowler, S. Devitt, and R. V. Meter, Surface code quantum computing by lattice surgery, New Journal of Physics **14**, 123011 (2012).
[50] D. Litinski, A game of surface codes: Large-scale quantum computing with lattice surgery, Quantum **3**, 128 (2019).
[51] H. Bombin and M. A. Martin-Delgado, Optimal resources for topological two-dimensional stabilizer codes: Comparative study, Phys. Rev. A **76**, 012305 (2007).
[52] H. Bombin, Gauge color codes: Optimal transversal gates and gauge fixing in topological stabilizer codes (2015), arXiv:1311.0879 [quant-ph].
[53] M. Vasmer and D. E. Browne, Three-dimensional surface codes: Transversal gates and fault-tolerant architectures, Phys. Rev. A **100**, 012312 (2019).
[54] A. Kubica and M. E. Beverland, Universal transversal gates with color codes: A simplified approach, Phys. Rev. A **91**, 032330 (2015).
[55] S. Bravyi and J. Haah, Magic-state distillation with low overhead, Phys. Rev. A **86**, 052329 (2012).
[56] C. Nayak, S. H. Simon, A. Stern, M. Freedman, and S. Das Sarma, Non-abelian anyons and topological quantum computation, Rev. Mod. Phys. **80**, 1083 (2008).
[57] P. Bonderson, D. J. Clarke, C. Nayak, and K. Shtengel, Implementing arbitrary phase gates with ising anyons, Phys. Rev. Lett. **104**, 180505 (2010).
[58] P. Dulian and A. Sawicki, A random matrix model for random approximate t-designs, IEEE Transactions on Information Theory **70**, 2637 (2024).
[59] O. Parzanchevski and P. Sarnak, Super-golden-gates for PU(2), Advances in Mathematics **327**, 869–901 (2018).
[60] `https://github.com/pdulian/qco`.
[61] A. O. Barut and R. Rączka, *Theory of group representations and applications* (World Scientific Publishing Co Pte Ltd., 1986).
[62] G. Benkart, M. Chakrabarti, T. Halverson, R. Leduc, C. Lee, and J. Stroomer, Tensor product representations of general linear groups and their connections with Brauer algebras, J. Algebra **166**, 529–567 (1994).

Appendix A: Unitary channels and the projective group

The unitary channel $\mathbf{U}$ acting on a Hilbert space $\mathcal{H} \cong \mathbb{C}^d$ is the CPTP map defined via $\mathbf{U}(\rho) = U\rho U^\dagger$, for any quantum state $\rho : \mathcal{H} \to \mathcal{H}$ and some fixed unitary representative U from $\mathrm{U}(d)$. Since two unitaries U, V which differ by a phase $U = Ve^{i\phi}$ define the same unitary channel, the group of all unitary channels $\mathbf{U}(d)$ can be identified with the projective unitary group $\mathrm{PU}(d) = \mathrm{U}(d)/U(1)$, where the canonical projection $\pi : \mathrm{U}(d) \to \mathbf{U}(d)$ is mapping the unitaries to the corresponding unitary channels $U \mapsto \mathbf{U}$.

In practice, one is often interested in the closeness of different unitary channels. Various norms (and induced metrics) can be used to quantify it. A prominent example is the diamond norm $||\cdot||_\diamond$ and the induced metric $d_\diamond(\mathbf{U}, \mathbf{V}) = ||\mathbf{U} - \mathbf{V}||_\diamond$. The diamond metric has a clear operational meaning in terms of the statistical distinguishability of two channels. The relationship between $d_\diamond$ and our metric d (1) is given by $d(\mathbf{U}, \mathbf{V}) \leq d_\diamond(\mathbf{U}, \mathbf{V}) \leq 2\, d(\mathbf{U}, \mathbf{V})$ [36].

Appendix B: Approximate t-designs and ϵ-nets

The balanced polynomials of degree t are homogeneous polynomials with degree t in using matrix elements $u_{i,j}$ and degree t in $\overline{u}_{i,j}$. Notice that such polynomials are well-defined on $\mathbf{U}(d)$ as they are not sensitive to the global phase factors. We denote the space of all such polynomials of degree t by $\mathcal{H}_t$. The space $\mathcal{H}_t$ is spanned by the entries of $U^{t,t} := U^{\otimes t} \otimes \bar{U}^{\otimes t}$ thus in general, each polynomial $f_t(U) \in \mathcal{H}_t$ can be expressed as

$$f_t(U) = \mathrm{Tr}\left(A\left(U^{\otimes t} \otimes \bar{U}^{\otimes t}\right)\right)$$

for some matrix A. Let μ be the normalized Haar measure on $\mathbf{U}(d)$, $\mu(\mathbf{U}(d)) = 1$. The Haar measure provides us with a notion of a uniform density on $\mathbf{U}(d)$.

A t-design is a probability measure ν on $\mathbf{U}(d)$ which yields the same averaging outcome as the Haar measure average for all polynomials $f_t(U) \in \mathcal{H}_t$

$$\int_{\mathbf{U}(d)} d\nu(U) f_t(U) = \int_{\mathbf{U}(d)} d\mu(U) f_t(U). \tag{B1}$$

The case in which the measure ν is supported on a finite number of points $\{\nu_i, U_i\}$ is of utmost practical importance. In such a case, the left-hand side integral of (B1) can be written as a sum

$$\sum_{U_i \in \mathcal{S}} \nu_i f_t(U_i) = \int_{\mathbf{U}(d)} d\mu(U) f_t(U), \tag{B2}$$

where $\mathcal{S}$ denotes a finite set supporting the measure ν.

We are mostly interested in a case of uniform t-designs, i.e., the ones for which all $\nu_i = 1/|\mathcal{S}|$, and denote such a measure as $\nu_{\mathcal{S}}$. Hence, by $\mathcal{S} \subset \mathbf{U}(d)$ being a t-design, we understand that the corresponding uniform discrete probability measure $\nu_{\mathcal{S}}$ is a t-design. Using the t-moment operators, the deviation from ν being a t-design (B1) can be measured as the difference in the operator norm $\delta(\nu, t)$ (see (5) and the formula above). This way, we can consider the cases where the condition (B1) is satisfied only approximately, which leads to the definition of a δ-approximate t-design. We say that ν is a δ-approximate t-design if $\delta(\nu, t) < 1$. In particular, the value $\delta(\nu, t) = 0$ corresponds to (an ideal) t-design.

Appendix C: Optimal spectral gap and Kesten-McKay measure

Below, we discuss the applicability of the optimal value (8) and the related measure in various settings considered in this paper.

For a symmetric (i.e., inverse-closed) gate set $\mathcal{S}$, the t-moment operator (5) is a bounded self-adjoint operator with a well-defined spectrum. Its spectral measure $\sigma_{\mathcal{S},t}$ is compactly supported and hence, determined by its moments $\sigma^{(m)}_{\mathcal{S},t}$. The asymptotic behavior of such moments, i.e., the limit $\lim_{t\to\infty} \sigma^{(m)}_{\mathcal{S},t}$ is determined by the number of length m spellings of identity and was provided in [39] in the case of $\mathcal{S}$ generating a free group. Moreover, it was shown in [39], that in this case there exists a measure $\sigma_{\mathcal{S}}$, such that $\sigma^{(m)}_{\mathcal{S}} = \lim_{t\to\infty} \sigma^{(m)}_{\mathcal{S},t}$, known as the Kesten-McKay or Plancherel measure

$$d\sigma_{\mathcal{S}}(x) = \frac{|\mathcal{S}|\sqrt{\delta^2_{\mathrm{opt}}(\mathcal{S}) - x^2}}{2\pi(1-x^2)} \mathbf{1}_{[-\delta_{\mathrm{opt}(\mathcal{S})}, \delta_{\mathrm{opt}(\mathcal{S})}]} dx, \tag{C1}$$

where $\delta_{\mathrm{opt}}(\mathcal{S})$ is the optimal value (8). This implies that $\sigma_{\mathcal{S},t}$ converge weakly to $\sigma_{\mathcal{S}}$ in the limit $t \to \infty$ (see [58] for details). Furthermore, analogous results can be obtained for any (i.e., not necessarily inverse-closed) finite $\mathcal{S}$, for which $\mathcal{S} \cup \mathcal{S}^{-1}$ generates a free group [58]. However, since in this setting the t-moment operator does not need to be self-adjoint, by the Kesten-McKay measure we understand the spectral measure of $\sqrt{T_{\nu_{\mathcal{S}},t} T^*_{\nu_{\mathcal{S}},t}}$ as $t \to \infty$, or equivalently the measure describing the singular values

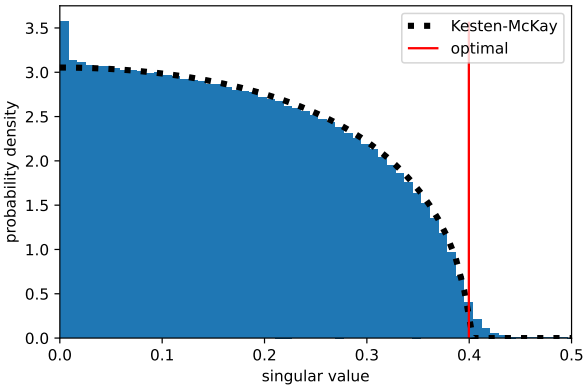


FIG. 7. The probability density of the singular values of the t-moment operator for a derived ensemble of type $\mathcal{C}_{\mu,8}$ with $\approx$ 20 gate sets for $t = 500$. The dotted line denotes the Kesten-McKay measure and the solid line denotes the corresponding optimal value.

of $T_{\nu_{\mathcal{S}},t}$ as $t \to \infty$, given by

$$\frac{|\mathcal{S}|\sqrt{\delta^2_{\mathrm{opt}}(\mathcal{S}) - x^2}}{\pi(1-x^2)} \mathbf{1}_{[0,\, \delta_{\mathrm{opt}(\mathcal{S})}]} dx. \tag{C2}$$

Thus, such a Kesten-McKay measure can be applied in the setting of Haar random gate sets $\mathcal{S}$, since then $\mathcal{S} \cup \mathcal{S}^{-1}$ generates a free group with probability 1.

Crucially, the Kesten-McKay measure can also be applied in the setting of T-QCO (15), when the additional gate T is of infinite order (e.g. Haar random). This follows from the fact that in this case the derived gate set construction (16), which is used to upper bound the T-QCO (17), does not change the number of spellings of identity, compared to the free group case. For a Haar-random gate T of fixed finite order, the number of spellings of identity is increased, which implies that the (even) spectral measure moments are larger than the moments of the Kesten-McKay measure. As a consequence, the support of the Kesten-McKay measure is contained in the support of such a spectral measure and the bound (8) can be applied. However, it was not clear how tight such a bound is with respect to the actual cut-off of the bulk spectrum. To verify it, we checked the distribution of the singular values of t-moments for (derived) ensembles of type $C_{\mu,r}$ with finite r. The resulting distributions are close to the Kesten-McKay distribution, with the support of the latter contained in that of the former quite tightly (see Fig. 7 and Fig. 8). Thus, the optimal value (8) is relevant in all cases considered in this paper.

Appendix D: T-Quantum Circuit Overhead

The useful property of a derived set $\mathcal{S}_T$ (16) is that the T-complexity of a fixed unitary with respect to

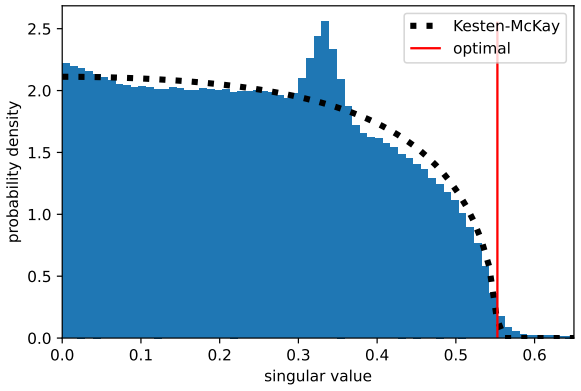


FIG. 8. The probability density of the singular values of the t-moment operator for a derived ensemble of type $\mathcal{H}_{\mu,2}$ with $\approx$ 20 gate sets for $t = 500$. The dotted line denotes the Kesten-McKay measure and the solid line denotes the corresponding optimal value.

$\mathcal{S}_T$ is equal to its complexity (for the same precision). This allows us to lower bound the optimal T-complexity by $\ell_{\text{opt}}(|\mathcal{S}_T|, \epsilon)$. Moreover, for every unitary U constructible using $\mathcal{S}$ with a non-zero T-complexity for precision ϵ, there exists a unitary U_T constructible using $\mathcal{S}_T$ with the same T-complexity for the same precision (and vice-versa). Indeed, each such unitary U can be ϵ-approximated by the reduced word over $\mathcal{S}$ of the form

$$U \approx_\epsilon c_{i_1} w_1 c_{i_2} w_2 \dots c_{i_p} w_p c_{i_{p+1}}, \tag{D1}$$

where each w_j is a word in $T_1, \dots, T_n$, the elements c_{i_j} belong to C and c_{i_1} and $c_{i_{p+1}}$ may be missing. For simplicity, let us assume we have only one costly gate T, the element c_{i_1} is present and $c_{i_{p+1}}$ is missing, so that $w_j = T^{k_j}$ for some integer k_j and the total T-count $\sum_{i=1}^{p} k_i$ is equal to said T-complexity [7]. Choosing the elements of S_T as $g_j \coloneqq d_j T d_j^\dagger$, where $d_j \coloneqq c_{i_1} c_{i_2} \dots c_{i_j}$, we have

$$U \approx_\epsilon g_1^{k_1} g_2^{k_2} \dots g_p^{k_p} d_{p+1} \tag{D2}$$

and $U_T = U d_{p+1}^\dagger$ is ϵ-approximated by the word over S_T of the form $g_1^{k_1} g_2^{k_2} \dots g_p^{k_p}$. It is easy to see that such a form needs to have the lowest possible T-count, so that U and U_T have the same T-complexity. Indeed, otherwise U could be ϵ-approximated by a word with the T-count smaller than that of (D1). Similarly, for other cases and vice versa. Hence, the supremum of T-complexities over all operations U in $\mathbf{U}(d)$ is the same for $\mathcal{S}$ and $\mathcal{S}_T$ and equals $\ell(\mathcal{S}_T, \epsilon)$. Thus, the T-QCO of a finite $\mathcal{S}$ can be bounded as

$$\frac{\ell(\mathcal{S}_T, \epsilon)}{\ell_{\text{opt}}(|\mathcal{S}_T|, \epsilon)} \lesssim Q(\mathcal{S}_T, \epsilon), \tag{D3}$$

where

$$Q(\mathcal{S}_T, \epsilon) = \frac{\log(|C|)}{\log\left(1/\delta(\nu_{\mathcal{S}_T}, t(\epsilon))\right)}, \tag{D4}$$

and $t(\epsilon)$ is the bound stemming from the ϵ-net t-design correspondence of type (9).

Appendix E: Numerical experiments - methods

In order to obtain the value of $Q(\mathcal{S}, \epsilon)$, one needs to computate the norm $\delta(\nu_{\mathcal{S}}, t) = \|T_{\nu_{\mathcal{S}},t} - T_{\mu,t}\|_\infty$ (see (5) and equation above). In a naive approach, one could compute $U^{t,t} = U^{\otimes t} \otimes \bar{U}^{\otimes t}$ for each U in $\mathcal{S}$, but performing such calculation is exponentialy hard in t.
This problem can be avoided by noticing that the mapping $U \mapsto U^{t,t}$ is a representation of the $SU(d)$ group onto $\mathbb{C}^{2dt}$. Every representation of $SU(d)$ can be expressed as a block diagonal matrix, where each block is some irreducible representation (irrep) of $SU(d)$ [61]. In our case, it reads

$$U^{t,t} = \begin{bmatrix} \pi_{\lambda_1}(U) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \pi_{\lambda_2}(U) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \pi_{\lambda_k}(U) \end{bmatrix}, \tag{E1}$$

where π_λ is an irrep with label λ (more on that later). It follows that the t-moment operators are block diagonal as well, and their blocks are given by $T_{\nu,\lambda} = \int_G d\nu(U) \pi_\lambda(U)$. Furthermore, by the orthogonality of irreps [61], the Haar measure blocks $T_{\mu,\lambda}$ are equal to zero for all irreps π_λ, except the trivial one $\pi_0(U) = 1$. In summary, the value of $\delta(\nu_{\mathcal{S}}, t)$ can be computed as

$$\max_\lambda \|T_{\nu_{\mathcal{S}},\lambda} - T_{\mu,\lambda}\|_\infty = \max_{\lambda \neq 0} \|T_{\nu_{\mathcal{S}},\lambda}\|_\infty, \tag{E2}$$

where maximization is performed over all unique irreps appearing in the decomposition of $U^{t,t}$. In the simplest case, $d = 2$, these are all $SU(2)$ representations with integer spin quantum number $s \leq t$. For $d \geq 2$, the irreps are labeled by the $d-1$-dimensional generalizations of a spin number (e.g. the Young tableaus), and thus, more complicated conditions are required [48, 58, 61, 62]. In either case, the dimensions of π_λ are $\mathcal{O}(t^{d(d-1)/2})$ and thus the norms $\|T_{\nu_{\mathcal{S}},\lambda}\|_\infty$ can be computed efficiently.

[7] The general case can be proved analogously.

Chapter 6

Summary and future directions

6.1 Summary

In this thesis, we focused on one of the most fundamental aspects in the quantum computing theory - the efficiency of various discrete universal quantum gate sets. Our strategy focused on the Solovay-Kitaev-like (SKL) theorems based on the spectral gap of averaging operators on compact groups, especially the finite-scale spectral gaps related to unitary t-designs and their relation with ϵ-nets in the projective unitary group. Crucially, we were interested in obtaining such SKL theorems using formulas with explicit or effective constants (i.e. with all the constants known or at least computable in principle). We identified three related aspects of such considerations, each addressed in a separate paper.

The first aspect was the derivation of the poly-logarithmic bound on the spectral gap decay. Such a bound allows one to use the computed value of the gap at a given scale t_0 to bound the gap for $t \geq t_0$. As a consequence, it can be used in conjunction with the SKL theorems based on the finite-scale spectral gap to obtain an SKL theorem with explicit ϵ-dependence at the cost of the worse constant c in $\log^c(1/\epsilon)$. This aspect was addressed in Paper I, where we provided a simple proof for the explicit and essentially calculable poly-logarithmic lower bounds on the finite-scale spectral gap decay for gate sets satisfying a specific condition (satisfied by generic gate sets). This result was supplemented by the numerical simulations for a single qubit. Additionally, we formulated an alternative proof for the upper bound on the efficiency of finite gate sets with the spectral gap.

The second aspect was the derivation of the SKL theorems based on the finite-scale spectral gaps. This aspect was addressed in Paper II, where we introduced and characterized

a new type of polynomial approximate identity on the projective unitary group via trimming the fundamental solution to the natural and well-known object - a heat kernel. We then proved that such a natural approximate identity can be used to improve the known correspondence between unitary δ-approximate t-designs and ϵ-nets in the space of unitary quantum channels. Namely, we achieved better scaling of δ while maintaining the scaling of t essentially unchanged. We then explained how such a correspondence can be used in areas such as the inverse-free SKL theorems, quantum complexity, black hole physics, and Quantum Circuit Overhead. We also suggested the use of the trimmed heat kernels in the derivation of poly-logarithmic bounds on the spectral gap decay.

The third and final aspect was to find a proper way to compare the efficiency of various gate sets. Although the SKL theorems based on the finite-scale spectral gap seem to be a reasonable way to bound the efficiency of various gate sets, their dependence on the number of elementary gates used makes the comparison more complicated. This aspect was addressed in Paper III, where we introduced a notion of Quantum Circuit Overhead (QCO) and a related notion of T-Quantum Circuit Overhead (T-QCO). We demonstrated that the (T-)QCO can be upper bounded via a quantity Q/Q_T given by a simple formula involving the spectral gap at a scale $t(\epsilon)$ stemming from the δ-approximate t-designs and ϵ-nets correspondence. We discussed the applicability of both overheads as reasonable proxies for the overall cost-effectiveness of various gate sets in different quantum computing architectures, including NISQ and fault-tolerant. To demonstrate that (T-)QCO can be calculated in practice, we performed extensive numerical simulations for various ensembles of universal finite gate-sets, including Haar-random ones and the random completions of single-qubit Clifford and Hurwitz groups. The most interesting conclusion from such experiments is that, regarding the upper bound Q_T on the T-QCO, the famous T gate (also known as $P(\pi/4)$ gate) is a fairly non-optimal choice for completing the Clifford group gate set. We also found the optimal completions for the Clifford and Hurwitz groups, in terms of Q/Q_T bounds. Interestingly, such bounds are close to optimal for the Super Golden Gates in the case of the Hurwitz group but not for the Clifford group. Our analysis shows that the optimal completions for the Clifford group are given by the gates of the form $UP(3\pi/4)U^\dagger$, where U is a Bloch sphere rotation around any axis $(x, y, 0)$ with $|x| \neq |y|$ by an angle in $[\pi/8, \pi/2]$.

6.2 Future directions

Below, we provide a list of future research problems, either a simple continuation of our work or a (possibly challenging) open problem.

1. (Paper II) Can our refined unitary δ-approximate t-designs and ϵ-nets correspondence be used to improve the existing results concerning the saturation and recurrence of the complexity of black holes? If yes, then to what extent?

2. (Paper II) Are our scaling of t and δ optimal? If not, then what is the optimal scaling? (see also II b. and II c. from Section 2.4.5)

3. (Paper III) Perform the numerical experiments for (T-)QCO in slightly higher dimensions, e.g. for the entangling gates (see also III c. from Section 2.4.5).

4. (Paper III) Assess the tightness of our Q/Q_T formula by finding the true value of (T-)QCO numerically for specific gate sets. What is the correlation between the true value and Q/Q_T?

5. (Paper III) Can a representative from the family of optimal (in Q_T) completions of the one-qubit Clifford group be realized fault-tolerantly with a cost comparable to that of the $P(\pi/4)$ gate?

6. (Paper I and II) Derive the explicit poly-logarithmic bounds on the finite-scale spectral gap decay with the logarithm exponent at least as good as Varju's. Can the trimmed heat kernel construction or the Fejér kernel be useful to obtain such bounds? (see also I b. and I c. from Section 2.4.5)

7. Prove that any universal discrete gate set is asymptotically optimally efficient, i.e. $\ell = \Theta(\log(1/\epsilon))$. For example, one may prove that each such gate set has a spectral gap.

8. Assuming that a universal gate set is asymptotically optimally efficient $\ell = \Theta(\log(1/\epsilon))$, can we find the explicit formula?

In terms of the future outlook, the ultimate goal for the unitary ϵ-nets and t-designs correspondence would be to prove the best possible scaling of t and δ, e.g., in the best-case scenario, the scaling of t equal to the lower bound $\simeq d^2/\epsilon$. In terms of scaling δ, any reasonable upper bound would represent significant progress.

In the case of the optimality of quantum gates, the (close to) ultimate goal would be to prove what is already widely accepted in the community. However, the proof is still lacking, namely that every universal discrete gate set has a spectral gap. Ideally, one should provide an explicit bound on the complexity, i.e. the bound of the form $\ell = \Theta(\log(1/\epsilon))$ with computable constants, depending on $\mathcal{S}$. A relaxed version of this problem would be to prove the poly-logarithmic spectral gap decay, say at least as good as Varjú's (but desirably better due to recent improvements in the SKL-like theorems breaking the cubic barrier [33]) with explicit constants c and especially r_0. Preferably, one would like to prove that r_0 can be taken to be as small as possible, e.g. just small enough to test the universality of the gate set, i.e. in the language of t-designs, $t = 6$ for $d = 2$ and $t = 4$ for $d \geq 3$ [52].

List of Publications

A complete list of the author's publications. Publications 5, 6, and 7 fall within the popular science category.

Publications from the thesis

1. O. Słowik, P. Dulian, A. Sawicki, "Quantum Circuit Overhead", *arXiv:2505.00683 (2025).*

2. O. Słowik, O. Reardon-Smith, A. Sawicki, "Fundamental solutions of the heat equation on unitary groups establish an improved relation between ϵ-nets and approximate unitary t-designs", *J. Phys. A: Math. Theor. 58 445301 (2025).*

3. O. Słowik, A. Sawicki, "Calculable lower bounds on the efficiency of universal sets of quantum gates", *J. Phys. A: Math. Theor. 56 115304 (2023).*

Other publications

4. J. Tuziemski, F. B. Maciejewski, J. Majsak, O. Słowik, M. Kotowski, K. Kowalczyk-Murynka, P. Podziemski, M. Oszmaniec, "Efficient reconstruction, benchmarking and validation of cross-talk models in readout noise in near-term quantum devices", *arXiv:2311.10661 (2023).*

5. G. H. Kasprowicz, M. Sowiński, K. Poźniak, J. Szmidt, T. Przywózki, P. Kulik, M. Kuś, Z. M. Wawrzyniak, P. Szczepański, O. Słowik, M. Życzkowski, P. Marć, A. Pakuła, A. Nawrat, K. Daniec, K. Wereszczyński, T. Kuczerski, R. Kawałek, M. Sadowski, W. Suleja, P. Witkowski, "Electronic control system for the quantum computer infrastructure in the MIKOK project", *Elektronika - konstrukcje, technologie, zastosowania Vol. 64, nr 8, p. 72–75 (2023).*

6. O. Słowik, G. Kasprowicz, "The quantum computing stack: from algorithms to ions", *Elektronika - konstrukcje, technologie, zastosowania Vol. 64, nr 8, p. 25–29 (2023).*

7. P. Gawron, K. Kara, M. Cholewa, O. Słowik, K. Hendzel, M. Stefaniak, P. Biskupski, "Rewolucja stanu - fantastyczne wprowadzenie do informatyki kwantowej.", *wydanie 2 poszerzone, quantumz.io, Warszawa (2021).*

8. O. Słowik, A. Sawicki, T. Maciążek, "Designing locally maximally entangled quantum states with arbitrary local symmetries", *Quantum 5, 450 (2021).*

9. J. Chojnacki, J. Krajecka, J. H. Kwapisz, O. Słowik, A. Strąg, "Is asymptotically safe inflation eternal?", *JCAP04(2021)076.*

10. O. Słowik, M. Hebenstreit, B. Kraus, A. Sawicki, "A link between symmetries of critical states and the structure of SLOCC classes in multipartite systems", *Quantum 4, 300 (2020).*

11. O. Słowik, K. Orłowska, D. Kopiec, P. Janus, P. Grabiec, T. Gotszalk, "Quantum mechanical aspects in the MEMS/NEMS technology", *Measurement Automation Monitoring, Mar. 2016, no. 03, vol. 62.*

Bibliography

[1] P. Benioff, "The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines," *Journal of Statistical Physics*, vol. 22, pp. 563–591, May 1980.

[2] Y. I. Manin, *Vychislimoe i nevychislimoe (in Russian); eng. Computable and Non-computable*. Soviet Radio, pp. 13-15, 1980.

[3] R. P. Feynman, "Simulating physics with computers," *Int. J. Theor. Phys*, vol. 21, no. 6/7, 1982.

[4] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

[5] D. Deutsch, "Quantum theory, the Church-Turing principle and the universal quantum computer," *Proceedings of the Royal Society of London Series A*, vol. 400, pp. 97–117, July 1985.

[6] D. Simon, "On the power of quantum computation," in *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pp. 116–123, 1994.

[7] P. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," in *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pp. 124–134, 1994.

[8] L. K. Grover, "A fast quantum mechanical algorithm for database search," in *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '96, (New York, NY, USA), p. 212–219, Association for Computing Machinery, 1996.

[9] A. W. Harrow, A. Hassidim, and S. Lloyd, "Quantum algorithm for linear systems of equations," *Phys. Rev. Lett.*, vol. 103, p. 150502, Oct 2009.

[10] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, "A variational eigenvalue solver on a photonic quantum processor," *Nature Communications*, vol. 5, p. 4213, Jul 2014.

[11] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," 2014.

[12] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi, *Classical and Quantum Computation*. USA: American Mathematical Society, 2002.

[13] V. Gheorghiu, J. Huang, S. M. Li, M. Mosca, and P. Mukhopadhyay, "Reducing the CNOT count for Clifford+T circuits on NISQ architectures," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, p. 1873–1884, June 2023.

[14] J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, Aug. 2018.

[15] K. Noh, L. Jiang, and B. Fefferman, "Efficient classical simulation of noisy random quantum circuits in one dimension," *Quantum*, vol. 4, p. 318, Sept. 2020.

[16] Y. Ge, W. Wenjie, C. Yuheng, P. Kaisen, L. Xudong, Z. Zixiang, W. Yuhan, W. Ruocheng, and Y. Junchi, "Quantum circuit synthesis and compilation optimization: Overview and prospects," 2024.

[17] T. Häner, D. S. Steiger, K. Svore, and M. Troyer, "A software methodology for compiling quantum programs," *Quantum Science and Technology*, vol. 3, p. 020501, Feb. 2018.

[18] M. E. Beverland, A. Kubica, and K. M. Svore, "Cost of universality: A comparative study of the overhead of state distillation and code switching with color codes," *PRX Quantum*, vol. 2, p. 020341, Jun 2021.

[19] V. Gheorghiu, M. Mosca, and P. Mukhopadhyay, "T-count and T-depth of any multi-qubit unitary," *npj Quantum Information*, vol. 8, Nov. 2022.

[20] F. J. R. Ruiz, T. Laakkonen, J. Bausch, M. Balog, M. Barekatain, F. J. H. Heras, A. Novikov, N. Fitzpatrick, B. Romera-Paredes, J. van de Wetering, A. Fawzi, K. Meichanetzidis, and P. Kohli, "Quantum circuit optimization with AlphaTensor," *Nature Machine Intelligence*, vol. 7, pp. 374–385, Mar 2025.

[21] D. Gosset, V. Kliuchnikov, M. Mosca, and V. Russo, "An algorithm for the T-count," *Quantum Info. Comput.*, vol. 14, p. 1261–1276, Nov. 2014.

[22] V. Vandaele, "Lower T-count with faster algorithms," 2024.

[23] H. Zhou, C. Zhao, M. Cain, D. Bluvstein, C. Duckering, H.-Y. Hu, S.-T. Wang, A. Kubica, and M. D. Lukin, "Algorithmic fault tolerance for fast quantum computing," 2024.

[24] L. Heyfron and E. T. Campbell, "An efficient quantum compiler that reduces T count," 2018.

[25] S. Lloyd, *Programming the Universe: A Quantum Computer Scientist Takes On the Cosmos*. Knopf Publishing Group, 2006.

[26] A. R. Brown and L. Susskind, "Second law of quantum complexity," *Phys. Rev. D*, vol. 97, p. 086015, Apr 2018.

[27] T. Tokusumi, A. Matsumura, and Y. Nambu, "Quantum circuit model of black hole evaporation," *Classical and Quantum Gravity*, vol. 35, p. 235013, Nov. 2018.

[28] M. P. Fisher, V. Khemani, A. Nahum, and S. Vijay, "Random quantum circuits," *Annual Review of Condensed Matter Physics*, vol. 14, p. 335–379, Mar. 2023.

[29] P. W. Claeys, M. Henry, J. Vicary, and A. Lamacraft, "Exact dynamics in dual-unitary quantum circuits with projective measurements," *Physical Review Research*, vol. 4, Dec. 2022.

[30] P. Hayden and J. Preskill, "Black holes as mirrors: quantum information in random subsystems," *Journal of High Energy Physics*, vol. 2007, p. 120–120, Sept. 2007.

[31] M. Oszmaniec, M. Kotowski, M. Horodecki, and N. Hunter-Jones, "Saturation and recurrence of quantum complexity in random local quantum dynamics," *Phys. Rev. X*, vol. 14, p. 041068, Dec 2024.

[32] C. M. Dawson and M. A. Nielsen, "The Solovay-Kitaev algorithm," 2005.

[33] G. Kuperberg, "Breaking the cubic barrier in the Solovay-Kitaev algorithm," 2023.

[34] A. Bocharov, Y. Gurevich, and K. M. Svore, "Efficient decomposition of single-qubit gates into V basis circuits," *Physical Review A*, vol. 88, 2013.

[35] N. J. Ross, "Optimal ancilla-free Clifford+V approximation of z-rotations," 2015.

[36] V. Kliuchnikov, A. Bocharov, M. Roetteler, and J. Yard, "A framework for approximating qubit unitaries," *arXiv:1510.03888*, 2015.

[37] V. Kliuchnikov, D. Maslov, and M. Mosca, "Practical approximation of single-qubit unitaries by single-qubit quantum Clifford and T circuits," *IEEE Transactions on Computers*, vol. 65, pp. 161–172, 2016.

[38] D. A. Kazhdan, "Connection of the dual space of a group with the structure of its close subgroups," *Functional Analysis and Its Applications*, vol. 1, pp. 63–65, Jan 1967.

[39] D. Dolgopyat, "On mixing properties of compact group extensions of hyperbolic systems," *Israel Journal of Mathematics*, vol. 130, pp. 157–205, Dec 2002.

[40] P. P. Varjú, "Random walks in compact groups," 2015.

[41] M. Oszmaniec, A. Sawicki, and M. Horodecki, "Epsilon-nets, unitary designs, and random quantum circuits," *IEEE Transactions on Information Theory*, vol. 68, no. 2, pp. 989–1015, 2022.

[42] S. J. Szarek, "Metric entropy of homogeneous spaces," *Banach Center Publications*, vol. 43, 1998.

[43] W. Fulton and J. Harris, *Representation Theory: A First Course*. Graduate Texts in Mathematics, volume 129, New York: Springer-Verlag, 1991.

[44] B. Hall, *Lie Groups, Lie Algebras and Representations: Elementary Introduction*. Graduate Texts in Mathematics, volume 222, Springer, 2015.

[45] A. Kirillov, *An Introduction to Lie Groups and Lie Algebras*. Cambridge studies in advanced mathematics, volume 113, Cambridge University Press, 2008.

[46] T. Brocker and T. tom Dieck, *Representations of Compact Lie Groups*. Graduate Texts in Mathematics, Springer-Verlag, Corr. 2nd print ed., 1985.

[47] G. B. Folland, *A Course in Abstract Harmonic Analysis (2nd ed.)*. New York: Chapman and Hall/CRC, 2015.

[48] P. Dulian and A. Sawicki, "A random matrix model for random approximate t-designs," *IEEE Transactions on Information Theory*, vol. 70, no. 4, pp. 2637–2654, 2024.

[49] J. Faraut, *Analysis on Lie Groups: An Introduction*. Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2008.

[50] M. Sugiura, "Fourier series of smooth functions on compact Lie groups," *Osaka Journal of Mathematics*, vol. 8, no. 1, pp. 33 – 47, 1971.

[51] W. Jaworski and C. R. E. Raja, "The choquet-deny theorem and distal properties of totally disconnected locally compact groups of polynomial growth," 2007.

[52] A. Sawicki, L. Mattioli, and Z. Zimborás, "Universality verification for a set of quantum gates," *Physical Review A*, vol. 105, May 2022.

[53] B. Bekka, P. de la Harpe, and A. Valette, *Kazhdan's Property (T)*. New Mathematical Monographs, Cambridge University Press, 2008.

[54] O. Parzanchevski and P. Sarnak, "Super-golden-gates for PU(2)," *Advances in Mathematics*, vol. 327, p. 869–901, Mar. 2018.

[55] G. Benkart, M. Chakrabarti, T. Halverson, R. Leduc, C. Lee, and J. Stroomer, "Tensor product representations of general linear groups and their connections with Brauer algebras," *Journal of Algebra*, vol. 166, pp. 529–567, June 1994.

[56] L. Grafakos, *Classical Fourier Analysis*. Graduate Texts in Mathematics, volume 249, Springer, 2014.

[57] E. B. Davies, *Heat Kernels and Spectral Theory*. Cambridge Tracts in Mathematics, Cambridge University Press, 1989.

[58] H. Urakawa, "The heat equation on a compact Lie group," *Osaka Journal of Mathematics*, vol. 2, pp. 285–297, 1974.

[59] D. G. Maher, "Wrapping Brownian motion and heat kernels I: compact Lie groups," 2010.

[60] G. Alexopoulos and N. Lohoué, "Riesz means on Lie groups and riemannian manifolds of nonnegative curvature," *Bulletin de la Société Mathématique de France*, vol. 122, no. 2, pp. 209–223, 1994.

[61] P. P. Varjú, "Random walks in compact groups," *Documenta Mathematica*, vol. 18, pp. 1137–1175, 2013.

[62] J. Watrous, *The Theory of Quantum Information*. Cambridge University Press, 2018.

[63] S. Lloyd, "Universal quantum simulators," *Science*, vol. 273, no. 5278, pp. 1073–1078, 1996.

[64] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, and S. Zhu, "Theory of trotter error with commutator scaling," *Physical Review X*, vol. 11, Feb. 2021.

[65] L. Piroli, B. Bertini, J. I. Cirac, and T. Prosen, "Exact dynamics in dual-unitary quantum circuits," *Physical Review B*, vol. 101, Mar. 2020.

[66] A. Nahum, J. Ruhman, S. Vijay, and J. Haah, "Quantum entanglement growth under random unitary dynamics," *Physical Review X*, vol. 7, July 2017.

[67] L. Susskind, "Computational complexity and black hole horizons," 2014.

[68] M. Oszmaniec, M. Kotowski, M. Horodecki, and N. Hunter-Jones, "Saturation and recurrence of quantum complexity in random local quantum dynamics," *Phys. Rev. X*, vol. 14, p. 041068, Dec 2024.

[69] D. Janzing, P. Wocjan, and T. Beth, "Identity check is QMA-complete," 2003.

[70] A. Sawicki and K. Karnas, "Criteria for universality of quantum gates," *Physical Review A*, vol. 95, June 2017.

[71] A. Sawicki and K. Karnas, "Universality of single-qudit gates," *Annales Henri Poincaré*, vol. 18, pp. 3515–3552, Nov 2017.

[72] M. Kuranishi, "On everywhere dense imbedding of free groups in Lie groups," *Nagoya Mathematical Journal*, vol. 2, pp. 63–71, Jan 1951.

[73] T. Häner, D. S. Steiger, K. Svore, and M. Troyer, "A software methodology for compiling quantum programs," *Quantum Science and Technology*, vol. 3, p. 020501, Feb. 2018.

[74] B. Eastin and E. Knill, "Restrictions on transversal encoded quantum gate sets," *Phys. Rev. Lett.*, vol. 102, p. 110502, Mar 2009.

[75] M. P. Woods and Á. M. Alhambra, "Continuous groups of transversal gates for quantum error correcting codes from finite clock reference frames," *Quantum*, vol. 4, p. 245, Mar. 2020.

[76] P. Faist, S. Nezami, V. V. Albert, G. Salton, F. Pastawski, P. Hayden, and J. Preskill, "Continuous symmetries and approximate quantum error correction," *Physical Review X*, vol. 10, Oct. 2020.

[77] D. Gottesman, "Quantum error correction and fault-tolerance," 2005.

[78] Y.-H. Luo, M.-C. Chen, M. Erhard, H.-S. Zhong, D. Wu, H.-Y. Tang, Q. Zhao, X.-L. Wang, K. Fujii, L. Li, N.-L. Liu, K. Nemoto, W. J. Munro, C.-Y. Lu, A. Zeilinger, and J.-W. Pan, "Quantum teleportation of physical qubits into logical code spaces,"

Proceedings of the National Academy of Sciences, vol. 118, no. 36, p. e2026250118, 2021.

[79] B. Eastin and E. Knill, "Restrictions on transversal encoded quantum gate sets," *Physical Review Letters*, vol. 102, Mar. 2009.

[80] C. Dankert, R. Cleve, J. Emerson, and E. Livine, "Exact and approximate unitary 2-designs and their application to fidelity estimation," *Phys. Rev. A*, vol. 80, p. 012304, Jul 2009.

[81] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter, "The mother of all protocols: restructuring quantum information's family tree," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 465, p. 2537–2563, June 2009.

[82] J. J. Wallman and S. T. Flammia, "Randomized benchmarking with confidence," *New Journal of Physics*, vol. 16, p. 103032, Oct. 2014.

[83] J. M. Epstein, A. W. Cross, E. Magesan, and J. M. Gambetta, "Investigating the limits of randomized benchmarking protocols," *Physical Review A*, vol. 89, June 2014.

[84] A. J. Scott, "Optimizing quantum process tomography with unitary 2-designs," *Journal of Physics A: Mathematical and Theoretical*, vol. 41, p. 055308, Jan. 2008.

[85] J. Helsen and M. Walter, "Thrifty shadow estimation: Reusing quantum circuits and bounding tails," *Physical Review Letters*, vol. 131, Dec. 2023.

[86] D. Gross, F. Krahmer, and R. Kueng, "A partial derandomization of phaselift using spherical designs," *Journal of Fourier Analysis and Applications*, vol. 21, p. 229–266, Oct. 2014.

[87] O. Szehr, F. Dupuis, M. Tomamichel, and R. Renner, "Decoupling with unitary approximate two-designs," *New Journal of Physics*, vol. 15, p. 053022, May 2013.

[88] J. Bae, B. C. Hiesmayr, and D. McNulty, "Linking entanglement detection and state tomography via quantum 2-designs," *New Journal of Physics*, vol. 21, p. 013012, Jan. 2019.

[89] P. Sen, "Random measurement bases, quantum state distinction and applications to the hidden subgroup problem," 2005.

[90] J. Czartowski, D. Goyeneche, M. Grassl, and K. Życzkowski, "Isoentangled mutually unbiased bases, symmetric quantum measurements, and mixed-state designs," *Phys. Rev. Lett.*, vol. 124, p. 090503, Mar 2020.

[91] A. Roy and A. J. Scott, "Unitary designs and codes," *Designs, Codes and Cryptography*, vol. 53, p. 13–31, Apr. 2009.

[92] H.-Y. Huang, R. Kueng, and J. Preskill, "Predicting many properties of a quantum system from very few measurements," *Nature Physics*, vol. 16, p. 1050–1057, June 2020.

[93] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki, "Local random quantum circuits are approximate polynomial-designs," *Communications in Mathematical Physics*, vol. 346, p. 397–434, Aug. 2016.

[94] L. Masanes, A. J. Roncaglia, and A. Acín, "Complexity of energy eigenstates as a mechanism for equilibration," *Physical Review E*, vol. 87, Mar. 2013.

[95] M. Oszmaniec, R. Augusiak, C. Gogolin, J. Kołodyński, A. Acín, and M. Lewenstein, "Random bosonic states for robust quantum metrology," *Phys. Rev. X*, vol. 6, p. 041044, Dec 2016.

[96] D. A. Roberts and B. Yoshida, "Chaos and complexity by design," *Journal of High Energy Physics*, vol. 2017, Apr. 2017.

[97] Y. Nakata, C. Hirche, M. Koashi, and A. Winter, "Efficient quantum pseudorandomness with nearly time-independent Hamiltonian dynamics," *Physical Review X*, vol. 7, Apr. 2017.

[98] F. G. Brandão, W. Chemissany, N. Hunter-Jones, R. Kueng, and J. Preskill, "Models of quantum complexity growth," *PRX Quantum*, vol. 2, July 2021.

[99] L. Susskind, "Three lectures on complexity and black holes," 2018.

[100] J. Haferkamp, P. Faist, N. B. T. Kothakonda, J. Eisert, and N. Yunger Halpern, "Linear growth of quantum circuit complexity," *Nature Physics*, vol. 18, p. 528–532, Mar. 2022.

[101] D. Hangleiter, J. Bermejo-Vega, M. Schwarz, and J. Eisert, "Anticoncentration theorems for schemes showing a quantum speedup," *Quantum*, vol. 2, p. 65, May 2018.

[102] M. Yoganathan, R. Jozsa, and S. Strelchuk, "Quantum advantage of unitary clifford circuits with magic state inputs," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 475, p. 20180427, May 2019.

[103] A. W. Harrow and A. Montanaro, "Quantum computational supremacy," *Nature*, vol. 549, p. 203–209, Sept. 2017.

[104] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, "Characterizing quantum supremacy in near-term devices," *Nature Physics*, vol. 14, p. 595–600, Apr. 2018.

[105] F. Arute et al., "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, p. 505–510, Oct. 2019.

[106] D. Gross, K. Audenaert, and J. Eisert, "Evenly distributed unitaries: On the structure of unitary designs," *Journal of Mathematical Physics*, vol. 48, May 2007.

[107] A. Kaposi, Z. Kolarovszki, A. Solymos, and Z. Zimborás, "Generalized group designs: overcoming the 4-design-barrier and constructing novel unitary 2-designs in arbitrary dimensions," 2024.

[108] J. Bourgain and A. Gamburd, "On the spectral gap for finitely-generated subgroups of SU(2)," *Inventiones mathematicae*, vol. 171, pp. 83–121, 2007.

[109] J. Bourgain and A. Gamburd, "A spectral gap theorem in SU(d)," 2011.

[110] Y. Benoist and N. de Saxcé, "A spectral gap theorem in simple Lie groups," 2014.

[111] A. W. Harrow, B. Recht, and I. L. Chuang, "Efficient discrete approximations of quantum gates," *Journal of Mathematical Physics*, vol. 43, p. 4445–4451, Sept. 2002.

[112] A. Lubotzky, R. Phillips, and P. Sarnak, "Hecke operators and distributing points on the sphere I," *Communications on Pure and Applied Mathematics. Supplement: Proceedings of the Symposium on Frontiers of the Mathematical Sciences: 1985.*, vol. 39, pp. S149–S186, 1986.

[113] A. Lubotzky, R. Phillips, and P. Sarnak, "Hecke operators and distributing points on S^2. II," *Communications on Pure and Applied Mathematics*, vol. 40, pp. 401–420, 1987.

[114] P. Sarnak, "Letter to Scott Aaronson and Andy Pollington on the Solovay-Kitaev theorem," 2015.

[115] P. Selinger, "Efficient Clifford+T approximation of single-qubit operators," *Quantum Information and Computation*, vol. 15, pp. 159–180, 2015.

[116] V. Kliuchnikov, D. Maslov, and M. Mosca, "Fast and efficient exact synthesis of single qubit unitaries generated by clifford and t gates," 2013.

[117] V. Kliuchnikov, A. Bocharov, M. Roetteler, and J. Yard, "A framework for approximating qubit unitaries," *arXiv:1510.03888*, 2015.

[118] R. Dalal, S. Evra, and O. Parzanchevski, "Multi-qubit golden gates," 2025.

[119] A. Gamburd, D. Jakobson, and P. Sarnak, "Spectra of elements in the group ring of SU(2)," *Journal of the European Mathematical Society*, 1999.

[120] A.Bouland and T. Giurgica-Tiron, "Efficient universal quantum compilation: An inverse-free Solovay-Kitaev algorithm," 2021.

[121] D. Belkin, J. Allen, S. Ghosh, C. Kang, S. Lin, J. Sud, F. T. Chong, B. Fefferman, and B. K. Clark, "Approximate t-designs in generic circuit architectures," *PRX Quantum*, vol. 5, p. 040344, Dec 2024.

[122] M. B. Hastings and A. W. Harrow, "Classical and quantum tensor product expanders," 2008.

[123] A. W. Harrow and R. A. Low, *Efficient Quantum Tensor Product Expanders and k-Designs*, p. 548–561. Springer Berlin Heidelberg, 2009.

[124] P. Dulian and A. Sawicki, "Matrix concentration inequalities and efficiency of random universal sets of quantum gates," *Quantum*, vol. 7, p. 983, Apr. 2023.

[125] H. Kesten, "Symmetric random walks on groups," *Transactions of the American Mathematical Society*, vol. 92, pp. 336–354, 1959.

[126] K. Życzkowski, K. A. Penson, I. Nechita, and B. Collins, "Generating random density matrices," *Journal of Mathematical Physics*, vol. 52, June 2011.