Explicit orthogonal and unitary designs

Ryan O'Donnell*Ⓕ Rocco A. Servedio†Ⓕ Pedro Paredes‡Ⓕ

ⒻAuthor ordering randomized.

June 6, 2023

Abstract

We give a strongly explicit construction of ϵ-approximate k-designs for the orthogonal group $\mathrm{O}(N)$ and the unitary group $\mathrm{U}(N)$, for $N = 2^n$. Our designs are of cardinality $\mathrm{poly}(N^k/\epsilon)$ (equivalently, they have seed length $O(nk + \log(1/\epsilon))$); up to the polynomial, this matches the number of design elements used by the construction consisting of completely random matrices.

1 Introduction

The main new result in our work is the following:

Theorem 1.1. *Let $N = 2^n$ and let $\mathrm{G}(n)$ denote either the orthogonal group $\mathrm{O}(N)$ or the unitary group $\mathrm{U}(N)$. Then for any $k = k(n)$, there is an explicit ϵ-approximate k-design for $\mathrm{G}(n)$ of cardinality $\mathrm{poly}(N^k/\epsilon)$; i.e., samplable using a seed of just $O(nk + \log(1/\epsilon))$ truly random bits. Moreover, these designs are strongly explicit in the following sense: (i) each output matrix is given by an n-qubit circuit consisting of $S = \mathrm{poly}(nk)\log(1/\epsilon)$ gates, each gate being either* CNOT *or one of a few fixed and explicitly specified 1-qubit gates; (ii) the algorithm that takes as input a seed and outputs the associated circuit runs in deterministic $\mathrm{poly}(S)$ time.*

In the unitary case, similar results in the literature only discuss the regime $k \leq \mathrm{poly}(n)$ [HL09, Sen18], or have polynomially worse seed length [BHH16, Haf22]. In contrast, our result holds for all k (even exponentially large as a function of n, or larger), and achieves a seed length which matches, up to constant factors, that of a random construction. The original motivation for our work was the orthogonal case, where the only prior works we know of are [KM15, HHJ21], which we discuss below. Our Theorem 1.1 provides the efficient orthogonal designs needed for Kothari and Meka's near-optimal pseudorandom generators for spherical caps [KM15].

Let us now discuss the general context for our result.

Derandomization. Let $\mathcal{G}$ be a class of objects, and assume informally that each object has "size" $N^{\Theta(1)}$ (think, e.g., of strings of length N, or $N \times N$ matrices). To choose an object from the uniform probability distribution on $\mathcal{G}$ typically requires using $\Omega(N)$ truly random bits. A broad goal in derandomization is to identify a useful notion of "pseudorandomness" for probability distributions on $\mathcal{G}$, and then to show that one can sample from such a distribution using just $r \ll N$ truly random bits.[1] An additional goal is for the sampling algorithm to be *efficient*; i.e., the sampled object should be produced by a deterministic $\mathrm{poly}(r)$-time algorithm, given the truly random seed of length r. In this case, since the sampler has

*Computer Science Department, Carnegie Mellon University.
†Department of Computer Science, Columbia University.
‡Department of Computer Science, Princeton University.

[1] In this introduction, we refer to sampling uniformly from a set of size R as "using $\log_2 R$ truly random bits".

only 2^r possible outcomes yet the total number of objects is exponential in N, it must be the case that the sampler represents the output objects in a "succinct" way. Informally, if it is possible to efficiently compute with objects represented in this succinct way, the sampler is said to be "strongly explicit".

Exact k-wise independence. One of the most common and useful notions of pseudorandomness is that of *bounded independence*. For random objects with $N^{\Theta(1)}$ "entries" ("coordinates"/"dimensions"), it often suffices for applications if the objects are merely "k-wise independent" for some $k \ll N$. This means that the object looks truly random whenever only k entries are inspected. In this case one may hope that the object can be sampled using a random seed of length just $O(k \log N)$ bits.

The paradigmatic example of this comes from k-wise independent length-N Boolean strings. Using results from coding theory [ABI86], it has long been known that $O(k \log N)$ random bits suffice to efficiently sample a precisely k-wise independent string $\boldsymbol{x} \in \{0,1\}^N$ (meaning that $(\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_k})$ is perfectly uniformly distributed on $\{0,1\}^k$ for any $i_1, \ldots, i_k$).

For other kinds of random objects, obtaining *exact* k-wise independence seems extremely difficult. Take the case of random permutations, where $\boldsymbol{\pi} \in S_N$ is said to be k-wise independent if $(\boldsymbol{\pi}(i_1), \ldots, \boldsymbol{\pi}(i_k))$ is uniformly distributed on $\binom{[N]}{k}$ for any distinct $i_1, \ldots, i_k$. While simple efficient methods for generating 2- and 3-wise independent permutations using $O(\log N)$ random bits are known, for any constant $k \geq 4$ the best known efficient construction uses $\Theta(N)$ random bits [FPY15]. The situation is similar for random unitary matrices, where $\boldsymbol{U} \in \mathrm{U}(N)$ is said to be drawn from a k-design if $\mathbf{E}[\boldsymbol{U}_{i_1 j_1} \cdots \boldsymbol{U}_{i_k j_k}]$ is equal to what it would be if $\boldsymbol{U}$ were Haar-distributed on $\mathrm{U}(N)$ (and similarly if any subset of the entries $\boldsymbol{U}_{i_t j_t}$ in the product were replaced with their complex conjugates). Here it is known how to efficiently construct exact 2-designs using $O(\log N)$ bits [DLT02], and exact 3-designs using $O(\log^2 N)$ bits [Web16], but good constructions of exact k-designs for $k \geq 4$ are lacking (see, e.g., [BNZZ19]).

Approximate k-wise independence. Given these issues, it is natural to seek ϵ-*approximate* k-wise independence (k-designs). Here it is important to carefully define the precise notion of "approximate", as different natural notions are often only equivalent if one is willing to change ϵ by a factor that is exponential in k. For example, in the context of Boolean strings in $\{\pm 1\}^N$, a weak notion of (ϵ, k)-wise independence is that $|\mathbf{E}[\boldsymbol{x}_{i_1} \cdots \boldsymbol{x}_{i_k}]| \leq \epsilon$ for all k-tuples of distinct values $i_1, \ldots, i_k$. Naor and Naor [NN93] showed that $O(\log(nk/\epsilon))$ random bits suffice to explicitly generate such a distribution, where we write $n = \log_2 N$. However, to get the stronger guarantee that every k bit positions are ϵ-close to the uniform distribution in statistical distance, one needs $(\epsilon 2^{-k}, k)$-wise independence (see, e.g., [AGHP92]), and hence the number of random bits used in known constructions is $O(k + \log(n/\epsilon))$. In general, for q-ary rather than 2-ary (Boolean) strings, the seed-length penalty becomes $O(k \log q)$. So if, e.g., one wants a distribution on $\mathbb{Z}_N^N$ in which every k coordinates have statistical difference ϵ from uniform where $N = 2^n$,[2] then the best known explicit constructions use $O(kn + \log(1/\epsilon))$ random bits.

In this work, we give a common framework for randomness-efficient generation of approximately k-wise independent distributions over *groups*, particularly subgroups of the unitary group. Our framework applies to, e.g., the group of q-ary strings $\mathbb{Z}_q^N$ (realized as diagonal matrices with qth roots of unity as the diagonal entries), the permutation group S_N (realized as $N \times N$ permutation matrices), the orthogonal group $\mathrm{O}(N)$, and the unitary group $\mathrm{U}(N)$ (with $N = 2^n$). We will not discuss strings further in this work, as they are already very well studied. We first describe prior work on the other three groups, and then explain our new general method.

Permutations. Explicit approximate k-wise independent permutations have found a wide variety of applications; e.g., in cryptography [KNR09], hashing/dimensionality reduction [LK10, KN14], and explicit constructions of expanders [MOP22]. One method for creating them was initiated by Gowers [Gow96], who showed that a random n-qubit circuit composed of $\mathrm{poly}(n,k)\log(1/\epsilon)$ "classical" 3-qubit gates (i.e., permutations on $\{0,1\}^3$) yields an ϵ-approximate k-wise independent permutation on S_{2^n}. (Note that since the circuit size is polynomial rather than linear in nk, the randomness-efficiency of [Gow96] is

[2] Cf. achieving ϵ-approximate k-wise independent permutations from S_N.

polynomially worse than the $O(nk + \log(1/\epsilon))$ random bits needed by a non-explicit random construction.) Gowers's technique was to lower-bound the spectral gap of the random walk on a related graph by $1/\operatorname{poly}(n,k)$. (See [HMMR05, BH08] for improvement of the spectral gap to $1/\widetilde{O}(k^2n^2)$.) Subsequently, using techniques related to space-bounded walks in graphs [Rei08], Kaplan–Naor–Reingold [KNR09] derandomized this "truly random walk" to achieve efficient ϵ-approximate k-wise independent permutations on S_{2^n} with seed length $O(kn + \log(1/\epsilon))$, matching the (inexplicit) random bound. Around the same time, Kassabov [Kas07] got the same seed length (without requiring N to be a power of 2) via a sophisticated construction of a constant-size generating set for any S_N that makes the resulting Cayley graph an expander.

Unitary matrices. Introduced to the quantum computing literature in [DCEL09], explicit ϵ-approximate k-designs for the unitary group have had a wide variety of applications, from randomized benchmarking of quantum gate sets [ZZP17], to efficient state and process tomography [HKOT23], to understanding quantum state and unitary complexity [RY17, BCHJ+21]. Previously, works on constructing approximate unitary designs have chiefly focused on achieving "strongly explicitness" rather than on randomness-efficiency. In particular, the goal has been to show that a *truly* random n-qubit quantum circuit composed of $S = \operatorname{poly}(n,k)\log(1/\epsilon)$ gates (i.e. each gate is a Haar random unitary operator on a constant number of uniformly randomly chosen qubits) constitutes an ϵ-approximate k-design for $\mathrm{U}(2^n)$. The breakthrough in this area came from the work of Brandão, Harrow, and Horodecki [BHH16], who showed that $S = O(n^2k^{10.5}\log(1/\epsilon))$ suffices for $k \le 2^{\Omega(n)}$. (See also [HL09] for an earlier construction using $\operatorname{poly}(n,k)\log(1/\epsilon)$ gates when $k = O(n/\log n)$, and [Sen18] for a construction in the $k = \operatorname{poly}(n)$ regime.) Further work has been done on improving the circuit depth and the exponent on k; see [Haf22]. As far as we are aware, ours is the first work to derandomize these results and achieve a seed length that is *linear* rather than polynomial in n and k and works for all k, thus matching the non-explicit random construction. As an example application of our result for unitary matrices, by applying [BCHJ+21] we get an efficient deterministic procedure for outputting $2^{O(nk)}$ n-qubit unitary circuits such that at least $2^{\Omega(nk)}$ of them have quantum circuit complexity $\Omega(\frac{n}{\log n}k)$ (provided $k \le 2^{\Omega(n)}$).

Orthogonal Matrices. It is natural to think that designs for $\mathrm{O}(N)$ and $\mathrm{U}(N)$ should be related (and indeed orthogonal designs have played a role in randomized benchmarking for quantum circuits [HFGW18]). However there is no obvious reduction between the tasks of constructing ϵ-approximate k-designs for the two groups. The first paper we are aware of that attempts to explicitly construct approximate orthogonal designs is [KM15]. That work used explicit orthogonal designs with $O(kn + \log(1/\epsilon))$ seed length as the core pseudorandom object underlying its state-of-the-art pseudorandom generator for linear threshold functions on $\mathbb{S}^{n-1}$. Unfortunately, there was an error in their construction of these designs.[3] Fixing this error was a key motivation for the present work, and indeed our Theorem 1.1 provides the crucial ingredient needed for the pseudorandom generators of [KM15].

Some of our technical ideas for handling the orthogonal group are drawn from the work of Haferkamp and Hunter-Jones, who showed (Theorem 9 of [HHJ21]) that truly random local orthogonal n-qudit circuits of size $\operatorname{poly}(n,k)\log(q/\epsilon)$ constitute ϵ-approximate k-designs for $\mathrm{O}(q^n)$, provided $q \ge 8k^2$. This result has suboptimal randomness complexity because of the polynomial rather than linear dependence on n and k, and only gives approximate k-designs for small values of k.

1.1 Our framework

As stated earlier, we are interested in k-wise independent distributions over groups, particularly the symmetric, orthogonal, and unitary groups. For each such group G, the notion of "k-wise independence" is defined through a certain *representation* ρ^k of the group. Informally, we say a distribution $\mathcal{P}$ on G is

[3] The error is in the interpretation of the main result of [BG12] that is used to establish Corollary 6.1 of [KM15]. Corollary 6.1 claims that the spectral gap established by [BG12] for $\mathrm{SU}(N)$ is independent of N, but this is in error [Kot22]; indeed, as noted in [BHH16] after their Corollary 7, "the proof [in [BG12]] does not give any estimate of the dependency of the spectral gap on N."

approximately k-wise independent if

$$\mathop{\mathbf{E}}_{\boldsymbol{g}\sim\mathcal{P}}[\rho^k(\boldsymbol{g})] \approx \mathop{\mathbf{E}}_{\boldsymbol{g}\sim \mathrm{G}}[\rho^k(\boldsymbol{g})], \tag{1}$$

where on the right-hand side $\boldsymbol{g}$ is drawn from the Haar distribution on G.[4]

Let us consider our three example groups G, starting with the orthogonal group $\mathrm{O}(N)$. In this case, the associated representation ρ^k is on $(\mathbb{C}^N)^{\otimes k}$, and it maps $R \in \mathrm{O}(N)$ to $R^{\otimes k}$. In other words, specialized to the orthogonal group, Equation (1) asserts that $\mathcal{P}$ is an approximate k-design on $\mathrm{O}(N)$ provided

$$\mathop{\mathbf{E}}_{\boldsymbol{R}\sim\mathcal{P}}[\boldsymbol{R}^{\otimes k}] \approx \mathop{\mathbf{E}}_{\boldsymbol{R}\sim \mathrm{O}(N)}[\boldsymbol{R}^{\otimes k}]. \tag{2}$$

As matrices, the entries of $\rho^k(\boldsymbol{R}) = \boldsymbol{R}^{\otimes k}$ are degree-k monomials in the entries of $\boldsymbol{R}$, and thus Equation (1) (qualitatively) implies that any degree-k polynomial in the entries of $\boldsymbol{R}$ has approximately the same expectation under $\mathcal{P}$ as it has under the Haar distribution. This is the usual meaning of approximate k-wise independence in theoretical computer science, and is often how the notion is used in applications. For the unitary matrices U we wish to consider polynomials in both the entries of U and their complex conjugates; thus the appropriate representation of $\mathrm{U}(N)$ is $\rho^{k,k}$ on $(\mathbb{C}^N)^{\otimes 2k}$ defined by

$$\rho^{k,k}(U) = U^{\otimes k} \otimes \overline{U}^{\otimes k}. \tag{3}$$

Actually, to unify notation we will work with $\rho^{k,k}$ even when studying the orthogonal group $\mathrm{O}(N) \leq \mathrm{U}(N)$; in this case of course $\rho^{k,k}$ is equivalent to ρ^{2k}, and we won't be concerned with the difference between k and $2k$. (Note that if k is odd then the expectation of any degree-k monomial in the entries of $\boldsymbol{R}$, $\boldsymbol{R} \sim \mathrm{O}(N)$, is trivially 0.) Finally, for the symmetric group $S_N \leq \mathrm{U}(N)$ we could again use $\rho^{k,k}$, but previous work has (implicitly) used an alternative representation, which we'll call $\mathcal{W}^k$. To define it, let $[N]_{(k)}$ denote the set of sequences of distinct indices $i_1, \dots, i_k \in [N]$ and let $\mathbb{C}^{[N]_{(k)}}$ denote the (complex) vector space with orthonormal basis vectors $|i_1 \cdots i_k\rangle$. Then the representation $\mathcal{W}^k$ is defined on $\pi \in S_N$ via $\mathcal{W}^k(\pi)\,|i_1 \cdots i_k\rangle = |\pi(i_1) \cdots \pi(i_k)\rangle$. This representation $\mathcal{W}^k$ is the one usually associated to k-wise independence on S_N, with the analogue of Equation (1) asserting that $\mathbf{E}_{\boldsymbol{\pi}\sim\mathcal{P}}[(\boldsymbol{\pi}(i_1), \dots, \boldsymbol{\pi}(i_k))]$ is close to being uniformly distributed on $[N]_{(k)}$ for each $(i_1, \dots, i_k) \in [N]_{(k)}$.

A first way to try to achieve approximate k-wise independence on $\mathrm{G} \in \{S_N, \mathrm{O}(N), \mathrm{U}(N)\}$ is through a Markov chain. Suppose $P \subset \mathrm{G}$ is a set (closed under inverses) of size $\mathrm{poly}(n)$, where $n = \log_2 N$. Consider the random walk on G that starts at $\mathbb{1}$ and multiplies by a uniformly random element of P at each step. We may hope that after, say, $\mathrm{poly}(n, k)\log(1/\epsilon)$ steps, the resulting distribution $\mathcal{P}$ on G is close enough to the Haar distribution on G that Equation (1) holds. As alluded to earlier, results of this form were previously shown for $\mathrm{G} = S_{2^n}$ (starting with [Gow96]) and for $\mathrm{G} = \mathrm{U}(2^n)$ (starting with [BHH16]). One significant contribution of the present work is to generalize the latter to apply also to $\mathrm{O}(2^n)$ (or, more precisely and essentially equivalently, its connected subgroup $\mathrm{SO}(2^n)$). Specifically, in Sections 3 to 5, our goal will essentially be to show the following:

Theorem 1.2. *Fix $n \geq 4$ and let $P_n \subset \mathrm{SO}(2^n)$ denote the $O(n^2)$-sized multiset of all n-qubit circuits consisting of 1 gate, either* CNOT *(on 2 qubits) or* $\mathrm{Q} = \begin{bmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{bmatrix}$ *on 1 qubit, and then closed under negation and inverses. Then for any $k \geq 1$,*

$$\left\| \mathop{\mathbf{E}}_{\boldsymbol{g}\sim P_n}[\rho^{k,k}(\boldsymbol{g})] - \mathop{\mathbf{E}}_{\boldsymbol{g}\sim \mathrm{SO}(2^n)}[\rho^{k,k}(\boldsymbol{g})] \right\|_{\mathrm{op}} \leq 1 - \frac{1}{n \cdot \mathrm{poly}(k)}. \tag{4}$$

A similar statement holds for $\mathrm{SU}(2^n)$ *with the* 1*-qubit* H, S, *and* T *gates replacing* Q.

[4] Here and throughout, whenever G is a compact Lie group we write $\boldsymbol{g} \sim \mathrm{G}$ to denote that $\boldsymbol{g}$ is drawn according to the Haar distribution; in particular, this is the uniform distribution if G is finite.

(See Theorem 3.1 for more details. In Section 2 we will pass from $\mathrm{SO}(2^n)$ and $\mathrm{SU}(2^n)$ to $\mathrm{O}(2^n)$ and $\mathrm{U}(2^n)$; our analysis in Sections 3 to 5 is carried out in the "special" versions of these groups for technical reasons which will become clear in Section 3, specifically Section 3.1.) As we discuss in Section 3.1, the high-level approach we take to establish Theorem 1.2 extends an approach from [HHJ21].

Given this result, we could improve the right-hand side to $\epsilon/2^{nk}$ by forming $\boldsymbol{g}$ as a product of $n \cdot \mathrm{poly}(k) \cdot \log(2^{nk}/\epsilon) = \mathrm{poly}(n,k) \cdot \log(1/\epsilon)$ uniformly randomly elements from P_n. (See Definition 2.2, where we give a precise definition of an "ϵ-approximate k-design," for a discussion of why $\epsilon/2^{nk}$ is the right bound for the right-hand-side.) The resulting distribution on $\mathrm{O}(2^n)$ would be an ϵ-approximate k-design, but unfortunately, drawing from this distribution would require a seed of $\mathrm{poly}(n,k) \cdot \log(1/\epsilon)$ truly random bits, which leaves something to be desired from the standpoint of randomness-efficiency.

To improve this and match the randomness-efficiency of the random construction, one may attempt to apply the method of "pseudorandom walks on consistently labeled graphs" from [Rei08, RTV06], or "derandomized squaring" from [RV05]. This is the approach taken in [KNR09] for the symmetric group, where the evolving value of $\mathcal{W}^k(\boldsymbol{\pi})\,|i_1 \cdots i_k\rangle$ can be thought of as a random walk on a graph with vertex set $[N]_{(k)}$. In the setting of Theorem 1.2 there is no graph. Nevertheless, in Section 6 we will show how derandomized squaring can be slightly generalized to obtain the following result (a similar generalization appeared recently in [JMRW22]):

Theorem 1.3. *(Abbreviated version of Theorem 6.21.) Given c, δ, ϵ, there is a strongly explicit deterministic algorithm that outputs a sequence $\mathcal{P}$ of $O(c/\mathrm{poly}(\delta\epsilon))$ "monomials" over the symbols $u_1, \ldots, u_c, u_1^\dagger, \ldots, u_c^\dagger$, each of length $O(\log(1/\epsilon)/\mathrm{poly}(\delta))$, such that $\|\mathrm{avg}_{\boldsymbol{m} \in \mathcal{P}}\{\boldsymbol{m}(\mathcal{U})\}\|_{\mathrm{op}} \leq \epsilon$ whenever $\mathcal{U} = (U_1, \ldots, U_c)$ is a sequence of unitaries with $\left\|\mathrm{avg}_{i \in [c]}\{U_i\}\right\|_{\mathrm{op}} \leq 1 - \delta$. (Here $m(\mathcal{U})$ denotes the product of U_i's and $U_i^\dagger$'s obtained by substituting $u_i = U_i$ in m.)*

Taking the δ of Theorem 1.3 to be the $1/\mathrm{poly}(n,k)$ of Theorem 1.2, and the unitaries $\mathcal{U} = (U_1, \ldots, U_c)$ to correspond to the 1-gate circuits of Theorem 1.2, we obtain strongly explicit ϵ-approximate k-designs, as described in Theorem 1.1, for the special orthogonal and special unitary groups. A simple modification gives corresponding designs for the unitary and orthogonal groups, thus yielding Theorem 1.1.

1.2 Organization of this paper

In Section 2 we give the detailed argument explaining how an initial spectral gap of the sort given by Theorem 1.2 and the generalized "derandomized squaring" result given by Theorem 1.3 together yield efficient explicit approximate designs for the orthogonal and unitary groups. The rest of the paper is devoted to establishing the two necessary ingredients Theorem 1.2 and Theorem 1.3. Section 3 gives our general framework for establishing the initial spectral gap for the special unitary and special orthogonal groups; as we explain there, a crucial step in this framework is establishing a spectral gap for a certain "auxiliary" m-qubit random walk which was inspired by the analysis of [HHJ21]. Similar to [HHJ21], it turns out that to analyze this auxiliary random walk, two quite different technical arguments are required depending on whether the tensor power k is "large" or "small" compared to the number of qubits m; we give these two arguments in Section 4 and Section 5 respectively. Finally, we provide the necessary analysis of the generalized "derandomized pseudorandom walks" in Section 6.

1.3 Notation and preliminaries

To give our constructions it will be convenient for us to use some of the language of quantum computing. We will generally consider operators on $\mathbb{C}^N$, where $N = 2^n$ for some $n \in \mathbb{N}^+$. We identify $\mathbb{C}^N = (\mathbb{C}^2)^{\otimes n}$ and think of the tensor factors as corresponding to n "qubits".

Notation 1.4. Let $g \in \mathrm{U}(2^\ell)$, thought of as an ℓ-qubit "gate" and let $e = (i_1, \ldots, i_\ell)$ be a sequence of ℓ distinct elements of $[n]$, i.e. $e \in [n]_\ell$. (Here ℓ should be thought of as "much less than n", in particular we will be interested in constant ℓ.) We use the notation g_e for the operator in $\mathrm{U}(N)$ defined by applying g on qubits (i.e., tensor factors) $i_1, \ldots, i_\ell$ (in that order) and applying the identity operator on the remaining

$n-\ell$ qubits. When $e \in \binom{[n]}{\ell}$ is a set rather than a sequence, we assume the increasing order on its elements.

We write $A^\dagger$ to denote the conjugate transpose of a complex matrix A, $\|A\|_{\mathrm{op}}$ to denote the operator norm and $\|A\|_1$ to denote its Schatten 1-norm. We use bold font to denote random variables.

2 A general framework: Explicit k-wise independent permutations, orthogonal designs, and unitary designs

Let $\mathrm{G}(n)$ be a subgroup of $\mathrm{U}(n)$ (the key examples to keep in mind are the group of permutations on 2^n elements, the 2^n-dimensional orthogonal group, the 2^n-dimensional unitary group itself, and the "special" versions of the latter two). In light of Theorem 1.3, given a probability distribution on a subset of $\mathrm{G}(n)$, we would like to understand how fast the associated random walk mixes vis-a-vis a particular representation, namely the k-wise tensor product representation (since that representation corresponds to k-wise independence).

Let $\mathcal{P}$ be a probability distribution on $\mathrm{G}(n)$ that is symmetric (meaning that $\boldsymbol{g}^{-1} = \boldsymbol{g}^\dagger$ is distributed as $\mathcal{P}$ when $\boldsymbol{g}$ is), and let ρ be a unitary representation of $\mathrm{G}(n)$. Note that since ρ is unitary, $\mathbf{E}_{\boldsymbol{g}\sim\mathcal{P}}[\rho(\boldsymbol{g})]$ is a Hermitian operator with real eigenvalues lying in $[-1,1]$. Since our goal is k-wise independence, the representations that are of interest to us are k-wise tensor product representations:

Notation 2.1 (k-wise tensor product representations)**.** For any $k \in \mathbb{N}^+$, we will write $\rho_{2^n}^{k,k}$ for the (complex) representation of $\mathrm{G}(n)$ defined by

$$\rho_{2^n}^{k,k}(g) = g^{\otimes k,k} \coloneqq g^{\otimes k} \otimes \overline{g}^{\otimes k}, \tag{5}$$

where $\overline{g}$ denotes the complex conjugation of matrix g.

There are several different definitions of ϵ-approximate k-designs in the literature, all of which are equivalent if one is willing to lose factors of 2^{nk} on ϵ. For definiteness, we choose the 1-norm definition from [HL09]. (One could also equivalently use the notion from Kothari–Meka [KM15], again up to 2^{nk} factors.)

Definition 2.2. A distribution $\mathcal{P}$ on a finite subset of matrices from $\mathrm{G}(n)$ is an *ϵ-approximate k-design for* $\mathrm{G}(n)$ if

$$\left\| \mathop{\mathbf{E}}_{\boldsymbol{g}\sim\mathcal{P}}[\rho_{2^n}^{k,k}(\boldsymbol{g})] - \mathop{\mathbf{E}}_{\boldsymbol{g}\sim\mathrm{G}(n)}[\rho_{2^n}^{k,k}(\boldsymbol{g})] \right\|_1 \leq \epsilon \tag{6}$$

(where $\|\cdot\|_1$ denotes the Schatten 1-norm). We remark that the above condition is implied by the following operator-norm bound:

$$\left\| \mathop{\mathbf{E}}_{\boldsymbol{g}\sim\mathcal{P}}[\rho_{2^n}^{k,k}(\boldsymbol{g})] - \mathop{\mathbf{E}}_{\boldsymbol{g}\sim\mathrm{G}(n)}[\rho_{2^n}^{k,k}(\boldsymbol{g})] \right\|_{\mathrm{op}} \leq \epsilon/2^{nk}, \tag{7}$$

and indeed we will establish our approximate design results by going through the operator norm.

Often we will study the operator $\mathbf{E}_{\boldsymbol{g}\sim\mathcal{P}}[\rho^k(\boldsymbol{g})]$ through its "Laplacian", which we define as follows:

Definition 2.3. We define the "Laplacian"

$$L_{\mathcal{P}}(\rho) = \mathbb{1} - \mathop{\mathbf{E}}_{\boldsymbol{g}\sim\mathcal{P}}[\rho(\boldsymbol{g})], \tag{8}$$

a self-adjoint (since $\mathcal{P}$ is symmetric) operator satisfying the following inequalities (in the PSD order):

$$0 \leq L_{\mathcal{P}}(\rho) \leq 2 \cdot \mathbb{1}. \tag{9}$$

Notation 2.4. In the preceding definition, we abuse notation as follows: In place of $\mathcal{P}$ we may write a finite (multi)set $P \subset \mathrm{G}(n)$, in which case the uniform distribution on P is understood. We may also write "$\mathrm{G}(n)$" in place of $\mathcal{P}$, in which case the uniform (Haar) distribution is understood. Finally, if $\mathcal{P}$ now denotes a distribution on $\mathrm{G}(\ell)$, and $E \subseteq [n]_\ell$, we write $\mathcal{P} \times E$ for the distribution on $\mathrm{G}(n)$ given by choosing $\boldsymbol{g} \sim \mathcal{P}$, independently choosing $\boldsymbol{e} \sim E$ (uniformly), and finally forming $\boldsymbol{g}_{\boldsymbol{e}}$.

Definition 2.5. Given a symmetric probability distribution $\mathcal{P}$ as in Definition 2.3, we define its "lazy" version, $\widetilde{\mathcal{P}}$, to be the distribution which is an equal mixture of $\mathcal{P}$ and the point distribution supported on the identity element $\mathbb{1}$ (note that $\widetilde{\mathcal{P}}$ is also a symmetric distribution). Similar to Definition 2.3, we have that $\mathbf{E}_{\boldsymbol{g} \sim \widetilde{\mathcal{P}}}[\rho(\boldsymbol{g})]$ is a Hermitian operator but now with real eigenvalues lying in $[0, 1]$, and we have the PSD inequalities

$$0 \leq L_{\widetilde{\mathcal{P}}}(\rho) \leq \mathbb{1}. \tag{10}$$

Fact 2.6. *In the setting of Definitions 2.3 and 2.5, $L_{\mathrm{G}(n)}(\rho)$ is an orthogonal projection operator, and for any symmetric $\mathcal{P}$ we have that*

$$\ker L_{\mathrm{G}(n)}(\rho) \subseteq \ker L_{\mathcal{P}}(\rho) \tag{11}$$

always holds (because for every *g_0 in the support of $\mathcal{P}$ we have $\rho(g_0)\Pi = \Pi$, where $\Pi = \mathbf{E}_{\boldsymbol{g} \sim \mathrm{G}(n)}[\rho(\boldsymbol{g})]$). From this, and Inequalities (9) and (10), we also get*

$$L_{\mathrm{G}(n)}(\rho) \geq \tfrac{1}{2} \cdot L_{\mathcal{P}}(\rho), \tag{12}$$

$$L_{\mathrm{G}(n)}(\rho) \geq L_{\widetilde{\mathcal{P}}}(\rho). \tag{13}$$

As Inequalities (12) and (13) contain a surfeit of symbols, one may wish to read them respectively as

$$\text{"(randomizing } n \text{ qubits)} \geq \tfrac{1}{2} \cdot (\mathcal{P}\text{-pseudorandomizing } n \text{ qubits}) \quad [\text{vis-a-vis } \rho]", \tag{14}$$

$$\text{"(randomizing } n \text{ qubits)} \geq (\widetilde{\mathcal{P}}\text{-pseudorandomizing } n \text{ qubits}) \quad [\text{vis-a-vis } \rho]", \tag{15}$$

with the "$\geq \frac{1}{2}\cdot$" part pronounced "is at least $\frac{1}{2}$ as good as".

It will be convenient to use the Laplacian operator in some of the steps in the following sections, even though we ultimately want statements about the expectation operator. To convert between the two we will use the following:

Fact 2.7. *For any unitary representation ρ, $L_{\mathcal{P}}(\rho) \leq \epsilon \cdot L_{\mathrm{G}(n)}(\rho)$ is equivalent to*

$$\left\| \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathcal{P}}[\rho(\boldsymbol{g})] - \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(n)}[\rho(\boldsymbol{g})] \right\|_{\mathrm{op}} \leq \epsilon. \tag{16}$$

2.1 Initial spectral gaps for S_N, $\mathrm{SO}(N)$ and $\mathrm{SU}(N)$

Here we summarize all of the non-trivial spectral gaps which we will amplify using Theorem 1.3.

Theorem 2.8 ([BH08])**.** *For any $k \geq 1$, there is a (multi)set $P_{S_{2^n}}$ of cardinality $O(n^3)$ such that*

$$\left\| \mathop{\mathbf{E}}_{\boldsymbol{g} \sim P_{S_{2^n}}}[\mathcal{W}^k_{2^n}(\boldsymbol{g})] - \mathop{\mathbf{E}}_{\boldsymbol{g} \sim S_{2^n}}[\mathcal{W}^k_{2^n}(\boldsymbol{g})] \right\|_{\mathrm{op}} \leq 1 - \frac{1}{\widetilde{O}(k^2 n^2)}. \tag{17}$$

Recall that the representation $\mathcal{W}^k_{2^n}$ is defined on $g \in S_{2^n}$ via $\mathcal{W}^k_{2^n}(g)\,|i_1 \cdots i_k\rangle = |g(i_1) \cdots g(i_k)\rangle$. The set $P_{S_{2^n}}$ mentioned above is the set of "simple 3-bit permutations". This is the set of permutations $f_{i,j_1,j_2,h}$, where $i, j_1, j_2 \in [n]$ are all distinct, and h is a Boolean function on $\{0,1\}^2$, which maps $(x_1, \ldots, x_n) \in \{0,1\}^n$ to $(x_1, \ldots, x_{i-1}, x_i \oplus h(x_{j_1}, x_{j_2}), x_{i+1} \ldots, x_n)$.

We establish the following in Sections 3 to 5.

Theorem 2.9 (Theorem 3.1 restated)**.** *For* $\mathrm{G}(n) \in \{\mathrm{SO}(2^n), \mathrm{SU}(2^n)\}$*, and any* $k \geq 1$*, there is a (multi)set* P_G *of cardinality* $O(n^2)$ *such that*

$$\left\| \mathop{\mathbf{E}}_{\boldsymbol{g} \sim P_G} [\rho_{2^n}^{k,k}(\boldsymbol{g})] - \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(n)} [\rho_{2^n}^{k,k}(\boldsymbol{g})] \right\|_{\mathrm{op}} \leq 1 - \frac{1}{n \cdot \mathrm{poly}(k)}. \tag{18}$$

The sets P_G for $\mathrm{SO}(2^n)$ and $\mathrm{SU}(2^n)$ are described in Section 3.4.1.

Our proof of Theorem 2.9 is itself a general framework that could potentially be used to obtain similar results for other subgroups of the unitary group (for example, the sympletic group), even though we only carry out the calculations for $\mathrm{SO}(2^n)$ and $\mathrm{SU}(2^n)$.

2.2 Explicit k-wise independent permutations, orthogonal designs, and unitary designs

We can finally apply Theorem 1.3, so let's write our above results in the notation of this theorem. Fix $k \geq 1$ and consider any of the P (multi)sets described in Theorems 2.8 and 2.9.

Let $\mathcal{U} = (\rho(g) - \mathbf{E}_{\boldsymbol{g} \sim \mathrm{G}(n)}[\rho(\boldsymbol{g})] : g \in P)$ be a sequence of unitaries, where ρ is the appropriate unitary representation ($\mathcal{W}_{2^n}^k$ for S_{2^n} and $\rho_{2^n}^{k,k}$ for $\mathrm{SO}(2^n)$ and $\mathrm{SU}(2^n)$). For this choice of $\mathcal{U}$ we have $c = |P| = \mathrm{poly}(n)$. Notice that $\left\|\mathrm{avg}_{i \in [c]}\{U_i\}\right\|_{\mathrm{op}}$ is exactly the left hand side of the equations in Theorems 2.8 and 2.9, so we know that this average is at most $1 - \delta$ for $\delta = 1/\mathrm{poly}(n, k)$ (as observed, this is actually $1/\widetilde{O}(k^2 n^2)$ for S_{2^n} and $1/(n\,\mathrm{poly}(k))$ for $\mathrm{SO}(2^n)$ and $\mathrm{SU}(2^n)$). Given $\epsilon > 0$, applying Theorem 1.3 (with its "ϵ" parameter set to $\epsilon/2^{nk}$) we obtain a sequence $\mathcal{P}$ of cardinality $\mathrm{poly}(2^{nk}/\epsilon)$ that satisfies $\|\mathrm{avg}_{U \in \mathcal{P}}\{U\}\|_{\mathrm{op}} \leq \epsilon$. Additionally, $U \in \mathcal{P}$ is a product of at most $\mathrm{poly}(nk)\log(1/\epsilon)$ elements of $\mathcal{U}$, and so it can be written as $\rho(g) - \mathbf{E}_{\boldsymbol{g} \sim \mathrm{G}(n)}[\rho(\boldsymbol{g})]$, where g is a product of at most $\mathrm{poly}(nk)\log(1/\epsilon)$ elements of P. This follows since for any $g, g' \in P$,

$$\left(\rho(g) - \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(n)}[\rho(\boldsymbol{g})]\right)\left(\rho(g') - \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(n)}[\rho(\boldsymbol{g})]\right) = \left(\rho(g \cdot g') - \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(n)}[\rho(\boldsymbol{g})]\right), \tag{19}$$

where we use the fact that $\mathbf{E}_{\boldsymbol{g} \sim \mathrm{G}(n)}[\rho(\boldsymbol{g})]$ is an orthogonal projection operator, and that ρ is a representation. We combine all of this in the following theorems:

Theorem 2.10. *Let* $\epsilon > 0$*. Then for any* $k = k(n)$*, there is a set* $\mathcal{P}_{S_{2^n}}$ *that satisfies:*

$$\left\| \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathcal{P}_{S_{2^n}}} [\mathcal{W}^k(\boldsymbol{g})] - \mathop{\mathbf{E}}_{\boldsymbol{g} \sim S_{2^n}} [\mathcal{W}^k(\boldsymbol{g})] \right\|_{\mathrm{op}} \leq \epsilon/2^{nk}. \tag{20}$$

Additionally, this set satisfies the following properties:

- *Its cardinality is* $\mathrm{poly}(2^{nk}/\epsilon)$*.*
- *Each element of* $\mathcal{P}_{S_{2^n}}$ *is given by an* n*-qubit circuit consisting of* $S = \mathrm{poly}(nk)\log(1/\epsilon)$ *gates, which are elements of* P_G*.*
- *The algorithm that takes as input a seed and outputs the associated circuit runs in deterministic* $\mathrm{poly}(S)$ *time.*

Theorem 2.11. *Let* $\mathrm{G}(n) \in \{\mathrm{SO}(2^n), \mathrm{SU}(2^n)\}$ *and* $\epsilon > 0$*. Then for any* $k = k(n)$*, there is a set* $\mathcal{P}_G$ *that satisfies:*

$$\left\| \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathcal{P}_G} [\rho_{2^n}^{k,k}(\boldsymbol{g})] - \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(n)} [\rho_{2^n}^{k,k}(\boldsymbol{g})] \right\|_{\mathrm{op}} \leq \epsilon/2^{nk}. \tag{21}$$

Additionally, this set satisfies the following properties:

- *Its cardinality is* $\mathrm{poly}(2^{nk}/\epsilon)$*.*

- *Each element of $\mathcal{P}_G$ is given by an n-qubit circuit consisting of $S = \mathrm{poly}(nk)\log(1/\epsilon)$ gates, which are elements of P_G.*
- *The algorithm that takes as input a seed and outputs the associated circuit runs in deterministic $\mathrm{poly}(S)$ time.*

Ultimately we want designs for $\mathrm{O}(2^n)$ and $\mathrm{U}(2^n)$; we obtain them from the above via the following simple corollary.

Corollary 2.12 (Theorem 1.1 restated)**.** *Let $\mathrm{G}(n) \in \{\mathrm{O}(2^n), \mathrm{U}(2^n)\}$ and $\epsilon > 0$. Then for any $k = k(n)$, there is a set $\mathcal{P}_G$ that satisfies the conditions of Theorem 2.11.*

Proof. To obtain the result for $\mathrm{O}(2^n)$, after sampling from $\mathcal{P}_{\mathrm{SO}}$ one samples $\boldsymbol{b}$ as a uniformly random ± 1 and multiplies the first column of the sampled matrix by $\boldsymbol{b}$ (which only changes the cardinality of the resulting $\mathcal{P}_{\mathrm{O}}$ by a factor of 2). For $\mathrm{U}(2^n)$, after sampling from $\mathcal{P}_{\mathrm{SU}}$ we sample a uniform random unit-norm complex value $\boldsymbol{u}$ from the $(2^{O(nk)}/\epsilon)$-element set $\{1, e^{i\epsilon/2^{O(nk)}}, e^{i2\epsilon/2^{O(nk)}}, \ldots\}$ and multiply the first column of the sampled matrix by $\boldsymbol{u}$ (which only changes the cardinality of the resulting $\mathcal{P}_{\mathrm{U}}$ by a factor of $2^{O(nk)}/\epsilon$.) □

It is of note that we can apply the framework of this section, through Theorem 1.3, to obtain explicit designs of any subgroup of the unitary group using any unitary representation, as long as one establishes an initial gap first, like in Section 2.1.

3 Establishing an initial spectral gap for special orthogonal and unitary groups

In the rest of the paper we consider a sequence of groups $(\mathrm{G}(n))_{n\geq 1}$ which is either $(\mathrm{SO}(2^n))_{n\geq 1}$ or $(\mathrm{SU}(2^n))_{n\geq 1}$. We recall (see e.g. [Mec19, Section 1.3]) that these groups have associated Lie algebras $\mathfrak{g}_n$, where

$$\text{for } \mathrm{G}(n) = \mathrm{SO}(2^n),\ \mathfrak{g}_n = \{H \in \mathbb{R}^{2^n \times 2^n} : H \text{ skew-symmetric}\}, \tag{22}$$

$$\text{for } \mathrm{G}(n) = \mathrm{SU}(2^n),\ \mathfrak{g}_n = \{H \in \mathbb{C}^{2^n \times 2^n} : H \text{ skew-Hermitian}, \mathrm{tr}(H) = 0\}. \tag{23}$$

When we need to specialize our discussion to a particular one of these two cases, we will do so explicitly; most of our arguments go through for both settings (and many go through for the more general setting in which $\mathrm{G}(n)$ is any compact connected Lie group).

As discussed in Section 2, given Theorem 6.21, in order to construct an explicit k-design for $\mathrm{G}(n)$ it suffices to construct an explicit sequence $\mathcal{U} = (U_1, \ldots, U_c)$ of $2^n \times 2^n$ matrices from $\mathrm{G}(n)$ satisfying $\left\|\rho_{2^n}^{k,k}(U_i)\right\|_{\mathrm{op}} \leq 1$ for all i and $\left\|\mathrm{avg}(\rho_{2^n}^{k,k}(\mathcal{U}))\right\|_{\mathrm{op}} \leq 1 - \frac{1}{n\cdot\mathrm{poly}(k)}$ (in fact, a spectral gap of $\mathrm{poly}(n,k)$ would also be sufficient). Constructing such a sequence for $\mathrm{G}(n)$ as described above is the main goal of this section and is accomplished in the following theorem:

Theorem 3.1. *Let $(\mathrm{G}(n))_{n\geq 1} \in \{(\mathrm{SO}(2^n))_{n\geq 1}, (\mathrm{SU}(2^n))_{n\geq 1}\}$. There is a fixed positive integer $n_0 = 4$ and a finite multiset $P_{n_0} \subset \mathrm{G}(n_0)$ such that for all sufficiently large n and all $k \geq 1$, we have*

$$\forall k \in \mathbb{N}^+, \quad L_{\widetilde{P}_{n_0} \times \binom{[n]}{n_0}}(\rho_{2^n}^{k,k}) \geq \frac{1}{n \cdot \mathrm{poly}(k)} \cdot L_{\mathrm{G}(n)}(\rho_{2^n}^{k,k}). \tag{24}$$

(We note that even without using the "pseudorandom walks" machinery of Section 6, as discussed in Section 1.1, since Theorem 3.1 establishes an initial spectral gap of $1 - \frac{1}{n\cdot\mathrm{poly}(k)}$, simply taking a product of $n \cdot \mathrm{poly}(k) \cdot \log(2^{nk}/\epsilon)$ uniform random draws from $\widetilde{P}_{n_0}$ would yield an ϵ-approximate k-design for $\mathrm{G}(n)$ with seed length $\mathrm{poly}(n,k) \cdot \log(1/\epsilon)$. By combining Theorem 3.1 with Theorem 6.21 (i.e. using pseudorandom walks) we are able to improve this to seed length $O(nk + \log(1/\epsilon))$, thus matching the random construction.)

3.1 Overview of the proof of Theorem 3.1

Our proof of Theorem 3.1 refines and extends an approach from [HHJ21], and combines it with arguments from [BHH16]. In this subsection we give a high-level overview of the structure of the proof, and in the next subsection we give (a modular version of) the actual proof. Establishing the various modular pieces will comprise the rest of the paper after Section 3.2.

In Theorem 4 of [HHJ21], Haferkamp and Hunter-Jones establish a spectral gap for non-local random quantum circuits with truly (Haar) random two-qudit unitary gates over the unitary group. This is done by analyzing Haar random unitary gates over $m-1$ randomly chosen qubits from an m-qubit system; this enables them to establish a recurrence relation which lets them bound the spectral gap of k-qudit Haar random unitary gates in terms of the spectral gap of $(k+1)$-qudit Haar random unitary gates. Our Lemma 3.2 below is a generalization and rephrasing of their recurrence relation for the special[5] unitary and special orthogonal groups; it essentially says that if truly randomizing (a randomly chosen) $m-1$ out of m qubits is "not too much worse" than truly randomizing all m qubits, then truly randomizing only a constant number of (randomly chosen) qubits out of m qubits is also not too much worse than truly randomizing all m qubits. Given this, the remaining tasks are (1) to show that indeed truly randomizing (a randomly chosen) $m-1$ out of m qubits is "not too much worse" than truly randomizing all m qubits; and (2) to show that at the bottom level of the argument, it suffices to *pseudorandomize* a constant number of (randomly chosen) qubits out of m qubits.

Task (1) requires a significant amount of technical work and is the subject of Section 4 and Section 5. We follow the high-level approach of [HHJ21] by breaking the analysis into two sub-cases (Theorem 3.3 and Theorem 3.4) depending on the relative sizes of k and m. In each of these sub-cases we adapt and generalize the analysis of [HHJ21] (we note that the "small-m" case of [HHJ21], for the unitary group, was based in turn on [BHH16]) in a way which permits a unified treatment of both the special orthogonal group and the special unitary group.

Task (2) is necessary because our ultimate goal statement, Theorem 3.1, requires the randomly chosen non-local gates to be drawn from a *finite* ensemble of gates rather than being Haar random ("truly random") gates over n_0 qubits. For this step (made formal in Lemma 3.6), following [BHH16] we use a deep result of Bourgain and Gamburd (subsequently generalized by Benoiste and de Saxcé [BdS16]) to pass from the Haar distribution over $\mathrm{G}(n_0)$ to a uniform distribution over an explicit finite ensemble of n_0-qubit gates; see Section 3.4. The [BdS16] results require that the Lie groups in question be compact and simple; this requirement is why we need to work with the special versions of the unitary and orthogonal groups (indeed, in the special orthogonal case we need to further pass to the projective special orthogonal group; see the proof of Corollary 3.10).

3.2 Proof of Theorem 3.1

In order to establish the lower bound of Inequality (24) we will need to chain together some statements that go in the opposite direction from Inequality (14) and Inequality (15). We do this via the following lemma, which we prove in Section 3.3.

Lemma 3.2. *Fix a positive integer constant $n_0 \geq 4$. Suppose that for $n_0 < m \leq n$ we have*

$$\forall k \in \mathbb{N}^+, \quad L_{\mathrm{G}(m-1)\times\binom{[m]}{m-1}}(\rho_{2^m}^{k,k}) \geq \tau_{k,m} \cdot L_{\mathrm{G}(m)}(\rho_{2^m}^{k,k}). \tag{25}$$

Then

$$\forall k \in \mathbb{N}^+, \quad L_{\mathrm{G}(n_0)\times\binom{[n]}{n_0}}(\rho_{2^n}^{k,k}) \geq \left(\prod_{n_0 < m \leq n} \tau_{k,m}\right) \cdot L_{\mathrm{G}(n)}(\rho_{2^n}^{k,k}). \tag{26}$$

We remark that Lemma 3.2 only deals with "truly" (Haar) random gates; later we will move from n_0-arity "truly random" gates to "pseudorandom" gates, which are drawn uniformly at random from

[5] At the end of this subsection we explain why, even though our ultimate goal is to obtain results for the orthogonal and unitary groups, we need to work with the special versions of these groups at this point in the argument.

a finite multiset. It may be helpful to think of the lemma's conclusion (Inequality (26)) as intuitively saying that truly randomizing only constantly many (randomly chosen) qubits is "not too much worse" than truly randomizing all n qubits, vis-a-vis the k-wise tensor product representation.

With the above lemma in hand, proving Theorem 3.1 breaks down naturally into two steps.

First step: Passing from truly random m-qubit gates to truly random $(m-1)$-qubit gates. In other words, lower-bounding $\tau_{k,m}$ for $m = n_0 + 1, \ldots, n$. This is the main technical task where the bulk of our work is required. The analysis is done separately for "large m" and "small m" cases, similar to Lemmas 6 and 7 of [HHJ21], respectively.

Section 4 lower bounds $\tau_{k,m}$ for "large m":

Theorem 3.3. *For all $k \leq \frac{1}{\sqrt{10}m^2} 2^{m/2}$ we have that Inequality (25) holds with $\tau_{k,m} \geq 1 - \frac{1}{m} - \frac{\sqrt{10}km}{2^{m/2}}$.*

Section 5 gives a lower bound on $\tau_{k,m}$ which will be useful for "small m":

Theorem 3.4. *For all $m \geq 4$ and all $k \in \mathbb{N}^+$, we have that Inequality (25) holds with $\tau_{k,m} \geq .04$.*

(We note that Theorem 3.4's requirement that $m \geq 4$ is why we take $n_0 = 4$ in Theorem 3.1.)

Given Theorem 3.3 and Theorem 3.4, we get the desired lower bound on $\tau_{k,n_0+1} \cdots \tau_{k,n}$ from a routine computation:

Lemma 3.5. *For any constant $n_0 \geq 4$, for all n and all $k \in \mathbb{N}^+$ we have $\tau_{k,n_0+1} \cdots \tau_{k,n} \geq \frac{1}{n \cdot \mathrm{poly}(k)}$.*

Proof. Fix $n_0 \geq 4$ and take any $n, k \geq 1$. Defining $\ell = \lfloor 4 \log_2(60k) \rfloor \geq 20$, by Theorem 3.4 we have

$$\tau_{k,n_0+1} \cdots \tau_{k,\ell} \geq (.04)^{\ell} = (.04)^{O(\log k)} \geq \frac{1}{\mathrm{poly}(k)}. \tag{27}$$

This proves the result if $n \leq \ell$. Otherwise, it remains to show that

$$\tau_{k,\ell+1} \cdots \tau_{k,n} \geq 1/n. \tag{28}$$

For $m \geq \ell + 1$ we have $k \leq \frac{1}{60} 2^{m/4} \leq \frac{1}{\sqrt{10}m^2} 2^{m/2}$, so we are eligible to use the bound from Theorem 3.3. Then using

$$\frac{\sqrt{10}km}{2^{m/2}} \leq \frac{\sqrt{10}m}{60 \cdot 2^{m/4}} \leq 2^{-m/5}, \quad 1 - \frac{1}{m} - 2^{-m/5} \geq \left(1 - \frac{1}{m}\right) \exp(-2^{1-m/5}) \tag{29}$$

(the last inequality using $m \geq \ell \geq 20$), we conclude

$$\tau_{k,\ell+1} \cdots \tau_{k,n} \geq \prod_{m=\ell+1}^{n} \left(1 - \frac{1}{m}\right) \exp(-2^{1-m/5}) = \frac{\ell}{n} \exp\left(-\sum_{m=\ell+1}^{n} 2^{1-m/5}\right) \geq \frac{1}{n} \tag{30}$$

(using $\ell \geq 20$), confirming Inequality (28). □

Second step: From "truly random" non-local n_0-qubit gates to "pseudorandom" non-local n_0-qubit gates. The next lemma, proved in Section 3.4, may be viewed as saying that (suitably) *pseudo*-randomizing constantly many randomly chosen qubits is "not much worse" than *truly* randomizing those qubits.

Lemma 3.6. *There is an absolute constant $n_0 = 4$ such that for $n \geq n_0 + 1$, we have*

$$\forall k \in \mathbb{N}^+, \quad L_{\widetilde{P}_{n_0} \times \binom{[n]}{n_0}}(\rho_{2^n}^{k,k}) \geq \kappa_{n_0} \cdot L_{\mathrm{G}(n_0) \times \binom{[n]}{n_0}}(\rho_{2^n}^{k,k}),$$

where κ_{n_0} is an absolute constant (depending only on n_0).

Theorem 3.1 follows from Lemma 3.2, Lemma 3.5 and Lemma 3.6.

3.3 Proof of Lemma 3.2

Lemma 3.7 (Restatement of Lemma 3.2)**.** *Fix a positive integer $n_0 \geq 4$. Suppose that for $n_0 < m \leq n$ we have*

$$\forall k \in \mathbb{N}^+, \quad L_{\mathrm{G}(m-1)\times\binom{[m]}{m-1}}(\rho_{2^m}^{k,k}) \geq \tau_{k,m} \cdot L_{\mathrm{G}(m)}(\rho_{2^m}^{k,k}). \tag{31}$$

Then

$$\forall k \in \mathbb{N}^+, \quad L_{\mathrm{G}(n_0)\times\binom{[n]}{n_0}}(\rho_{2^n}^{k,k}) \geq \left(\prod_{n_0<m\leq n} \tau_{k,m}\right) \cdot L_{\mathrm{G}(n)}(\rho_{2^n}^{k,k}). \tag{32}$$

Proof. For readability we simply write τ_i in this proof to stand for $\tau_{k,i}$. Also for readability we express the lemma as

$$\begin{aligned} &(\text{randomizing } m-1 \text{ out of } m \text{ qubits}) \geq \tau_m \cdot (\text{randomizing all } m \text{ qubits}) \quad \forall\ n_0 < m \leq n \\ \implies \quad &(\text{randomizing } n_0 \text{ out of } n \text{ qubits}) \geq \tau_{n_0+1} \cdots \tau_n \cdot (\text{randomizing all } n \text{ qubits}), \end{aligned} \tag{33}$$

with the modifier "vis-as-vis all $\rho_{2^m}^{k,k}$" being implied. The $m = n_0 + 1$ case of Inequality (33) is

$$(\text{randomizing } n_0 \text{ out of } n_0+1 \text{ qubits}) \geq \tau_{n_0+1} \cdot (\text{randomizing all } n_0+1 \text{ qubits}). \tag{34}$$

From this, by adding an ignored (n_0+2)th qubit, we are able to conclude

$$\begin{aligned} &(\text{randomizing } n_0 \text{ out of the first } n_0+1 \text{ of } n_0+2 \text{ qubits}) \\ &\qquad \geq \tau_{n_0+1} \cdot (\text{randomizing the first } n_0+1 \text{ of } n_0+2 \text{ qubits}). \end{aligned} \tag{35}$$

To derive this implication more formally, start with Inequality (34), which says that for all $k \in \mathbb{N}^+$,

$$\mathop{\mathbf{E}}_{\substack{\boldsymbol{g}\sim\mathrm{G}(n_0)\\ \boldsymbol{e}\sim\binom{[n_0+1]}{n_0}}} [\mathbb{1} - \boldsymbol{g}_{\boldsymbol{e}}^{\otimes k,k}] \geq \tau_{n_0+1} \cdot \mathop{\mathbf{E}}_{\boldsymbol{h}\sim\mathrm{G}(n_0+1)} [\mathbb{1} - \boldsymbol{h}^{\otimes k,k}]. \tag{36}$$

We now consider tacking on a (n_0+2)th tensor factor that is ignored by both $\boldsymbol{g}_e$ and by $\boldsymbol{h}$. Since $A \geq B \implies A \otimes \mathbb{1} \geq B \otimes \mathbb{1}$, we can tensor-product both sides of Inequality (36) by $\mathbb{1}^{\otimes k,k}$ (where $\mathbb{1}$ denotes the 2×2 identity matrix) to conclude

$$\mathop{\mathbf{E}}_{\substack{\boldsymbol{g}\sim\mathrm{G}(n_0)\\ \boldsymbol{e}\sim\binom{[n_0+1]}{n_0}\in[n_0+2]_{n_0}}} [\mathbb{1} - \boldsymbol{g}_{\boldsymbol{e}}^{\otimes k,k}] \geq \tau_{n_0+1} \cdot \mathop{\mathbf{E}}_{\substack{\boldsymbol{h}\sim\mathrm{G}(n_0+1)\\ f:=[n_0+1]\in[n_0+2]_{n_0+1}}} [\mathbb{1} - \boldsymbol{h}_f^{\otimes k,k}], \tag{37}$$

and this is the meaning of Inequality (35). Indeed, we can insert the ignored (n_0+2)th qubit at any position, not just the last one; i.e., for any $j \in [n_0+2]$,

$$\mathop{\mathbf{E}}_{\substack{\boldsymbol{g}\sim\mathrm{G}(n_0)\\ \boldsymbol{e}\sim\binom{[n_0+2]\setminus j}{n_0}}} [\mathbb{1} - \boldsymbol{g}_{\boldsymbol{e}}^{\otimes k,k}] \geq \tau_{n_0+1} \cdot \mathop{\mathbf{E}}_{\substack{\boldsymbol{h}\sim\mathrm{G}(n_0+1)\\ f:=[n_0+2]\setminus j}} [\mathbb{1} - \boldsymbol{h}_f^{\otimes k,k}.] \tag{38}$$

If we now average the above (PSD-order) inequality over $\boldsymbol{j} \sim [n_0+2]$ we get

$$\mathop{\mathbf{E}}_{\substack{\boldsymbol{g}\sim\mathrm{G}(n_0)\\ \boldsymbol{e}\sim\binom{[n_0+2]}{n_0}}} [\mathbb{1} - \boldsymbol{g}_{\boldsymbol{e}}^{\otimes k,k}] \geq \tau_{n_0+1} \cdot \mathop{\mathbf{E}}_{\substack{\boldsymbol{h}\sim\mathrm{G}(n_0+1)\\ \boldsymbol{f}\sim\binom{[n_0+2]}{n_0+1}}} [\mathbb{1} - \boldsymbol{h}_{\boldsymbol{f}}^{\otimes k,k}], \tag{39}$$

which we would express as

$$(\text{randomizing } n_0 \text{ out of } n_0+2 \text{ qubits}) \geq \tau_{n_0+1} \cdot (\text{randomizing } n_0+1 \text{ out of } n_0+2 \text{ qubits}). \tag{40}$$

But the $m = n_0 + 2$ case of our hypothesis Inequality (33) is

$$(\text{randomizing } n_0+1 \text{ out of } n_0+2 \text{ qubits}) \geq \tau_{n_0+2} \cdot (\text{randomizing all } n_0+2 \text{ qubits}), \tag{41}$$

so chaining this together with Inequality (40) (using the PSD-ordering fact $A \geq B, B \geq C \implies A \geq C$) gives

$$(\text{randomizing } n_0 \text{ out of } n_0+2 \text{ qubits}) \geq \tau_{n_0+1} \cdot \tau_{n_0+2} \cdot (\text{randomizing all } n_0+2 \text{ qubits}). \tag{42}$$

Iterating this argument completes the proof of the lemma. □

3.4 Proof of Lemma 3.6

An ingredient we need for Lemma 3.6 is the existence of a suitable finite "gate set" with useful properties. This is provided by the following lemma, which follows from known universality results in quantum computing (see Section 3.4.1):

Lemma 3.8. *There is an absolute constant $n_0 = 4$ for which there is a finite multiset $P_{n_0} \subset \mathrm{SO}(2^{n_0})$, closed under negations and inverses, with two properties:*

(A) (There is a basis in which) every matrix in P_{n_0} has algebraic entries.

(B) Finite products of elements of P_{n_0} are dense in $\mathrm{SO}(2^{n_0})$.

The same statement is true for $\mathrm{SU}(2^{n_0})$ (also with $n_0 = 4$).

Lemma 3.8 allows us to use a deep result of Benoist and de Saxcé [BdS16], which extended earlier work of Bourgain–Gamburd [BG12] (for the case of the special unitary group) to a broader range of groups. The main result of [BdS16] is as follows:

Theorem 3.9. *([BdS16, Consequence of Theorem 1.2].) For $n \geq 1$ let $\mathrm{G}(n) \subseteq \mathrm{SU}(2^n)$ be a connected compact simple Lie group. Fix a positive integer n_0 and suppose that $P_{n_0} \subset \mathrm{G}(n_0)$ satisfies properties (A) and (B) of Lemma 3.8. Then there exists a constant $\kappa > 0$ such that*

$$\left\| \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \widetilde{P_{n_0}}} [\mathrm{reg}(\boldsymbol{g})] - \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(n_0)} [\mathrm{reg}(\boldsymbol{g})] \right\|_{\mathrm{op}} \leq 1 - \kappa, \tag{43}$$

where reg *denotes the regular representation of $\mathrm{G}(n_0)$. Equivalently, $L_{\widetilde{P_{n_0}}}(\mathrm{reg}) \geq \kappa \cdot L_{\mathrm{G}(n_0)}(\mathrm{reg})$, or*

$$(\widetilde{P_{n_0}}\text{-pseudorandomizing in } 2^{n_0} \text{ dimensions}) \geq \kappa \cdot (\text{randomizing } 2^{n_0} \text{ dimensions}) \quad [\text{vis-a-vis reg}]. \tag{44}$$

We remark that (as noted by [BHH16]) a weaker form of Equation (43), with the k-wise tensor product representation in place of the regular representation and κ depending on k, has been known at least since [AK63]; however, the stronger quantitative bound of Equation (43) is essential for our purposes.

Theorem 3.9 yields the following useful corollary:

Corollary 3.10. *For $n_0 = 4$, $\mathrm{G}(n_0) = \mathrm{SO}(2^{n_0})$, and $P_{n_0} \subset \mathrm{G}(n_0)$ satisfying properties (A) and (B) of Lemma 3.8, there is a constant $\kappa > 0$ such that for all $k \in \mathbb{N}^+$ we have $L_{\widetilde{P_{n_0}}}(\rho_{2^{n_0}}^{k,k}) \geq \kappa \cdot L_{\mathrm{G}(n_0)}(\rho_{2^{n_0}}^{k,k})$. That is, vis-a-vis any $\rho_{2^{n_0}}^{k,k}$, we have*

$$(\widetilde{P_{n_0}}\text{-pseudorandomizing } n_0 \text{ qubits}) \geq \kappa \cdot (\mathrm{G}(n_0)\text{-randomizing } n_0 \text{ qubits}). \tag{45}$$

The same is true for $n_0 = 4$, $\mathrm{G}(n_0) = \mathrm{SU}(2^{n_0})$.

Proof. We first note that since all irreducible representations appear in the regular representation[6], the conclusion of Theorem 3.9 also holds for any $\rho_{2^\ell}^{k,k}$ representation. Since the special unitary group is connected, compact, and simple[7], this immediately gives Corollary 3.10 in the case $\mathrm{G}(n_0) = \mathrm{SU}(2^{n_0})$.

For the special orthogonal case, while $\mathrm{G}(n_0) = \mathrm{SO}(2^{n_0})$ is not simple, the projective special orthogonal group $\mathrm{PSO}(2^{n_0}) = \mathrm{SO}(2^{n_0})/\{\pm 1\}$ is a connected compact simple Lie group. Writing P'_{n_0} to denote the multiset of elements of $\mathrm{PSO}(2^{n_0})$ corresponding to P_{n_0}, Theorem 3.9 gives us that

$$\left\| \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \widetilde{P'_{n_0}}} [\rho_{2^{n_0}}^{k,k}(\boldsymbol{g})] - \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{PSO}(2^{n_0})} [\rho_{2^{n_0}}^{k,k}(\boldsymbol{g})] \right\|_{\mathrm{op}} \leq 1 - \kappa. \tag{46}$$

[6] For a concrete proof in the case of $\mathrm{G}(\ell) = \mathrm{SO}(2^\ell)$, see e.g. [KM15, Lem. 6.1].

[7] Recall that the Lie algebra of the special unitary group is simple, and that Benoiste and de Saxcé remark, following their Theorem 1.2 in [BdS16], that "*For us, a compact simple Lie group will be a compact real Lie group whose Lie algebra is simple.*"

Now recalling that $\rho_{2^{n_0}}^{k,k}(g) = g^{\otimes k} \otimes g^{\otimes k}$, since P_{n_0} is closed under negation there is no need to distinguish between $\mathrm{PSO}(2^{n_0})$ and $\mathrm{SO}(2^{n_0})$ in either of the expectations appearing in Equation (46), i.e. we have

$$\mathop{\mathbf{E}}_{\boldsymbol{g}\sim\widetilde{P'_{n_0}}}[\rho_{2^{n_0}}^{k,k}(\boldsymbol{g})] = \mathop{\mathbf{E}}_{\boldsymbol{g}\sim\widetilde{P_{n_0}}}[\rho_{2^{n_0}}^{k,k}(\boldsymbol{g})], \qquad \mathop{\mathbf{E}}_{\boldsymbol{g}\sim\mathrm{PSO}(2^{n_0})}[\rho_{2^{n_0}}^{k,k}(\boldsymbol{g})] = \mathop{\mathbf{E}}_{\boldsymbol{g}\sim\mathrm{SO}(2^{n_0})}[\rho_{2^{n_0}}^{k,k}(\boldsymbol{g})], \tag{47}$$

which gives Corollary 3.10 for the case $\mathrm{G}(n_0) = \mathrm{SO}(2^{n_0})$. □

With Corollary 3.10 in hand, now we are ready to prove Lemma 3.6:

Proof of Lemma 3.6. By Corollary 3.10, we have $L_{\widetilde{P_{n_0}}}(\rho_{2^{n_0}}^{k,k}) \geq \kappa_{n_0} \cdot L_{\mathrm{G}(n_0)}(\rho_{2^{n_0}}^{k,k})$, i.e.

$$\mathbb{1} - \mathop{\mathbf{E}}_{\boldsymbol{h}\sim\widetilde{P_{n_0}}}[\rho(\boldsymbol{h})] \geq \kappa_{n_0}\left(\mathbb{1} - \mathop{\mathbf{E}}_{\boldsymbol{g}\sim\mathrm{G}(n_0)}[\rho(\boldsymbol{g})]\right). \tag{48}$$

We consider tacking on $n - n_0$ tensor factors that are ignored by both $\boldsymbol{g}$ and by $\boldsymbol{h}$. Since $A \geq B \implies A \otimes \mathbb{1} \geq B \otimes \mathbb{1}$, we can tensor-product both sides of Equation (48) by the identity to conclude

$$\mathbb{1} - \mathop{\mathbf{E}}_{\boldsymbol{h}\sim\widetilde{P_{n_0}}}[\rho(\boldsymbol{h}_{[n_0]})] \geq \kappa_{n_0}\left(\mathbb{1} - \mathop{\mathbf{E}}_{\boldsymbol{g}\sim\mathrm{G}(n_0)}[\rho(\boldsymbol{g}_{[n_0]})]\right). \tag{49}$$

We can insert the ignored $n - n_0$ qubits at any positions, not just the last one; averaging the resulting inequalities, we get

$$\frac{1}{\binom{n}{n-n_0}} \sum_{1\leq i_1<\cdots<i_{n_0}\leq n}\left(\mathbb{1} - \mathop{\mathbf{E}}_{\boldsymbol{h}\sim\widetilde{P_{n_0}}}[\rho(\boldsymbol{h}_{(i_1,\ldots,i_{n_0})})]\right) \geq \kappa_{n_0} \cdot \frac{1}{\binom{n}{n-n_0}} \sum_{1\leq i_1<\cdots<i_{n_0}\leq n}\left(\mathbb{1} - \mathop{\mathbf{E}}_{\boldsymbol{g}\sim\mathrm{G}(n_0)}[\rho(\boldsymbol{g}_{(i_1,\ldots,i_{n_0})})]\right), \tag{50}$$

which is what Lemma 3.6 asserts. □

3.4.1 Proof of Lemma 3.8

We first consider $\mathrm{SO}(2^4)$; so we must show that there is a finite multiset $P_4 \subset \mathrm{SO}(2^4)$, closed under inverses, that satisfies conditions (A) and (B) of Lemma 3.8.

Define the 1- and 2-qubit gates

$$\mathrm{Q} \coloneqq \begin{bmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{bmatrix} \quad \text{and} \quad \mathrm{CNOT} \coloneqq \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{51}$$

and let P_4 be the following finite subset[8] of $\mathrm{SO}(2^4)$:

$$P_4 \coloneqq \text{the closure of } \{\mathrm{Q}_{(j)} : j \in [4]\} \cup \{\mathrm{CNOT}_{(i,j)} : i, j \in [4], i \neq j\} \text{ under inverses and negations.} \tag{52}$$

Clearly P_4 satisfies (A), and (B) follows from the following result from [Shi02, Thm. 3.1]:

Fact 3.11. *The 1- and 2-qubit gates*

$$\mathrm{Q} \coloneqq \begin{bmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{bmatrix} \quad \textit{and} \quad \mathrm{CNOT} \coloneqq \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{53}$$

are together universal for quantum computing with real amplitudes. More precisely, recalling Equation (52), we have that finite products of elements of P_4 are dense in $\mathrm{SO}(2^4)$.

[8] Recall that $\mathrm{CNOT} \notin \mathrm{SO}(2^2)$, but $\mathrm{CNOT} \otimes \mathbb{1}_{4\times 4} \in \mathrm{SO}(16)$.

Next we turn to $\mathrm{SU}(2^4)$. Define the 1-qubit Hadamard gate (denoted H), phase gate (denoted S), and "$\pi/8$ gate" (denoted T) respectively as

$$\mathrm{H} := \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}, \quad \mathrm{S} := \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}, \qquad \text{and} \quad \mathrm{T} := \begin{bmatrix}1 & 0\\ 0 & e^{i\pi/4}\end{bmatrix}, \tag{54}$$

and recall the definition of CNOT from Equation (51). Now let P_4' be the closure of $\{\mathrm{H}_{(j)} : j \in [4]\} \cup \{\mathrm{S}_{(j)} : j \in [4]\} \cup \{\mathrm{T}_{(j)} : j \in [4]\} \cup \{\mathrm{CNOT}_{(i,j)} : i, j \in [4], i \neq j\}$ under inverses and negations. It is clear that P_4' is a finite set of elements of $\mathrm{U}(2^4)$, closed under inverses, satisfying (A). The fact that P_4' satisfies (B) follows from the well-known fact (see e.g. [NC10, Sec. 4.5.3]) that H, S, T and CNOT together are universal for quantum computing. Finally, we obtain the desired set of elements $P_4 \subset \mathrm{SU}(2^4)$ by multiplying elements of P_4' by suitable complex values of unit norm to have determinant one.

4 Lower bounding τ_m for large m

In this section we prove Theorem 3.3, restated below, using simplifications of techniques introduced in [HHJ21]:

Theorem 4.1 (Restatement of Theorem 3.3)**.** *Let the sequence of groups $(\mathrm{G}(n))_{n\geq 1}$ be either $(\mathrm{SO}(2^n))_{n\geq 1}$ or $(\mathrm{SU}(2^n))_{n\geq 1}$. Define the following operators on $(\mathbb{C}^{2k})^{\otimes m}$:*

$$\Pi^{(m)} = \mathop{\mathbf{E}}_{\boldsymbol{g}\sim \mathrm{G}(m)}[\rho_{2^m}^{k,k}(\boldsymbol{g})], \quad \Pi_{[m]\setminus i} \otimes \mathbb{1}_i = (\mathbb{1}_{2k\times 2k} \text{ on the } i\text{th tensor factor, } \Pi^{(m-1)} \text{ on the remainder}). \tag{55}$$

Then for all $k \leq \frac{1}{\sqrt{10}m^2} 2^{m/2}$ we have

$$\left\| \mathop{\mathrm{avg}}_{i=1}^{m} \{\Pi_{[m]\setminus i} \otimes \mathbb{1}_i\} - \Pi^{(m)} \right\|_{\mathrm{op}} \leq \frac{1}{m} + \frac{\sqrt{10}km}{2^{m/2}}; \tag{56}$$

equivalently, in the notation of Theorem 3.3, $\tau_m \geq 1 - (\frac{1}{m} + \frac{\sqrt{10}km}{2^{m/2}})$.

We observe:

Fact 4.2. *$\mathrm{Im}\,\Pi^{(m)}$ is a subspace of $\mathrm{Im}(\Pi_{[m]\setminus i} \otimes \mathbb{1}_i)$ for all i.*

4.1 Identifying the projectors

To prove Theorem 4.1, we will need to have a description of the projection operator $\Pi^{(m)}$; luckily, this is provided by known representation theory. To state the results we need some notation.

Notation 4.3. If $X \in \mathbb{C}^{r\times r}$ is a matrix, we write $\mathrm{vec}(X) \in \mathbb{C}^r \otimes \mathbb{C}^r$ for its vectorization; here vec is the linear map that takes $|i\rangle\langle j|$ to $|ij\rangle$.

Fact 4.4. *For matrices $R_0, R_1, S \in \mathbb{C}^{r\times r}$ it holds that $(R_0 \otimes R_1)\mathrm{vec}(S) = \mathrm{vec}(R_0 S R_1^\top)$.*

Notation 4.5. Having fixed some $D = 2^m \in \mathbb{N}^+$, we write

$$|\Phi\rangle = D^{-1/2} \sum_{a=1}^{D} |a\rangle \otimes |a\rangle = D^{-1/2}\mathrm{vec}(\mathbb{1}_{D\times D}) \tag{57}$$

for the maximally entangled state on $\mathbb{C}^D \otimes \mathbb{C}^D$.

Notation 4.6. For $k \in \mathbb{N}^+$, let $\mathcal{M}_{2k}$ denote the set of all perfect matchings on $[2k]$, and let $\mathcal{M}_{2k}^{\mathrm{bip}}$ denote the subset of all "bipartite" perfect matchings, meaning that each pair in the matching can be written as $\{i, j\}$ with $i \leq k$ and $j > k$.

Notation 4.7. For $M \in \mathcal{M}_{2k}$, we introduce the unit vector

$$|\Phi_M\rangle = \bigotimes_{\{i,j\}\in M} |\Phi\rangle_{ij} \in (\mathbb{C}^D)^{\otimes 2k}, \tag{58}$$

where we abuse notation slightly by writing $|\Phi\rangle_{ij}$ for the maximally entangled state on the ith and jth tensor components.

Let us give two examples. First, with $k = 3$:

$$M = \{\{1,2\},\{3,6\},\{4,5\}\} \implies |\Phi_M\rangle = D^{-k/2} \sum_{a,b,c=1}^{D} |aabccb\rangle = D^{-k/2} \cdot \sum_{\substack{\chi:[2k]\to[D] \\ \text{all edges of } M \text{ monochromatic} \\ \text{for vertex-coloring } \chi}} |\chi\rangle\,. \tag{59}$$

As a second example, with general k:

$$M_0 = \{\{1,k+1\},\{2,k+2\},\ldots,\{k,2k\}\} \implies |\Phi_{M_0}\rangle = D^{-k/2}\mathrm{vec}(\mathbb{1}_{D^k\times D^k}). \tag{60}$$

It is not hard to show that every $|\Phi_M\rangle$ with $M \in \mathcal{M}_{2k}$ (respectively, $M \in \mathcal{M}_{2k}^{\mathrm{bip}}$) is fixed by every $\rho_D^{k,k}(g)$ for $g \in \mathrm{SO}(D)$ (respectively, $g \in \mathrm{SU}(D)$). To illustrate this for the particular $M_0 \in \mathcal{M}_{2k}^{\mathrm{bip}} \subseteq \mathcal{M}_{2k}$ from Equation (60), we have that for $g \in \mathrm{SO}(D) \leq \mathrm{SU}(D)$,

$$g^{\otimes k} \otimes \overline{g}^{\otimes k}\, |\Phi_{M_0}\rangle = \frac{g^{\otimes k} \otimes \overline{g}^{\otimes k}\mathrm{vec}(\mathbb{1}_{D^k\times D^k})}{D^{k/2}} = \frac{\mathrm{vec}(g^{\otimes k}\mathrm{vec}(\mathbb{1}_{D^k\times D^k})(\overline{g}^{\otimes k})^\top)}{D^{k/2}} = \frac{\mathrm{vec}(\mathbb{1}_{D^k\times D^k})}{D^{k/2}} = |\Phi_{M_0}\rangle\,, \tag{61}$$

where we used Fact 4.4 and $\overline{g}^\top = g^\dagger = g^{-1}$. Given this fact, each $|\Phi_M\rangle$ must be fixed by the average representation $\Pi^{(m)}$, and thus be in $\mathrm{Im}\,\Pi^{(m)}$. On the other hand, it is elementary to show (e.g., [BC20, Prop. 1]) that $\mathrm{Im}\,\Pi^{(m)}$ is *precisely* the set of vectors fixed by every operator in $\{\rho_D^{k,k}(g) : g \in \mathrm{G}(m)\}$ (recall that $D = 2^m$). In turn, these are precisely the vectorizations of all matrices in the *commutant* (centralizer) of $\mathcal{A} = \{g^{\otimes k} : g \in \mathrm{G}(m)\}$. Finally, the commutants of tensor product representations of our groups have been identified under the umbrella of *Schur–Weyl duality*.

Theorem 4.8. *By Schur–Weyl duality for* $\mathrm{U}(D)$ *[Sch01, Wey39, Yua12],* $D = 2^m$*, when* $\mathrm{G}(m) = \mathrm{U}(2^m)$ *the projector* $\Pi^{(m)}$ *has image equal to the span of* $|\Phi_M\rangle$ *for* $M \in \mathcal{M}_{2k}^{\mathrm{bip}}$*. The same is true when* $\mathrm{G}(m) = \mathrm{SU}(2^m)$*, since* $\Pi^{(m)}$ *is unchanged in this case.*[9]

By Schur–Weyl duality for $\mathrm{SO}(D)$ *[Bra37, Gro99],* $D = 2^m$*, when* $\mathrm{G}(m) = \mathrm{SO}(2^m)$ *and* $k < 2^{m-1}$ *the projector* $\Pi^{(m)}$ *has image equal to the span of* $|\Phi_M\rangle$ *for* $M \in \mathcal{M}_{2k}$*.*

Remark 4.9. The condition $k < 2^{m-1}$ in the previous theorem cannot be dropped. For example,

$$\mathop{\mathbf{E}}_{\boldsymbol{g}\sim\mathrm{O}(2)}[\rho_2^{1,1}(\boldsymbol{g})] = \text{projection onto } \tfrac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \tag{62}$$

but

$$\mathop{\mathbf{E}}_{\boldsymbol{g}\sim\mathrm{SO}(2)}[\rho_2^{1,1}(\boldsymbol{g})] = \text{projection onto } \mathrm{span}\{\tfrac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \tfrac{1}{\sqrt{2}}(|01\rangle - |10\rangle)\}. \tag{63}$$

We have now identified a spanning set for $\mathrm{Im}\,\Pi^{(m)}$, but working with it is complicated by the fact that it is not an orthonormal basis. It is, however, relatively "close" to being so, as we now show (following and simplifying some arguments from [BHH16, Lem. 17] and [HHJ21, Lem. 9]). First, an elementary lemma in linear algebra:

[9] Observe that because of the conjugation in the definition of $\rho_{2^n}^{k,k}$, the expectation $\Pi^{(m)}$ is the same whether the expectation is taken over $\boldsymbol{g} \sim \mathrm{G}(n) = \mathrm{SU}(2^n)$ or $\boldsymbol{g} \sim \mathrm{G}(n) = \mathrm{U}(2^n)$.

Lemma 4.10. *Let $W \in \mathbb{C}^{d\times t}$ have unit vector columns $|w_1\rangle, \ldots, |w_t\rangle$, and suppose their Gram matrix $W^\dagger W \in \mathbb{C}^{t\times t}$ is close to the identity, in the sense that $E = W^\dagger W - \mathbb{1}$ has $\|E\|_{\mathrm{op}} \le \kappa < 1$. (For example, this would hold if*

$$\|E\|_{1\mapsto 1} = \max_{j\in[t]} \sum_{i\neq j} |\langle w_i|w_j\rangle| \le \kappa, \tag{64}$$

since generally $\|E\|_{1\mapsto 1} \ge \rho(E) = \|E\|_{\mathrm{op}}$, as E is Hermitian.) Then $WW^\dagger = \sum_i |w_i\rangle\langle w_i|$ satisfies

$$WW^\dagger \overset{\kappa}{\approx} \Pi_T, \tag{65}$$

where Π_T is the projector onto $T = \mathrm{span}\{|w_1\rangle, \ldots, |w_t\rangle\}$, and $X \approx^\kappa Y$ denotes $\|X - Y\|_{\mathrm{op}} \le \kappa$.

Proof. By hypothesis, all eigenvalues λ of $W^\dagger W$ satisfy $|\lambda - 1| \le \kappa < 1$. Hence $WW^\dagger$ also has these t (nonzero) λ's within κ of 1 as eigenvalues (associated to eigenvectors in T), plus possibly additional eigenvalues of 0 (outside T). This confirms Inequality (65). □

Theorem 4.11. *In the setting of $\mathrm{G}(m) = \mathrm{SO}(D)$, $D = 2^m$ and provided $k^2 \le \frac{1}{9}D$, we have*

$$\sum_{M\in\mathcal{M}_{2k}} |\Phi_M\rangle\langle\Phi_M| \overset{\kappa_m}{\approx} \Pi^{(m)}, \tag{66}$$

where $\kappa_m := \frac{10}{9}\frac{k^2}{D}$. In the setting of $\mathrm{G}(m) = \mathrm{SU}(2^m)$, the same is true with $\mathcal{M}_{2k}$ replaced by $\mathcal{M}_{2k}^{\mathrm{bip}}$ (and one could replace κ_m by $\frac{5}{9}\frac{k^2}{D}$, but we won't).

Proof. The result for $\mathrm{U}(2^m)$ (hence $\mathrm{SU}(2^m)$) appears in [BHH16], and for $\mathrm{O}(2^m)$ in [HHJ21], but we present here a representation theory-free proof, focusing on the $\mathrm{SO}(2^m)$ case.

We will employ Lemma 4.10, with the $|w_i\rangle$'s being the $|\Phi_M\rangle$'s, $M \in \mathcal{M}_{2k}$. In particular, we will establish the premise in Inequality (64) with $\kappa = \kappa_m$. By symmetry of all matchings in $\mathcal{M}$, the quantity inside the maximum is the same for every "$|w_j\rangle$"; thus, we need only bound it for one particular choice, say the M_0 from Equation (60). Thus we need to establish

$$\sum_{M\in\mathcal{M}_{2k}} |\langle\Phi_M|\Phi_{M_0}\rangle| = 1 + \sum_{M\neq M_0} |\langle\Phi_M|\Phi_{M_0}\rangle| \le 1 + \kappa_m. \tag{67}$$

In computing $\langle\Phi_M|\Phi_{M_0}\rangle$, it is easy to see (e.g., from Equation (59)) we get a contribution of D^{-k} from every vertex-coloring $\chi : [2k] \to [D]$ that makes all edges of M and M_0 monochromatic. Since $M \cup M_0$ is a union of cycles, this is equivalent to a contribution of $D^{\mathrm{cc}(M\cup M_0)}$, where $\mathrm{cc}(\cdot)$ denotes the number of connected components. Thus (cf. [HHJ21, (B10)])

$$D^k \cdot \sum_{\text{matchings } M} |\langle\Phi_M|\Phi_{M_0}\rangle| = D^k \cdot \sum_{\text{matchings } M} \langle\Phi_M|\Phi_{M_0}\rangle = \sum_M D^{\mathrm{cc}(M\cup M_0)}. \tag{68}$$

The summation on the right is just the generating function (with "indeterminate" D) for the number of connected components obtained when placing a matching (initially: M) onto the endpoints of k labeled paths (initially: M_0). But this is a very simple exercise. Take the first labeled path, with endpoints x, y, and consider the vertex z to which x is matched. There are $2k-1$ possibilities for z, with one of them ($z = y$) increasing the component count by 1, and the other $2k-2$ increasing the count by 0. Thus the generating function picks up a factor of $(D^1 + (2k-2)\cdot D^0)$, and we reduce k to $k-2$. We conclude that (cf. [HHJ21, (B12)])

$$\sum_M D^{\mathrm{cc}(M\cup M_0)} = (D + (2k-2))(D + (2k-4))\cdots(D+2)D \tag{69}$$

and hence

$$\sum_M |\langle\Phi_M|\Phi_{M_0}\rangle| = (1)\big(1+\tfrac{2}{D}\big)\big(1+\tfrac{4}{D}\big)\cdots\big(1+\tfrac{2k-2}{D}\big) \le \exp\big(\tfrac{k(k-1)}{D}\big) \le 1 + \tfrac{10}{9}\tfrac{k^2}{D} = 1 + \kappa_m, \tag{70}$$

the last inequality holding because we have assumed $k^2 \le \frac{1}{9}D$. Thus we have indeed verified Inequality (67).

The case of $\mathrm{G}(m) = \mathrm{U}(2^m)$ is similar; we just need to compute the generating function for bipartite matchings, meaning $\mathcal{M}_{2k}^{\mathrm{bip}}$ replaces $\mathcal{M}$. The bound for κ_m becomes $(1)(1+\frac{1}{D})(1+\frac{2}{D})\cdots(1+\frac{k-1}{D}) - 1$, which is only smaller (by a factor of about $\frac{1}{2}$). □

4.2 Proof of Theorem 4.1

In this section we establish Theorem 4.1. We begin by proving some general facts about projectors that are nearly orthogonal to each other.

Lemma 4.12. *Let $P_1, \dots, P_m$ be orthogonal projections, and write $A = \operatorname{avg}_{i=1}^m \{P_i\}$. Then*

$$\|P_i P_j\|_{\mathrm{op}} \le \epsilon \ \ \forall\, i \ne j \quad \Longrightarrow \quad \|A\|_{\mathrm{op}} \le \tfrac{1}{m} + \min\{\sqrt{\epsilon}, m\epsilon\}. \tag{71}$$

Proof. We have

$$A^2 = \frac{1}{m}A + \frac{1}{m^2}\sum_{i\ne j} P_i P_j; \quad \Longrightarrow \quad \|A\|_{\mathrm{op}}^2 \le \frac{1}{m}\|A\|_{\mathrm{op}} + \frac{m(m-1)}{m^2}\epsilon \le \frac{1}{m}\|A\|_{\mathrm{op}} + \epsilon. \tag{72}$$

Solving the quadratic inequality yields $\|A\|_{\mathrm{op}} \le \frac{1}{2m} + \sqrt{\frac{1}{4m^2} + \epsilon}$, from which the result follows. □

Corollary 4.13. *In the setting of Lemma 4.12, let P be an orthogonal projection with $\operatorname{Im} P \le \operatorname{Im} P_i$ for all i. Then Inequality (71) holds with each instance of P_i replaced by $\widetilde{P}_i = P_i - P$.*

Proof. It suffices to note that $\widetilde{P}_i^2 = \widetilde{P}_i$, since $P_i \cdot P = P \cdot P_i = P$. □

Remark 4.14. The identity used in the proof easily extends to $\widetilde{P}_{i_1}\widetilde{P}_{i_2}\cdots\widetilde{P}_{i_k} = P_{i_1}P_{i_2}\cdots P_{i_k} - P$. Also, this identity remains true if any set of tildes is removed from the LHS (except for the set of all k).

Let us now study the particular orthogonal projectors involved in Theorem 4.1. We wish to employ Corollary 4.13 with

$$P_i \coloneqq \Pi_{[m]\setminus i} \otimes \mathbb{1}_i, \quad i = 1 \dots m, \qquad P \coloneqq \Pi^{(m)}. \tag{73}$$

Fact 4.2 tells us Corollary 4.13's hypothesis is satisfied. We thus obtain

$$\left\| \operatorname*{avg}_{i=1}^m \{P_i\} - \Pi^{(m)} \right\|_{\mathrm{op}} \le \frac{1}{m} + \min\{\sqrt{\epsilon}, m\epsilon\}, \quad \text{for } \epsilon = \max_{i\ne j}\left\{ \left\| \widetilde{P}_i \widetilde{P}_j \right\|_{\mathrm{op}} \right\}. \tag{74}$$

By symmetry of the m tensor factors, we have $\epsilon = \left\| \widetilde{P}_1 \widetilde{P}_m \right\|_{\mathrm{op}}$, and hence

$$\epsilon^2 = \left\| (\widetilde{P}_1\widetilde{P}_m)^\dagger (\widetilde{P}_1\widetilde{P}_m) \right\|_{\mathrm{op}} = \left\| P_m \widetilde{P}_1 P_m \right\|_{\mathrm{op}}, \tag{75}$$

where we used Remark 4.14 to get $\widetilde{P}_m\widetilde{P}_1\widetilde{P}_1\widetilde{P}_m = P_m\widetilde{P}_1 P_m$.

Our goal will be to use Theorem 4.11 (recall its κ_m notation) to establish the following:

$$\textbf{Claim:} \quad \epsilon^2 = \left\| P_m \widetilde{P}_1 P_m \right\|_{\mathrm{op}} \le \kappa_{m-2} + 2\kappa_{m-1} + \kappa_m \tag{76}$$

$$= \tfrac{10}{9}k^2(2^{2-m} + 2\cdot 2^{1-m}\cdot 2^{-m}) = 10k^2 2^{-m} =: \delta. \tag{77}$$

We will apply Theorem 4.11 for $m-2, m-1, m$; its hypothesis will be satisfied even for $m-2$, since we have $k^2 \le \frac{1}{9}2^{m-2}$ by virtue of the assumption $k^2 \le \frac{1}{10m^4}2^m$ in the theorem we're proving. Moreover, this assumption implies that $\delta^{1/4} \le 1/m$, meaning that Inequality (74) gives us the bound

$$\left\| \operatorname*{avg}_{i=1}^m \{P_i\} - \Pi^{(m)} \right\|_{\mathrm{op}} \le \frac{1}{m} + \min\{\delta^{1/4}, m\delta^{1/2}\} = \frac{1}{m} + m\delta^{1/2} = \frac{1}{m} + \frac{\sqrt{10}km}{2^{m/2}}, \tag{78}$$

verifying Inequality (56) and completing the proof of Theorem 4.1. Thus it remains to establish Inequality (76).

To establish the claim, let us write $\mathcal{M}$ for either $\mathcal{M}_{2k}$ or $\mathcal{M}_{2k}^{\text{bip}}$ (depending on $\mathrm{G}(m)$); and, for $M \in \mathcal{M}$ let us write

$$J_M = |\phi_M\rangle\langle\phi_M|\,, \quad \text{where } |\phi_M\rangle \text{ is the } D = 2 \text{ case of } |\Phi_M\rangle \text{ from Notations 4.5 and 4.7.} \tag{79}$$

Then (up to tensor factoring reordering) we Have $J_M^{\otimes m} = |\Phi_M\rangle$, and hence Theorem 4.11 tells us

$$\sum_{M\in\mathcal{M}} J_M^{\otimes m} \overset{\kappa_m}{\approx} \Pi^{(m)}. \tag{80}$$

We will also use this to derive

$$\sum_{M\in\mathcal{M}} J_M^{\otimes(m-1)} \overset{\kappa_{m-1}}{\approx} \Pi^{(m-1)} \quad \Longrightarrow \quad \sum_{M\in\mathcal{M}} \mathbb{1}_1 \otimes J_M^{\otimes(m-1)} \overset{\kappa_{m-1}}{\approx} P_1, \tag{81}$$

where the implication is by tensoring with $\mathbb{1}_1$ (which doesn't change operator norm differences). Using Inequality (80) again, and the triangle inequality, we reach

$$\widetilde{P}_1 = P_1 - P \overset{\kappa_{m-1}+\kappa_m}{\approx} \sum_{M\in\mathcal{M}} \mathbb{1}_1 \otimes J_M^{\otimes(m-1)} - \sum_{M\in\mathcal{M}} J_M^{\otimes m} = \sum_{M\in\mathcal{M}} \overline{J}_M \otimes J_M^{\otimes(m-1)}, \tag{82}$$

where $\overline{J}_M \coloneqq \mathbb{1} - J_M$. Since $\|P_m\|_{\text{op}} \leq 1$, we can further conclude

$$P_m \widetilde{P}_1 P_m \overset{\kappa_{m-1}+\kappa_m}{\approx} P_m \left(\sum_{M\in\mathcal{M}} \overline{J}_M \otimes J_M^{\otimes(m-1)} \right) P_m \tag{83}$$

$$= (\Pi^{(m-1)} \otimes \mathbb{1}_m) \left(\sum_{M\in\mathcal{M}} \overline{J}_M \otimes J_M^{\otimes(m-2)} \otimes J_M \right) (\Pi^{(m-1)} \otimes \mathbb{1}_m) \tag{84}$$

$$= \sum_{M\in\mathcal{M}} \left(\Pi^{(m-1)} (\overline{J}_M \otimes J_M^{\otimes(m-2)}) \Pi^{(m-1)} \right) \otimes J_M. \tag{85}$$

Writing

$$Z_M \coloneqq \Pi^{(m-1)} (\overline{J}_M \otimes J_M^{\otimes(m-2)}) \Pi^{(m-1)}, \tag{86}$$

we can put Inequality (85) into Equation (75) to obtain

$$\epsilon^2 \leq \kappa_{m-1} + \kappa_m + \left\| \sum_{M\in\mathcal{M}} Z_M \otimes J_M \right\|_{\text{op}}. \tag{87}$$

Now Z_M is PSD, being a conjugation (by $\Pi^{(m-1)}$) of a PSD matrix: the tensor product of projections J_M and $\overline{J}_M$. Since $0 \leq J_M \leq \mathbb{1}$, we therefore conclude $0 \leq Z_M \otimes J_M \leq Z_M \otimes \mathbb{1}_m$. Summing this over M yields

$$0 \leq \sum_{M\in\mathcal{M}} Z_M \otimes J_M \leq \sum_{M\in\mathcal{M}} Z_M \otimes \mathbb{1}_m = \left(\sum_{M\in\mathcal{M}} Z_M \right) \otimes \mathbb{1}_m, \tag{88}$$

and hence (from Inequality (87))

$$\epsilon^2 \leq \kappa_{m-1} + \kappa_m + \left\| \sum_{M\in\mathcal{M}} Z_M \right\|_{\text{op}} = \kappa_{m-1} + \kappa_m + \left\| \Pi^{(m-1)} \left(\sum_{M\in\mathcal{M}} \overline{J}_M \otimes J_M^{\otimes(m-2)} \right) \Pi^{(m-1)} \right\|_{\text{op}}. \tag{89}$$

We have effectively now reduced from m tensor components to $m-1$. Indeed, suppose we had defined the "$m-1$" analogues of $P_1, P_2, \dots$ and P, calling them $P_1^{(m-1)}, P_2^{(m-1)}, \dots$ and $P^{(m-1)} = \Pi^{(m-1)}$. Then Inequality (82) would tell us

$$\widetilde{P}_1^{(m-1)} = P_1^{(m-1)} - P^{(m-1)} \overset{\kappa_{m-2}+\kappa_{m-1}}{\approx} \sum_{M \in \mathcal{M}} \overline{J}_M \otimes J_M^{\otimes(m-2)}, \tag{90}$$

and putting this into Inequality (89) (using $\left\|P^{(m-1)}\right\|_{\mathrm{op}} \le 1$) yields

$$\epsilon^2 \le \kappa_{m-2} + 2\kappa_{m-1} + \kappa_m + \left\| P^{(m-1)} \widetilde{P}_1^{(m-1)} P^{(m-1)} \right\|_{\mathrm{op}}. \tag{91}$$

But $P^{(m-1)} \widetilde{P}_1^{(m-1)} P^{(m-1)}$ is in fact 0! (In the notation of Corollary 4.13 this would be "$P \cdot \widetilde{P}_1 \cdot P = 0$".) Thus we have established the claim, Inequality (76).

5 Lower bounding τ_m for small m

In this section we prove Theorem 3.4, restated below:

Theorem 5.1 (Restatement of Theorem 3.4)**.** *Let the sequence of groups $(\mathrm{G}(n))_{n\ge1}$ be either $(\mathrm{SO}(2^n))_{n\ge1}$ or $(\mathrm{SU}(2^n))_{n\ge1}$. For any $m \ge 4$ we have that*

$$\forall k \in \mathbb{N}^+, \qquad \left\| \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(m-1)^{\times \binom{[m]}{m-1}}} [\rho_{2^m}^{k,k}(\boldsymbol{g})] - \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(m)} [\rho_{2^m}^{k,k}(\boldsymbol{g})] \right\|_{\mathrm{op}} \le \left(1 - (1 - \tfrac{1}{m}) \tfrac{1-2^{2-m}}{4-2^{3-m}}\right)^{1/4} \le .96; \tag{92}$$

equivalently, in the notation of Theorem 3.4, $\tau_m \ge .04$.

5.1 Metrics

As discussed in Section 3, for $\mathrm{G}(m) = \mathrm{SO}(2^m)$ or $\mathrm{G}(m) = \mathrm{SU}(2^m)$ we have that $\mathrm{G}(m) \subseteq \mathrm{U}(2^m)$ is a compact connected Lie group with associated Lie algebra $\mathfrak{g}_m$, where

$$\text{for } \mathrm{G}(m) = \mathrm{SO}(2^m),\ \mathfrak{g}_m = \{H \in \mathbb{R}^{2^m \times 2^m} : H \text{ skew-symmetric}\}, \tag{93}$$

$$\text{for } \mathrm{G}(m) = \mathrm{SU}(2^m),\ \mathfrak{g}_m = \{H \in \mathbb{C}^{2^m \times 2^m} : H \text{ skew-Hermitian},\ \operatorname{tr} H = 0\}. \tag{94}$$

As per [Tao14, Prop. 2.11.1], $\mathrm{G}(m)$ can be given the structure of a Riemannian manifold with a bi-invariant metric. Moreover, $\mathrm{G}(m)$ is totally geodesic within $\mathrm{U}(2^m)$, hence the exponential map $\exp : \mathfrak{g}_m \to \mathrm{G}(m)$ is surjective and Riemannian distance d_{Rie} within $\mathrm{G}(m)$ coincides with Riemannian distance within $\mathrm{U}(2^m)$. This distance can be computed straightforwardly (see, e.g., [Mec19, within Lem. 1.3]), as follows:

- The Riemannian distance is bi-invariant, so $d_{\mathrm{Rie}}(X, Y) = d_{\mathrm{Rie}}(\mathbb{1}, Z)$ for $Z = YX^{-1}$.
- Given $Z \in \mathrm{G}(m)$, we can choose a unique $H \in \mathfrak{g}_m$ with $\exp(H) = Z$ such that the eigenvalues of H are of the form $\mathrm{i}\theta_j$ for $\theta_j \in (-\pi, \pi]$. We write $H = \log Z$ for this choice of H.
- Then $d_{\mathrm{Rie}}(\mathbb{1}, Z) = \|H\|_{\mathrm{Fro}} = (\sum_j \theta_j^2)^{1/2}$.

In other words,

$$d_{\mathrm{Rie}}(X, Y) = \|\log(YX^{-1})\|_{\mathrm{Fro}}. \tag{95}$$

For the sake of computation it will be convenient to work not just with the Riemannian distance d_{Rie} on $\mathrm{G}(m)$, but also the (very similar) Frobenius distance d_{Fro}, where $d_{\mathrm{Fro}}(X, Y)$ denotes $\|X - Y\|_{\mathrm{Fro}}$. In the above setup, now using bi-invariance of d_{Fro}, we evidently have

$$d_{\mathrm{Fro}}(X, Y) = \|\mathbb{1} - Z\|_{\mathrm{Fro}} = \left(\textstyle\sum_j |1 - \exp(\mathrm{i}\theta_j)|^2\right)^{1/2} = \left(\textstyle\sum_j (2\sin(\theta_j/2))^2\right)^{1/2}. \tag{96}$$

For some constant $c < .4 \leq 1$ we have the following numerical inequality (for $|\theta| \leq \pi$):

$$(2\sin(\theta/2))^2 \leq \theta^2 \leq (2\sin(\theta/2))^2 + c(2\sin(\theta/2))^4. \tag{97}$$

Using just $c \leq 1$, we may conclude[10]

$$d_{\mathrm{Fro}}(X,Y)^2 \leq d_{\mathrm{Rie}}(X,Y)^2 \leq d_{\mathrm{Fro}}(X,Y)^2 + d_{\mathrm{Fro}}(X,Y)^4. \tag{98}$$

Finally, we will also use the operator-norm distance, $d_{\mathrm{op}}(X,Y) = \|X-Y\|_{\mathrm{op}}$, which satisfies $d_{\mathrm{op}}(X,Y) \leq d_{\mathrm{Fro}}(X,Y)$.

We now move on to considering (Borel) probability measures on metric spaces (always assumed to be complete and separable). First we recall some basic definitions:

Definition 5.2. A pair of jointly distributed random variables $(\boldsymbol{X}, \boldsymbol{Y})$ is a *coupling* of probability distributions ν_1, ν_2 if $\boldsymbol{X}$ (respectively, $\boldsymbol{Y}$) has marginal distribution ν_1 (respectively, ν_2).

Definition 5.3. On the metric space (M,d), the *L^p-Wasserstein distance* between two measures ν_1 and ν_2 is

$$W_{d,p}(\nu_1,\nu_2) = \inf\Big\{\mathbf{E}[d(\boldsymbol{X},\boldsymbol{Y})^p]^{1/p} \ : \ (\boldsymbol{X},\boldsymbol{Y}) \text{ is a coupling of } (\nu_1,\nu_2)\Big\}. \tag{99}$$

Notation 5.4. If ν is a probability measure on metric space M and K is a Markov transition kernel on M, we write $K^\ell\nu$ for the probability measure on M resulting from starting with probability measure ν and taking $\ell \in \mathbb{N}$ steps according to K.

5.2 Oliveira's theorem and its consequences

We now state a key result of Oliveira [Oli09] that says that on any *length space* (see e.g. [BH99]), L^2-Wasserstein local contraction implies global contraction. As we only need the result in the particular case of compact, connected Lie groups (which are finite-diameter complete Riemannian manifolds), we state it only in this simpler context:

Theorem 5.5. *(Implied by [Oli09, Thm. 3].) Let (M,d) be a finite-diameter complete Riemannian manifold, and let K be a Markov transition kernel on M satisfying the following:*

$$W_{d,2}(K\delta_X, K\delta_Y) \leq (\eta + o(1))d(X,Y), \quad \text{with respect to } d(X,Y) \to 0. \tag{100}$$

(Here δ_Z denotes the measure that puts all of its probability mass on $Z \in M$.) Then for all probability measures ν_1, ν_2 on M it holds that

$$W_{d,2}(K\nu_1, K\nu_2) \leq \eta \cdot W_{d,2}(\nu_1,\nu_2). \tag{101}$$

Iterating this yields the following:

Corollary 5.6. *In the setting of Theorem 5.5, for any $\ell \in \mathbb{N}^+$ we have*

$$W_{d,2}(K^\ell\nu_1, K^\ell\nu_2) \leq \eta^\ell \cdot W_{d,2}(\nu_1,\nu_2) \leq D\eta^\ell, \tag{102}$$

where D is an upper bound on the diameter of M.

We now specialize this corollary to the case where (M,d) is $(\mathrm{G}(m), d_{\mathrm{Rie}})$; combining it with Definition 5.3 and using also $W_{d_{\mathrm{op}},1} \leq W_{d_{\mathrm{Fro}},1} \leq W_{d_{\mathrm{Fro}},2} \leq W_{d_{\mathrm{Rie}},2}$, we may conclude:

Corollary 5.7. *Let $\mathrm{G}(m)$ be a compact connected Lie group, and let K be a Markov transition kernel on $\mathrm{G}(m)$ such that Inequality (100) holds for d_{Rie} with constant η. Then for any probability measures ν_1, ν_2 on $\mathrm{G}(m)$, and any $\ell \in \mathbb{N}^+$, there is a coupling $(\boldsymbol{X},\boldsymbol{Y})$ of the measures $K^\ell\nu_1, K^\ell\nu_2$ under which*

$$\mathbf{E}[\|\boldsymbol{X}-\boldsymbol{Y}\|_{\mathrm{op}}] \leq 2D\eta^\ell \tag{103}$$

(where D is a bound on the d_{Rie}-diameter of $\mathrm{G}(m)$, and the factor 2 accounts for the inf*).*

[10] Here we are clarifying slightly the deduction of [BHH16, eq. (112a)].

Our next step is to get rid of the coupling in Corollary 5.7. To do this, we first observe that the representation $\rho_{2^m}^{k,k}$ is uniformly continuous on $\mathrm{G}(m)$ with respect to the operator-norm distance. Concretely, from the identity

$$g_1 \otimes \cdots \otimes g_K - h_1 \otimes \cdots \otimes h_K = \sum_{i=1}^{K} g_1 \otimes \cdots \otimes g_{i-1} \otimes (g_i - h_i) \otimes h_{i+1} \otimes \cdots \otimes h_K \tag{104}$$

and $\|X\|_{\mathrm{op}}, \left\|\overline{X}\right\|_{\mathrm{op}} = 1$ for $X \in \mathrm{G}(m)$, as well as multiplicativity of d_{op} with respect to tensor products, we may conclude that

$$\left\|\rho_{2^m}^{k,k}(X) - \rho_{2^m}^{k,k}(Y)\right\|_{\mathrm{op}} \leq 2k\|X - Y\|_{\mathrm{op}} \tag{105}$$

for any $X, Y \in \mathrm{G}(m)$. Using this, as well as the triangle inequality for d_{op}, in Corollary 5.7 yields:

Corollary 5.8. *In the setting of Corollary 5.7,*

$$\left\| \mathop{\mathbf{E}}_{\boldsymbol{X} \sim K^\ell \nu_1} [\rho_{2^m}^{k,k}(\boldsymbol{X})] - \mathop{\mathbf{E}}_{\boldsymbol{Y} \sim K^\ell \nu_2} [\rho_{2^m}^{k,k}(\boldsymbol{Y})] \right\|_{\mathrm{op}} \leq 4kD\eta^\ell. \tag{106}$$

(Note that in contrast with Corollary 5.7, here Corollary 5.8 does not feature any coupling between $K^\ell \nu_1$ and $K^\ell \nu_2$.)

Now we further specialize by taking $\nu_1 = \delta_{\mathbb{1}}$ (the measure with all probability on the identity element $\mathbb{1} \in \mathrm{G}(m)$), taking ν_2 to be Haar measure, and specifying that

$$K \text{ arises from left-multiplying by a random } \boldsymbol{g} \sim \mathcal{P}, \tag{107}$$

where $\mathcal{P}$ is some symmetric probability distribution on $\mathrm{G}(m)$ as in Definition 2.3. Note that, whatever $\mathcal{P}$ is, we have $K^\ell \nu_2 = \nu_2$ (Haar measure), and

$$\mathop{\mathbf{E}}_{\boldsymbol{X} \sim K^\ell \nu_1} [\rho_{2^m}^{k,k}(\boldsymbol{X})] = \mathop{\mathbf{E}}_{\substack{\boldsymbol{g}_1, \ldots, \boldsymbol{g}_\ell \sim \mathcal{P} \\ \text{independent}}} [\rho_{2^m}^{k,k}(\boldsymbol{g}_\ell \cdots \boldsymbol{g}_1)] = \mathbf{E}[\rho_{2^m}^{k,k}(\boldsymbol{g}_\ell) \cdots \rho_{2^m}^{k,k}(\boldsymbol{g}_1)] = \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathcal{P}} [\rho_{2^m}^{k,k}(\boldsymbol{g})]^\ell. \tag{108}$$

From this and Corollary 5.8 we conclude the following:

Corollary 5.9. *Let $\mathcal{P}$ be a symmetric probability distribution on $\mathrm{G}(m)$. Given $X, Y \in \mathrm{G}(m)$, write $\mathcal{P}^{(X)}$ (respectively, $\mathcal{P}^{(Y)}$) for the distribution of $\boldsymbol{g}X$ (respectively, $\boldsymbol{g}Y$) when $\boldsymbol{g} \sim \mathcal{P}$. Then supposing*

$$W_{d_{\mathrm{Rie}},2}(\mathcal{P}^{(X)}, \mathcal{P}^{(Y)}) \leq (\eta + o(1)) d_{\mathrm{Rie}}(X, Y) \quad \text{with respect to } d_{\mathrm{Rie}}(X, Y) \to 0, \tag{109}$$

it follows that for any $\ell, k \in \mathbb{N}^+$ we have

$$\left\| \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathcal{P}} [\rho_{2^m}^{k,k}(\boldsymbol{g})]^\ell - \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(m)} [\rho_{2^m}^{k,k}(\boldsymbol{g})] \right\|_{\mathrm{op}} \leq 4kD \cdot \eta^\ell. \tag{110}$$

Our goal for the next section will be to establish the following:

Theorem 5.10. *Let ν_m denote the distribution $\mathrm{G}(m-1) \times \binom{[m]}{m-1}$ on $\mathrm{G}(m)$, thought of as inducing a Markov chain on $\mathrm{G}(m)$ via left-multiplication. Fix any $X, Y \in \mathrm{G}(m)$ with $d_{\mathrm{Rie}}(X, Y) = \epsilon \leq 1$, and let $\boldsymbol{X}''$ (respectively, $\boldsymbol{Y}''$) denote the result of taking* two *independent steps from X (respectively, Y) according to ν_m. Then there is a coupling of $\boldsymbol{X}'', \boldsymbol{Y}''$ under which*

$$\mathbf{E}[d_{\mathrm{Rie}}(\boldsymbol{X}'', \boldsymbol{Y}'')^2] \leq (1 - \gamma_m)\epsilon^2 + O_m(\epsilon^3), \tag{111}$$

where $\gamma_m = (1 - \frac{1}{m})\gamma_m'$ with $\gamma_m' = \frac{1 - 2^{2-m}}{4 - 2^{3-m}}$, and the $O_m(\cdot)$ hides a constant depending only on m.

This theorem establishes the hypothesis of Corollary 5.9 with $\mathcal{P} = \nu_m * \nu_m$ and $\eta = \sqrt{1 - \gamma_m}$. We can therefore easily derive the following (where the equality uses the fact that $\mathbf{E}_{\boldsymbol{g} \sim \mathrm{G}(m)} [\rho_{2^m}^{k,k}(\boldsymbol{g})]$ is a projection operator):

$$\left\| \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \nu_m} [\rho_{2^m}^{k,k}(\boldsymbol{g})]^{2\ell} - \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(m)} [\rho_{2^m}^{k,k}(\boldsymbol{g})] \right\|_{\mathrm{op}} = \left\| \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \nu_m} [\rho_{2^m}^{k,k}(\boldsymbol{g})] - \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(m)} [\rho_{2^m}^{k,k}(\boldsymbol{g})] \right\|_{\mathrm{op}}^{2\ell} \leq 4kD \cdot (1 - \gamma_m)^{\ell/2}. \tag{112}$$

Taking (2ℓ)th roots and then $\ell \to \infty$ thus yields Theorem 5.1.

5.3 Proof of Theorem 5.10

We begin by describing the needed coupling. First, we use the same randomness to take one step from each of X, Y; that is, we define

$$\boldsymbol{X}' = \boldsymbol{g}_{[m]\setminus \boldsymbol{i}} \cdot X, \qquad \boldsymbol{Y}' = \boldsymbol{g}_{[m]\setminus \boldsymbol{i}} \cdot Y, \tag{113}$$

where $\boldsymbol{i} \sim [m]$, $\boldsymbol{g} \sim \mathrm{G}(m-1)$ are uniformly random and independent. To take the second steps, we first draw $\boldsymbol{j} \sim [m]$. Then, based on the outcomes $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{g}$, we will deterministically define some

$$\boldsymbol{h} = h(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{g}) \in \mathrm{G}(m-1) \tag{114}$$

and then take

$$\boldsymbol{X}'' = (\widetilde{\boldsymbol{g}}\boldsymbol{h})_{[m]\setminus \boldsymbol{j}} \cdot \boldsymbol{X}', \qquad \boldsymbol{Y}'' = \widetilde{\boldsymbol{g}}_{[m]\setminus \boldsymbol{j}} \cdot \boldsymbol{Y}', \tag{115}$$

where $\widetilde{\boldsymbol{g}} \sim \mathrm{G}(m-1)$ is drawn uniformly and independently of all other random variables. This is a valid coupling, since for every outcome of $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{g}$ the distributions of $\widetilde{\boldsymbol{g}}\boldsymbol{h}$ and $\widetilde{\boldsymbol{g}}$ are identical. Then

$$d_{\mathrm{Rie}}(\boldsymbol{X}'', \boldsymbol{Y}'') = d_{\mathrm{Rie}}(\boldsymbol{h}_{[m]\setminus \boldsymbol{j}} \cdot \boldsymbol{X}', \boldsymbol{Y}'), \tag{116}$$

since $d_{\mathrm{Rie}}(\cdot, \cdot)$ is unitarily invariant. In case $\boldsymbol{i} = \boldsymbol{j}$, we will "give up" and simply define $\boldsymbol{h} = \mathbb{1}$, in which case we get $d_{\mathrm{Rie}}(\boldsymbol{X}'', \boldsymbol{Y}'') = d_{\mathrm{Rie}}(\boldsymbol{X}', \boldsymbol{Y}') = d_{\mathrm{Rie}}(X, Y) = \epsilon$. Thus we have

$$\mathbf{E}[d_{\mathrm{Rie}}(\boldsymbol{X}'', \boldsymbol{Y}'')^2] = \frac{1}{m}\epsilon^2 + \left(1 - \frac{1}{m}\right) \operatorname*{avg}_{\boldsymbol{i} \neq \boldsymbol{j}} \left\{ \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(m-1)} \left[d_{\mathrm{Rie}}\left(\boldsymbol{h}_{[m]\setminus \boldsymbol{j}} \cdot \boldsymbol{X}', \boldsymbol{Y}'\right)^2 \right] \right\}. \tag{117}$$

To complete the definition of $\boldsymbol{h}$, we specify the function h:

$$\text{for } i \neq j, \text{ we define } h = h(i, j, g) \text{ to minimize } d_{\mathrm{Fro}}\left(h_{[m]\setminus j} \cdot g_{[m]\setminus i} \cdot X, g_{[m]\setminus i} \cdot Y\right)^2; \tag{118}$$

in other words, $\boldsymbol{h} = h(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{g})$ minimizes $d_{\mathrm{Fro}}\left(\boldsymbol{h}_{[m]\setminus \boldsymbol{j}} \cdot \boldsymbol{X}', \boldsymbol{Y}'\right)$. With this choice of h, note that we have $d_{\mathrm{Fro}}\left(\boldsymbol{h}_{[m]\setminus \boldsymbol{j}} \cdot \boldsymbol{X}', \boldsymbol{Y}'\right) \leq \epsilon \leq 1$ for every outcome of $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{g}$, since $\boldsymbol{h} = \mathbb{1}$ is always an option (and $d_{\mathrm{Fro}}(\boldsymbol{X}', \boldsymbol{Y}') = d_{\mathrm{Fro}}(X, Y) \leq d_{\mathrm{Rie}}(X, Y) = \epsilon$). Thus employing Inequality (98) we may conclude

$$\mathbf{E}[d_{\mathrm{Rie}}(\boldsymbol{X}'', \boldsymbol{Y}'')^2] \leq \frac{1}{m}\epsilon^2 + \left(1 - \frac{1}{m}\right) \operatorname*{avg}_{\boldsymbol{i} \neq \boldsymbol{j}} \left\{ \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(m-1)} \left[d_{\mathrm{Fro}}\left(\boldsymbol{h}_{[m]\setminus \boldsymbol{j}} \cdot \boldsymbol{X}', \boldsymbol{Y}'\right)^2 \right] \right\} + \epsilon^4. \tag{119}$$

Thus to complete the proof of Theorem 5.10, it suffices to establish the following:

$$\forall i \neq j, \qquad \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(m-1)} \left[\min_{h \in \mathrm{G}(m-1)} \left\{ d_{\mathrm{Fro}}\left(h_{[m]\setminus j}, \boldsymbol{Y}' \cdot (\boldsymbol{X}')^{-1}\right)^2 \right\} \right] \leq (1 - \gamma'_m)\epsilon^2 + O_m(\epsilon^3). \tag{120}$$

(Here we used $d_{\mathrm{Fro}}\left(h_{[m]\setminus j} \cdot \boldsymbol{X}', \boldsymbol{Y}'\right) = d_{\mathrm{Fro}}\left(h_{[m]\setminus j}, \boldsymbol{Y}' \cdot (\boldsymbol{X}')^{-1}\right)$.) Our proof of this will not have any particular dependence on i, j, so without loss of generality let us fix $i = 1$ and $j = m$. We establish Inequality (120) via the below two lemmas. (Here and subsequently the notation "$\mathrm{tr}_i X$" below denotes the partial trace corresponding to tracing out the ith qubit of X.)

Lemma 5.11. *Fix any $Z \in \mathrm{G}(m)$ with $d_{\mathrm{Rie}}(\mathbb{1}, Z) = \epsilon$. Then*

$$\min_{h \in \mathrm{G}(m-1)} \left\{ d_{\mathrm{Fro}}(h \otimes \mathbb{1}, Z)^2 \right\} \leq (1 - \tfrac{1}{2}\|\mathrm{tr}_m B\|_{\mathrm{Fro}}^2)\epsilon^2 + O_m(\epsilon^3), \tag{121}$$

where $B = \frac{1}{\epsilon} \log Z \in \mathfrak{g}_m$ satisfies $\|B\|_{\mathrm{Fro}} = 1$.

Lemma 5.12. *For $m \geq 2$ and any $A \in \mathfrak{g}_m$, writing $\delta = 2^{2-m}$, we have*

$$\mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(m-1)} [\|\mathrm{tr}_m((\mathbb{1} \otimes \boldsymbol{g}_{[m]\setminus 1})A(\mathbb{1} \otimes \boldsymbol{g}_{[m]\setminus 1}^\dagger))\|_{\mathrm{Fro}}^2] \geq \frac{1 - \delta}{2 - \delta}\|A\|_{\mathrm{Fro}}^2. \tag{122}$$

To see how the above two lemmas imply Inequality (120) (in the case $i = 1$, $j = m$), we first apply Lemma 5.11 with Z being the outcome of $\boldsymbol{Y}' \cdot (\boldsymbol{X}')^{-1}$. Writing $\boldsymbol{B} = \frac{1}{\epsilon}\log(\boldsymbol{Y}'(\boldsymbol{X}')^{-1})$ (recall that from Equation (113) this is a random matrix depending on $\boldsymbol{g}$), Lemma 5.11 tells us that

$$\mathop{\mathbf{E}}_{\boldsymbol{g}\sim \mathrm{G}(m-1)}\left[\min_{h\in \mathrm{G}(m-1)}\left\{d_{\mathrm{Fro}}\big(h_{[m]\setminus j}\otimes \mathbb{1}_j, \boldsymbol{Y}'\cdot(\boldsymbol{X}')^{-1}\big)^2\right\}\right] \le (1-\tfrac{1}{2}\,\mathbf{E}[\|\operatorname{tr}_m \boldsymbol{B}\|_{\mathrm{Fro}}^2])\epsilon^2 + O_m(\epsilon^3). \tag{123}$$

But, for $\boldsymbol{g}\sim \mathrm{G}(m-1)$, we have

$$\boldsymbol{B} = \tfrac{1}{\epsilon}\log\Big((\mathbb{1}\otimes \boldsymbol{g}_{[m]\setminus 1})YX^{-1}(\mathbb{1}\otimes \boldsymbol{g}^\dagger_{[m]\setminus 1})\Big) = (\mathbb{1}\otimes \boldsymbol{g}_{[m]\setminus 1})\big(\tfrac{1}{\epsilon}\log(YX^{-1})\big)(\mathbb{1}\otimes \boldsymbol{g}^\dagger_{[m]\setminus 1}). \tag{124}$$

The result now follows by applying Lemma 5.12 with $A = \frac{1}{\epsilon}\log(YX^{-1})$, which has $\|A\|_{\mathrm{Fro}} = 1$ since $d_{\mathrm{Rie}}(X,Y) = \epsilon$.

5.3.1 Proof of Lemma 5.11

To prove Lemma 5.11, it suffices to show that the particular choice

$$h \coloneqq \exp(-\tfrac{1}{2}\epsilon \operatorname{tr}_m B) \tag{125}$$

satisfies Inequality (121). We observe that since $\mathrm{G}(m)$ is either $\mathrm{SO}(2^m)$ or $\mathrm{SU}(2^m)$, recalling Equations (93) and (94) we have that $\operatorname{tr}_m B \in \mathfrak{g}_{m-1}$ since $B \in \mathfrak{g}_m$ (note that $\operatorname{tr} B = 0$ implies $\operatorname{tr}(\operatorname{tr}_m B) = 0$), and hence indeed $h \in \mathrm{G}(m-1)$ as required. Now we must bound

$$d_{\mathrm{Fro}}(h\otimes\mathbb{1}, Z)^2 = \langle h\otimes\mathbb{1} - Z, h\otimes\mathbb{1} - Z\rangle = 2\operatorname{tr}\mathbb{1} - \langle h\otimes\mathbb{1}, Z\rangle - \langle Z, h\otimes\mathbb{1}\rangle = 2\operatorname{tr}\mathbb{1} - 2\Re\langle h\otimes\mathbb{1}, Z\rangle. \tag{126}$$

Recalling $Z = \exp(\epsilon B)$ where $\|B\|_{\mathrm{Fro}} = 1$, we abuse notation slightly by writing

$$Z = 1 + \epsilon B + \epsilon^2 B^2/2 + O_m(\epsilon^3), \tag{127}$$

where "$O_m(\epsilon^3)$" stands for some matrix E satisfying $\|E\|_{\mathrm{Fro}} \le C\epsilon^3$, with C a constant depending only on m that may change from line to line.

We may similarly expand $h\otimes\mathbb{1} = \exp(-\frac{1}{2}\epsilon\operatorname{tr}_m B)\otimes\mathbb{1}$, and upon substituting into Equation (126) and simplifying, we obtain

$$(126) = \Re\operatorname{tr}(T-2B)\epsilon + \Re\operatorname{tr}\big(TB - \tfrac{1}{4}T^2 - B^2\big)\epsilon^2 + O_m(\epsilon^3), \qquad T \coloneqq (\operatorname{tr}_m B)\otimes\mathbb{1}. \tag{128}$$

Now B, T are both in the Lie algebra for $\mathrm{G}(m)$; i.e., they are skew-symmetric in the case $\mathrm{G}(m) = \mathrm{SO}(2^m)$, and traceless skew-Hermitian in the case $\mathrm{G}(m) = \mathrm{SU}(2^m)$. Thus both have purely imaginary trace, meaning $\Re\operatorname{tr}(T-2B) = 0$. Moreover, B skew-Hermitian implies $\operatorname{tr} B^2 = \operatorname{tr} B(-B^\dagger) = -\langle B, B\rangle = -\|B\|_{\mathrm{Fro}}^2 = -1$, and similarly $\operatorname{tr} T^2 = -\|T\|_{\mathrm{Fro}}^2 = -2\|\operatorname{tr}_m B\|_{\mathrm{Fro}}^2$. Finally,

$$\operatorname{tr}(TB) = -\langle T, B\rangle = -\langle(\operatorname{tr}_m B)\otimes\mathbb{1}, B\rangle = -\langle \operatorname{tr}_m B, \operatorname{tr}_m B\rangle = -\|\operatorname{tr}_m B\|_{\mathrm{Fro}}^2. \tag{129}$$

Putting these deductions into Equation (128) yields

$$(126) = (1 - \tfrac{1}{2}\|\operatorname{tr}_m B\|_{\mathrm{Fro}}^2)\epsilon^2 + O_m(\epsilon^3), \tag{130}$$

completing the proof of Lemma 5.11.

5.3.2 Proof of Lemma 5.12

Given m qubits, we'll write $L = \{1,\dots,m-1\}$ for the system defined by the first $m-1$ of them, and (slightly abusing notation) write m for the system defined by the mth one. We will also write L' and m' for duplicate copies of these systems, and given a subset $S \subseteq [m]$, we write $\mathrm{SWAP}_{S,S'}$ to denote the

operator that swaps the qubits in S with the corresponding subset of qubits $1', \dots, m'$. Now for any $(m-1)$-qubit operator C we have

$$\|C\|_{\text{Fro}}^2 = \operatorname{tr}(C^\dagger C) = \operatorname{tr}((C_L^\dagger \otimes C_{L'}) \cdot \text{SWAP}_{L,L'}). \tag{131}$$

In turn, if $C = \operatorname{tr}_m B$ for some operator B on m qubits, we conclude

$$\|\operatorname{tr}_m B\|_{\text{Fro}}^2 = \operatorname{tr}(\operatorname{tr}_{m,m'}(B^\dagger \otimes B) \cdot \text{SWAP}_{L,L'}) = \operatorname{tr}((B^\dagger \otimes B) \cdot (\text{SWAP}_{L,L'} \otimes \mathbb{1}_{m,m'})). \tag{132}$$

Next, if $B = HAH^\dagger$ for unitary H, we may use the cyclic property of trace to conclude

$$\|\operatorname{tr}_m B\|_{\text{Fro}}^2 = \operatorname{tr}((A^\dagger \otimes A) \cdot W) = \langle A \otimes A^\dagger, W \rangle, \qquad W \coloneqq (H \otimes H)(\text{SWAP}_{L,L'} \otimes \mathbb{1}_{m,m'})(H^\dagger \otimes H^\dagger). \tag{133}$$

(The above formula, specialized to $m = 3$, essentially appears as [BHH16, Eqn. (103)].) Finally, suppose $H = \mathbb{1} \otimes g$ for some $(m-1)$-qubit unitary g. For notational clarity we break up the system L into subsystems "1" and $K = \{2, \dots, m-1\}$, writing $H = \mathbb{1}_1 \otimes g_{K,m}$ and

$$H \otimes H = \mathbb{1}_{1,1'} \otimes (g_{K,m} \otimes g_{K',m'}). \tag{134}$$

Putting this into the definition of W, we see that the two qubits labeled 1 and $1'$ are simply swapped by W, and we have

$$W = \text{SWAP}_{1,1'} \cdot \widehat{W}, \qquad \widehat{W} \coloneqq (g_{K,m} \otimes g_{K',m'}) S (g_{K,m}^\dagger \otimes g_{K',m'}^\dagger), \qquad S \coloneqq (\text{SWAP}_{K,K'} \otimes \mathbb{1}_{m,m'}). \tag{135}$$

Recalling Fact 4.4, we see that

$$\operatorname{vec}(\widehat{W}) = \rho_{2^{m-1}}^{2,2}(g) \cdot \operatorname{vec}(S). \tag{136}$$

In other words, $\widehat{W}$ is the action of g on S under representation $\rho_{2^{m-1}}^{2,2}$, when we suitably use the "matricized" interpretation of this representation. Finally, we are interested in the case that $\boldsymbol{g} \sim \mathrm{G}(m-1)$ is chosen "uniformly" (Haar measure on $\mathrm{G}(m-1)$); then we conclude from the above equations that

$$\mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(m-1)} [\|\operatorname{tr}_m((\mathbb{1} \otimes \boldsymbol{g}) A (\mathbb{1} \otimes \boldsymbol{g}^\dagger))\|_{\text{Fro}}^2] = \langle A \otimes A^\dagger, \text{SWAP}_{1,1'} \cdot S_0 \rangle, \qquad \operatorname{vec}(S_0) \coloneqq \mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(m-1)} [\rho_{2^{m-1}}^{2,2}(\boldsymbol{g})] \cdot \operatorname{vec}(S). \tag{137}$$

We now compute S_0 (we note that a similar calculation for $\mathrm{G}(m-1) = \mathrm{U}(2^{m-1})$ is given in [HHJ21, Eqn. (61)]):

Proposition 5.13. *Let $D = 2^{m-1}$, and define the following operators acting across systems $K \cup \{m\}$, $K' \cup \{m'\}$:*

$$Q_2 = D \cdot |\Phi\rangle\langle\Phi|, \quad Q_3 = \mathbb{1}, \quad Q_4 = \text{SWAP} \tag{138}$$

(where $|\Phi\rangle = D^{-1/2} \sum_{a \in \{0,1\}^{m-1}} |a\rangle \otimes |a\rangle$ is the maximally entangled state). Then

$$\mathrm{G}(m-1) = \mathrm{SO}(D) \implies S_0 = c_2 \cdot Q_2 + c_3 \cdot Q_3 + c_4 \cdot Q_4, \tag{139}$$

$$\mathrm{G}(m-1) = \mathrm{SU}(D) \implies S_0 = c_3' \cdot Q_3 + c_4' \cdot Q_4, \tag{140}$$

where the non-negative constants $c_2, c_3, c_4, c_3', c_4'$ are given by

$$c_2 = \frac{D/2 - 1}{(D-1)(D+2)}, \quad c_3 = \frac{3D/2 + 1}{(D-1)(D+2)}, \quad c_4 = \frac{(D/2-1)(D+3)}{(D-1)(D+2)}, \tag{141}$$

$$c_3' = \frac{3D/2}{(D-1)(D+1)}, \quad c_4' = \frac{D^2/2 - 2}{(D-1)(D+1)} \geq c_4. \tag{142}$$

Proof. We recall from Theorem 4.8 that[11]

$$\mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(m-1)} [\rho_{2^{m-1}}^{2,2}(\boldsymbol{g})] = \text{projection onto the span of } \{|\varphi_M\rangle : M \in \mathcal{M}\}, \tag{143}$$

[11] Note that here we are using $m \geq 4$.

where we use the following notation:

$$M_{12} = \{\{1,2\},\{3,4\}\}, \quad M_{13} = \{\{1,3\},\{2,4\}\}, \quad M_{14} = \{\{1,4\},\{2,3\}\}; \tag{144}$$

$$|\varphi_{M_{12}}\rangle = \mathrm{vec}(Q_2) = \sum_{x,y\in\{0,1\}^{m-1}} |x,x,y,y\rangle\,, \tag{145}$$

$$|\varphi_{M_{13}}\rangle = \mathrm{vec}(Q_3) = \sum_{x,y} |x,y,x,y\rangle\,, \quad |\varphi_{M_{14}}\rangle = \mathrm{vec}(Q_4) = \sum_{x,y} |x,y,y,x\rangle\,; \tag{146}$$

$$\mathrm{G}(m-1) = \mathrm{SO}(2^{m-1}) \implies \mathcal{M} = \{M_{12}, M_{13}, M_{14}\}, \qquad \mathrm{G}(m-1) = \mathrm{SU}(2^{m-1}) \implies \mathcal{M} = \{M_{13}, M_{14}\}. \tag{147}$$

Let us further define

$$|\psi_{10}\rangle = \sum_{x\in\{0,1\}^{m-1}} |x,x,x,x\rangle \quad \text{and} \quad |\psi_{1j}\rangle = |\varphi_{M_{1j}}\rangle - |\psi_{10}\rangle\,, \tag{148}$$

so that the $|\psi_{1j}\rangle$'s are pairwise orthogonal, with $\langle\psi_{10}|\psi_{10}\rangle = D$ and $\langle\psi_{1j}|\psi_{1j}\rangle = D(D-1)$ for $j > 1$. Then, since from Equation (135) we have

$$\mathrm{vec}(S) = \sum_{\substack{x=(x',a)\in\{0,1\}^{m-2}\times\{0,1\}\\ y=(y',b)\in\{0,1\}^{m-2}\times\{0,1\}}} |(x',a),(y',b),(y',a),(x',b)\rangle\,, \tag{149}$$

we can easily compute

$$\langle\psi_{10}|\,\mathrm{vec}(S) = D, \quad \langle\psi_{12}|\,\mathrm{vec}(S) = 0, \quad \langle\psi_{13}|\,\mathrm{vec}(S) = D, \quad \langle\psi_{14}|\,\mathrm{vec}(S) = D(D/2-1). \tag{150}$$

From this we conclude that the projection of $\mathrm{vec}(S)$ onto the span of the four $|\psi_{1j}\rangle$'s (which is also the span of $|\psi_{10}\rangle$ and the three $|\varphi_{M_{1j}}\rangle$'s) is

$$|\sigma\rangle \coloneqq |\psi_{10}\rangle + \frac{1}{D-1}\,|\psi_{13}\rangle + \frac{D/2-1}{D-1}\,|\psi_{14}\rangle\,. \tag{151}$$

Now one may easily verify that the following vector $|\tau\rangle$ is orthogonal to each $|\varphi_{M_{1j}}\rangle = |\psi_{10}\rangle + |\psi_{1j}\rangle$:

$$|\tau\rangle = -(D-1)\,|\psi_{10}\rangle + |\psi_{12}\rangle + |\psi_{13}\rangle + |\psi_{14}\rangle\,. \tag{152}$$

Thus we can bring $|\sigma\rangle$ into the span of the three $|\varphi_{M_{1j}}\rangle$'s by adding a suitable multiple of $|\tau\rangle$ to zero out the $|\psi_{10}\rangle$ component as follows:

$$|\sigma\rangle + c\,|\tau\rangle = (1-(D-1)c)\,|\psi_{10}\rangle + c\,|\psi_{12}\rangle + \left(\frac{1}{D-1}+c\right)|\psi_{13}\rangle + \left(\frac{D/2-1}{D-1}+c\right)|\psi_{14}\rangle \tag{153}$$

$$= \left(\frac{D/2-1}{D-1} - (D+2)c\right)|\psi_{10}\rangle + c\,|\varphi_{M_{12}}\rangle + \left(\frac{1}{D-1}+c\right)|\varphi_{M_{13}}\rangle + \left(\frac{D/2-1}{D-1}+c\right)|\varphi_{M_{14}}\rangle\,, \tag{154}$$

and taking $c = \frac{D/2-1}{(D-1)(D+2)}$ we finally get that

$$\text{the projection of } \mathrm{vec}(S) \text{ onto the span of } |\varphi_{M_{12}}\rangle, |\varphi_{M_{13}}\rangle, |\varphi_{M_{14}}\rangle \text{ is } c_2\,|\varphi_{M_{12}}\rangle + c_3\,|\varphi_{M_{13}}\rangle + c_4\,|\varphi_{M_{14}}\rangle\,. \tag{155}$$

One can repeat the above using $|\tau'\rangle = -(D-1)\,|\psi_{10}\rangle + |\psi_{13}\rangle + |\psi_{14}\rangle$ in place of $|\tau\rangle$ to similarly deduce

$$\text{the projection of } \mathrm{vec}(S) \text{ onto the span of } |\varphi_{M_{13}}\rangle, |\varphi_{M_{14}}\rangle \text{ is } c_3'\,|\varphi_{M_{13}}\rangle + c_4'\,|\varphi_{M_{14}}\rangle\,. \tag{156}$$

The proof is complete. □

Now we compute:

$$\langle A \otimes A^\dagger, \mathrm{SWAP}_{1,1'} \cdot Q_4 \rangle = \langle A \otimes A^\dagger, \mathrm{SWAP}_{[m],[m]'} \rangle = \|A\|_{\mathrm{Fro}}^2 \tag{157}$$

(similar to Equation (131)), and

$$\langle A \otimes A^\dagger, \mathrm{SWAP}_{1,1'} \cdot Q_3 \rangle = \langle A \otimes A^\dagger, \mathrm{SWAP}_{1,1'} \cdot \mathbb{1}_{[m]\setminus 1,[m]'\setminus 1'} \rangle = \| \operatorname{tr}_{[m]\setminus 1} A \|_{\mathrm{Fro}}^2 \geq 0 \tag{158}$$

(similar to Equation (132)). Finally, since (as can be easily verified)

$$\mathrm{SWAP}_{1,1'} \cdot Q_2 = \sum_{\substack{a,b \in \{0,1\} \\ x,y \in \{0,1\}^{m-1}}} |(a,x),(b,x)\rangle\langle(b,y),(a,y)|\,, \tag{159}$$

we may conclude that

$$\langle A \otimes A^\dagger, \mathrm{SWAP}_{1,1'} \cdot Q_2 \rangle = \sum_{a,b,x,y} \langle (b,y)|A^\dagger|(a,x)\rangle \, \langle (a,y)|A|(b,x)\rangle \geq -\|A\|_{\mathrm{Fro}}^2 \tag{160}$$

by Cauchy–Schwarz. Putting these conclusions together with Equation (137) and Proposition 5.13, we get that for both $\mathrm{G}(m-1) = \mathrm{SO}(2^{m-1})$ and $\mathrm{G}(m-1) = \mathrm{SU}(2^{m-1})$ it holds that

$$\mathop{\mathbf{E}}_{\boldsymbol{g} \sim \mathrm{G}(m-1)} [\|\operatorname{tr}_m((\mathbb{1} \otimes \boldsymbol{g}) A (\mathbb{1} \otimes \boldsymbol{g}^{-1}))\|_{\mathrm{Fro}}^2] \geq (c_4 - c_2)\|A\|_{\mathrm{Fro}}^2 = \frac{D/2-1}{D-1}\|A\|_{\mathrm{Fro}}^2 = \frac{1-2^{2-m}}{2-2^{2-m}}\|A\|_{\mathrm{Fro}}^2, \tag{161}$$

completing the proof of Lemma 5.12.

6 Pseudorandom products of operators

In this section we generalize the "derandomized squaring" technique of Rozenman and Vadhan [RV05] so that it may be applied to random walks on groups, where the goal is to show rapid mixing of a particular representation. We remark that the proofs are not really different from those in [RV05], and that a similar generalization appeared recently in [JMRW22].

Notation 6.1. Throughout we will be considering noncommutative polynomials, with real coefficients, over symbols $u_1, \ldots, u_c, u_1^\dagger, \ldots, u_c^\dagger$. (These symbols will eventually be substituted by square matrices.) If p is such a polynomial, its *adjoint* $p^\dagger$ is formed in the natural way (i.e., $(u_i^\dagger)^\dagger = u_i$ and $(u_i u_j)^\dagger = u_j^\dagger u_i^\dagger$, etc.), and we call p *self-adjoint* if $p^\dagger = p$.

Notation 6.2. We will also consider *polynomial sequences* $S = (s_1, \ldots, s_m)$, where each s_i is a polynomial in the u_j's. (Usually s_i will in fact be a *monomial*.)

Notation 6.3. If $\mathcal{U} = (U_1, \ldots, U_c)$ is a sequence of matrices, we write $S(\mathcal{U}) = (s_1(\mathcal{U}), \ldots, s_m(\mathcal{U}))$, where $s_j(\mathcal{U})$ is the matrix resulting from substituting $u_i = U_i$ for each $i \in [c]$.

Notation 6.4. Given a polynomial sequence S we write $\mathrm{avg}(S)$, or $\mathrm{avg} \circ S$, for the polynomial $\frac{1}{m}\sum_{j=1}^m s_j$.

Definition 6.5. If p is a polynomial over $u_1, \ldots, u_c$, we define

$$\|p\| = \sup_r \{\|p(\mathcal{U})\|_{\mathrm{op}} : \mathcal{U} = (U_1, \ldots, U_c),\ U_j \in \mathbb{C}^{r \times r},\ \|U_j\|_{\mathrm{op}} \leq 1\ \forall j\}, \tag{162}$$

the largest operator norm that p can achieve when $u_1, \ldots, u_c$ are substituted with square matrices of bounded operator norm. More generally, if $S = (s_1, \ldots, s_m)$ is a sequence of polynomials we write $\|S\| = \max(\|s_1\|, \ldots, \|s_m\|)$.

Definition 6.6. A *directed graph* $G = (V, E)$ will consist of a finite *sequence* of vertices V, and a finite *sequence* of edges E from $V \times V$ (so parallel edges and self-loops are allowed). Such a graph is *undirected* if E can be partitioned into pairs of the form $\{(i,j),(j,i)\}$. We say G is *d-out-regular* if for each $i \in V$ we have exactly d elements of the form (i,j) in E; one can analogously define in-regularity, and the two concepts are the same for undirected graphs. Note that if G is an undirected d-regular graph, then $|E| = d|V|$ (contrary to usual convention, as E is still composed of directed edges).

Definition 6.7. Given a graph $G = (V, E)$, where $V = (1, 2, \ldots, m)$, and given a polynomial sequence $S = (s_1, \ldots, s_m)$, we define $q_G \circ S$ to be the polynomial sequence[12]

$$(s_j^\dagger s_i)_{(i,j)\in E}. \tag{163}$$

Remark 6.8. If G is undirected, the polynomial $\mathrm{avg}(q_G \circ S)$ is self-adjoint.

Fact 6.9. *We always have $\|q_G \circ S\| \leq \|S\|^2$, and hence $\|S\| \leq 1 \implies \|q_G \circ S\| \leq 1$.*

Definition 6.10. If $G = (V, E)$ is a d-out-regular directed graph with $V = (1, 2, \ldots, c)$, then the normalized adjacency matrix of G is

$$A_G \coloneqq \frac{1}{d} \sum_{(i,j)\in E} |j\rangle\langle i| = c \cdot \mathrm{avg}(q_G \circ \mathcal{U}), \quad \text{where } \mathcal{U} = (\langle 1|, \ldots, \langle c|). \tag{164}$$

Fact 6.11. *Let $G = (V, E)$ be an out-regular directed graph with $V = (1, 2, \ldots, m)$ and let $\mathcal{W} = (W_1, \ldots W_m)$ be a sequence of matrices from $\mathbb{C}^{r\times r'}$.*[13] *Then*

$$\mathrm{avg}(q_G(\mathcal{W})) = \frac{1}{m}\mathcal{W}^\dagger (A_G \otimes \mathbb{1}_{r\times r})\mathcal{W}, \tag{165}$$

where we identify $\mathcal{W}$ with $\sum_{j=1}^m |j\rangle \otimes W_j$, the $mr \times r'$ matrix formed by stacking the W_j's into a column. (For $r' = r$, this identity is essentially the formula $w^\dagger A w = \sum_{ij} w_i^\dagger A_{ij} w_j$, but with entries from the ring $\mathbb{C}^{r\times r}$.)

Notation 6.12. We let K_m denote the complete (regular) undirected graph with self-loops on m vertices, which has $V = (1, 2, \ldots, m)$ and $E = ((1,1),(1,2),\ldots,(1,m),(2,1),\ldots,(m,m))$. We may write K in place of K_m if the context is clear.

Fact 6.13. *If $S = (s_1, \ldots, s_m)$ is a polynomial sequence,*

$$\mathrm{avg}(q_{\mathrm{K}_m} \circ S) = \mathrm{avg}(S)^\dagger \mathrm{avg}(S), \tag{166}$$

the Hermitian-square of $\mathrm{avg}(S)$. *Hence if $\mathcal{U}$ is a sequence of matrices, $\|\mathrm{avg}(q_{\mathrm{K}} \circ S(\mathcal{U}))\|_{\mathrm{op}} = \|\mathrm{avg}(S(\mathcal{U}))\|_{\mathrm{op}}^2$.*

Definition 6.14. Recall that a regular undirected graph G is said to be a *(2-sided) μ-expander* if $\|A_G - A_{\mathrm{K}}\|_{\mathrm{op}} \leq \mu$.

Fact 6.15. *Since A_{K} is the projection onto the 1-dimensional subspace spanned by $\sum_j |j\rangle$, and since A_G also fixes this subspace, G being a μ-expander is equivalent to $\|A_G - (1-\mu)A_{\mathrm{K}}\|_{\mathrm{op}} \leq \mu$.*

The following result is essentially the same as [RV05, Thm. 4.4]:

Proposition 6.16. *Let G be a μ-expander on vertex set $V = (1, 2, \ldots, m)$, let $S = (s_1, \ldots, s_m)$ be a polynomial sequence with $\|S\| \leq 1$, and let $\mathcal{U} = (U_1, \ldots, U_c)$ be a sequence of matrices in $\mathbb{C}^{r\times r}$ with $\|U_j\|_{\mathrm{op}} \leq 1$ for all j. Then*

$$\|\mathrm{avg}(q_G \circ S(\mathcal{U}))\|_{\mathrm{op}} \leq (1-\mu)\|\mathrm{avg}(S(\mathcal{U}))\|_{\mathrm{op}}^2 + \mu. \tag{167}$$

[12] One would have hoped for the more natural-looking ordering $s_i^\dagger s_j$, but alas we are forced to follow standard conventions: the length-2 path (i,j) means "first i, then j"; but, with operators acting on the left, $s_i^\dagger s_j$ means "first do s_j, then do $s_i^\dagger$".

[13] We only really care about $r' = r$, but we allow $r' \neq r$ for the sake of comparison with Equation (164), where $r = 1$ and $r' = c$.

Proof. Write $\mathcal{W} = S(\mathcal{U}) = (W_1, \dots, W_m)$, so $\|W_j\|_{\mathrm{op}} \le 1$ for all j. Using Fact 6.11 twice, we derive

$$\|\mathrm{avg}(q_G(\mathcal{W})) - (1-\mu)\mathrm{avg}(q_K(\mathcal{W}))\|_{\mathrm{op}} = \frac{1}{m}\left\|\mathcal{W}^\dagger(\Delta \otimes \mathbb{1}_{r\times r})\mathcal{W}\right\|_{\mathrm{op}}, \quad \text{where } \Delta = A_G - (1-\mu)A_K. \tag{168}$$

We have $\|\Delta\|_{\mathrm{op}} \le \mu$ by Fact 6.15, and $\|\mathcal{W}\|_{\mathrm{op}} \le \sqrt{\sum_j \|W_j\|_{\mathrm{op}}} \le \sqrt{m}$. So by submultiplicativity of operator norm, the right-hand side above is at most μ, and the proof is complete from Fact 6.13 and the triangle inequality. □

Iterating this, and using Fact 6.9 to conclude that $\|q_{G_t} \circ \cdots \circ q_{G_1} \circ S\| \le 1$ whenever $\|S\| \le 1$, yields:

Proposition 6.17. *Let $S = (s_1, \dots, s_m)$ be a polynomial sequence with $\|S\| \le 1$ and let $\mathcal{U} = (U_1, \dots, U_c)$ be a matrix sequence with $\|U_j\|_{\mathrm{op}} \le 1$ for all j. Moreover, let $G_1, G_2, \dots, G_t$ be a sequence of regular graphs, where $G_i = (V_i, E_i)$ is a μ_i-expander with $V_{i+1} = E_i$ (and $V_1 = (1, 2, \dots, m)$). Then*

$$\left\|\mathrm{avg}(q_{G_t} \circ q_{G_{t-1}} \circ \cdots \circ q_{G_1} \circ S(\mathcal{U}))\right\|_{\mathrm{op}} \le f_{\mu_t} \circ f_{\mu_{t-1}} \circ \cdots \circ f_{\mu_1}(\|\mathrm{avg}(S(\mathcal{U}))\|_{\mathrm{op}}), \tag{169}$$

where $f_\mu(\lambda) = (1-\mu)\lambda^2 + \mu$. In particular, if $m = c$, $S = (u_1, \dots, u_c)$, and we write $Q = q_{G_t} \circ \cdots \circ q_{G_1}$ and $F_{(\mu_1,\dots,\mu_t)} = f_{\mu_t} \circ \cdots \circ f_{\mu_1}$, then

$$\|\mathrm{avg}(Q \circ \mathcal{U})\|_{\mathrm{op}} \le F_{(\mu_1,\dots,\mu_t)}(\|\mathrm{avg}(\mathcal{U})\|_{\mathrm{op}}). \tag{170}$$

The work [RV05] also contains calculations very similar to the following (wherein the special number .11 is chosen due to certain explicit expander constructions):

Proposition 6.18. *For $0 < \delta, \epsilon \le 1$, we have $F_{\vec{\mu}}(1-\delta) \le \epsilon$ for any sequence $\vec{\mu}$ that entrywise satisfies*

$$(0, \dots, 0) \le \vec{\mu} \le (\vec{\mu}^{(1)}, \vec{\mu}^{(2)}), \qquad \vec{\mu}^{(1)} \coloneqq (\underbrace{.11, \dots, .11}_{\ell_1 \text{ times}}), \qquad \vec{\mu}^{(2)} \coloneqq \tfrac{1}{4}(2^{-2}, 2^{-4}, 2^{-8}, \dots, 2^{-2^{\ell_2}}), \tag{171}$$

where $\ell_1 \ge \log_{2^{.8}}(1/\delta) + 3$ (note: $2^{.8} \approx 1.74$) and $\ell_2 \ge \log_2 \log_2(1/\epsilon)$.

Proof. Since $f_\mu(\lambda)$ is nondecreasing on $[0,1]$ for both μ and λ, it suffices to analyze all upper bounds as if they were equalities. It is easy to check that $f_{.11}(1-\delta) \le 1 - 2^{.8}\delta$ for all $0 \le \delta \le .03$, and hence

$$\ell \ge \log_{2^{.8}}(1/\delta) - 6 \quad \Longrightarrow \quad f_{.11}^{\circ \ell}(1-\delta) \le 1 - .03/1.75 \le .985. \tag{172}$$

Also, $f_{.11}^{\circ 9}(.985) \le 1/4$, and hence $F_{\vec{\mu}^{(1)}}(1-\delta) \le 1/4$. The proof is now complete by observing that $F_{\vec{\mu}^{(2)}}(1/4) \le \frac{1}{2}2^{-2^{-\ell_2}}$. □

Regarding explicit construction of expander graphs, taking $p = 29$ and 509 in [Alo21, Thm. 1.2] and adding a self-loop to every vertex yields:

Theorem 6.19. *For $(d, \mu) = (32, .45)$ and also $(d, \mu) = (512, .11)$, there is a strongly explicit algorithm for constructing n-vertex, d-regular, μ-expander graphs (for all sufficiently large n).*

By repeatedly squaring the 32-regular graphs above, one can also conclude the following (in which it is possible that $d = d(n) > n$):

Corollary 6.20. *For any easy-to-compute $j = j(n) \in \mathbb{N}$, there is a strongly explicit ($\mathrm{polylog}(n, d)$ time) algorithm for constructing n-vertex, d-regular, μ-expander graphs (for all sufficiently large n) where, for $k = 2^j$, we have $d = 32^k$ and $\mu = \mu(n) = .45^k \le \frac{1}{4}2^{-k} = \frac{1}{4}d^{-1/5}$ (the inequality holding provided $j \ge 4$).*

Putting together Corollary 6.20, Proposition 6.18, and Proposition 6.17 yields the following:

Theorem 6.21. *There is a strongly explicit, space-minimal algorithm with the following behavior on inputs c and $0 < \delta, \epsilon < 1$ (where we assume $c = 2^{i_1}$, $\delta = 16^{-i_2}$, and $\epsilon = 2^{-2^{i_3}}$ for some $i_1, i_2, i_3 \in \mathbb{N}$ sufficiently large). The algorithm outputs a sequence Q of $N = O(c/(\delta^{11.25}\epsilon^{10}))$ monomials over symbols $u_1, \dots, u_c$ and $u_1^\dagger, \dots, u_c^\dagger$, each of length $L = 8\log_2(1/\epsilon)/\delta^{1.25}$, with the following property:*

For any sequence $\mathcal{U} = (U_1, \dots, U_c)$ of matrices in $\mathbb{C}^{r\times r}$ satisfying $\|U_i\|_{\mathrm{op}} \le 1$ for all i and $\|\mathrm{avg}(\mathcal{U})\|_{\mathrm{op}} \le 1 - \delta$, it holds that $\|\mathrm{avg}(Q \circ \mathcal{U})\|_{\mathrm{op}} \le \epsilon$.

Here "strongly explicit and space-minimal" means that, given a monomial index $i \in [N]$ and a monomial position index $j \in [L]$, the algorithm runs in deterministic $\mathrm{polylog}(c/\delta\epsilon)$ time and $O(\log(c/\delta\epsilon))$ space and outputs the jth symbol of the ith monomial in Q.

Proof. Given c, δ, ϵ, the desired Q is $q_{G_t} \circ \cdots \circ q_{G_1} \circ (u_1, \dots, u_c)$, where $G_1, \dots, G_t$ is a sequence as in Proposition 6.17, with:

- $\ell_1 = \log_{2^{.8}}(1/\delta) + 3 = \frac{5}{4}\log_2(1/\delta) + 3$, $\ell_2 = \log_2\log_2(1/\epsilon)$, and $t = \ell_1 + \ell_2$;
- $G_1, \dots, G_{\ell_1}$ are 512-regular .11-expanders, with G_j on $512^{j-1}c$ vertices, as in Theorem 6.19;
- $G_{\ell_1+1}, \dots, G_{\ell_1+\ell_2}$ are as in Corollary 6.20, with G_{ℓ_1+j} being a 32^k-regular, $\frac{1}{4}2^{-k}$-expander (for $k = 2^{\min(j,4)}$) on $32^{k+32}N_0$ vertices (once $j \ge 4$), where $N_0 = 512^{\ell_1}c$ is $|E(G_{\ell_1})|$.

The length of Q is

$$N = |E(G_t)| = 32^{2^{\ell_2+1}+32}N_0 = 2^{160}\cdot 2^{5\cdot\log_2(1/\epsilon)\cdot 2}\cdot 2^{9(\log_{2^{4/5}}(1/\delta)+3)} = 2^{187}\cdot c/\delta^{11.25}\epsilon^{10}, \tag{173}$$

and each monomial in Q has length $2^t = 8\log_2(1/\epsilon)/\delta^{1.25}$. The desired bound $\|\mathrm{avg}(Q(\mathcal{U}))\|_{\mathrm{op}} \le \epsilon$ follows from Propositions 6.17 and 6.18. Finally, the time and space bounds are easy to verify, as computation of the jth symbol of the ith monomial of Q simply amounts to determining the ith edge of G_t, and then following a path down a binary tree of height t, where at each node one has to compute the the ath edge of a particular G_b. □

Remark 6.22. As in [RV05, Thm. 5.8], if δ is not small but is rather already of the form $\delta = 1 - \lambda$ for small λ, one can retain only the last $\ell_2 - \log_2\log_2(1/\lambda)$ or so expanders and obtain $L = O(\log(1/\epsilon)/\log(1/\lambda))$; we omit details.

When using Theorem 6.21, we will often want to disregard a certain "trivial" subspace; we will then employ the following simple observation:

Fact 6.23. *In the setting of Theorem 6.21, say each U_j may be written as $U_j = R_j \oplus U_j'$, where R_j acts on subspace T and U_j' acts on its orthogonal complement $T^\perp$ in $\mathbb{C}^r$. Then $\mathrm{avg}(Q(\mathcal{U})) = \mathrm{avg}(Q(\mathcal{R})) \oplus \mathrm{avg}(Q(\mathcal{U}'))$, where $\mathcal{R} = (R_1, \dots, R_c)$ and $\mathcal{U}' = (U_1', \dots, U_c')$.*

For example, suppose $G = (V, E)$ is a d-regular undirected graph on $V = (1, 2, \dots, n)$ with normalized adjacency matrix expressed as

$$A_G = \mathrm{avg}(P_1, \dots, P_d), \tag{174}$$

where $P_1, \dots, P_d$ are $n \times n$ permutation matrices. Each P_i and $P_i^\dagger$ has operator norm 1 and fixes the one-dimensional space $T = \mathrm{span}\{|1\rangle + \cdots + |n\rangle\}$. If we write $P_i = \mathrm{proj}_T \oplus U_i'$ where U_i' is the action of P_i on $T^\perp$, then

$$A_G = \mathrm{proj}_T \oplus \mathrm{avg}(U_1', \dots, U_d') \tag{175}$$

and we are in a position to apply Fact 6.23 and Theorem 6.21 together. The result is a sequence Q of "walks", each of the form $P_{i_L}^\dagger P_{i_{L-1}} \cdots P_{i_2}^\dagger P_{i_1}$. Applying one such walk to any starting vertex $|v\rangle$ leads to a valid walk of length L in G (with the steps $P_i^\dagger$ being valid since G is undirected). If we write $\widetilde{G}$ for the $|Q|$-regular undirected graph on V wherein each $v \in V$ has an edge to all its walk outcomes, the result is that

$$A_{\widetilde{G}} = \mathrm{avg}(Q \circ (P_1, \dots, P_d)) = \mathrm{proj}_T \oplus \mathrm{avg}(Q(U_1', \dots, U_d')). \tag{176}$$

Hence if G is a $(1-\delta)$-expander, we obtain that $\widetilde{G}$ is an ϵ-expander with $|Q| = O(d/(\delta\epsilon)^{O(1)})$ and walks of length $O(\log(1/\epsilon)/\delta^{O(1)})$. As shown in [Rei08, RV05], given any simple, connected, n-vertex, undirected graph, there is a very simple transformation preserving connectivity that produces a 4-regular undirected graph (together with the associated $P_1, \ldots, P_4$ as in Equation (174)) that has $\delta \geq 1/\operatorname{poly}(n)$; by taking $\epsilon = 1/\operatorname{poly}(n)$, one can use these pseudorandom walks to establish Reingold's Theorem $\mathsf{SL} = \mathsf{L}$ [Rei08].

References

[ABI86] Noga Alon, László Babai, and Alon Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *Journal of Algorithms*, 7(4):567–583, 1986. 1

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992. 1

[AK63] Vladimir Arnol'd and Alexander Krylov. Uniform distribution of points on a sphere and certain ergodic properties of solutions of linear ordinary differential equations in a complex domain. *Doklady Akademii Nauk SSSR*, 148:9–12, 1963. 3.4

[Alo21] Noga Alon. Explicit expanders of every degree and size. *Combinatorica*, 41(4):447–463, 2021. 6

[BC20] Charles Bordenave and Benoît Collins. Strong asymptotic freeness for independent uniform variables on compact groups associated to non-trivial representations. Technical Report 2012.08759, arXiv, 2020. 4.1

[BCHJ+21] Fernando Brandão, Wissam Chemissany, Nicholas Hunter-Jones, Richard Kueng, and John Preskill. Models of quantum complexity growth. *PRX Quantum*, 2(3):030316, 2021. 1

[BdS16] Yves Benoist and Nicolas de Saxcé. A spectral gap theorem in simple Lie groups. *Inventiones Mathematicae*, 205(2):337–361, 2016. 3.1, 3.4, 3.9, 7

[BG12] Jean Bourgain and Alex Gamburd. A spectral gap theorem in $\mathrm{SU}(d)$. *Journal of the European Mathematical Society (JEMS)*, 14(5):1455–1511, 2012. 3, 3.4

[BH99] Martin Bridson and André Haefliger. *Length Spaces*, pages 32–46. Springer Berlin Heidelberg, 1999. 5.2

[BH08] Alex Brodsky and Shlomo Hoory. Simple permutations mix even better. *Random Structures & Algorithms*, 32(3):274–289, 2008. 1, 2.8

[BHH16] Fernando Brandão, Aram Harrow, and Michał Horodecki. Local random quantum circuits are approximate polynomial-designs. *Communications in Mathematical Physics*, 346(2):397–434, 2016. 1, 1, 3, 1.1, 3.1, 3.4, 4.1, 4.1, 10, 5.3.2

[BNZZ19] Eiichi Bannai, Mikio Nakahara, Da Zhao, and Yan Zhu. On the explicit constructions of certain unitary t-designs. *Journal of Physics. A.*, 52(49):495301, 17, 2019. 1

[Bra37] Richard Brauer. On algebras which are connected with the semisimple continuous groups. *Annals of Mathematics*, 38(4):857–872, 1937. 4.8

[DCEL09] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Physical Review A*, 80(1):012304, 2009. 1

[DLT02] David DiVincenzo, Debbie Leung, and Barbara Terhal. Quantum data hiding. *Transactions on Information Theory*, 48(3):580–598, 2002. 1

[FPY15] Hilary Finucane, Ron Peled, and Yariv Yaari. A recursive construction of t-wise uniform permutations. *Random Structures & Algorithms*, 46(3):531–540, 2015. 1

[Gow96] W. Timothy Gowers. An almost m-wise independent random permutation of the cube. *Combinatorics, Probability and Computing*, 5(2):119–130, 1996. 1, 1.1

[Gro99] Cheryl Grood. Brauer algebras and centralizer algebras for $\mathrm{SO}(2n, \mathbf{C})$. *Journal of Algebra*, 222(2):678–707, 1999. 4.8

[Haf22] Jonas Haferkamp. Random quantum circuits are approximate unitary t-designs in depth $o(nt^{5+o(1)})$. *Quantum*, 6:795, 2022. 1, 1

[HFGW18] Anna-Lena Hashagen, Steven Flammia, David Gross, and Joel Wallman. Real randomized benchmarking. *Quantum*, 2:85, 2018. 1

[HHJ21] Jonas Haferkamp and Nicholas Hunter-Jones. Improved spectral gaps for random quantum circuits: large local dimensions and all-to-all interactions. *Physical Review A*, 104(2):Paper No. 022417, 18, 2021. 1, 1, 1.1, 1.2, 3.1, 3.2, 4, 4.1, 4.1, 4.1, 4.1, 5.3.2

[HKOT23] Jeongwan Haah, Robin Kothari, Ryan O'Donnell, and Ewin Tang. Query-optimal estimation of unitary channels in diamond distance. Technical Report 2302.14066, arXiv, 2023. 1

[HL09] Aram Harrow and Richard Low. Efficient quantum tensor product expanders and k-designs. In *Proceedings of the 2009 International Workshop on Approximation, Randomization, and Combinatorial Optimization (APPROX)*, pages 548–561. Springer, 2009. 1, 1, 2

[HMMR05] Shlomo Hoory, Avner Magen, Steven Myers, and Charles Rackoff. Simple permutations mix well. *Theoretical Computer Science*, 348(2-3):251–261, 2005. 1

[JMRW22] Fernando Granha Jeronimo, Tushant Mittal, Sourya Roy, and Avi Wigderson. Almost Ramanujan expanders from arbitrary expanders via operator amplification. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science—FOCS 2022*, pages 378–388. IEEE Computer Soc., Los Alamitos, CA, [2022] ©2022. 1.1, 6

[Kas07] Martin Kassabov. Symmetric groups and expander graphs. *Inventiones Mathematicae*, 170(2):327–354, 2007. 1

[KM15] Pravesh Kothari and Raghu Meka. Almost optimal pseudorandom generators for spherical caps. In *Proceedings of the 2015 Symposium on the Theory of Computing (STOC)*, pages 247–256. ACM, 2015. 1, 1, 3, 2, 6

[KN14] Daniel Kane and Jelani Nelson. Sparser Johnson–Lindenstrauss transforms. *Journal of the ACM*, 61(1):Art. 4, 23, 2014. 1

[KNR09] Eyal Kaplan, Moni Naor, and Omer Reingold. Derandomized constructions of k-wise (almost) independent permutations. *Algorithmica*, 55(1):113–133, 2009. 1, 1.1

[Kot22] Kothari, Pravesh. Personal communication. 2022. 3

[LK10] Ping Li and Christian König. b-Bit minwise hashing. In *Proceedings of the 19th Annual International Conference on World Wide Web*, pages 671–680, 2010. 1

[Mec19] Elizabeth Meckes. *The random matrix theory of the classical compact groups*, volume 218. Cambridge University Press, 2019. 3, 5.1

[MOP22] Sidhanth Mohanty, Ryan O'Donnell, and Pedro Paredes. Explicit near-Ramanujan graphs of every degree. *SIAM Journal on Computing*, 51(3):STOC20–1–STOC20–23, 2022. 1

[NC10] Michael Nielsen and Isaac Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 10th anniversary edition edition, 2010. 3.4.1

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, 1993. 1

[Oli09] Roberto Imbuzeiro Oliveira. On the convergence to equilibrium of Kac's random walk on matrices. *Ann. Appl. Probab.*, 19(3):1200–1231, 2009. 5.2, 5.5

[Rei08] Omer Reingold. Undirected connectivity in log-space. *Journal of the ACM*, 55(4):Art. 17, 24, 2008. 1, 1.1, 6

[RTV06] Omer Reingold, Luca Trevisan, and Salil Vadhan. Pseudorandom walks on regular digraphs and the RL vs. L problem. In *Proceedings of the 2006 Symposium on the Theory of Computing (STOC)*, pages 457–466. ACM, 2006. 1.1

[RV05] Eyal Rozenman and Salil Vadhan. Derandomized squaring of graphs. Electronic Colloquium on Computational Complexity, TR05-092, 2005. 1.1, 6, 6, 6, 6.22, 6

[RY17] Daniel Roberts and Beni Yoshida. Chaos and complexity by design. *Journal of High Energy Physics*, 2017(4):1–64, 2017. 1

[Sch01] Issai Schur. *Über eine Klasse von Matrizen, die sich einer gegebenen Matrix zuordnen lassen.* PhD thesis, Universität Berlin, 1901. 4.8

[Sen18] Pranab Sen. Efficient quantum tensor product expanders and unitary t-designs via the zigzag product. Technical Report 1808.10521, arXiv, 2018. 1, 1

[Shi02] Yaoyun Shi. Both Toffoli and Controlled-NOT need little help to do universal quantum computation. Technical Report quant-ph/0205115, arXiv, 2002. 3.4.1

[Tao14] Terence Tao. *Hilbert's fifth problem and related topics*, volume 153. American Mathematical Society, 2014. 5.1

[Web16] Zak Webb. The Clifford group forms a unitary 3-design. *Quantum Information & Computation*, 16(15-16):1379–1400, 2016. 1

[Wey39] Hermann Weyl. *The Classical Groups. Their Invariants and Representations.* Princeton University Press, 1939. 4.8

[Yua12] Qiaochu Yuan. Four flavors of Schur–Weyl duality, 2012. https://qchu.wordpress.com/2012/11/13/four-flavors-of-schur-weyl-duality/. 4.8

[ZZP17] Linxi Zhang, Chuanghua Zhu, and Changxing Pei. Randomized benchmarking using unitary t-design for average fidelity estimation of practical quantum circuit. Technical Report 1711.08098, arXiv, 2017. 1