Matrix concentration inequalities and efficiency of random universal sets of quantum gates

Piotr Dulian[1,2] and Adam Sawicki[1]

[1]Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland
[2]Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland

For a random set $\mathcal{S} \subset U(d)$ of quantum gates we provide bounds on the probability that $\mathcal{S}$ forms a δ-approximate t-design. In particular we have found that for $\mathcal{S}$ drawn from an exact t-design the probability that it forms a δ-approximate t-design satisfies the inequality $\mathbb{P}(\delta \geq x) \leq 2D_t \frac{e^{-|\mathcal{S}|x\,\mathrm{arctanh}(x)}}{(1-x^2)^{|\mathcal{S}|/2}} = O\left(2D_t \left(\frac{e^{-x^2}}{\sqrt{1-x^2}}\right)^{|\mathcal{S}|}\right)$, where D_t is a sum over dimensions of unique irreducible representations appearing in the decomposition of $U \mapsto U^{\otimes t} \otimes \bar{U}^{\otimes t}$. We use our results to show that to obtain a δ-approximate t-design with probability P one needs $O(\delta^{-2}(t\log(d) - \log(1-P)))$ many random gates. We also analyze how δ concentrates around its expected value $\mathbb{E}\delta$ for random $\mathcal{S}$. Our results are valid for both symmetric and non-symmetric sets of gates.

1 Introduction and summary of main results

Practical realisations of quantum computers are constricted by noise and decoherence that affect large-scale quantum systems. Although quantum error correction codes can overcome those effects they require usage of thousands of physical qubits to implement a single logical noiseless fault-tolerant qubit [1]. This is clearly out of reach for contemporary quantum computers with the number of physical qubits of the order of hundreds [2]. Hence, in the near future we are forced to work with noisy intermediate-scale quantum devices (NISQ) [2–4]. Moreover, currently the best error rates per gate are the order of 0.1% [5, 6] which implies we cannot build circuits much longer than thousand [2]. The length of a circuit is also limited by the coherence time and the time of execution of a single gate. It is noteworthy that it is hard to make gates that are both fast and have low error rates [5]. Clearly, there is a great need for quantum computation using as few gates as possible. One way of achieving this is by using universal sets of gates (gate-sets) of high *efficiency*, i.e. such that can approximate any unitary with circuits of minimal length. This idea is also connected to complexity of unitaries (see [7] for more details).

The efficiency of a universal set $\mathcal{S}$ (see [8–10] for criteria that allow deciding universality) is typically measured by the length of a circuit needed to approximate any quantum transformation with a given precision ϵ. The Solovay-Kitaev theorem states that all symmetric universal sets [1] are roughly the same efficient. More precisely, the length of a circuit that ϵ-approximates any $U \in SU(d)$ is bounded by $A(\mathcal{S})\log^c(1/\epsilon)$ [11], where $c \geq 1$. There have been recently some new developments connected to the Solovay-Kitaev theorem for

[1]A set $S \subset U(d)$ is symmetric if for any $U \in \mathcal{S}$ the inverse $U^{-1} \in \mathcal{S}$

gate-sets without inverses. First, in [12, 13] it was shown that an ϵ-approximate poly-log length circuit exists even if one drops the assumption that a set $\mathcal{S}$ is symmetric. Moreover in [14] an algorithm implementing this sequence was given. To estimate the value of $A(\mathcal{S})$ one can use the concept of δ-approximate t-designs [13, 15]. To this end let $\{\mathcal{S}, \nu_{\mathcal{S}}\}$ be an ensemble of quantum gates, where $\mathcal{S}$ is a finite subset of $U(d)$ and $\nu_{\mathcal{S}}$ is a probability measure on $\mathcal{S}$. Such an ensemble is called $\delta(\nu_{\mathcal{S}}, t)$-approximate t-design if and only if

$$\delta(\nu_{\mathcal{S}}, t) \coloneqq \|T_{\nu_{\mathcal{S}},t} - T_{\mu,t}\| < 1 \ ,$$

where $\|\cdot\|$ is the operator norm and for any measure ν (in particular for the Haar measure μ) we define a *moment operator*

$$T_{\nu,t} \coloneqq \int_{U(d)} d\nu(U) U^{t,t}, \text{ where } U^{t,t} = U^{\otimes t} \otimes \bar{U}^{\otimes t},$$

where $\bar{U}$ is entry-wise conjugation of U. When $\delta(\nu_{\mathcal{S}}, t) = 0$ we say that $\{\mathcal{S}, \nu_{\mathcal{S}}\}$ is an *exact t-design*. Thus (approximate) t-design are ensembles of unitaries that (approximately) recover Haar averages of polynomials in entries of unitaries up to the order t. More precisely any balanced polynomial of degree t on $U(d)$ can be written as $f_A(U) = \mathrm{Tr}\left(AU^{t,t}\right)$, where A is a fixed matrix of size $d^{2t} \times d^{2t}$. Assuming that $\{\mathcal{S}, \nu_{\mathcal{S}}\}$ is a δ-approximate t-design we have

$$\left| \int_{U(d)} d\nu_{\mathcal{S}}(U) f_A(U) - \int_{U(d)} d\mu(U) f_A(U) \right| = \left| \mathrm{Tr}\left(A\left(T_{\nu_{\mathcal{S}},t} - T_{\mu}, t\right)\right) \right| \leq \|A\|_1 \delta(\nu_{\mathcal{S}}, t), \quad (1)$$

where $\|A\|_1 = \mathrm{Tr}\sqrt{AA^\dagger}$. Thus $\delta(\nu_{\mathcal{S}}, t)$ controls the error we make when taking average of f_A with respect to $\nu_{\mathcal{S}}$ instead of the Haar measure. Unitaries constituting δ-approximate t-design form ϵ-nets for $t \simeq \frac{d^{5/2}}{\epsilon}$ and $\delta \simeq \left(\frac{\epsilon^{3/2}}{d}\right)^{d^2}$ [13]. As a direct consequence of property (1) δ-approximate t-designs find numerous applications throughout quantum information, including randomized benchmarking [16, 17], information transmission [18], quantum state discrimination [19], criteria for universality of quantum gates [10] and complexity growth [7, 20–22]. It is also known that the constant $A(\mathcal{S})$ from the Solovay-Kitaev theorem is inversely proportional to $1 - \delta(\nu_{\mathcal{S}})$, where $\delta(\nu_{\mathcal{S}}) \coloneqq \sup_t \delta(\nu_{\mathcal{S}}, t)$, whenever $\delta(\nu_{\mathcal{S}}) < 1$ [15]. Determining the value of $\delta(\nu_{\mathcal{S}})$ requires computation of the norm of an infinite number of operators $T_{\nu_{\mathcal{S}},t}$ which is analytically and numerically intractable. It is known, however, that $\delta(\nu_{\mathcal{S}}) < 1$ under the additional assumption that gates have algebraic entries [23, 24]. In this case also the constant c in the Solovay-Kitaev theorem is equal to 1. Recent results [25–29] based on some number theoretic constructions give examples of universal sets with $c = 1$ and the smallest possible value of $\delta(\nu_{\mathcal{S}})$. The approach presented in these contributions has been unified in [29] where the author pointed out the connection of these new results to the seminal work concerning distributions of points on the sphere S^2 [30].

In contrast to the above mentioned contributions, in this work, we do not focus on concrete gate-sets but instead we aim to answer the natural question of how likely it is that a set of gates has high efficiency. In order to achieve this goal we need to characterize efficiency of random universal gate-sets. Calculation of $\delta(\nu_{\mathcal{S}})$ is of course intractable. Therefore we follow the approach of [12, 13] and consider $\delta(\nu_{\mathcal{S}}, t)$ for some fixed t (which is determined by a precision ϵ). The results of [12, 13] ensure that for a given precision ϵ the constant $A(\mathcal{S})$ in the Solovay-Kitaev theorem is inversely proportional to $1 - \delta(\nu_{\mathcal{S}}, t(\epsilon))$, where $t(\epsilon) = O(\epsilon^{-1})$, and the constant $c = 1$. Therefore, in order to characterize efficiency of random universal sets of gates we need to characterize a probability distribution of $\delta(\nu_{\mathcal{S}}, t)$.

What remains is to make precise what kind of random gate-sets we want to consider. As there is a natural uniform measure on the unitary group, i.e. the Haar measure one can consider two types of gate-sets:

1. $\mathcal{S} = \{U_1, \ldots, U_n\}$, where U_k's are independent and Haar random unitaries from $U(d)$. Such $\mathcal{S}$ will be called *Haar random gate-set*.

2. $\mathcal{S} = \{U_1, \ldots, U_n\} \cup \{U_1^{-1}, \ldots, U_n^{-1}\}$, where U_k's are independent and Haar random unitaries from $U(d)$. Such $\mathcal{S}$ will be called *symmetric Haar random gate-set*.

Another choice would be to start with an ensemble $\{\mathcal{D}, \nu_{\mathcal{D}}\}$ that forms an exact t-design and consider two sets:

1. $\mathcal{S} = \{U_1, \ldots, U_n\} \subset \mathcal{D}$, where U_k's are independent and distributed according to $\nu_{\mathcal{D}}$. Such $\mathcal{S}$ will be called *t-random gate-set*.

2. $\mathcal{S} = \{U_1, \ldots, U_n\} \cup \{U_1^{-1}, \ldots, U_n^{-1}\}$, where $\{U_1, \ldots, U_n\} \subset \mathcal{D}$ and U_k's are independent and distributed according to $\nu_{\mathcal{D}}$. Such $\mathcal{S}$ will be called *symmetric t-random gate-set*.

We note that putting $\mathcal{D} = U(d)$ and $\nu_{\mathcal{D}} = \mu$, where μ is the Haar measure on $U(d)$, we get that a (symmetric) Haar random gate-set is a (symmetric) t-random gate set for any t. Thus (symmetric) Haar random gate-sets are (symmetric) ∞-random gate-sets and all the results that we prove for the (symmetric) t-random gate-sets are automatically true for (symmetric) Haar random gate-sets.

In order to simplify the notation we will often denote the cardinality of $\mathcal{S}$ by $\mathcal{S}$ (instead of $|\mathcal{S}|$). Our main results are given in a form of bounds on the probability $\mathbb{P}\left(\delta(\nu_{\mathcal{S}}, t) \geq \delta\right)$, where $\mathcal{S}$ is a (symmetric) t-random gate set and $\nu_{\mathcal{S}}$ is a uniform measure on $\mathcal{S}$. In order to obtain them we first show that $T_{\nu_{\mathcal{S}},t}$ decomposes as a direct sum over irreducible representations of the unitary group $U(d)$ that can be labeled by elements subset $\lambda \in \Lambda_t \subset \mathbb{Z}^d$ (see formula (23)). The blocks appearing in this decomposition, that we denote by $T_{\nu_{\mathcal{S}},\lambda}$, have dimensions d_λ given by the Weyl dimension formula (20). Using the union bound

$$\mathbb{P}\left(\delta(\nu_{\mathcal{S}}, t) \geq \delta\right) \leq \sum_{\lambda \in \Lambda_t} \mathbb{P}\left(\delta(\nu_{\mathcal{S}}, \lambda) \geq \delta\right), \tag{2}$$

we reduce the problem to finding bounds on $\mathbb{P}\left(\delta(\nu_{\mathcal{S}}, \lambda) \geq \delta\right)$. We achieve this combining recently derived matrix concentration inequalities [31] with the recent result concerning calculation of higher Frobenius-Schur indicators for semisimple Lie algebras [32], that we further improve. Our main results are Theorems 1, 2 that give concrete calculable bounds on $\mathbb{P}\left(\delta(\nu_{\mathcal{S}}, t) \geq \delta\right)$.

Theorem 1. *Let $\mathcal{S}$ be a t-random gate-set and $\nu_{\mathcal{S}}$ a uniform measure. Then for any $\delta < 1$*

$$\mathbb{P}\left(\delta(\nu_{\mathcal{S}}, t) \geq \delta\right) \leq \frac{2e^{-\delta \mathcal{S} \operatorname{arctanh}(\delta)}}{(1-\delta^2)^{\frac{\mathcal{S}}{2}}} \sum_{\lambda \in \Lambda_t} d_\lambda, \tag{3}$$

where, Λ_t is given by (23) and d_λ is given by (20).

Theorem 2. *Let $\mathcal{S}$ be a symmetric Haar random gate-set. Then for any $\delta < 1$*

$$\mathbb{P}\left(\delta(\nu_{\mathcal{S}}, t) \geq \delta\right) \leq \sum_{\lambda \in \Lambda_t} d_\lambda e^{-\frac{\mathcal{S}\delta^2}{\sqrt{1-\delta^2}}} \left[F\left(\frac{\mathcal{S}\delta}{\sqrt{1-\delta^2}}, \lambda, \mathcal{S}\right) + F\left(-\frac{\mathcal{S}\delta}{\sqrt{1-\delta^2}}, \lambda, \mathcal{S}\right)\right], \tag{4}$$

where Λ_t is given by (23), d_λ is given by (20) and $F(\cdot,\cdot,\cdot)$ is given by (33).

These theorems are then used to obtain bounds on the size of a t-random gate-set needed to form, with a probability P, a δ-approximate t-designs, for various t's and δ's. We show that this size is of the order $O(\delta^{-2}(t\log(d)-\log(1-P)))$. Moreover, we compare the number of independent t-random n-qubit gates, $\mathcal{S}_n$, needed to form 0.01-approximate 2-design with the probability 0.99 and the size of the n-qubit Clifford group, $\mathcal{C}_n$, that is known to be an exact 2-design. The ratio of $\mathcal{C}_n/\mathcal{S}_n$ turns out to grow exponentially with the number of qubits.
In Section 6.3 we also show that Theorems 1 and 2 can be easily generalised to a scenario where instead of t-random set of gates we have a set of random quantum circuits composed of t-random independent gates.

Theorem 1 can be used when one needs to calculate average of any (t,t)-polynomial over $U(d)$. Such polynomials arise, for example, in the randomized benchmarking protocols where one is interested in assessing quality of quantum gates implementation and considers the k-th moments of the fidelity [33–35].

$$\int_{U(d)} d\mu(U)\mathrm{Tr}\left(\rho U^{-1}\Phi(U\rho U^{-1})U\right)^k, \tag{5}$$

where Φ is a quantum channel that represents gate independent noise and for perfect implementation of U is the identity. Of course one can replace the Haar measure in (5) by an exact $2k$-design as the integrated function is $(2k,2k)$-polynomial. Using our results we can replace an exact $2k$-design, which as shown in [35] has exponential size in $n\sqrt{2k}$, where n is a number of qubits, by a δ-approximate t-designs of size (see Section 7)

$$\mathcal{S} \geq \frac{2\left(2t\log\left(d\right)+\log(2)-\log\left(1-P\right)\right)}{\log\left(\left(1+\delta\right)^{1+\delta}\left(1-\delta\right)^{1-\delta}\right)}.$$

Moreover we can control the error using (1).

Finally we analyze concentration properties of $\delta(\nu_{\mathcal{S}},t)$ around its mean value $\mathop{\mathbb{E}}_{U\sim\mu}\delta(\nu_{\mathcal{S}},t)$. Our main result is

Theorem 3. *Let $\mathcal{S}$ be a Haar random gate-set. Then*

$$\mathbb{P}\left(\delta(\nu_{\mathcal{S}},t)\geq \mathop{\mathbb{E}}_{U\sim\mu}\delta(\nu_{\mathcal{S}},t)+\alpha\right)\leq \exp\left(\frac{-d\mathcal{S}\alpha^2}{32t^2}\right). \tag{6}$$

The methods behind its proof [36] can be extended to Haar random gate-sets with particular structure or architecture. Following this observation we analyze efficiency of random d-mode circuits build from 2-mode beamsplitters. More precisely we consider the Hilbert space $\mathcal{H}=\mathcal{H}_1\oplus\ldots\oplus\mathcal{H}_d$, where $\mathcal{H}_k\simeq\mathbb{C}$, $d>2$ and we call spaces $\mathcal{H}_k$ modes. For a matrix $B\in SU(2)$, which we call a 2-mode beamsplitter, we define matrices B^{ij}, $i\neq j$, to be the matrices that act on a 2-dimensional subspace $\mathcal{H}_i\oplus\mathcal{H}_j\subset\mathcal{H}$ as B and on the other components of $\mathcal{H}$ as the identity. This way a matrix $B\in SU(2)$ gives $d(d-1)$ matrices in $SU(d)$. Applying this procedure to a Haar random gate-set set $S\subset SU(2)$ we obtain random gate-set S^d (see [37, 38]) and it is natural to ask about its efficiency. Our main conclusion here is that a Haar random gate-set $\mathcal{S}\subset SU(2)$ gives the gate-set $S^d\subset SU(d)$ for which $\delta(\nu_{\mathcal{S}^d},t)$ has the same concentration rate around the mean as a Haar random gate-set $\mathcal{S}'\subset SU(d)$ of size: $\mathcal{S}'=\frac{2\mathcal{S}}{d}$, i.e.

Theorem 4. *Let $\mathcal{S}\subset SU(2)$ be a Haar random gate-set and $\mathcal{S}^d\subset SU(d)$ the corresponding d-mode gate-set. Then*

$$\mathbb{P}\left(\delta(\nu_{\mathcal{S}^d},t)\geq \mathop{\mathbb{E}}_{U\sim\mu}\delta(\nu_{\mathcal{S}^d},t)+\alpha\right)\leq \exp\left(\frac{-\mathcal{S}\alpha^2}{16t^2}\right). \tag{7}$$

Using similar methods one can show the concentration around the mean value of any function

$$F : U(d)^{\times n} \ni (U_1, ..., U_n) \to F(U_1, ..., U_n) \in \mathbb{R},$$

for U_k's independent and Haar random as long as F is L-Lipschitz:

$$|F(U_1, ..., U_n) - |F(V_1, ..., V_n)| \leq L \left(\sum_{k=1}^{n} \|U_k - V_k\|_F^2 \right)^{\frac{1}{2}},$$

where $\|\cdot\|_F$ is a Frobenius (Hilbert-Schmidt) norm. We explain this in detail in Section 2. A function F can be, for example, given by the k-th moment of the fidelity of a quantum circuit of the length n (see [35] for more details).

The paper is organized as follows. In Section 2 we present a short review of matrix concentration inequalities that will play a central role in our setting. Next, in Section 3 we provide necessary information concerning irreducible representations of unitary groups. Then in Section 4 we introduce notion of moment operators. In Section 5 we explain the role of Frobenius-Schur indicators and how to compute them. The main results of these paper are then showed in Section 6, while Section 7 contains analysis of the results and applications.

2 Short review of relevant matrix concentration inequalities

In this section we review known upper bounds on the probability that $\mathbb{P}(F(X) \geq \delta)$, for classes of random matrices X and real valued functions F that are relevant in our setting. An interested reader is referred to [31, 36] for more details. The first class of inequalities concerns a situation when $X = \sum X_k$, where $X_k \in \mathrm{Mat}(d, \mathbb{C})$ are independent, random, Hermitian matrices and the function F is the operator norm of X, $F(X) = \|X\|$. Thus we are looking for an upper bound for $\mathbb{P}(\|X\| \geq \delta)$. The line of reasoning is as follows. Let $\lambda_{\max}(X)$ and $\lambda_{\min}(X)$ denote the largest and the smallest eigenvalues of X respectively. In the first step one uses the exponential Markov inequality, i.e.

$$\mathbb{P}(Y \geq t) = \mathbb{P}\left(e^{\theta Y} \geq e^{\theta t}\right) \leq e^{-\theta t} \mathbb{E} e^{\theta Y},$$

for any $\theta > 0$. Taking $Y = \lambda_{\max}(X)$ and using the fact that $e^{\theta \lambda_{\max}(X)} \leq \mathrm{tr} e^{\theta X}$ we get

$$\mathbb{P}(\lambda_{\max}(X) \geq \delta) \leq \inf_{\theta > 0} e^{-\theta \delta} \mathbb{E} \mathrm{tr} e^{\theta X}. \tag{8}$$

Next we note that $\mathbb{P}(\lambda_{\min}(X) \leq \delta) = \mathbb{P}(-\lambda_{\min}(X) \geq -\delta) = \mathbb{P}(\lambda_{\max}(-X) \geq -\delta)$. Thus

$$\mathbb{P}(\lambda_{\min}(X) \leq \delta) \leq \inf_{\theta > 0} e^{\theta \delta} \mathbb{E} \mathrm{tr} e^{-\theta X}. \tag{9}$$

Next, using the Lieb theorem [39] (one can alternatively use the Golden-Thomson inequality [40, 41] but the resulting bound is in general weaker [31]) we obtain

$$\mathbb{E} \mathrm{tr} e^{\theta \sum_k X_k} \leq \mathrm{tr} \exp\left(\sum_k \log \mathbb{E} e^{\theta X_k} \right), \tag{10}$$

for any $\theta \in \mathbb{R}$. Combining (8) and (9) with (10) we get

$$\mathbb{P}\left(\lambda_{\max}(X) \geq \delta\right) \leq \inf_{\theta>0} e^{-\theta\delta} \operatorname{tr} \exp\left(\sum_k \log \mathbb{E} e^{\theta X_k}\right), \tag{11}$$

$$\mathbb{P}\left(\lambda_{\min}(X) \leq \delta\right) \leq \inf_{\theta>0} e^{\theta\delta} \operatorname{tr} \exp\left(\sum_k \log \mathbb{E} e^{-\theta X_k}\right). \tag{12}$$

Finally,

$$\begin{aligned}\mathbb{P}\left(\|X\| \geq \delta\right) &= \mathbb{P}\left(\max\{\lambda_{\max}(X), -\lambda_{\min}(X)\} \geq \delta\right) \leq \\ &\leq \mathbb{P}\left(\lambda_{\max}(X) \geq \delta\right) + \mathbb{P}\left(\lambda_{\min}(X) \leq -\delta\right).\end{aligned}$$

Fact 1 (Master bound)**.** *Let $X = \sum X_k$, where $X_k \in \mathrm{Mat}(d, \mathbb{C})$ are independent, random, Hermitian matrices. Then*

$$\mathbb{P}\left(\|X\| \geq \delta\right) \leq \inf_{\theta>0} e^{-\theta\delta} \operatorname{tr} \exp\left(\sum_k \log \mathbb{E} e^{\theta X_k}\right) + \inf_{\theta>0} e^{-\theta\delta} \operatorname{tr} \exp\left(\sum_k \log \mathbb{E} e^{-\theta X_k}\right).$$

A master bound can be also derived for the sum of non-Hermitian random matrices. To this end, following [31], we consider the Hermitian dilation map

$$\mathcal{H} : \mathrm{Mat}(d, \mathbb{C}) \to \mathbb{H}(2d),$$

where $\mathbb{H}(2d)$ is the space of $2d \times 2d$ Hermitian matrices given by

$$\mathcal{H}(X) = \begin{pmatrix} 0 & X \\ X^\dagger & 0 \end{pmatrix}.$$

This map is clearly a real linear map. One can also show (cf. [31]) that $\|X\| = \|\mathcal{H}(X)\| = \lambda_{\max}\left(\mathcal{H}(X)\right)$. Making use of (11) we get:

Fact 2. *Let $X = \sum X_k$, where $X_k \in \mathrm{Mat}(d, \mathbb{C})$ are independent random matrices. Then*

$$\mathbb{P}\left(\|X\| \geq \delta\right) \leq \inf_{\theta>0} e^{-\theta\delta} \operatorname{tr} \exp\left(\sum_k \log \mathbb{E} e^{\theta \mathcal{H}(X_k)}\right).$$

The upper bounds in Facts 1 and 2 can be further simplified by finding a majorization of $\log \mathbb{E} e^{\theta X_k}$ in terms of moments $\mathbb{E} X_k^n$ that allow analytic optimization over θ (see chapter 6 of [31] for more details). Usage of moments up to order two leads to the matrix Bernstein inequality.

Fact 3 (Matrix Bernstein inequality)**.** *Let $X = \sum X_k$, where $X_k \in \mathrm{Mat}(d, \mathbb{C})$ are independent, random matrices such that:*

$$\forall_k \quad \mathbb{E} X_k = 0 \quad \text{and} \quad \|X_k\| \leq L.$$

Let v be the matrix variance statistic of the sum:

$$\begin{aligned} v &= \max\left\{\|\mathbb{E}(XX^\dagger)\|, \|\mathbb{E}(X^\dagger X)\|\right\}, \\ &= \max\left\{\|\sum_k \mathbb{E}(X_k X_k^\dagger)\|, \|\sum_k \mathbb{E}(X_k^\dagger X_k)\|\right\}. \end{aligned}$$

Then for all $\delta \geq 0$

$$\mathbb{P}\left(\|X\| \geq \delta\right) \leq 2d \exp\left(\frac{-\delta^2/2}{v + L\delta/3}\right).$$

2.1 Bounds for Haar random matrices

The second type of inequalities we will consider in this paper rely on the fact that the randomness comes from the Haar measure. The interested reader is referred to chapter 5 of [36] for detailed discussion. Here we only give the main result and mention that its proof is based on the fact that, by Bakery-Émery curvature criterion, the (special)unitary group equipped with the Haar measure satisfies the so-called logarithmic Sobolev inequality which, via the Herbst argument, leads to concentration of measure for Lipschitz functions.

Fact 4. *Let $G^{\times \mathcal{S}}$, where G is $SU(d)$ or $U(d)$, be equipped with the metric given by the L_2-sum of Hilbert-Schmidt (Frobenius) metrics on the group G, i.e. the distance between $(U_1, \ldots, U_{\mathcal{S}}) \in G^{\times \mathcal{S}}$ and $(V_1, \ldots, V_{\mathcal{S}}) \in G^{\times \mathcal{S}}$ is*

$$\left(\sum_{k=1}^{\mathcal{S}} \|U_k - V_k\|_F^2 \right)^{\frac{1}{2}}.$$

Suppose that

$$F : G^{\times \mathcal{S}} \ni (U_1, ..., U_{\mathcal{S}}) \to F(U_1, ..., U_{\mathcal{S}}) \in \mathbb{R},$$

is L-Lipschitz, i.e.

$$|F(U_1, ..., U_{\mathcal{S}}) - F(V_1, ..., V_{\mathcal{S}})| \leq L \left(\sum_{k=1}^{\mathcal{S}} \|U_k - V_k\|_F^2 \right)^{\frac{1}{2}},$$

and that $U_1, ..., U_{\mathcal{S}}$ are independent and chosen according to the Haar measure on G. Then for any $\alpha > 0$

$$\mathbb{P}\left(F(U_1, ..., U_{\mathcal{S}}) \geq \mathbb{E}F(U_1, ..., U_{\mathcal{S}}) + \alpha\right) \leq \exp\left(-\frac{d\alpha^2}{4CL^2}\right).$$

where C is equal to 2 for $SU(d)$ and 6 for $U(d)$.

3 Irreducible representations of $U(d)$ and $SU(d)$

In this section we recall some basic facts about Lie groups, Lie algebras and their representations in the context of groups $G = U(d)$ and $G = SU(d)$. In Table 1 we summarize information about those groups, where we used $M_d^0(\mathbb{C}) := \{X \in M_d(\mathbb{C}) \,|\, \mathrm{Tr}(X) = 0\}$. We will denote by $\mathfrak{g}$ the Lie algebra of G. The Lie algebra $\mathfrak{su}(d)$ has no non-trivial ideals and thus is semisimple. On the other hand the algebra $\mathfrak{u}(d)$ has an abelian ideal consisting of matrices proportional to the identity and is not semisimple. We note, however, that $[\mathfrak{u}(d), \mathfrak{u}(d)] = \mathfrak{su}(d)$. We will call a Lie group semisimple if its Lie algebra is semisimple. Other relevant for us algebras are Lie algebra complexification $\mathfrak{g}_{\mathbb{C}} := \mathfrak{g} + i\mathfrak{g}$, the Lie algebra of the maximal torus T in G - $\mathfrak{t}$ and the Cartan subalgebra (CSA) - $\mathfrak{h} := \mathfrak{t} + i\mathfrak{t}$.

The functional $\alpha \in \mathfrak{h}^*$ is called a *root* of $\mathfrak{g}$ if and only if there exists $X_\alpha \in \mathfrak{g}_{\mathbb{C}}$ such that:

$$\forall_{H \in \mathfrak{h}} \; [H, X_\alpha] = \alpha(H) X_\alpha. \tag{13}$$

We denote the set of all roots by Φ and call it *the root system*. For a given root α the subspace of all X_α satisfying (13) is called a root subspace of α and denoted by $\mathfrak{g}_\alpha$. The algebra $\mathfrak{g}_{\mathbb{C}}$ decomposes

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

G	$U(d)$	$SU(d)$
$\mathfrak{g}$	$\mathfrak{u}(d) := \left\{X \in M_d(\mathbb{C}) \,\middle|\, X^\dagger = -X\right\}$	$\mathfrak{su}(d) := \mathfrak{u}(d) \cap M_d^0(\mathbb{C})$
semi-simple	no	yes
$\mathfrak{g}_\mathbb{C}$	$\mathfrak{u}(d)_\mathbb{C} \cong \mathfrak{gl}(d,\mathbb{C}) := M_d(\mathbb{C})$	$\mathfrak{su}(d)_\mathbb{C} \cong \mathfrak{sl}(d,\mathbb{C}) := M_d^0(\mathbb{C})$
$\mathfrak{t}$	$\mathfrak{t} := \{X \in \mathfrak{u}(d) \,\vert\, X \text{ diagonal}\}$	$\mathfrak{t}_0 := \mathfrak{t} \cap M_d^0(\mathbb{C})$
$\mathfrak{h}$	$\mathfrak{h} := \{X \in \mathfrak{gl}(d,\mathbb{C}) \,\vert\, X \text{ diagonal}\}$	$\mathfrak{h}_0 := \mathfrak{h} \cap M_d^0(\mathbb{C})$

Table 1: The comparison of groups $U(d)$ and $SU(d)$ in view of the Lie group theory. We used $M_d^0(\mathbb{C})$ to denote $\{X \in M_d(\mathbb{C}) \,|\, \mathrm{Tr}(X) = 0\}$.

If we define $L_i, \alpha_{i,j} \in \mathfrak{h}^*$ as

$$L_i\left(\begin{pmatrix} a_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & a_d \end{pmatrix}\right) := a_i, \tag{14}$$

$$\alpha_{i,j} := L_i - L_j, \tag{15}$$

then the root system for $U(d)$ and $SU(d)$ is

$$\Phi = \{\alpha_{i,j} \,|\, 1 \leq i, j \leq d\}\,.$$

Among roots we distinguish positive roots

$$\Phi^+ := \{\alpha_{i,j} \,|\, 1 \leq i < j \leq d\}\,,$$

and positive simple roots

$$\Delta := \{\alpha_{i,i+1} \,|\, 1 \leq i \leq d-1\}\,.$$

For $\alpha, \beta \in \mathfrak{h}^*$ we say that α is higher (lower) than β iff $\alpha - \beta$ is a linear combination of positive simple roots with non-negative (non-positive) coefficients and we denote it by $\alpha > \beta$ ($\alpha < \beta$).
We introduce the inner product on $\mathfrak{h}^*$ defined by

$$\langle L_i | L_j \rangle := \delta_{i,j}.$$

The inner product gives us an isomorphism $\mathfrak{h}^* \ni L \mapsto \langle \bar{L}| \in \mathfrak{h}$ hence further in the text we will identify $\mathfrak{h}^*$ with $\mathfrak{h}$. For any $\alpha \in \Phi$ the root system is preserved under the reflection about the hyper-plane perpendicular to α:

$$s_\alpha : \mathfrak{h} \ni H \mapsto H - \frac{2\,\langle H|\alpha\rangle}{\langle \alpha|\alpha\rangle}\alpha.$$

The group $\mathcal{W} :=< s_\alpha \,|\, \alpha \in \Phi >$ is called the Weyl group of Φ. In our case $\mathcal{W}$ is isomorphic to the group of permutations S_d and action of $\sigma \in \mathcal{W}$ on $\mathfrak{h}$ is given by:

$$\sigma \cdot L_i := L_{\sigma^{-1}(i)}.$$

An element $H \in \mathfrak{h}$ is called:

- integral iff $\forall_{\alpha\in\Phi}\ \frac{2\langle H|\alpha\rangle}{\langle\alpha|\alpha\rangle} \in \mathbb{Z}$,
- analytically integral iff for all $X \in \mathfrak{h}$ such that $e^{2\pi i X} = \mathbb{1}$ there holds $\langle X|H\rangle \in \mathbb{Z}$,

- dominant iff $\forall_{\alpha\in\Delta}\ \langle\alpha|H\rangle \geq 0$.

For every finite dimensional representation $\pi : G \to GL(V)$ of G there exists representation $\rho : \mathfrak{g} \to \mathfrak{gl}(V)$ such that for any $X \in \mathfrak{g}$ there holds

$$\pi\left(e^X\right) = e^{\rho(X)}.$$

Let us define the complexification of ρ to be $\rho_{\mathbb{C}} : \mathfrak{g}_{\mathbb{C}} \ni X + iY \mapsto \rho(x) + i\rho(Y) \in \mathfrak{gl}(V, \mathbb{C})$. Then the following are equivalent:

- π is irreducible,
- ρ is irreducible,
- $\rho_{\mathbb{C}}$ is irreducible.

Further in the text we will slightly abuse notation and we will use ρ also for $\rho_{\mathbb{C}}$. A weight of ρ is an integral element μ such that there exists a non-zero vector $v_\mu \in \rho$ satisfying:

$$\forall_{H\in\mathfrak{h}}\ \rho(H)v_\mu = \langle\mu|H\rangle\, v_\mu. \tag{16}$$

Subspace of all v_μ satisfying (16) is called the weight space of μ and we denote it by ρ^μ or π^μ. The multiplicity m^μ (or $m(\mu)$) of μ is the dimension of its weight space. Every irreducible representation is a direct sum of its weight spaces. The notions of weight space and root space are closely connected. Indeed for $v_\mu \in \rho^\mu$, $X_\alpha \in \mathfrak{g}_\alpha$ and $H \in \mathfrak{h}$ we have:

$$\begin{aligned}\rho\left(H\right)\rho\left(X_\alpha\right)v_\mu = \left(\rho([H, X_\alpha]) + \rho\left(X_\alpha\right)\rho\left(H\right)\right)v_\mu = \\ = \left(\langle\alpha|H\rangle + \langle\mu|H\rangle\right)\rho\left(X_\alpha\right)v_\mu = \langle\mu+\alpha|H\rangle\,\rho\left(X_\alpha\right)v_\mu,\end{aligned}$$

thus $\rho\left(X_\alpha\right)v_\mu$ is either 0 or in $\rho^{\mu+\alpha}$. The highest weight λ is a weight in ρ that is higher than all other weights in ρ. Now, we can state the theorem of the highest weight.

Fact 5 (Theorem of the highest weight)**.** *For semi-simple complex Lie algebra $\mathfrak{g}_{\mathbb{C}}$ and its finite-dimensional representation ρ we have:*

1. *ρ has unique highest weight λ,*
2. *λ is dominant,*
3. *if $\tilde{\rho}$ is another representation of $\mathfrak{g}_{\mathbb{C}}$ with highest weight λ then $\tilde{\rho}$ is isomorphic to ρ,*
4. *if λ is dominant and integral then there exists finite-dimensional irreducible representation of $\mathfrak{g}_{\mathbb{C}}$ with highest weight λ.*

Fact 6. *If G is compact and connected the analogous theorem holds with the only difference that the highest weight λ has to be analytically integral.*

In case of $U(d)$ we will identify highest weights λ with sequences $(\lambda_1, ..., \lambda_d)$ where $\lambda_i := \langle\lambda|L_i\rangle$ and in case of $SU(d)$ we will identify highest weights λ^s with sequences $(\lambda^s_1, ..., \lambda^s_{d-1})$ where $\lambda^s_i := \langle\lambda^s|\alpha_{i,i+1}\rangle$. From the second point of Fact 5 we have that for $i < j$ there holds $\lambda_i \geq \lambda_j$ and from analytical integrality $\lambda_i \in \mathbb{Z}$. Moreover, from the fourth point of Fact 5 every sequence from $\mathbb{Z}^d$ satisfying those conditions uniquely defines the highest weight λ and the associated representation. Analogously for $SU(d)$ the condition is $\lambda^s_i \in \mathbb{Z}_+$ and any element of $\mathbb{Z}^{d-1}_+$ defines the highest weight. In both cases

we use π_λ and ρ_λ to denote the irreducible finite-dimensional representations with highest weight λ.

Since $SU(d)$ is a subgroup of $U(d)$ any representation π_λ of $U(d)$ can be restricted to a representation $\pi'_{\lambda^s} := \pi_\lambda|_{SU(d)}$ of $SU(d)$. The question arises what is the relation between λ and λ^s. On the other hand, irreducible representations of $SU(d)$ are often labeled by the Young diagrams instead of highest weights. In the following lemma we explore the relations between highest weights of $U(d)$, highest weights of $SU(d)$ and Young diagrams.

Lemma 5. *Let λ be a highest weight of representation π_λ of $U(d)$. Then the corresponding representation of $SU(d)$ has a Young diagram given by $(\lambda_1 - \lambda_d, \ldots, \lambda_{d-1} - \lambda_d)$ and the standard highest weight of this representation is given by $(\lambda_1 - \lambda_2, \ldots, \lambda_{d-2} - \lambda_{d-1}, \lambda_{d-1} - \lambda_d)$.*

Proof. By the definition of λ we have that for the highest weight vector $v_\lambda \in \rho^\lambda_\lambda$ and any $H \in \mathfrak{h}$ it holds:

$$\rho_\lambda(H)v = \langle \lambda | H \rangle \, v_\lambda.$$

Then for the representation $\rho'_\lambda := \rho_\lambda|_{\mathfrak{sl}(d,\mathbb{C})}$ and its any subspace V we have:

$$\forall_{X\in\mathfrak{su}(d)}\, \rho'_\lambda(X)V \subset V \Rightarrow \forall_{X\in\mathfrak{su}(d),\, \phi\in\mathbb{R}}\, (\rho'_\lambda(X) + i\phi\mathbb{1})V \subset V \Rightarrow$$
$$\Rightarrow \forall_{X\in\mathfrak{su}(d),\, \phi\in\mathbb{R}}\, \rho_\lambda(X + i\phi\mathbb{1})V \subset V \Rightarrow \forall_{X\in\mathfrak{u}(d)}\, \rho_\lambda(X)V \subset V.$$

Hence irreducibility of ρ_λ implies the irreducibility of ρ'_λ.
Note that $L_1 - L_d, ..., L_{d-1} - L_d$ is a basis of $\mathfrak{h}_0$ and for $i \in \{1, ..., d-1\}$:

$$\rho'_\lambda(L_i - L_d)v_\lambda = \rho_\lambda(L_i - L_d)v_\lambda = \langle \lambda | L_i - L_d \rangle \, v_\lambda = (\lambda_i - \lambda_d)v_\lambda.$$

Since v_λ is the highest weight vector the sequence $\lambda^Y := (\lambda_1 - \lambda_d, ..., \lambda_{d-1} - \lambda_d)$ uniquely determines the $\mathfrak{sl}(d,\mathbb{C})$ representation and the associated $SU(d)$ representation. Moreover, from the relations $\lambda_1 \geq ... \geq \lambda_d$ we have that λ^Y is a sequence of non-negative, descending integers and as such can be interpreted as a Young diagram. If we choose $L_1 - L_2, ..., L_{d-1} - L_d$ as a basis of $\mathfrak{h}_0$, we analogously obtain the sequence $\lambda^s := (\lambda_1 - \lambda_2, ..., \lambda_{d-1} - \lambda_d)$ that is a highest weight of $SU(d)$ representation. □

To simplify the description of highest weights we introduce the following notation. For any $\lambda = (\lambda_1, \ldots, \lambda_d)$ in $\mathbb{Z}^d$ we will denote its length by $l(\lambda) = d$. By λ_+ we will denote the subsequence of λ of positive integers. Moreover

$$\Sigma(\lambda) := \sum_{k=1}^{d} \lambda_k, \tag{17}$$

$$\|\lambda\|_1 := \sum_{k=1}^{d} |\lambda_k|, \tag{18}$$

$$d_\lambda := \dim \pi_\lambda \tag{19}$$

The value of d_λ is determined by the Weyl dimension formula

$$d_\lambda = \prod_{1\leq i<j\leq d} \frac{\lambda_i - \lambda_j + j - i}{j - i}. \tag{20}$$

Fact 7. *[42] Suppose that π_λ is an irreducible representation of semi-simple Lie group G then $\mu \in \mathfrak{h}$ is a weight of π_λ if and only if the following two conditions are satisfied:*

1. *μ is contained in the convex hull of the orbit of λ under the Weyl group,*

2. *$\lambda - \mu$ is a linear combination of positive simple roots with integer coefficients.*

Corollary 6. *Since $SU(d)$ is semi-simple and $SU(d)$ and $U(d)$ have the same roots and root spaces the above fact also applies to $U(d)$.*

We use Fact 7 to prove the following lemma.

Lemma 7. *For any weight μ of the representation π_λ we have $\|\mu\|_1 \leq \|\lambda\|_1$ and $\Sigma(\mu) = \Sigma(\lambda)$.*

Proof. The first condition of Fact 7 reads

$$\exists\, \{t_\sigma \in [0,1] \mid \sigma \in S_d\} \text{ such that } \sum_{\sigma\in S_d} t_\sigma = 1 \text{ and } \mu = \sum_{\sigma\in S_d} t_\sigma \sigma\cdot\lambda,$$

hence

$$\|\mu\|_1 = \sum_{i=1}^{d} |\sum_{\sigma\in S_d} t_\sigma \lambda_{\sigma(i)}| \leq \sum_{i=1}^{d}\sum_{\sigma\in S_d} t_\sigma |\lambda_{\sigma(i)}| = \sum_{\sigma\in S_d} t_\sigma \|\lambda\|_1 = \|\lambda\|_1,$$

which proves the first part of the lemma. The second condition from Fact 7 implies

$$\lambda - \mu = \sum_{i=1}^{d-1} c_i (\underbrace{0,...,0}_{i-1}, 1, -1, \underbrace{0,...,0}_{d-i-1}).$$

By acting Σ on both sides we obtain

$$\Sigma(\mu - \lambda) = 0 \quad \Rightarrow \quad \Sigma(\mu) = \Sigma(\lambda).$$

□

Fact 8 (Kostant formula)**.** *Let $p : \mathfrak{h}_0 \to \mathbb{N}$ be such that $p(\mu)$ is equal to the number of ways μ can be expressed as a linear combinations of positive simple roots with non-negative integer coefficients, $m_{\lambda^s}(\mu)$ a multiplicity of μ in π_{λ^s} and ρ the half-sum of positive roots*[2] *then*

$$m_{\lambda^s}(\mu) = \sum_{\sigma\in S_d} \operatorname{sgn}(\sigma)\, p(\sigma\cdot(\lambda^s + \rho) - (\mu + \rho)) \tag{21}$$

4 Moment operators

Let $\{\mathcal{S}, \nu_\mathcal{S}\}$ be an ensemble of quantum gates, where $\mathcal{S}$ is a subset of $U(d)$ and $\nu_\mathcal{S}$ is a probability measure on $\mathcal{S}$. Such an ensemble is called $\delta(\nu_\mathcal{S}, t)$-approximate t-design if and only if

$$\delta(\nu_\mathcal{S}, t) \coloneqq \|T_{\nu_\mathcal{S},t} - T_{\mu,t}\| < 1\ ,$$

where $\|\cdot\|$ is the operator norm and for any measure ν (in particular for the Haar measure μ) we define a *moment operator*

$$T_{\nu,t} \coloneqq \int_{U(d)} d\nu(U) U^{t,t}, \text{ where } U^{t,t} = U^{\otimes t} \otimes \bar{U}^{\otimes t},$$

where $\bar{U}$ is entry-wise conjugation of U.

[2]Note that the corresponding Lie algebra representation is denoted by ρ_{λ^s}

One can easily show that $0 \leq \delta(\nu_{\mathcal{S}}, t) \leq 1$ [13]. When $\delta(\nu_{\mathcal{S}}, t) = 0$ we say that $\mathcal{S}$ is *an exact t-design* and when $\delta(\nu_{\mathcal{S}}, t) = 1$ we say that $\mathcal{S}$ is *not a t-design*.

Let us note that the entries of $U^{t,t}$ are monomials of the order t in the entries of U and of the order t in the entries of $\bar{U}$. We will call them (t,t)-monomials. The space of (t,t)-polynomials, which we denote by $\mathcal{H}_t$, is defined as the linear span of (t,t)-monomials. One can write any element $f_A \in \mathcal{H}_t$ as

$$f_A(U) = \mathrm{Tr}\left(AU^{t,t}\right),$$

where A is a $d^{2t} \times d^{2t}$ matrix. Assuming that $\{\mathcal{S}, \nu_{\mathcal{S}}\}$ is a δ-approximate t-design we have

$$\left|\int_{U(d)} d\nu_{\mathcal{S}}(U) f_A(U) - \int_{U(d)} d\mu(U) f_A(U)\right| = \left|\mathrm{Tr}\left(A\left(T_{\nu_{\mathcal{S}},t} - T_{\mu}, t\right)\right)\right| \leq \|A\|_1 \delta(\nu_{\mathcal{S}}, t),$$

where $\|A\|_1 = \mathrm{Tr}\sqrt{AA^{\dagger}}$. Thus $\delta(\nu_{\mathcal{S}}, t)$ controls the error we make when taking average of f_A with respect to $\nu_{\mathcal{S}}$ instead of the Haar measure.

A map $U \mapsto U^{t,t}$ is a representation of the unitary group $U(d)$. This representation is reducible and decomposes into some irreducible representations $U(d)$.

Fact 9. *([43]) Irreducible representations that appear in the decomposition of $U \mapsto U^{\otimes t} \otimes \bar{U}^{\otimes t}$ are π_λ with $l(\lambda) = d$, $\Sigma(\lambda) = 0$ and $\Sigma(\lambda_+) \leq t$. That is we have*

$$U^{\otimes t} \otimes \bar{U}^{\otimes t} \simeq \mathbb{1}^{\oplus m_0} \oplus \bigoplus_{\lambda \in \Lambda_t} \pi_\lambda(U)^{\oplus m_\lambda} \simeq \left(U \otimes \bar{U}\right)^{\otimes t}, \tag{22}$$

where

$$\Lambda_t = \left\{\lambda = (\lambda_1, \ldots, \lambda_d) |\, \lambda \in \mathbb{Z}^d, \lambda \neq 0, \forall_k\, \lambda_k \geq \lambda_{k+1}, \Sigma(\lambda) = 0, \Sigma(\lambda_+) \leq t\right\}, \tag{23}$$

and $\mathbb{1}$ stands for the trivial representation and m_0 is its multiplicity and m_λ is the multiplicity of π_λ and $\simeq$ stands for a unitary equivalence of representations.

The representations occurring in decomposition (22) are in fact irreducible representation of the projective unitary group, $PU(d) = U(d)/\sim$, where $U \sim V$ iff $U = e^{i\phi}V$. One can show that every irreducible representation of $PU(d)$ arises this way for some, possibly large, t [44]. For $t = 1$ decomposition (22) is particularly simple and reads $U \otimes \bar{U} \simeq \mathrm{Ad}_U \oplus \mathbb{1}$, where $\mathbb{1}$ stands for the trivial representation and Ad_U is the adjoint representation of $U(d)$ and $PU(d) \simeq \mathrm{Ad}_{U(d)}$[3].

For any irreducible representation π_λ, $\lambda \in \Lambda_t$ we define

$$T_{\nu_{\mathcal{S}},\lambda} = \int_{U(d)} d\nu_{\mathcal{S}}(U) \pi_\lambda(U).$$

Next we define $\delta(\nu_{\mathcal{S}}, \lambda) \coloneqq \|T_{\nu_{\mathcal{S}},\lambda}\|$. It follows directly from the definitions and discussion above that

$$T_{\nu_{\mathcal{S}},t} \simeq \bigoplus_{\lambda \in \Lambda_t} \left(T_{\nu_{\mathcal{S}},\lambda}\right)^{\oplus m_\lambda}, \ \delta(\nu_S, t) = \sup_{\lambda \in \Lambda_t} \delta(\nu_{\mathcal{S}}, \lambda). \tag{24}$$

One can also define $\delta(\nu_{\mathcal{S}}) := \sup_t \delta(\nu_{\mathcal{S}}, t)$. It is known that for $\mathcal{S}$ finite and $\nu_{\mathcal{S}}$ uniform $\delta_{\mathrm{opt}}(\mathcal{S}) \leq \delta(\nu_{\mathcal{S}}) \leq 1$, where $\delta_{\mathrm{opt}} = \frac{2\sqrt{|\mathcal{S}|-1}}{|\mathcal{S}|}$ [29].

[3] By Ad_U we mean the matrix $\mathrm{Ad}_U : \mathfrak{su}(d) \to \mathfrak{su}(d)$, $\mathrm{Ad}_U(X) = UXU^{-1}$.

Lemma 8. *Assume $\lambda \in \Lambda_t$. Then $\|\lambda\|_1 = 2k$, where the integer k satisfies $1 \leq k \leq t$. Moreover the number of distinct irreducible representations π_λ with $\|\lambda\|_1 = 2k$ is given by*

$$\alpha_{2k} = \begin{cases} p(k)^2 & d \geq 2k, \\ \sum_{n=1}^{k} p_n(k)\tilde{p}_{d-n}(k) & k+1 \leq d < 2k, \\ \sum_{n=1}^{d-1} p_n(k)\tilde{p}_{d-n}(k) & 2 \leq d \leq k. \end{cases} \tag{25}$$

where $p_n(k)$ is the number of partitions of k with exactly n integers and $\tilde{p}_n(k)$ is the number of partitions of k with at most n integers. When $d \geq 2t$ formula (25) simplifies to

$$\alpha_{2k} = p(k)^2, \tag{26}$$

where $p(k)$ is number of all partitions of k.

Proof. First, we will prove $\|\lambda\|_1 = 2k$. We define λ_- to be the subsequence of λ of negative integers. Then from the condition $\Sigma(\lambda) = 0$ we have:

$$\Sigma(\lambda_-) = (\Sigma(\lambda_-) + \Sigma(\lambda_+)) - \Sigma(\lambda_+) = \Sigma(\lambda) - \Sigma(\lambda_+) = -\Sigma(\lambda_+).$$

Next, from the Eq. 18:

$$\|\lambda\|_1 = \sum_{k=1}^{d} |\lambda_k| = \Sigma(\lambda_+) - \Sigma(\lambda_-) = 2\Sigma(\lambda_+) \leq 2t.$$

Thus $\|\lambda\|_1 = 2k$ for $k = \Sigma(\lambda_+)$. Now let us put $n = l(\lambda_+)$ then the λ_+ is a decreasing sequence of n positive integers that sum up to $\Sigma(\lambda_+) = k$ so it is a partition of k with exactly n integers and $(-(\lambda_-)_{l(\lambda_-)}, ..., -(\lambda_-)_1)$ is a decreasing sequence of $l(\lambda_-) \leq d - n$ positive integers that sum up to $-\Sigma(\lambda_-) = k$ so it is a partition of k with at most n integers.
On the other hand, if we take η to be a partition of k with exactly n integers and ζ to be a partition of k with at most n integers such that $d \geq n + l(\zeta)$ then they can be combined into a sequence:

$$\tilde{\lambda} = (\eta_1, ..., \eta_n, \underbrace{0, ..., 0}_{d-n-l(\zeta)}, -\zeta_{l(\zeta)}, ..., -\zeta_1),$$

that is clearly an element of Λ_t and $\|\tilde{\lambda}\|_1 = 2k$. Hence there is a one to one correspondence between sequences like λ and pairs of partitions like (η, ζ). Thus to prove the formula (25) the only thing left to do is to note that the equations $\Sigma(\lambda) = 0$ and $\lambda \neq 0$ imply inequalities $n \geq 1$ and $n \leq d - l(\lambda_-) \leq d - 1$.
The formula (26) results easily from the fact that for $n > k$ we have $p_n(k) = 0$ and $\tilde{p}_n(k) = p(k)$ for $n \geq k$. □

Lemma 9. *Let $U, V \in U(d)$. Assume $\lambda \in \Lambda_t$. Then*

$$\|\pi_\lambda(U) - \pi_\lambda(V)\| \leq \frac{\pi}{2}\|\lambda\|_1\|U - V\|.$$

Proof. As the operator norm is unitarily invariant it is enough to show that for a diagonal unitary matrix $U = \operatorname{diag}\left(e^{i\phi_1}, \ldots, e^{i\phi_d}\right) \in U(d)$, $\phi_k \in (-\pi, \pi]$ we have:

$$\|\pi_\lambda(U) - \pi_\lambda(\mathbb{1})\| \leq \frac{\pi}{2}\|\lambda\|_1\|U - \mathbb{1}\|.$$

One easily sees that

$$\|U-\mathbb{1}\| = 2\sup_{1\leq k\leq d}\left|\sin\left(\frac{\phi_k}{2}\right)\right|,$$
$$\frac{2}{\pi}\sup_{1\leq k\leq d}|\phi_k| \leq \|U-\mathbb{1}\| \leq \sup_{1\leq k\leq d}|\phi_k|.$$

The eigenvalues of $\pi_\lambda(U)$ are given by

$$e^{i\psi_k},\ \psi_k = \sum_{i=1}^{d}\lambda_{i,k}\phi_i,$$

where $\lambda_k = (\lambda_{1,k},\ldots,\lambda_{d,k})$ satisfies $\|\lambda_k\|_1 \leq \|\lambda\|_1$ (see Lemma 7). Thus

$$\|\pi_\lambda(U)-\mathbb{1}\| = 2\sup_{1\leq k\leq d_\lambda}\left|\sin\left(\frac{\psi_k}{2}\right)\right| \leq \sup_{1\leq k\leq d_\lambda}|\psi_k| \leq \|\lambda\|_1 \sup_{1\leq k\leq d}|\phi_k| \leq \frac{\pi}{2}\|\lambda\|_1\|U-\mathbb{1}\|.$$

□

4.1 Moment operators for quantum circuits

We can generalize the notion of moment operators to random quantum circuit in a natural way. Recall that a quantum circuit of depth m is a product of m quantum gates. Thus we can identify quantum circuits with elements of $U(d)^{\times m}$.

Consider an ensemble $\{\mathcal{R},\nu_{\mathcal{R}}\}$ where $\mathcal{R}\subset U(d)^{\times m}$ is a set of circuits of length m and $\nu_{\mathcal{R}}$ is a probability measure on $\mathcal{R}$. For $\mathbf{U}=(U_1,...,U_m)\in U(d)^{\times m}$ let us define

$$\mathbf{U}^{t,t} := U_1^{t,t}\otimes ...\otimes U_m^{t,t}. \tag{27}$$

Note that the entries of $\mathbf{U}^{t,t}$ form a basis of the space of all $(t,t)-$polynomials in entries of $U_1,...,U_m$ which we will call $\mathcal{H}_{t,m}$. The average of $\mathbf{U}^{t,t}$ over all circuits with the same probability is

$$T_{\mu^{\times m},t} := \int_{U(d)^m} d\mu^{\times m}(\mathbf{U})\mathbf{U}^{t,t} = \int_{U(d)} d\mu(U_1)... \int_{U(d)} d\mu(U_m)\mathbf{U}^{t,t} =$$
$$= \left(\int_{U(d)} d\mu(U_1)U_1^{t,t}\right)\otimes ...\otimes \left(\int_{U(d)} d\mu(U_m)U_m^{t,t}\right) = T_{\mu,t}^{\otimes m}.$$

where $\times$ is a product of measures. Using Fact 9 we get that:

$$\mathbf{U}^{t,t} \simeq \bigotimes_{i=1}^{m}\left(\mathbb{1}^{\oplus m_0}\oplus\bigoplus_{\lambda\in\Lambda_t}\pi_\lambda(U_i)^{\oplus m_\lambda}\right) \simeq \bigoplus_{\lambda_1,...,\lambda_m\in\tilde{\Lambda}_t}\bigotimes_{i=1}^{m}\pi_{\lambda_i}(U_i)^{\oplus m_{\lambda_i}},$$

where $\tilde{\Lambda}_t := \Lambda_t\cup\{0\}$. To simplify notation let us define:

$$\boldsymbol{\lambda} := (\lambda_1,...,\lambda_m)\in\tilde{\Lambda}_t^{\times m},$$
$$\Lambda_{t,m} := \tilde{\Lambda}_t^{\times m}\setminus\{(0,...,0)\},$$
$$\pi_{\boldsymbol{\lambda}}(\mathbf{U}) := \pi_{\lambda_1}(U_1)\otimes ...\otimes\pi_{\lambda_m}(U_m),$$
$$T_{\nu_{\mathcal{R}},t} := \int_{U(d)^m} d\nu_{\mathcal{R}}(\mathbf{U})\mathbf{U}^{t,t},$$
$$T_{\nu_{\mathcal{R}},\boldsymbol{\lambda}} := \int_{U(d)^m} d\nu_{\mathcal{R}}(\mathbf{U})\pi_{\boldsymbol{\lambda}}(\mathbf{U}),$$
$$\delta(\nu_{\mathcal{R}},\boldsymbol{\lambda}) := \|T_{\nu_{\mathcal{R}},\boldsymbol{\lambda}}\|.$$

It is easy to see that:

$$\delta(\nu_{\mathcal{R}}, t) := \|T_{\mu,t}^{\otimes m} - T_{\nu_{\mathcal{R}},t}\| = \sup_{\boldsymbol{\lambda}\in\Lambda_{t,m}} \|T_{\nu_{\mathcal{R}},\boldsymbol{\lambda}}\| = \sup_{\boldsymbol{\lambda}\in\Lambda_{t,m}} \delta(\nu_{\mathcal{R}}, \boldsymbol{\lambda}).$$

5 Frobenius-Schur indicators

In Section 2 we explained that to obtain bounds on the norm of the random matrix $X = \sum_k X_k$ one needs to compute all moments $\mathbb{E}\,(X_k^n)$. In this article we will be interested in estimating $\delta(\nu_{\mathcal{S}}, t_1)$ for $\nu_{\mathcal{S}}$ uniform and $\mathcal{S}$ finite with each element of $\mathcal{S}$ chosen from an exact t_2-design $\{\mathcal{D}, \nu_{\mathcal{D}}\}$. Thus in our scenario the role of X is played by

$$T_{\nu_{\mathcal{S}},\lambda} = \int_{U(d)} d\nu_{\mathcal{S}}(U)\pi_\lambda(U) = \frac{1}{|\mathcal{S}|}\sum_{U\in\mathcal{S}} \pi_\lambda(U),$$

X_k's are matrices proportional to $\pi_\lambda(U)$ and the average is taken over $\nu_{\mathcal{D}}$

$$\mathop{\mathbb{E}}_{U\sim\nu_{\mathcal{D}}} \pi_\lambda(U^n) := \int_{U(d)} d\nu_{\mathcal{D}}(U)\pi_\lambda(U)^n,$$

for $\lambda \in \Lambda_{t_1}$ and $n \in \mathbb{Z}$. In particular in Lemma 11 we show that this integral is proportional to $\mathbb{1}_{d_\lambda}$ and the proportionality constant is the Frobenius-Schur indicator $\int_{U(d)} d\mu(U)\chi_\lambda(U^n)$ divided by d_λ.

Lemma 10. *Consider an irreducible finite-dimensional representation π_λ of $U(d)$ with the highest weight $\lambda \in \Lambda_t$, an integer n and an exact $(|n|\cdot t)$-design $\{\mathcal{D}, \nu_{\mathcal{D}}\}$. Then the average of $\pi_\lambda(U)^n$ taken over $\nu_{\mathcal{D}}$ and μ is the same, that is*

$$\mathop{\mathbb{E}}_{U\sim\nu_{\mathcal{D}}} \pi_\lambda(U^n) = \int_{U(d)} d\nu_{\mathcal{D}}(U)\pi_\lambda(U)^n = \int_{U(d)} d\mu(U)\pi_\lambda(U)^n = \mathop{\mathbb{E}}_{U\sim\mu} \pi_\lambda(U^n)$$

Proof. If $n < 0$ we can use unitarity of π_λ to obtain that for any measure ν

$$\int_{U(d)} d\nu(U)\pi_\lambda(U)^n = \int_{U(d)} d\nu(U)\pi_\lambda(U)^{-|n|} = \left(\int_{U(d)} d\nu(U)\pi_\lambda(U)^{|n|}\right)^\dagger.$$

Therefore it is enough to prove the Lemma for $n \geq 0$. In such case note that

$$U^{\otimes nt} \otimes \bar{U}^{\otimes nt} \simeq \left[U^{\otimes t} \otimes \bar{U}^{\otimes t}\right]^{\otimes n} \simeq \left[\bigoplus_{\lambda\in\Lambda_t} \pi_\lambda(U)^{\oplus m^\lambda}\right]^{\otimes n}.$$

Entries of the right-hand side operator above are

$$\pi_{\lambda_1}(U)_{i_1 j_1} \cdot \pi_{\lambda_2}(U)_{i_2 j_2} \cdot \ldots \cdot \pi_{\lambda_n}(U)_{i_n j_n}, \tag{28}$$

for all possible $\lambda_1, \ldots, \lambda_n \in \Lambda_t$ and $1 \leq i_k, j_k \leq d_{\lambda_k}$. Since $\nu_{\mathcal{D}}$ is a $(n\cdot t)$-design integrating the expression (28) over $\nu_{\mathcal{D}}$ and μ gives the same result. Thus the Lemma follows from the fact that the entries of $\pi_\lambda(U)^n$ are linear combinations of expressions (28) with $\lambda_1 = \ldots = \lambda_n = \lambda$ and $j_k = i_{k+1}$. □

Lemma 11. *For any irreducible representation π_λ of a compact Lie group G and $n \in \mathbb{Z}$ we have*

$$\int_G \pi_\lambda(U^n)d\mu(U) = \delta_\lambda(n)\mathbb{1}_{d_\lambda},$$
$$\delta_\lambda(n) := \frac{1}{d_\lambda}\int_G \chi_\lambda(U^n)d\mu(U).$$

Proof. Assume that $n \geq 0$. For any $g \in G$ we have

$$\pi_\lambda(g) \int_G \pi_\lambda(U^n) d\mu(U) = \int_G \pi_\lambda(gU^n) d\mu(U) = \int_G \pi_\lambda(gVgV \dots gVg) d\mu(V),$$

where in the last equality we performed a change of variables $U = Vg$ and used the invariance of the Haar measure. Similarly

$$\left(\int_G \pi_\lambda(U^n) d\mu(U) \right) \pi_\lambda(g) = \int_G \pi_\lambda(U^n g) d\mu(U) = \int_G \pi_\lambda(gVgV \dots gVg) d\mu(V),$$

where in the last equality we performed a change of variables $U = gV$ and used the invariance of the Haar measure. In case $n < 0$ the argument is analogous but with the change of variables $U = g^{-1}V$ in the first equation and $U = Vg^{-1}$ in the second. Thus for any $g \in G$, the matrix $\pi^\lambda(g)$ commutes with the integral in (11). By Schur lemma this integral must be proportional to identity. The proportionality constant, $\delta_\lambda(n)$, can be establish by taking trace of both sides of equation (11). □

Corollary 12. *For any irreducible finite-dimensional representation π_λ of $U(d)$ with the highest weight $\lambda \in \Lambda_t$, an integer n and an exact $(|n| \cdot t)$-design $\{\mathcal{D}, \nu_\mathcal{D}\}$ we have*

$$\mathop{\mathbb{E}}_{U \sim \nu_\mathcal{D}} \pi_\lambda(U^n) = \delta_\lambda(n) \mathbb{1}_{d_\lambda}.$$

Proof. From Lemma 10 we know that

$$\mathop{\mathbb{E}}_{U \sim \nu_\mathcal{D}} \pi_\lambda(U^n) = \mathop{\mathbb{E}}_{U \sim \mu} \pi_\lambda(U^n),$$

and from Lemma 11 that

$$\mathop{\mathbb{E}}_{U \sim \mu} \pi_\lambda(U^n) = \delta_\lambda(n) \mathbb{1}_{d_\lambda}.$$

□

From the orthogonality of characters we know that for $\lambda \neq 0$

$$\int_G d\mu(U) \chi_\lambda(U) = \int_G d\mu(U)\, 1 \cdot \chi_\lambda(U) = \int_G d\mu(U) \overline{\chi_0(U)} \chi_\lambda(U) = 0,$$

thus $\delta_\lambda(\pm 1) = 0$. In case of $\delta_\lambda(\pm 2)$ we use a well known fact that

$$\int_G \chi_\lambda \left(U^{\pm 2} \right) d\mu(U)$$

is equal to 1, 0 or -1 for π_λ real, complex or quaternionic respectively. Thus

$$\delta_\lambda(\pm 2) = \begin{cases} \frac{1}{d_\lambda} & \text{if } \pi_\lambda \text{ real} \\ 0 & \text{if } \pi_\lambda \text{ complex} \\ -\frac{1}{d_\lambda} & \text{if } \pi_\lambda \text{ quaternionic} \end{cases}.$$

To calculate $\delta_\lambda(n)$ for $|n| > 2$ we will use the following result from [32].

Fact 10. *Let G be a finite dimensional semisimple Lie group. Let π_{λ^s} be an irreducible representation of G with highest weight λ^s, $\mathcal{W}$ the Weyl group of $\mathfrak{g}$ and ρ the half sum of positive roots. Then for $n \neq 0$*

$$\delta_{\lambda^s}(n) = \frac{1}{d_{\lambda^s}} \sum_{\sigma \in \mathcal{W}} \operatorname{sgn}(\sigma)\, m_{\lambda^s} \left(\frac{\rho - \sigma \cdot \rho}{n} \right).$$

Immediate conclusion from Fact 10 is that there is n_0 such that for $n \geq n_0$ elements $\frac{\rho-\sigma\cdot\rho}{n}$ are integral elements if and only if $\sigma \cdot \rho = \rho$. In the next lemma we calculate n_0 for the group $SU(d)$.

Lemma 13. *For the group $SU(d)$ the constant n_0 is equal to $d+1$ and for $n \geq n_0$ the formula (10) simplifies to:*

$$\delta_{\lambda^s}(n) = \frac{m_{\lambda^s}(0)}{d_{\lambda^s}}. \tag{29}$$

Proof. First, let us calculate ρ:

$$\rho = \frac{1}{2}\sum_{\alpha\in\Phi^+}\alpha = \frac{1}{2}\sum_{1\leq i<j\leq d}\alpha_{i,j} = \frac{1}{2}\sum_{1\leq i<j\leq d} L_i - L_j = \sum_{i=1}^{d}\left(\frac{d-1}{2} - i + 1\right)L_i.$$

Thus for $i \neq j$ we have $\rho_i \neq \rho_j$ and the only $\sigma \in S_d$ such that $\sigma\cdot\rho = \rho$ is the identity, which proves the simplified formula (29). In order to find minimal n_0 note first that

$$(\rho - \sigma\cdot\rho)_i = \rho_i - \rho_{\sigma(i)} = \sigma(i) - i,$$

and that for $\frac{\rho-\sigma\cdot\rho}{n}$ to be an integral element it is required that for all $1 \leq i \leq d-1$

$$\langle \rho - \sigma\cdot\rho | \alpha_{i,i+1}\rangle = (\rho-\sigma\cdot\rho)_i - (\rho-\sigma\cdot\rho)_{i+1} = \sigma(i) - \sigma(i+1) + 1,$$

is divisible by n or equivalently

$$\forall_{1\leq i\leq d-1}\exists_{C_i\in\mathbb{Z}} \quad \sigma(i+1) - \sigma(i) = C_i n + 1. \tag{30}$$

If $n = d$ then $\sigma(i) = i - 1 \bmod d$ satisfies this condition thus $n_0 > d$. On the other hand, for $n > d$ the condition $|\sigma(i+1) - \sigma(i)| \leq d-1$ implies that the constant C_i in (30) has to be zero for all i. Let us choose j such that $j \neq d$ and $\sigma(j) = d$ then

$$1 = C_j n + 1 = \sigma(j+1) - \sigma(j) = \sigma(j+1) - d \quad \Rightarrow \quad \sigma(j+1) = d+1,$$

what is an obvious contradiction and $\sigma^{-1}(d)$ has to be d but this implies $\sigma(d-1) = \sigma(d) - 1 = d - 1$ and analogously for all $1 \leq i \leq d$ we have $\sigma(i) = i$. Therefore for $n > d$ the only σ satisfying (30) is the trivial one. □

In the next lemma we show when the representation π_{λ^s} has the weight 0. In particular it implies that for every $\lambda \in \Lambda_t$ the representation π_{λ^s} has the weight 0.

Lemma 14. *Suppose π_{λ^s} is a representation of $SU(d)$ with the highest weight λ^s. Then the following are equivalent:*

1. *π_{λ^s} has weight 0,*
2. *$\frac{1}{d}\sum_{j=1}^{d-1} j\lambda_j^s$ is an integer,*
3. *there exists an irreducible finite-dimensional representation of $U(d)$ with highest weight λ - π_λ such that $\Sigma(\lambda) = 0$ and $\pi_\lambda|_{SU(d)} = \pi_{\lambda^s}$.*

Proof. We will start with 1) ⇒ 2).
From Fact 7 there exist $c_1, ..., c_{d-1} \in \mathbb{Z}$ such that

$$\lambda^s = \lambda^s - 0 = \sum_{i=1}^{d-1} c_i \alpha_{i,i+1},$$

thus

$$\lambda_j^s = \langle \lambda^s | \alpha_{j,j+1} \rangle = \langle \sum_{i=1}^{d-1} c_i \alpha_{i,i+1} | \alpha_{j,j+1} \rangle = \sum_{i=1}^{d-1} c_i \langle \alpha_{i,i+1} | \alpha_{j,j+1} \rangle = -c_{j-1} + 2c_j - c_{j+1},$$

where we assume $c_0 = 0 = c_d$. It follows

$$\frac{1}{d} \sum_{j=1}^{d-1} j \lambda_j^s = \frac{1}{d} \sum_{j=1}^{d-1} j(-c_{j-1} + 2c_j - c_{j+1}) =$$

$$= \frac{1}{d} \left\{ \sum_{j=1}^{d-2} \left[-(j-1) + 2j - (j+1) \right] c_j + \left[-(d-2) + 2(d-1) \right] c_{d-1} \right\} =$$

$$= \frac{d}{d} c_{d-1} = c_{d-1} \in \mathbb{Z}.$$

2) $\Rightarrow$ 3)
For any $m \in \mathbb{Z}$

$$\lambda = (m + \sum_{i=1}^{d-1} \lambda_i^s,\ m + \sum_{i=2}^{d-1} \lambda_i^s,\ \dots,\ m + \lambda_{d-1}^s,\ m),$$

satisfies $\lambda_i \in \mathbb{Z}$ and $i < j \Rightarrow \lambda_i \geq \lambda_j$. Therefore there exists an irreducible finite-dimensional representation of $U(d)$ with highest weight λ - π_λ. From Lemma 5 clearly $\pi_\lambda|_{SU(d)} = \pi_{\lambda^s}$ and for $m = -\frac{1}{d} \sum_{j=1}^{d-1} j \lambda_j^s$ we have $\Sigma(\lambda) = 0$.

3) $\Rightarrow$ 1)
Let us choose $\forall_{\sigma \in S_d}\ t_\sigma = \frac{1}{d!}$ then $\sum_{\sigma \in S_d} t_\sigma = 1$ and

$$\sum_{\sigma \in S_d} t_\sigma \sigma \cdot \lambda = \frac{1}{d!} (\sum_{\sigma \in S_d} \lambda_{\sigma(1)},\ \dots,\ \sum_{\sigma \in S_d} \lambda_{\sigma(d-1)}) = \frac{1}{d} (\Sigma(\lambda),\ \dots,\ \Sigma(\lambda)) = 0.$$

What is more for $c_1 = \lambda_1$ and $\forall_{2 \leq i \leq d-1}\ c_i = \lambda_{i-1} + \lambda_i$ we have

$$\sum_{i=1}^{d-1} c_i \alpha_{i,i+1} = \lambda = \lambda - 0.$$

Hence by Fact 7 weight 0 is in π_λ what implies that weight 0 is in π_{λ^s}.

□

As an example we next explicitly calculate the value of $\delta_\lambda(n)$ for $G = SU(2)$.

Lemma 15. *The constant in Lemma 11 for $SU(2)$ and non-trivial π_λ is given by*

$$\delta_\lambda(n) = \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{if } n = \pm 1, \\ \frac{1}{d_\lambda} & \text{if } n \in \mathbb{Z} \setminus \{-1, 0, 1\} \text{ and } \lambda^s\text{-even}, \\ -\frac{1}{d_\lambda} & \text{if } n = \pm 2 \text{ and } \lambda^s\text{-odd} \\ 0 & \text{if } n \in \mathbb{Z} \setminus \{-2, 0, 2\} \text{ and } \lambda^s\text{-odd} \end{cases}$$

Proof. The result for $n = 0$ and $n = 1$ follows from definition of δ_λ and irreducibility of π_λ. For remaining n's we will use Fact 10. First, let us notify that for $SU(2)$:

$$\Delta = \{\alpha_{12}\}, \qquad \alpha_{12} = 2, \qquad \rho = 1, \qquad [1\ 2] \cdot \mu = -\mu,$$

then from Kostant formula 8:

$$m_{\lambda^s}(\mu) = \sum_{\sigma \in S_2} \operatorname{sgn}(\sigma)\, p(\sigma \cdot (\lambda^s + \rho) - (\mu + \rho)) = p(\lambda^s - \mu) - p(-\lambda^s - \mu - 2). \tag{31}$$

Recall that $p(m)$ is the number of ways m can be expressed as a linear combinations of positive simple roots with non-negative integer coefficients so in our case

$$p(m) = \begin{cases} 1 & \text{if } m\text{-even and non-negative,} \\ 0 & \text{if } m\text{-odd or negative.} \end{cases}$$

This combined with the fact that $-\lambda^s - \mu - 2 > 0$ implies $\|\mu\|_1 > \|\lambda\|_1$ and Lemma 7 simplifies (31) to

$$m_{\lambda^s}(\mu) = p(\lambda^s - \mu) = \begin{cases} 1 & \text{if } (\lambda^s - \mu)\text{-even and non-negative,} \\ 0 & \text{if } (\lambda^s - \mu)\text{-odd or negative.} \end{cases}$$

Now, we are ready to use formula (10) which for $SU(2)$ has a form

$$\delta_{\lambda^s}(n) = \frac{1}{d_{\lambda^s}} \left(m_{\lambda^s}(0) - m_{\lambda^s}\left(\frac{2}{n}\right) \right).$$

Hence

$$\delta_{\lambda^s}(\pm 2) = \frac{1}{d_{\lambda^s}} \left(m_{\lambda^s}(0) - m_{\lambda^s}(\pm 1) \right) = \begin{cases} \frac{1}{d_{\lambda^s}} & \text{if } \lambda^s\text{-even,} \\ -\frac{1}{d_{\lambda^s}} & \text{if } \lambda^s\text{-odd.} \end{cases},$$

and for $|n| \geq 3$

$$\delta_{\lambda^s}(n) = \frac{1}{d_{\lambda^s}} m_{\lambda^s}(0) = \begin{cases} \frac{1}{d_{\lambda^s}} & \text{if } \lambda^s\text{-even,} \\ 0 & \text{if } \lambda^s\text{-odd.} \end{cases}$$

□

6 Main results

In this section we will consider two types of gate-sets:

1. $\mathcal{S} = \{U_1, \ldots, U_n\}$, where U_k's are independent random unitaries from an exact t-design $\mathcal{D}$ chosen according to the measure $\nu_{\mathcal{D}}$. Such $\mathcal{S}$ will be called *t-random gate-set*.

2. $\mathcal{S} = \{U_1, \ldots, U_n\} \cup \{U_1^{-1}, \ldots, U_n^{-1}\}$, where U_k's are independent random unitaries from an exact t-design $\mathcal{D}$ chosen according to the measure $\nu_{\mathcal{D}}$. Such $\mathcal{S}$ will be called *symmetric t-random gate-set*.

Note that in the symmetric case we have $U_k \in \mathcal{D}$ but not necessarily $U_k^{-1} \in \mathcal{D}$ so to obtain a symmetric t-random gate-set of size $2n$ we draw n gates from $\{\mathcal{D}, \nu_{\mathcal{D}}\}$ and then we add to the set their inverses. We also want to point out that since $\{U(d), \mu\}$ is an exact t-design for any t all theorems we present below apply to gate-sets where gates are Haar random or Haar random with inverses (we could call them ∞-random gate-sets or symmetric ∞-random gate-sets).
In order to simplify the notation we often denote the cardinality of $\mathcal{S}$ by $\mathcal{S}$ (instead of $|\mathcal{S}|$). The moment operators associated with the above two types of random sets of gates are random matrices. When $\mathcal{S}$ is symmetric they are actually random Hermitian matrices. Using inequalities listed in Section 2 we derive upper bounds on $\mathbb{P}\left(\delta(\nu_{\mathcal{S}}, \lambda) \geq \delta\right)$ and $\mathbb{P}\left(\delta(\nu_{\mathcal{S}}, t) \geq \delta\right)$. Before we proceed with concrete inequalities we note that

Lemma 16. *Assume that $\mathcal{S} \subset U(d)$ is a random set of quantum gates and that for every $\lambda \in \Lambda_t$ we have $\mathbb{P}\left(\delta(\nu_{\mathcal{S}}, \lambda) \geq \delta\right) \leq F(\delta, \lambda, \mathcal{S})$. Then*

$$\mathbb{P}\left(\delta(\nu_{\mathcal{S}}, t) \geq \delta\right) \leq \sum_{\lambda \in \Lambda_t} F(\delta, \lambda, \mathcal{S}).$$

Proof. By (24) and using the union bound for probabilities we have

$$\mathbb{P}\left(\delta(\nu_{\mathcal{S}}, t) \geq \delta\right) = \mathbb{P}\left(\sup_{\lambda \in \Lambda_t} \delta(\nu_{\mathcal{S}}, \lambda) \geq \delta\right) \leq \sum_{\lambda \in \Lambda_t} \mathbb{P}\left(\delta(\nu_{\mathcal{S}}, \lambda) \geq \delta\right) \leq \sum_{\lambda \in \Lambda_t} F(\delta, \lambda, \mathcal{S}).$$

□

Thus, to find a bound on $\mathbb{P}\left(\delta(\nu_{\mathcal{S}}, t) \geq \delta\right)$ it is enough to find bounds on $\mathbb{P}\left(\delta(\nu_{\mathcal{S}}, \lambda) \geq \delta\right)$, $\lambda \in \Lambda_t$, which we do in next sections.

6.1 Bernstein type bounds

In this section we derive bounds on $\mathbb{P}\left(\|T_{\nu_{\mathcal{S}}, \lambda}\| \geq \delta\right)$ using Bernstein inequality (see Section 2). It is worth to mention that computationally this is the simplest derivation presented in this paper as it requires only the knowledge of the second moments.

Theorem 17. *Let λ be an element of Λ_t and $\mathcal{S}$ be a t-random gate-set or a symmetric $(2t)$-random gate-set and $\nu_{\mathcal{S}}$ a uniform measure. Then*

$$\mathbb{P}\left(\delta(\nu_{\mathcal{S}}, \lambda) \geq \delta\right) \leq \begin{cases} B_1(\delta, \lambda, \mathcal{S}), & \text{when } \mathcal{S} - \textit{symmetric } (2t)\textit{-random gate-set} \\ B_2(\delta, \lambda, \mathcal{S}), & \text{when } \mathcal{S} - t\textit{-random gate-set} \end{cases}$$

where

$$B_1(\delta, \lambda, \mathcal{S}) = 2d_\lambda \exp\left(\frac{-3\mathcal{S}\delta^2}{6\left(1 + \delta_\lambda(2)\right) + 4\delta}\right),$$

$$B_2(\delta, \lambda, \mathcal{S}) = 2d_\lambda \exp\left(\frac{-3\mathcal{S}\delta^2}{6 + 2\delta}\right),$$

$$\delta_\lambda(2) = \begin{cases} \frac{1}{d_\lambda} & \text{if } \pi_\lambda \text{ real} \\ 0 & \text{if } \pi_\lambda \text{ complex} \\ -\frac{1}{d_\lambda} & \text{if } \pi_\lambda \text{ quaternionic} \end{cases}.$$

Proof. In this proof we will use Fact 3 where we take $X = \sum_k X_k$ to be equal $T_{\nu_{\mathcal{S}}, \lambda} = \frac{1}{\mathcal{S}} \sum_{U \in \mathcal{S}} \pi_\lambda(U)$.
Let us start with the Haar random case. In this case all elements of $\mathcal{S}$ are independent. Thus we will put $X_k = \frac{1}{\mathcal{S}} \pi_\lambda(U_k)$ for $k = 1, ..., \mathcal{S}$. Then from Corollary 12

$$\mathop{\mathbb{E}}_{U \sim \nu_{\mathcal{D}}} X_k = \mathop{\mathbb{E}}_{U \sim \nu_{\mathcal{D}}} \left(\frac{1}{\mathcal{S}} \pi_\lambda(U)\right) = \frac{\delta_\lambda(1)}{\mathcal{S}} \mathbb{1}_{d_\lambda} = 0.$$

Since π_λ is unitary we have

$$\|X_k\| = \frac{1}{\mathcal{S}} \|\pi_\lambda(U_k)\| = \frac{1}{\mathcal{S}},$$

$$X_k X_k^\dagger = X_k^\dagger X_k = \frac{1}{\mathcal{S}^2} \mathbb{1}_{\delta_\lambda}.$$

Hence $L = \frac{1}{\mathcal{S}}$ and $\nu = \| \sum_{k=1}^{\mathcal{S}} \mathop{\mathbb{E}}_{U\sim\nu_{\mathcal{D}}} (X_k X_k^\dagger) \| = \frac{1}{\mathcal{S}}$ satisfy conditions from Fact 3 and the result follows.

When $\mathcal{S}$ is symmetric we put $X_k = \frac{1}{\mathcal{S}} \left(\pi_\lambda (U) + \pi_\lambda \left(U^{-1} \right) \right)$. From Corollary 12 we have $\mathop{\mathbb{E}}_{U\sim\nu_{\mathcal{D}}} X_k = 0$. Then using triangle inequality we get

$$\|X_k\| = \frac{1}{\mathcal{S}} \|\pi_\lambda(U) + \pi_\lambda(U^{-1})\| \leq \frac{2}{\mathcal{S}} \|\pi_\lambda(U)\| = \frac{2}{\mathcal{S}},$$

and from hermiticity of X_k

$$X_k X_k^\dagger = X_k^\dagger X_k = \frac{1}{\mathcal{S}^2} \left(\pi_\lambda \left(U^2 \right) + 2\mathbb{1}_{d_\lambda} + \pi_\lambda \left(U^{-2} \right) \right).$$

Then using Corollary 12 once more we obtain

$$\mathop{\mathbb{E}}_{U\sim\nu_{\mathcal{D}}} \left(X_k X_k^\dagger \right) = \frac{2}{\mathcal{S}^2} \left(\delta_\lambda(2) + 1 \right) \mathbb{1}_{d_\lambda}.$$

Hence for $L = \frac{2}{\mathcal{S}}$ and

$$\nu = \| \sum_{k=1}^{\mathcal{S}/2} \mathop{\mathbb{E}}_{U\sim\nu_{\mathcal{D}}} (X_k X_k^\dagger) \| = \frac{\delta_\lambda(2) + 1}{\mathcal{S}},$$

the result follows from Fact 3.

□

6.2 Master bounds

In this section we derive Theorems 1 and 2 from Introduction using the formula for the master bound (see Section 2). This derivation requires knowledge of all moments and is much more accurate than the Bernstein bound given in the previous section.

6.2.1 Haar random gate-sets

For t-random gate-sets the derivation of the bound for $\mathbb{P} \left(\delta(\nu_{\mathcal{S}}, \lambda) \geq \delta \right)$ turns out to be relatively simple.

Theorem 18. *Let λ be an element of Λ_t and $\mathcal{S}$ be a t-random gate-set and $\nu_{\mathcal{S}}$ a uniform measure. Then for any $\delta < 1$*

$$\mathbb{P} \left(\delta(\nu_{\mathcal{S}}, \lambda) \geq \delta \right) \leq F(\delta, \lambda, \mathcal{S}),$$

where

$$F(\delta, \lambda, \mathcal{S}) = \frac{2d_\lambda}{(1+\delta)^{\frac{\mathcal{S}}{2}(1+\delta)} (1-\delta)^{\frac{\mathcal{S}}{2}(1-\delta)}} = \frac{2d_\lambda}{(1-\delta^2)^{\frac{\mathcal{S}}{2}}} e^{-\delta \mathcal{S} \operatorname{arctanh}(\delta)}.$$

Proof. To compute the master bound for non-Hermitian matrix (Fact 2) we need to compute $\mathop{\mathbb{E}}_{U\sim\nu_{\mathcal{D}}} \exp \left(\frac{\theta}{\mathcal{S}} \mathcal{H}(\pi_\lambda(U)) \right)$. First, let us note that for any unitary matrix $U \in U(d_\lambda)$:

$$\mathcal{H}(U)^2 = \begin{pmatrix} 0 & U \\ U^\dagger & 0 \end{pmatrix}^2 = \begin{pmatrix} UU^\dagger & 0 \\ 0 & U^\dagger U \end{pmatrix} = \mathbb{1}_{2d_\lambda}.$$

Since π_λ is unitary we have

$$\mathop{\mathbb{E}}_{U\sim\nu_{\mathcal{D}}} \exp\left(\frac{\theta}{\mathcal{S}}\mathcal{H}(\pi_\lambda(U))\right) = \cosh\left(\frac{\theta}{\mathcal{S}}\right)\mathbb{1}_{2d_\lambda} + \sinh\left(\frac{\theta}{\mathcal{S}}\right)\mathop{\mathbb{E}}_{U\sim\nu_{\mathcal{D}}}\mathcal{H}(\pi_\lambda(U)).$$

From Corollary 12 and the orthogonality of characters:

$$\mathop{\mathbb{E}}_{U\sim\nu_{\mathcal{D}}}\mathcal{H}(\pi_\lambda(U)) = \mathcal{H}\left(\mathop{\mathbb{E}}_{U\sim\nu_{\mathcal{D}}}\pi_\lambda(U)\right) = 0.$$

Hence the right hand side of the master bound is:

$$\inf_{\theta>0} e^{-\theta\delta}\operatorname{tr}\exp\left(\mathcal{S}\log\mathop{\mathbb{E}}_{U\sim\nu_{\mathcal{D}}} e^{\frac{\theta}{\mathcal{S}}\mathcal{H}(X_k)}\right) = \inf_{\theta>0} 2d_\lambda\, e^{-\theta\delta}\cosh^{\mathcal{S}}\left(\frac{\theta}{\mathcal{S}}\right).$$

By calculating derivative we easily get that the infimum is obtained for $\theta = \mathcal{S}\operatorname{arctanh}(\delta)$ and the bound is:

$$F(\delta,\lambda,\mathcal{S}) = \frac{2d_\lambda}{(1+\delta)^{\frac{\mathcal{S}}{2}(1+\delta)}(1-\delta)^{\frac{\mathcal{S}}{2}(1-\delta)}} = \frac{2d_\lambda}{(1-\delta^2)^{\frac{\mathcal{S}}{2}}}e^{-\delta\mathcal{S}\operatorname{arctanh}(\delta)}.$$

□

6.2.2 Symmetric Haar random gate-sets

In the following we develop new bounds which are based on the fact that we can calculate explicitly the expected values that appear in the master bound (see Section 2). In order to perform this calculation we need to know all moments $\delta_\lambda(n)$. To do this for $n \leq d$ we use an explicit formula from Fact 10 and for $n > d$ we use it simplified form from Lemma 13. We derive the bound in Theorem 19 which, using Fact 11, we next optimize to obtain the main result of this section, that is, Theorem 20.

Theorem 19. *Let $\mathcal{S}$ be a symmetric Haar random gate-set from $SU(d)$ and $\nu_{\mathcal{S}}$ a uniform measure. Then*

$$\mathbb{P}\left(\delta(\nu_{\mathcal{S}},\lambda^s)\| \geq \delta\right) \leq d_{\lambda^s}\left[\inf_{\theta>0} e^{-\theta\delta}F(\theta,\lambda^s,\mathcal{S}) + \inf_{\theta>0} e^{-\theta\delta}F(-\theta,\lambda^s,\mathcal{S})\right], \tag{32}$$

where

$$F(\theta,\lambda^s,\mathcal{S}) = \left[\frac{m^0_{\lambda^s}}{d_{\lambda^s}}e^{\frac{2\theta}{\mathcal{S}}} + \sum_{k=-d}^{d}\gamma_{\lambda^s}(k)I_{|k|}\left(\frac{2\theta}{\mathcal{S}}\right)\right]^{\frac{\mathcal{S}}{2}}, \tag{33}$$

$$\gamma_{\lambda^s}(k) = \begin{cases} 1 - \frac{m^0_{\lambda^s}}{d_{\lambda^s}} & \text{if } k = 0, \\ \frac{1}{d_{\lambda^s}}\sum_{\substack{\sigma\in S_d \\ \sigma\neq id}}\operatorname{sgn}(\sigma)m_{\lambda^s}\left(\frac{\rho-\sigma\cdot\rho}{k}\right) & \text{if } k \neq 0, \end{cases}$$

and $I_n(x)$ is n-th modified Bessel function of the first kind.

Proof. Clearly, the bound from Fact 1 is of the form

$$\inf_{\theta>0} e^{-\theta\delta}\tilde{F}(\theta,\lambda^s,\mathcal{S}) + \inf_{\theta>0} e^{-\theta\delta}\tilde{F}(-\theta,\lambda^s,\mathcal{S}),$$

where

$$\tilde{F}(\theta,\lambda^s,\mathcal{S}) = \operatorname{tr}\exp\left(\sum_{k=1}^{\mathcal{S}/2}\log \mathop{\mathbb{E}}_{U\sim\mu} e^{\frac{\theta}{\mathcal{S}}X_{U_k,\lambda^s}}\right),$$
$$X_{U,\lambda^s} = \pi_{\lambda^s}(U) + \pi_{\lambda^s}(U^{-1}).$$

Thus our main objective is to compute the $\mathop{\mathbb{E}}_{U\sim\mu} e^{\frac{\theta}{\mathcal{S}}X_{U,\lambda^s}}$. Using the binomial formula one easily gets:

$$\mathop{\mathbb{E}}_{U\sim\mu} X^n_{U,\lambda^s} = \sum_{k=0}^{n}\binom{n}{k}\mathop{\mathbb{E}}_{U\sim\mu}\pi_{\lambda^s}(U^{n-2k}) = \mathbb{1}_{\lambda^s}\left[\sum_{k=0}^{n}\binom{n}{k}\delta_{\lambda^s}(n-2k)\right]. \tag{34}$$

From Lemma 13 we know that for $|n-2k| > d$ the value of δ_{λ^s} is $\frac{m^0_{\lambda^s}}{d_{\lambda^s}}$. Thus Eq. (34) becomes

$$\mathop{\mathbb{E}}_{U\sim\mu} X^n_{U,\lambda^s} = \mathbb{1}_{\lambda^s}\left[\sum_{k=0}^{n}\binom{n}{k}\frac{m^0_{\lambda^s}}{d_{\lambda^s}} + \sum_{k=\lceil\frac{n-d}{2}\rceil}^{\lfloor\frac{n+d}{2}\rfloor}\binom{n}{k}\left(\delta_{\lambda^s}(n-2k) - \frac{m^0_{\lambda^s}}{d_{\lambda^s}}\right)\right] =$$
$$= \mathbb{1}_{\lambda^s}\left[\frac{m^0_{\lambda^s}\,2^n}{d_{\lambda^s}} + \sum_{k=\lceil\frac{n-d}{2}\rceil}^{\lfloor\frac{n+d}{2}\rfloor}\binom{n}{k}\gamma_{\lambda^s}(n-2k)\right],$$

where we assume $\binom{n}{k} = 0$ for $n < k$ or $k < 0$. We will consider separately cases for different parities of d and n:

- $d = 2p$ and $n = 2l$:

$$\mathop{\mathbb{E}}_{U\sim\mu} X^{2l}_{U,\lambda^s} = \mathbb{1}_{\lambda^s}\left[\frac{m^0_{\lambda^s}\,2^{2l}}{d_{\lambda^s}} + \sum_{k=l-p}^{l+p}\binom{2l}{k}\gamma_{\lambda^s}(2l-2k)\right] =$$
$$= \mathbb{1}_{\lambda^s}\left[\frac{m^0_{\lambda^s}\,2^{2l}}{d_{\lambda^s}} + \sum_{k=-p}^{p}\binom{2l}{l+k}\gamma_{\lambda^s}(2k)\right],$$

- $d = 2p$ and $n = 2l+1$:

$$\mathop{\mathbb{E}}_{U\sim\mu} X^{2l+1}_{U,\lambda^s} = \mathbb{1}_{\lambda^s}\left[\frac{m^0_{\lambda^s}\,2^{2l+1}}{d_{\lambda^s}} + \sum_{k=l+1-p}^{l+p}\binom{2l+1}{k}\gamma_{\lambda^s}(2l+1-2k)\right] =$$
$$= \mathbb{1}_{\lambda^s}\left[\frac{m^0_{\lambda^s}\,2^{2l+1}}{d_{\lambda^s}} + \sum_{k=-p+1}^{p}\binom{2l+1}{l+k}\gamma_{\lambda^s}(2k-1)\right].$$

Then we have:

$$\mathop{\mathbb{E}}_{U\sim\mu} e^{\frac{\theta}{\mathcal{S}}X_{U,\lambda^s}} = \mathbb{1}\left[\frac{m^0_{\lambda^s}}{d_{\lambda^s}}e^{\frac{2\theta}{\mathcal{S}}} + \sum_{k=-p}^{p}\gamma_{\lambda^s}(2k)I_{|2k|}\left(\frac{2\theta}{\mathcal{S}}\right) + \sum_{k=-p+1}^{p}\gamma_{\lambda^s}(2k-1)I_{|2k-1|}\left(\frac{2\theta}{\mathcal{S}}\right)\right] =$$
$$= \mathbb{1}\left[\frac{m^0_{\lambda^s}}{d_{\lambda^s}}e^{\frac{2\theta}{\mathcal{S}}} + \sum_{k=-d}^{d}\gamma_{\lambda^s}(k)I_{|k|}\left(\frac{2\theta}{\mathcal{S}}\right)\right]. \tag{35}$$

- $d = 2p+1$ and $n = 2l$:

$$\mathop{\mathbb{E}}_{U\sim\mu} X^{2l}_{U,\lambda^s} = \mathbb{1}_{\lambda^s}\left[\frac{m^0_{\lambda^s}\, 2^{2l}}{d_{\lambda^s}} + \sum_{k=l-p}^{l+p}\binom{2l}{k}\gamma_{\lambda^s}(2l-2k)\right] = \\ = \mathbb{1}_{\lambda^s}\left[\frac{m^0_{\lambda^s}\, 2^{2l}}{d_{\lambda^s}} + \sum_{k=-p}^{p}\binom{2l}{l+k}\gamma_{\lambda^s}(2k)\right],$$

- $d = 2p+1$ and $n = 2l+1$:

$$\mathop{\mathbb{E}}_{U\sim\mu} X^{2l+1}_{U,\lambda^s} = \mathbb{1}_{\lambda^s}\left[\frac{m^0_{\lambda^s}\, 2^{2l+1}}{d_{\lambda^s}} + \sum_{k=l-p}^{l+p+1}\binom{2l+1}{k}\gamma_{\lambda^s}(2l+1-2k)\right] = \\ = \mathbb{1}_{\lambda^s}\left[\frac{m^0_{\lambda^s}\, 2^{2l+1}}{d_{\lambda^s}} + \sum_{k=-p}^{p+1}\binom{2l+1}{l+k}\gamma_{\lambda^s}(2k-1)\right].$$

Then we have:

$$\mathop{\mathbb{E}}_{U\sim\mu} e^{\frac{\theta}{\mathcal{S}}X_{U,\lambda^s}} = \mathbb{1}\left[\frac{m^0_{\lambda^s}}{d_{\lambda^s}}e^{\frac{2\theta}{\mathcal{S}}} + \sum_{k=-p}^{p}\gamma_{\lambda^s}(2k)I_{|2k|}\left(\frac{2\theta}{\mathcal{S}}\right) + \sum_{k=-p}^{p+1}\gamma_{\lambda^s}(2k-1)I_{|2k-1|}\left(\frac{2\theta}{\mathcal{S}}\right)\right] = \\ = \mathbb{1}\left[\frac{m^0_{\lambda^s}}{d_{\lambda^s}}e^{\frac{2\theta}{\mathcal{S}}} + \sum_{k=-d}^{d}\gamma_{\lambda^s}(k)I_{|k|}\left(\frac{2\theta}{\mathcal{S}}\right)\right]. \tag{36}$$

Combining Eq. 35 and Eq. 36 with Fact 1 we get the desired result.

□

Note that from Kostant formula (8) the multiplicities $m^{\mu}_{\lambda^s}$ are polynomials in coefficients of λ^s with degree up to $\frac{(d-2)(d-1)}{2}$ while from Weyl dimension formula (20) the dimension d_{λ^s} is a polynomial of degree $\frac{(d-1)d}{2}$. Thus for λ^s with large coefficients the expression in bracket in Eq. (32) is approximately equal to $\gamma_{\lambda^s}(0)I_0(\frac{2\theta}{\mathcal{S}})$. In order to proceed we need the following property of the modified Bessel functions.

Fact 11. *[45] For any $n \geq 0$ we have the following upper and lower bounds on the ratio of the modified Bessel functions*

$$\frac{x}{n-\frac{1}{2}+\sqrt{\left(n+\frac{1}{2}\right)^2+x^2}} < \frac{I_n(x)}{I_{n-1}(x)} < \frac{x}{n-1+\sqrt{(n+1)^2+x^2}}.$$

Combining Fact 11 with Theorem 19 we arrive at

Theorem 20. *Let $\mathcal{S}$ be a symmetric Haar random gate-set from $SU(d)$ and $\nu_{\mathcal{S}}$ a uniform measure. Then*

$$\mathbb{P}\left(\delta(\nu_{\mathcal{S}},\lambda) \geq \delta\right) \leq d_{\lambda^s}\, e^{-\frac{\mathcal{S}\delta^2}{\sqrt{1-\delta^2}}}\left[F\left(\frac{\mathcal{S}\delta}{\sqrt{1-\delta^2}},\lambda,\mathcal{S}\right) + F\left(-\frac{\mathcal{S}\delta}{\sqrt{1-\delta^2}},\lambda,\mathcal{S}\right)\right],$$

where $F(\cdot,\cdot,\cdot)$ is given by (33).

Proof. In order to find the best bound we need to determine θ that realizes

$$\inf_{\theta>0} e^{-\theta\delta}\left[\frac{m^0_{\lambda^s}}{d_{\lambda^s}}e^{\frac{2\theta}{\mathcal{S}}}+\sum_{k=-d}^{d}\gamma_{\lambda^s}(k)I_{|k|}\left(\frac{2\theta}{\mathcal{S}}\right)\right]^{\frac{\mathcal{S}}{2}}.$$

As finding minimum of the above functions is analytically intractable, in both cases we look for θ that minimizes $e^{-\theta\delta}I_0\left(\frac{2\theta}{\mathcal{S}}\right)^{\frac{\mathcal{S}}{2}}$. Taking the derivative with respect to θ we get

$$\delta=\frac{I_1(x)}{I_0(x)},$$

where $x=\frac{2\theta}{\mathcal{S}}$. Using the upper bound for the ration of modified Bessel functions from Fact 11 we get

$$\delta=\frac{x}{\sqrt{4+x^2}}\implies x=\frac{2\delta}{\sqrt{1-\delta^2}}.$$

The result follows. □

6.3 Bernstein and master bounds for random circuits

Let us consider a scenario where we draw m independent gates from the exact t_0-design $\{\mathcal{D},\nu_{\mathcal{D}}\}$ and combine them into a circuit $\mathbf{U}=(U_1,...,U_m)$. We repeat this procedure some number of times and in this way we construct a set of random circuits $\mathcal{R}\subset\mathcal{D}^{\times m}$. Then for $f\in\mathcal{H}_{t,m}$ we ask how much averaging of f over $\mathcal{R}$ differs from averaging over all circuits $\mathcal{D}^{\times m}$. More precisely we want to bound the tail probability

$$\mathbb{P}(\|T_{\nu_{\mathcal{D}}^{\times m},t}-T_{\nu_{\mathcal{R}},t}\|\geq\delta),$$

for $\nu_{\mathcal{R}}$ uniform, $0<\delta<1$ and $t\leq t_0$. Note that since $\mathcal{D}$ is an exact t_0-design for any $t\leq t_0$ we have:

$$T_{\nu_{\mathcal{D}}^{\times m},t}=\mathop{\mathbb{E}}_{U_1\sim\nu_{\mathcal{D}}}...\mathop{\mathbb{E}}_{U_m\sim\nu_{\mathcal{D}}}\bigotimes_{i=1}^{m}U_i^{t,t}=\bigotimes_{i=1}^{m}\left(\mathop{\mathbb{E}}_{U_i\sim\nu_{\mathcal{D}}}U_i^{t,t}\right)=\bigotimes_{i=1}^{m}T_{\mu,t}=T_{\mu,t}^{\otimes m}.$$

Thus

$$\mathbb{P}(\|T_{\nu_{\mathcal{D}}^{\times m},t}-T_{\nu_{\mathcal{C}},t}\|\geq\delta)=\mathbb{P}(\delta(\nu_{\mathcal{C}},t)\geq\delta)\leq\sum_{\boldsymbol{\lambda}\in\Lambda_{t,m}}\mathbb{P}(\delta(\nu_{\mathcal{C}},\boldsymbol{\lambda})\geq\delta).$$

With this we can prove Bernstein and master bounds for random circuits in a very similar way to the case with random gates. For example, consider calculation of moments of $\pi_{\boldsymbol{\lambda}}(\mathbf{U})$ for $n\cdot t\leq t_0$:

$$\mathop{\mathbb{E}}_{U_1\sim\nu_{\mathcal{D}}}...\mathop{\mathbb{E}}_{U_m\sim\nu_{\mathcal{D}}}\pi_{\boldsymbol{\lambda}}(\mathbf{U})^n=\bigotimes_{i=1}^{m}\left(\mathop{\mathbb{E}}_{U_i\sim\nu_D}\pi_{\lambda_i}(U_i)^n\right)=\prod_{i=1}^{m}\delta_{\lambda_i}(n)\mathbb{1}.$$

Therefore Bernstein, symmetric Bernstein and master bounds for random circuits are almost exactly the same as for random gates with only small changes summarized in the below diagram:

$$\begin{array}{c}\mathcal{S}\longrightarrow\mathcal{C},\\ \sum_{\lambda\in\Lambda_t}\longrightarrow\sum_{\boldsymbol{\lambda}\in\Lambda_{t,m}},\end{array}$$

and then for $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_m) \in \Lambda_{t,m}$:

$$\delta_\lambda(n) \longrightarrow \delta_{\boldsymbol{\lambda}}(n) := \prod_{i=1}^{m} \delta_{\lambda_i}(n),$$
$$d_\lambda \longrightarrow d_{\boldsymbol{\lambda}} := \prod_{i=1}^{m} d_{\lambda_i}.$$

In case of symmetric master bound we have to additionally substitute

$$m^0_{\lambda^s} \longrightarrow m^0_{\boldsymbol{\lambda}^s} := \prod_{i=1}^{m} m^0_{\lambda^s_i},$$

and redefine $\gamma_{\lambda^s}(n)$ to

$$\gamma_{\boldsymbol{\lambda}^s}(k) = \begin{cases} 1 - \prod_{i=1}^{m} \frac{m^0_{\lambda^s_i}}{d_{\lambda^s_i}} & \text{if } k = 0, \\ \prod_{i=1}^{m} \left[\frac{1}{d_{\lambda^s_i}} \sum_{\sigma \in S_d} \operatorname{sgn}(\sigma) m_{\lambda^s_i} \left(\frac{\rho - \sigma \cdot \rho}{k} \right) \right] - \prod_{i=1}^{m} \frac{m^0_{\lambda^s_i}}{d_{\lambda^s_i}} & \text{if } k \neq 0. \end{cases}$$

6.4 Concentration of $\delta(\nu_{\mathcal{S}}, t)$ and $\delta(\nu_{\mathcal{S}}, \lambda)$ around their expected values

In this section we derive Theorems 3 and 4 from Introduction about the concentration properties of $\delta(\nu_{\mathcal{S}}, t)$ and $\delta(\nu_{\mathcal{S}}, \lambda)$ using Fact 4.

Theorem 21. *Let $\mathcal{S} \subset SU(d)$ be a Haar random gate-set. Then*

$$\mathbb{P}\left(\delta(\nu_{\mathcal{S}}, t) \geq \mathop{\mathbb{E}}_{U \sim \mu} \delta(\nu_{\mathcal{S}}, t) + \alpha \right) \leq \exp\left(\frac{-d\mathcal{S}\alpha^2}{32t^2} \right).$$

Proof. Let $\mathcal{S} \subset SU(d)$ be any set of cardinality $n > 1$. For the convenience we denote by $\mathcal{S}_1 = \{U_1, \ldots, U_n\}$, and $\mathcal{S}_2 = \{V_1, \ldots, V_n\}$ two exemplary sets $\mathcal{S}$. Let

$$F(U_1, \ldots, U_n) = \| T_{\nu_{\mathcal{S}_1}, t} - T_{\mu, t} \|.$$

We need to show that there is a constant L such that for any $\mathcal{S}_1$ and $\mathcal{S}_2$ we have

$$|F(U_1, \ldots, U_n) - F(V_1, \ldots, V_n)| \leq L \left(\sum_{i=1}^{n} \|U_i - V_i\|_F^2 \right)^{\frac{1}{2}}.$$

In order to determine the value of L we go through the following chain of inequalities

$$\Big| \|T_{\nu_{\mathcal{S}_1}, t} - T_{\mu, t}\| - \|T_{\nu_{\mathcal{S}_2}, t} - T_{\mu, t}\| \Big| \leq \|T_{\nu_{\mathcal{S}_1}, t} - T_{\nu_{\mathcal{S}_2}, t}\| \leq \frac{1}{n} \sum_{i=1}^{n} \|U_i^{t,t} - V_i^{t,t}\| \leq$$
$$\leq \frac{2t}{n} \sum_{i=1}^{n} \|U_i - V_i\| \leq \frac{2t}{n} \sum_{i=1}^{n} \|U_i - V_i\|_F \leq \frac{2t}{\sqrt{n}} \left(\sum_{i=1}^{n} \|U_i - V_i\|_F^2 \right)^{\frac{1}{2}}.$$

Thus the Lipschitz constant is $L = \frac{2t}{\sqrt{\mathcal{S}}}$. Knowing the value of L we use Fact 4 and obtain the result. □

Theorem 22. *Let $\mathcal{S} \subset SU(d)$ be a Haar random gate-set. Then*

$$\mathbb{P}\left(\delta(\nu_{\mathcal{S}}, \lambda) \geq \mathop{\mathbb{E}}_{U \sim \mu} \delta(\nu_{\mathcal{S}}, \lambda) + \alpha \right) \leq \exp\left(\frac{-d\mathcal{S}\alpha^2}{2\pi^2 \|\lambda\|_1^2} \right).$$

Proof. Let $\mathcal{S} \subset SU(d)$ be any set of cardinality $n > 1$. For the convenience we denote by $\mathcal{S}_1 = \{U_1, \ldots, U_n\}$, and $\mathcal{S}_2 = \{V_1, \ldots, V_n\}$ two exemplary sets $\mathcal{S}$. Let

$$F(U_1, \ldots, U_n) = \|T_{\nu_{\mathcal{S}_1}, \lambda}\|.$$

We need to show that there is a constant L such that for any $\mathcal{S}_1$ and $\mathcal{S}_2$ we have

$$|F(U_1, \ldots, U_n) - F(V_1, \ldots, V_n)| \leq L \left(\sum_{i=1}^{n} \|U_i - V_i\|_F^2 \right)^{\frac{1}{2}}.$$

In order to determine the value of L we go through the following chain of inequalities

$$\left| \|T_{\nu_{\mathcal{S}_1}, \lambda}\| - \|T_{\nu_{\mathcal{S}_2}, \lambda}\| \right| \leq \frac{1}{n} \sum_{i=1}^{n} \|\pi_\lambda(U_i) - \pi_\lambda(V_i)\| \leq$$

$$\leq \frac{\pi \|\lambda\|_1}{2n} \sum_{i=1}^{n} \|U_i - V_i\| \leq \frac{\pi \|\lambda\|_1}{2\sqrt{n}} \left(\sum_{i=1}^{n} \|U_i - V_i\|_F^2 \right)^{\frac{1}{2}},$$

where in the second inequality we used Lemma 9. Thus the Lipschitz constant is $L = \frac{\pi \|\lambda\|_1}{2\sqrt{\mathcal{S}}}$. Knowing the value of L we use Fact 4 and obtain the result. □

6.4.1 d-mode beamsplitters built from random 2-mode beamsplitters

In this section we consider the Hilbert space $\mathcal{H} = \mathcal{H}_1 \oplus \ldots \oplus \mathcal{H}_d$, where $\mathcal{H}_k \simeq \mathbb{C}$, $d > 2$. We will call spaces $\mathcal{H}_k$ modes. For a matrix $B \in SU(2)$, which we call a 2-mode beamsplitter, we define matrices B^{ij}, $i \neq j$, to be the matrices that act on a 2-dimensional subspace $\mathcal{H}_i \oplus \mathcal{H}_j \subset \mathcal{H}$ as B and on the other components of $\mathcal{H}$ as the identity. This way a matrix $B \in SU(2)$ gives $d(d-1)$ matrices in $SU(d)$. Applying this procedure to a Haar random gate-set set $S \subset SU(2)$ we obtain random gate-set S^d (see [37, 38]). We are interested in efficiency of $\mathcal{S}^d$.

Theorem 23. *Let $\mathcal{S} \subset SU(2)$ be a Haar random gate-set and $\mathcal{S}^d \subset SU(d)$ the corresponding d-mode gate-set. Then*

$$\mathbb{P}\left(\delta(\nu_{\mathcal{S}^d}, t) \geq \mathop{\mathbb{E}}_{U \sim \mu} \delta(\nu_{\mathcal{S}^d}, t) + \alpha \right) \leq \exp\left(\frac{-\mathcal{S}\alpha^2}{16t^2} \right).$$

Proof. Let $\mathcal{S}_1, \mathcal{S}_2 \subset SU(2)$ be two gate-sets of the same size n, i.e. $\mathcal{S}_1 = \{U_1, \ldots, U_n\}$, and $\mathcal{S}_2 = \{V_1, \ldots, V_n\}$. We denote by $\mathcal{S}_1^d$, and $\mathcal{S}_2^d$ the corresponding d-mode gate-sets

$$\mathcal{S}_1^d = \{U_k^{i,j} | k \in \{1, \ldots, n\},\, i, j \in \{1, \ldots, d\}, i \neq j\},$$
$$\mathcal{S}_2^d = \{V_k^{i,j} | k \in \{1, \ldots, n\},\, i, j \in \{1, \ldots, d\}, i \neq j\}.$$

Similarly as in the proofs of Theorems 21 and 22 we calculate,

$$\left| \|T_{\nu_{\mathcal{S}_1^d}, t} - T_{\mu, t}\| - \|T_{\nu_{\mathcal{S}_2^d}, t} - T_{\mu, t}\| \right| \leq \|T_{\nu_{\mathcal{S}_1^d}, t} - T_{\nu_{\mathcal{S}_2^d}, t}\| \leq$$

$$\leq \frac{1}{d(d-1)} \sum_{1 \leq i \neq j \leq d} \frac{1}{n} \sum_{k=1}^{n} \|(U_k^{i,j})^{t,t} - (V_k^{i,j})^{t,t}\| \leq \frac{1}{d(d-1)} \sum_{1 \leq i \neq j \leq d} \frac{2t}{n} \sum_{k=1}^{n} \|U_k - V_k\|_F \leq$$

$$\frac{1}{d(d-1)} \sum_{1 \leq i \neq j \leq d} \frac{2t}{\sqrt{n}} \left(\sum_{k=1}^{n} \|U_k - V_k\|_F^2 \right)^{\frac{1}{2}} = \frac{2t}{\sqrt{n}} \left(\sum_{k=1}^{n} \|U_k - V_k\|_F^2 \right)^{\frac{1}{2}}.$$

Thus the Lipschitz constant is $L = \frac{2t}{\sqrt{\mathcal{S}}}$. Knowing the value of L we use Fact 4 and obtain the result. □

As a conclusion we see that a Haar random gate-set $\mathcal{S} \subset SU(2)$ gives the gate-set $S^d \subset SU(d)$ for which $\delta(\nu_{\mathcal{S}^d}, t)$ has the same concentration rate around the mean as a Haar random gate-set $\mathcal{S}' \subset SU(d)$ of size:

$$\mathcal{S}' = \frac{2\mathcal{S}}{d}.$$

We note, however, that using this approach one cannot say anything about the relationship between $\mathop{\mathbb{E}}_{U\sim\mu} \delta(\nu_{\mathcal{S}^d}, t)$ and $\mathop{\mathbb{E}}_{U\sim\mu} \delta(\nu_{\mathcal{S}'}, t)$.

7 Comparison of bounds

In this section we use numerical results to compare derived bounds for various values of t (Fig. 1), d (Fig. 2) and $\mathcal{S}$ (Fig. 3 and Fig. 4). Throughout this section we will use the following convention:

- Bounds using Theorem 17 will be called Bernstein bounds and symmetric Bernstein bounds. They will be plotted with a dashed yellow line and a solid yellow line respectively.
- Bounds using Theorem 18 will be called master bounds and they will be plotted with a dashed blue line.
- Bounds using Theorem 19 will be called symmetric master bounds and they will be plotted with a solid blue line.
- Bounds using Theorem 20 will be called simplified symmetric master bound and they will be plotted with solid orange line with crosses.

We assume that gate-sets are $(c\cdot t)$-random with c chosen in such a way that the Theorems mentioned above can be applied to $\delta(\nu_{\mathcal{S}}, t)$, that is:

$$c = \begin{cases} 1 & \text{for Bernstein and master bounds,} \\ 2 & \text{for symmetric Bernstein bounds,} \\ \infty & \text{for symmetric master bounds.} \end{cases}$$

First conclusion from Figures 1, 2, 3 and 4 is that the (symmetric) master bound is tighter than the (symmetric) Bernstein bound. The difference gets more pronounced with bigger t and d and smaller $\mathcal{S}$. Next, note that simplified symmetric master bound is almost identical with symmetric master bound what implies that our guess in derivation of Theorem 20 was very close to optimal, even for small t. This can be explained by the fact that the functions under the infimum from (32) have the derivatives very close to zero in a quite wide interval near the minimum (see Fig. 5). Thus the range of close to optimal guesses for infimum is quite wide as well.

We also analyze what is (according to our bounds) the gain of adding inverses to random gate-sets i.e. by making them symmetric. Figures 1, 2 and 3 indicate that bounds for set with inverses are tighter but the difference decreases with growing $\mathcal{S}$. On the other hand, when we compare gate-sets of the same size (see Fig. 4) then bounds for sets without inverses are better. Thus, we conclude that new random gates improve $\delta(\nu_{\mathcal{S}}, t)$ more significantly than additional inverses of gates that were already in the set.

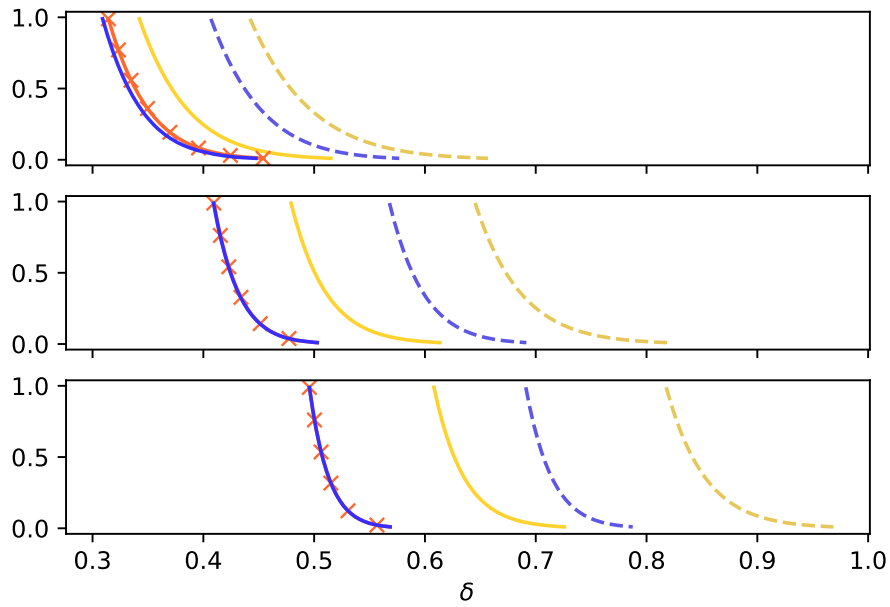

Figure 1: Upper bounds on $\mathbb{P}(\|T_{\nu_{\mathcal{S}},t}\| \geq \delta)$ for $d = 2$, $\mathcal{S} = 50$ or in symmetric case $\mathcal{S} = 2 \times 50$ and different t: top - 5, middle - 50, bottom - 500. Master - dashed blue, symmetric master - solid blue, simplified symmetric master - solid orange with x-markers, Bernstein - dashed yellow and symmetric Bernstein - solid yellow.

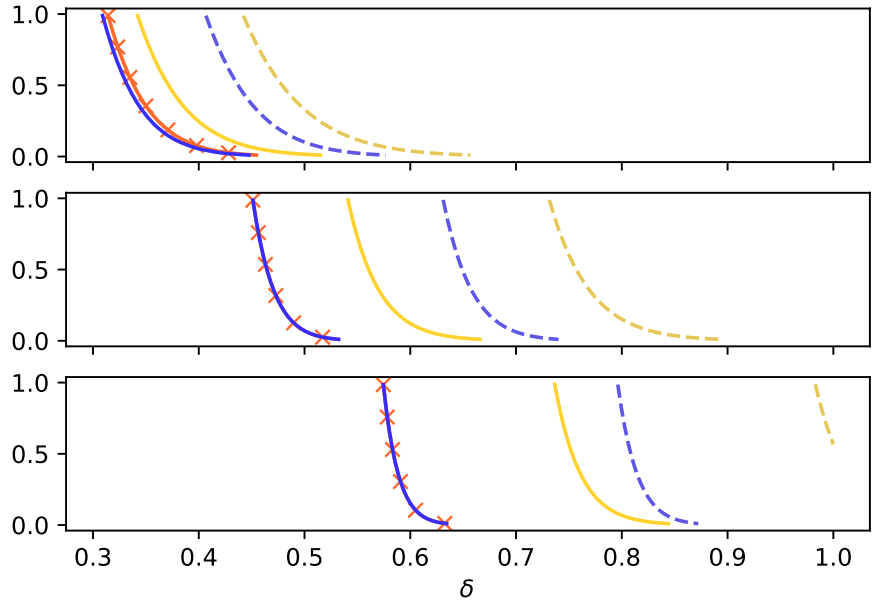

Figure 2: Upper bounds on $\mathbb{P}(\|T_{\nu_{\mathcal{S}},t}\| \geq \delta)$ for $t = 5$, $\mathcal{S} = 50$ or in symmetric case $\mathcal{S} = 2 \times 50$ and different d: top - 2, middle - 4, bottom - 8. Master - dashed blue, symmetric master - solid blue, simplified symmetric master - solid orange with x-markers, Bernstein - dashed yellow and symmetric Bernstein - solid yellow.

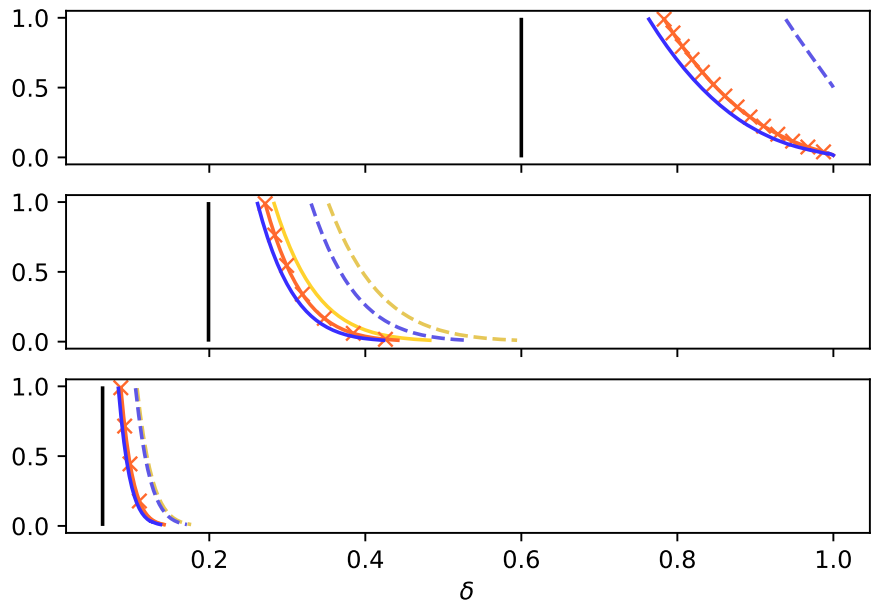

Figure 3: Upper bounds on $\mathbb{P}(\|T_{\nu_{\mathcal{S}},t}\| \geq \delta)$ for $d = 2$, $t = 2$ and different $\mathcal{S}$: top - 5, middle - 50, bottom - 500 or in symmetric case $\mathcal{S}$: top - 2×5, middle - 2×50, bottom - 2×500. Master - dashed blue, symmetric master - solid blue, simplified symmetric master - solid orange with x-markers, Bernstein - dashed yellow and symmetric Bernstein - solid yellow. Black vertical lines indicate the value of $\delta_{\text{opt}}(\mathcal{S})$.

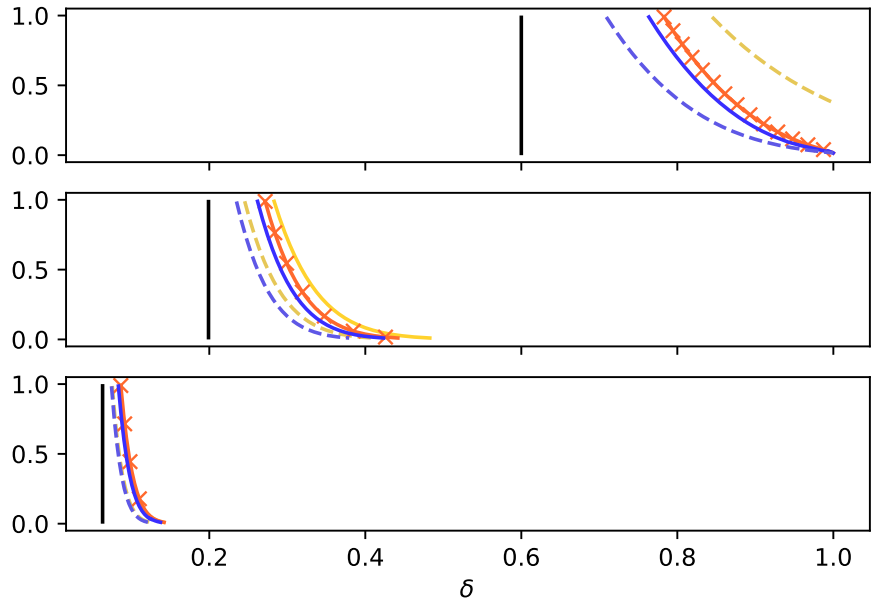

Figure 4: Upper bounds on $\mathbb{P}(\|T_{\nu_{\mathcal{S}},t}\| \geq \delta)$ for $d = 2$, $t = 2$ and different $\mathcal{S}$: top - 10, middle - 100, bottom - 1000 or in symmetric case $\mathcal{S}$: top - 2×5, middle - 2×50, bottom - 2×500. Master - dashed blue, symmetric master - solid blue, simplified symmetric master - solid orange with x-markers, Bernstein - dashed yellow and symmetric Bernstein - solid yellow. Black vertical lines indicate the value of $\delta_{\text{opt}}(\mathcal{S})$.

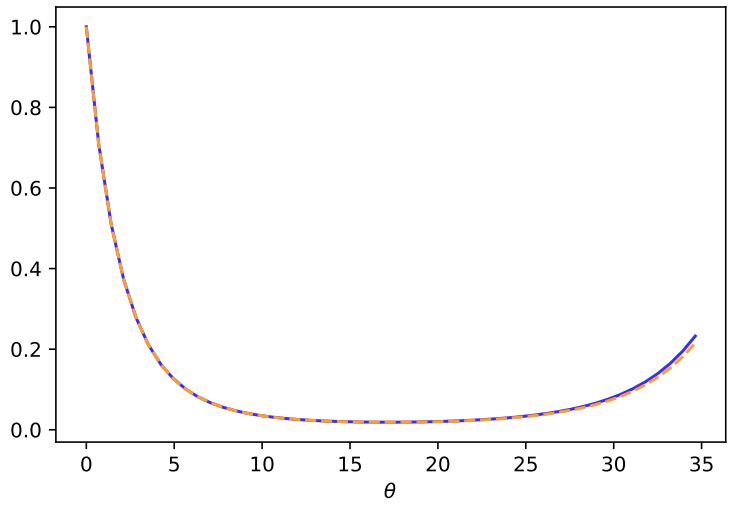

Figure 5: Functions under the infimum from Eq. (32) for $\mathcal{S} = 30$ and $\delta = 0.5$.

Next we check what are minimal sizes of $\mathcal{S}$ required to obtain δ-approximate t-design, with probability at least P according to the master bound. For $t-$random gate-sets the formula can be easily obtained from Theorem 1 and reads:

$$\mathcal{S} \geq \frac{2\left(\log\left(2\sum_{\lambda\in\Lambda_t} d_\lambda\right) - \log\left(1-P\right)\right)}{\log\left((1+\delta)^{1+\delta}\,(1-\delta)^{1-\delta}\right)}. \tag{37}$$

For symmetric Haar random gate-sets, however, it is much more difficult to obtain analogous formula using Theorem 2. Nevertheless the numerical calculations of Table 2 suggest that the required number of gates is around half of (37) plus inverses.

For $d = 2^n$ and $t = 2$, $P = 0.99$, $\delta = 0.01$, it is interesting to compare the number of gates given by (37) with the number of elements of the n-qubit Clifford group, $\mathcal{C}_n$, which is known to be an exact 2-design. Using Lemma 8 one easily checks that for $t = 2$ and $d \geq 4$ the set Λ_2 has exactly $p(2)^2 + p(1)^2 = 5$ distinct λ's. They are given by

$$\begin{aligned}
\lambda_1 &= (1, 0, \ldots, 0, -1),\ \|\lambda_1\|_1 = 2,\ d_{\lambda_1} = d^2 - 1,\\
\lambda_2 &= (2, 0, \ldots, 0, -2),\ \|\lambda_2\|_1 = 4,\ d_{\lambda_2} = \frac{d^2(d-1)(d+3)}{4},\\
\lambda_3 &= (2, 0, \ldots, 0, -1, -1),\ \|\lambda_3\|_1 = 4,\ d_{\lambda_3} = \frac{(d^2-1)(d^2-4)}{4},\\
\lambda_4 &= (1, 1, 0, \ldots, 0, -1, -1),\ \|\lambda_4\|_1 = 4,\ d_{\lambda_4} = \frac{d^2(d-3)(d+1)}{4},\\
\lambda_5 &= (1, 1, 0, \ldots, 0, -2),\ \|\lambda_5\|_1 = 4,\ d_{\lambda_5} = \frac{(d^2-1)(d^2-4)}{4}.
\end{aligned}$$

Thus $\sum_{\lambda\in\Lambda_2} d_\lambda = d^4 - 3d^2 + 1$. Using these results we find that t-random gate-set $\mathcal{S}_n \subset SU(2^n)$, $n \geq 2$, with

$$\mathcal{S}_n \geq \left\lceil 2\cdot 10^4 \left(\log\left(2^{4n+1} - 3\cdot 2^{2n+1} + 2\right) + 4.61\right)\right\rceil. \tag{38}$$

forms 0.01-approximate 2-design with the probability $P = 0.99$. On the other hand the cardinality of $\mathcal{C}_n$ is given by [11]

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

Figure 6 shows $\frac{|\mathcal{C}_n|}{\mathcal{S}_n}$ for up to 50 qubits. The ratio $\frac{|\mathcal{C}_n|}{\mathcal{S}_n}$ grows at least exponentially with n. In fact, one can easily see, that $\sum_{\lambda\in\Lambda_t} d_\lambda \leq d^{2t}$. Thus (37) grows in a logarithmic way

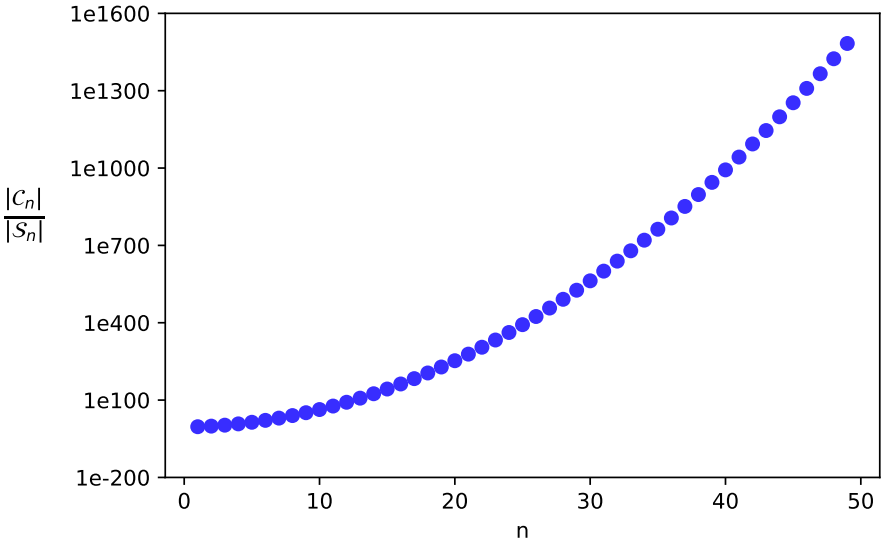

Figure 6: Ratio of the cardinality of the set of n-qubit Clifford gates $\mathcal{C}_n$ and the minimal cardinality of the t-random gate-set $\mathcal{S}_n \subset U(2^n)$ that forms a 0.01-approximate 2-design with the probability at least 0.99.

with d and linearly with t and t-random gate-set $\mathcal{S}$ with

$$\mathcal{S} \geq \frac{2\left(2t\log\left(d\right) + \log(2) - \log\left(1-P\right)\right)}{\log\left(\left(1+\delta\right)^{1+\delta}\left(1-\delta\right)^{1-\delta}\right)},$$

forms δ-approximate t-design with the probability P. Moreover, for a small δ we can further reduce this formula using $\log\left(\left(1+\delta\right)^{1+\delta}\left(1-\delta\right)^{1-\delta}\right) \simeq \delta^2$. Thus we get that S scales like $O(\delta^{-2}(t\log(d) - \log(1-P)))$.

We also note that if $\nu_{\mathcal{S}}$ is a δ-approximate t-design with probability bounded by $1-\epsilon$. Then $\nu_{\mathcal{S}}^{*l}$, whose support are all circuits of depth l built from gates form $\mathcal{S}$, is a δ^l-approximate t-design with probability bounded by $1-\epsilon$. Thus one can construct first $1/2$-approximate t-design and then by building all possible circuits of length l change it to $1/2^l$-approximate t-design. Table 2) presents numerical calculations for various t and d.

Acknowledgments

This research was funded by the National Science Centre, Poland under the grant OPUS: UMO-2020/37/B/ST2/02478 and supported in part by PLGrid Infrastructure.

		t	2	3	4	5	20	500	5000
$d=2$	Master	$\mathcal{S}$-symmetric	2×36	2×37	2×39	2×40	2×46	2×67	2×84
		$\mathcal{S}$	69	75	80	83	107	166	209
	Bernstein	$\mathcal{S}$-symmetric	2×47	2×50	2×52	2×53	2×64	2×95	2×120
$d=4$	Master	$\mathcal{S}$	82	96	108	118	195		
		$\mathcal{S}$-symmetric	2×41	2×47	2×53	2×58	2×95		
	Bernstein	$\mathcal{S}$	100	117	132	144	238		
		$\mathcal{S}$-symmetric	2×58	2×67	2×76	2×82	2×136		
$d=8$	Master	$\mathcal{S}$	104	129	154	175			
		$\mathcal{S}$-symmetric	2×51	2×63	2×75	2×85			
	Bernstein	$\mathcal{S}$	127	158	187	213			
		$\mathcal{S}$-symmetric	2×73	2×90	2×107	2×122			
$d=16$	Master	$\mathcal{S}$	126	162	197	229			
		$\mathcal{S}$	153	197	240	280			
	Bernstein	$\mathcal{S}$-symmetric	2×88	2×113	2×137	2×160			
$d=32$	Master	$\mathcal{S}$	147	194	239	282			
		$\mathcal{S}$	179	236	292	345			
	Bernstein	$\mathcal{S}$-symmetric	2×103	2×135	2×167	2×197			
$d=64$	Master	$\mathcal{S}$	168	226	282	336			
		$\mathcal{S}$	205	275	344	410			
	Bernstein	$\mathcal{S}$-symmetric	2×117	2×158	2×197	2×234			

Table 2: Minimal sizes of Haar random gate-sets $\mathcal{S}$ required to obtain $\frac{1}{2}$-approximate t-design with probability at least 0.99.

References

1. Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: Towards practical large-scale quantum computation. *Phys. Rev. A* **86,** 032324. DOI: 10.1103/PhysRevA.86.032324 (2012).
2. Preskill, J. Quantum Computing in the NISQ era and beyond. *Quantum* **2,** 79. DOI: 10.22331/q-2018-08-06-79 (2018).
3. Boixo, S. *et al.* Characterizing Quantum Supremacy in Near-Term Devices. *Nature Physics* **14,** 595–600. DOI: 10.1038/s41567-018-0124-x (2018).
4. Harrow, A. W. & Montanaro, A. Quantum Computational Supremacy. *Nature* **549,** 203–209. DOI: 10.1038/nature23458 (2017).
5. Ballance, C. J., Harty, T. P., Linke, N. M., Sepiol, M. A. & Lucas, D. M. High-fidelity quantum logic gates using trapped-ion hyperfine qubits. *Phys. Rev. Lett.* **117,** 060504. DOI: 10.1103/PhysRevLett.117.060504 (2016).
6. Barends, R. *et al.* Logic gates at the surface code threshold: Superconducting qubits poised for fault-tolerant quantum computing. *Nature* **508,** 500–503. DOI: 10.1038/nature13171 (2014).
7. Susskind, L. *Three Lectures on Complexity and Black Holes* DOI: 10.1007/978-3-030-45109-7 (Springer Cham, 2020).
8. Sawicki, A. & Karnas, K. Criteria for universality of quantum gates. *Physical Review A* **95,** 062303. DOI: 10.1103/physreva.95.062303 (2017).
9. Sawicki, A. & Karnas, K. Universality of Single-Qudit Gates. *Annales Henri Poincaré* **18,** 3515–3552. DOI: 10.1007/s00023-017-0604-z (2017).
10. Sawicki, A., Mattioli, L. & Zimborás, Z. Universality verification for a set of quantum gates. *Phys. Rev. A* **105,** 052602. DOI: 10.1103/PhysRevA.105.052602 (5 2022).
11. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information: 10th Anniversary Edition* DOI: 10.1017/CBO9780511976667 (Cambridge University Press, 2010).
12. Varjú, P. P. Random walks in compact groups. *Doc. Math.* **18,** 1137–1175. DOI: 10.4171/DM/423 (2013).
13. Oszmaniec, M., Sawicki, A. & Horodecki, M. Epsilon-Nets, Unitary Designs, and Random Quantum Circuits. *IEEE Transactions on Information Theory* **68,** 989–1015. DOI: 10.1109/TIT.2021.3128110 (2022).
14. Bouland, A. & Giurgica-Tiron, T. Efficient Universal Quantum Compilation: An Inverse-free Solovay-Kitaev Algorithm. *arXiv e-prints.* DOI: 10.48550/ARXIV.2112.02040 (2021).
15. Harrow, A. W., Recht, B. & Chuang, I. L. Efficient Discrete Approximations of Quantum Gates. *J. Math. Phys.* **43,** 4445. DOI: 10.1063/1.1495899 (2002).
16. Epstein, J. M., Cross, A. W., Magesan, E. & Gambetta, J. M. Investigating the limits of randomized benchmarking protocols. *Physical Review A* **89,** 062321. DOI: 10.1103/physreva.89.062321 (2014).
17. Dalzell, A. M., Hunter-Jones, N. & Brandão, F. G. S. L. Random quantum circuits transform local noise into global white noise. *arXiv e-prints.* DOI: 10.48550/ARXIV.2111.14907 (2021).

18. Abeyesinghe, A., Devetak, I., Hayden, P. & Winter, A. The mother of all protocols: restructuring quantum information's family tree. *Proceedings of the Royal Society of London Series A* **465,** 2537–2563. DOI: 10.1098/rspa.2009.0202 (2009).

19. Radhakrishnan, J., Rötteler, M. & Sen, P. Random measurement bases, quantum state distinction and applications to the hidden subgroup problem. *Algorithmica* **55,** 490–516. DOI: 10.1007/s00453-008-9231-x (2009).

20. Roberts, D. A. & Yoshida, B. Chaos and complexity by design. *Journal of High Energy Physics* **2017,** 121. DOI: 10.1007/JHEP04(2017)121 (2017).

21. Oszmaniec, M., Horodecki, M. & Hunter-Jones, N. Saturation and recurrence of quantum complexity in random quantum circuits. *arXiv e-prints.* DOI: 10.48550/ARXIV.2205.09734 (2022).

22. Haferkamp, J., Faist, P., Kothakonda, N. B. T., Eisert, J. & Yunger Halpern, N. Linear growth of quantum circuit complexity. *Nature Physics* **18,** 528–532. DOI: 10.1038/s41567-022-01539-6 (2022).

23. Bourgain, J. & Gamburd, A. A spectral gap theorem in SU(d). *J. Eur. Math. Soc.* **14,** 1455–1511. DOI: 10.4171/JEMS/337 (2012).

24. Bourgain, J & Gamburd, A. On the spectral gap for finitely-generated subgroups of SU(2). *Invent. math.* **171,** 83–121. DOI: 10.1007/s00222-007-0072-z (2008).

25. Bocharov, A., Gurevich, Y. & Svore, K. M. Efficient decomposition of single-qubit gates into V basis circuits. *Phys. Rev. A* **88,** 012313. DOI: 10.1103/physreva.88.012313 (2013).

26. Kliuchnikov, V., Bocharov, A., Roetteler, M. & Yard, J. A Framework for Approximating Qubit Unitaries. *arXiv e-prints.* DOI: 10.48550/arXiv.1510.03888 (2015).

27. Kliuchnikov, V., Maslov, D. & Mosca, M. Fast and efficient exact synthesis of single qubit unitaries generated by Clifford and T gates. *Quantum Information and Computation* **13,** 607–630. DOI: 10.26421/QIC13.7-8-4 (2013).

28. Selinger, P. Efficient Clifford+T approximation of single-qubit operators. *Quantum Information and Computation* **15,** 159–180. DOI: 10.26421/QIC15.1-2-10 (2015).

29. Sarnak, P. *Letter to Scott Aaronson and Andy Pollington on the Solovay-Kitaev theorem* 2015.

30. Lubotzky, A., Phillips, R. & Sarnak, P. Hecke operators and distributing points on S2. II. *Communications on Pure and Applied Mathematics* **40,** 401–420. DOI: 10.1002/cpa.3160400402 (1987).

31. Tropp, J. A. *An Introduction to Matrix Concentration Inequalities Foundations and Trends in Machine Learning* **1-2,** 1–230. DOI: 10.1561/2200000048 (Now Publishers Inc, 2015).

32. Abu-Hamed, M. & Gelaki, S. Frobenius-Schur indicators for semisimple Lie algebras. *Journal of Algebra* **315,** 178–191. DOI: 10.1016/j.jalgebra.2007.06.003 (2007).

33. Emerson, J., Alicki, R. & Życzkowski, K. Scalable noise estimation with random unitary operators. *Journal of Optics B: Quantum and Semiclassical Optics* **7,** S347. DOI: 10.1088/1464-4266/7/10/021 (2005).

34. Dankert, C., Cleve, R., Emerson, J. & Livine, E. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A* **80,** 012304. DOI: 10.1103/PhysRevA.80.012304 (1 2009).

35. Nakata, Y. *et al.* Quantum Circuits for Exact Unitary t-Designs and Applications to Higher-Order Randomized Benchmarking. *PRX Quantum* **2,** 030339. DOI: 10.1103/PRXQuantum.2.030339 (3 2021).

36. Meckes, E. S. *The Random Matrix Theory of the Classical Compact Groups* DOI: 10.1017/9781108303453 (Cambridge University Press, 2019).

37. Reck, M., Zeilinger, A., Bernstein, H. J. & Bertani, P. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.* **73,** 58–61. DOI: 10.1103/PhysRevLett.73.58 (1994).

38. Sawicki, A. Universality of beamsplitters. *Quantum Information and Computation* **16,** 291–312. DOI: 10.26421/QIC16.3-4-6 (2016).

39. Lieb, E. H. Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Advances in Mathematics* **11,** 267–288. DOI: 10.1016/0001-8708(73)90011-X (1973).

40. Golden, S. Lower Bounds for the Helmholtz Function. *Phys. Rev.* **137,** B1127–B1128. DOI: 10.1103/PhysRev.137.B1127 (1965).

41. Thompson, C. J. Inequality with Applications in Statistical Mechanics. *J. Math. Phys.* **6,** 1812–1813. DOI: 10.1063/1.1704727 (1965).

42. Hall, B. C. *Lie Groups Lie Algebras and Representations An Elementary Introduction* DOI: 10.1007/978-3-319-13467-3 (Springer-Verlag New York, 2004).

43. Benkart, G. *et al.* Tensor product representations of general linear groups and their connections with Brauer algebras. *J. Algebra* **166,** 529–567. DOI: 10.1006/jabr.1994.1166 (1994).

44. Bröcker, T. & Dieck, T. *Representations of Compact Lie Groups* DOI: 10.1007/978-3-662-12918-0 (Springer Berlin Heidelberg, 2003).

45. Ruiz-Antolin, D. & Segura, J. A new type of sharp bounds for ratios of modified Bessel functions. *J. Math. Anal. Appl.* **443,** 1232–1246. DOI: 10.1016/j.jmaa.2016.06.011 (2016).