ANNOTATIONS ON QUANTUM PSEUDORANDOMNESS AND k-DESIGNS

EDWARD KIM

ABSTRACT. We investigate constructions of efficient approximate unitary k-designs which are defined as probability distributions over the unitary group which model the Haar distribution up to the kth-moment. In particular, we review the paper by Harrow and Low as a directed reading conducted during the Spring 2020 semester.

CONTENTS

1. INTRODUCTION

The theory of randomness and its role in computation are cornerstones of contemporary computation theory as evidenced by decades of beautiful, celebrated results. Quantum Information has, by no means, become stranger to this trend. Numerous applications of random unitaries have been discovered in the areas of quantum tomography, quantum cryptography, and, rather naturally, quantum algorithms. To gain access to uniformly-sampled elements from $\mathbb{U}(d)$, one must sample from the Haar distribution over $\mathbb{U}(d)$ where $\mathbb{U}(d)$ denotes the group of unitary $d \times d$ matrices. One method of sampling from the Haar distribution is to apply local random unitaries sampled from distributions over lower-dimensional unitary groups. For example, if we consider an n-qubit system, one method of sampling from the Haar distribution over $\mathbb{U}(2^n)$ would be through the following scheme:

(1) Sample from the Haar distribution over $\mathbb{U}(4)$ to produce a random gate G
(2) Uniformly pick two distinct qubits and locally apply G to the pair.

However, this becomes computationally infeasible as even reasonably approximating the Haar distribution under this manner requires an exponential number of local 2-qubit gates [4]. This hurdle motivates us to study the construction of efficient k-designs i.e distributions over $\mathbb{U}(4)$ such that sampling over these distributions and applying our scheme above will converge to good approximations of the kth moments of the Haar distributon over $\mathbb{U}(2^n)$. To this end, we begin with annotations of the seminal paper by Harrow and Low [3]. We first introduce concepts germane to the area of efficient k-designs and summarize the techniques employed in the analyses presented in the paper.

2. Approximate 2-designs via Local Random Quantum Circuits

2.1. **Definitions and Motivations.** In [3], the authors tackle the problem of approximate 2-designs, specifically *unitary* 2-designs. The central result of the paper claims that the scheme of applying sampled 2-qubit unitaries from certain distributions with a k-gapped property converge to the second moment of the Haar distribution rapidly. The crux of the convergence argument relies on mapping the evolution of diagonal states to that of a Markov chain's. Consequently, the analysis becomes amenable to classic techniques created to study mixing times of Markov chains. To begin, we first state the relevant definitions as introduced by Harrow and Low.

Remark 2.1. As dissecting every proof in [3] may unnecessarily length this report, we occasionally omit proofs of lemmas. Some major theorems will have their proofs delegated to the appendix of this report.

Definition 2.2. An ensemble of operators $\mathcal{E} = \{p_i, U_i\}$ over $\mathbb{U}(d)$ is called an *(exact) unitary k-design* if $\mathcal{G}_W = \mathcal{G}_H$ where

$$(2.1) \qquad \mathcal{G}_W(\rho) = \sum_i p_i U_i^{\otimes k} \rho (U_i^\dagger)^{\otimes k}$$

$$(2.2) \qquad \mathcal{G}_H(\rho) = \int_{\mathbb{U}(d)} U^{\otimes k} \rho (U^\dagger)^{\otimes k} \, dU$$

We would naturally like to define some distance between $\mathcal{G}_W$ and $\mathcal{G}_H$. One formulation involves interpreting the two $\mathcal{G}_W, \mathcal{G}_H$ as quantum channels in order to utilize the diamond norm:

Definition 2.3. The diamond norm of an operator T is defined as

$$||T||_\diamond = \sup_d ||T \otimes I_d||_\infty = \sup_d \sup_{X \neq 0} \frac{||(T \otimes I_d)(X)||_1}{||X||_1}$$

Definition 2.4. $\mathcal{G}_W$ is an *ϵ-approximate k-design* if

$$
(2.3) \qquad\qquad ||\mathcal{G}_W - \mathcal{G}_H||_\diamond \leq \epsilon
$$

As discussed in the introduction above, the unitary constructed by sampling from a universal gate set and applying them to random pairs of qubits will converge to a unitary sampled from Haar distribution. However, we would like to find *efficient approximate k-*designs. Here, efficient refers to random circuits of asymptotic length polynomial in the number of qubits and $\log 1/\epsilon$. The authors actually prove some stronger, namely that *k-copy gapped* distributions suffice.

Definition 2.5. Let μ be a (possibly continuous) probability measure on $\mathbb{U}(d)$. If

$$
(2.4) \qquad\qquad \mathcal{G}_\mu = \int_{\mathbb{U}(d)} d\mu(U) U^{\otimes k} \otimes (U^*)^{\otimes k}
$$

then we say that $\mathcal{G}_\mu$ is *k-copy gapped* if $\mathcal{G}_\mu$ has only $k!$ eigenvectors with unit eigenvalues.

A more general definition for the universality of gatesets can be defined as well:

Definition 2.6. Let μ be a distribution over $\mathbb{U}(d)$. Then μ is deemed to be *universal* for $\mathbb{U}(d)$ if for any open ball $B \subset \mathbb{U}(d)$, there exists a sufficiently large integer ℓ such that $\mu^{\star\ell}(B) > 0$ where $\mu^{\star\ell}$ refers to the ℓ-fold *convolution* product of μ:

$$
(2.5) \qquad\qquad \mu^{\star\ell} = \int_{\mathbb{U}(d)} \delta_{U_1,\cdots,U_\ell} d\mu(U_1) \cdots d\mu(U_\ell)
$$

Remark 2.7. This is the convolution product in the context of *measures*. For example, if we let μ, σ be two finite Borel measures over $\mathbb{U}(d)$, then we have a new finite Borel measure on $\mathbb{U}(d)$, $\mu^{\star\ell}$:

$$
\mu^{\star\ell}(S) = \int_{\mathbb{U}(d)} \cdots \int_{\mathbb{U}(d)} \mathbb{I}_S(U_1 \cdots U_\ell) d\mu(U_1) \cdots d\mu(U_\ell)
$$

where $\mathbb{I}_S$ is the characteristic function for measurable set $S \subset \mathbb{U}(d)$.

Intuitively, universality implies that sampling enough times over μ is enough to approximate any gate in $\mathbb{U}(d)$. This aligns with the definition introduced for discrete measures μ over finite gate sets. We will prove later that any universal gate set will be k-copy gapped on $\mathbb{U}(4)$ (Theorem 2.16). For now, let us state the main theorem and its auxiliary lemma to guide our upcoming analysis:

Theorem 2.8. *Let μ be a 2-copy gapped distribution over $\mathbb{U}(4)$ and W a random circuit of length t drawn from sampling 2-qubit unitaries from μ and applying them to random pairs of of a system of n-qubits. Then there exists a constant C depending on μ such that for $\epsilon > 0$, $\mathcal{G}_W$ is an ϵ-approximate 2-design if $t \geq n(n + \log 1/\epsilon)$.*

Notation 2.9. Let ρ be an initial state of an $2n$-qubit system. We can expand ρ in terms of the Pauli basis $\{\sigma_i\}_{0 \leq i \leq 3}$ as follows:

$$(2.6) \qquad \rho = \frac{1}{2^n} \sum_{p_1, p_2} \gamma_0(p_1, p_2) \sigma_{p_1} \otimes \sigma_{p_2}, \quad \gamma_0(p_1, p_2) \in \mathbb{C}$$

where p_1, p_2 run over all strings of Pauli indices of length n, $\{0, 1, 2, 3\}^n$. Let $\gamma_W(p_1, p_2)$ be the coefficient of the term $\sigma_{p_1} \otimes \sigma_{p_2}$ for the output state after applying random circuit W, $\rho_W = W^{\otimes 2} \rho (W^{\dagger})^{\otimes 2}$. Extending this notation allows us to express the expected coefficient after t-samples from our distribution μ as $\gamma_t(p_1, p_2) = \mathbb{E}_W[\gamma_W(p_1, p_2)]$. Note that the expectation is taken over all t-length random circuits.

Lemma 2.10. *Let ρ be a an initial state such that $\gamma_0(p, p) \geq 0$ and $\sum_p \gamma_0(p, p) = 1$ with μ and W as in Theorem 2.8. Then there exists constant C such that for $\epsilon > 0$*

$$(2.7) \qquad \sum_{p_1, p_2 \neq 00} \left(\gamma_t(p_1, p_2) - \delta_{p_1 p_2} \frac{1}{2^n(2^n + 1)} \right)^2 \leq \epsilon$$

for $t \geq Cn \log 1/\epsilon$.

$$(2.8) \qquad \sum_{p_1, p_2 \neq 00} \left| \gamma_t(p_1, p_2) - \delta_{p_1 p_2} \frac{1}{2^n(2^n + 1)} \right| \leq \epsilon$$

for $t \geq Cn(n + \log 1/\epsilon)$

Corollary 2.10.1. *Let μ, W, γ_W be as above. Then for **any** initial state ρ, there exists constant C such that for $\epsilon > 0$,*

$$(2.9) \qquad \sum_{p_1, p_2 \neq 00} \left(\gamma_t(p_1, p_2) - \delta_{p_1 p_2} \frac{\sum_{p \neq 0} \gamma_0(p, p)}{4^n - 1} \right)^2 \leq \epsilon$$

$$(2.10) \qquad \sum_{p_1, p_2 \neq 00} \left| \gamma_t(p_1, p_2) - \delta_{p_1 p_2} \frac{\sum_{p \neq 0} \gamma_0(p, p)}{4^n - 1} \right| \leq \epsilon$$

both for $t \geq Cn(n + \log 1/\epsilon)$

2.2. **Analysis of Moments.** We turn back to the general case and begin to define concepts and definitions for group $\mathbb{U}(d)$ and for d-qudit states. Set a Hermitian basis of $d \times d$ to be $\sigma_0, \sigma_1, \cdots \sigma_{d^2-1}$ such that $\sigma_0 = I$. These basis elements span the d-qudit states. Here we stress that we sample our unitaries over the *Haar distribution* over $\mathbb{U}(d)$. We denote sampling from the Haar distribution with the notation $U \sim \mathbb{U}(d)$ when necessary. Furthermore, our Hermitian basis operators are such that $Tr(\sigma_p \sigma_q) = d \cdot \delta_{p,q}$ i.e orthogonal under the

Hilbert-Schmidt inner product.

Let $\mathbf{p} = (p_1, \cdots, p_k) \in \{0, 1, \cdots, d^2 - 1\}^k$ denote a k-string of basis indices. We will be interested in the quantity:

$$(2.11) \qquad T(\mathbf{p}) := \mathbb{E}_{U \sim \mathbb{U}(d)} \left[U^{\otimes k} \sigma_{p_1} \otimes \sigma_{p_2} \otimes \cdots \otimes \sigma_{p_k} (U^\dagger)^{\otimes k} \right]$$

This will be the expected state after applying a sampled unitary U from the Haar distribution over $\mathbb{U}(d)$. Naturally, we can reexpress this resulting state into the σ_p basis:

$$(2.12) \qquad T(\mathbf{p}) = \sum_{\mathbf{q}} \widehat{G}(\mathbf{p}, \mathbf{q}) \sigma_{q_1} \otimes \cdots \otimes \sigma_{q_k}$$

where the same is taken over all k-strings in $\{0, \cdots, d^2 - 1\}^k$. This allows us to define the matrix $\widehat{G}$ where the rows and columns are indiced by k-strings with entries $\widehat{G}(\mathbf{p}, \mathbf{q})$. By virtue of the invariance of the Haar distribution, this matrix turns out to be a *symmetric projector*.

Lemma 2.11. *The matrix $\widehat{G}$ above is symmetric i.e. $\widehat{G}(\boldsymbol{p}, \boldsymbol{q}) = \widehat{G}(\boldsymbol{q}, \boldsymbol{p})$*

Proof. By orthogonality:

$$(2.13) \qquad \widehat{G}(\mathbf{q}, \mathbf{p}) = d^{-k} \cdot Tr(\sigma_{q_1} \otimes \cdots \otimes \sigma_{q_k} T(\mathbf{p}))$$

$$(2.14) \qquad = d^{-k} \mathbb{E}_U \left[Tr(\sigma_{q_1} \otimes \cdots \otimes \sigma_{q_k}) U^{\otimes k} (\sigma_{p_1} \otimes \sigma_{p_2} \otimes \cdots \otimes \sigma_{p_k})(U^\dagger)^{\otimes k} \right]$$

$\square$

Note that the expectation and the trace operator commute by linearity of latter. The result follows from the cyclic invariance of trace.

Lemma 2.12. $\widehat{G}$ *fixes the permutation operators P_π, $\pi \in S_k$.*

These two results are all that is required to show that $\widehat{G}$ is indeed a projector.

Theorem 2.13. $\widehat{G}$ *is a projector i.e. $\widehat{G}^2 = \widehat{G}$*

Proof. Expanding the matrix product shows that

$$(2.15) \qquad \widehat{G}(\mathbf{p}, \mathbf{q}) = \sum_{\mathbf{q'}} \widehat{G}(\mathbf{p}, \mathbf{q'}) \widehat{G}(\mathbf{q'}, \mathbf{q}) =$$

$$(2.16) \qquad = \sum_{\mathbf{q'}} \widehat{G}(\mathbf{p}, \mathbf{q'}) \left(d^{-k} \cdot Tr(\sigma_{q'_1} \otimes \cdots \otimes \sigma_{q'_k} T(\mathbf{q})) \right)$$

By Schur-Weyl Duality, $T(\mathbf{q})$ must be a linear combination of permutation operators ($T(\mathbf{p})$ commutes with all unitaries of the form $U^{\otimes k}$ by the invariance of the Haar integral). However,

by Lemma (2.12) above, $\widehat{G}$ fixes the permuatation operators. To understand this more explicitly, rewrite the Lemmma (2.12) as follows:

$$(2.17) \qquad \sum_{\mathbf{q}} \widehat{G}(\mathbf{p}, \mathbf{q}) \cdot Tr(\sigma_{q_1} \otimes \cdots \otimes \sigma_{q_k} P_\pi) = Tr(\sigma_{p_1} \otimes \cdots \otimes \sigma_{p_k} P_\pi)$$

This simply expressing the equality $\widehat{G} P_\pi = P_\pi$ for permutation operator p_π, $\pi \in S_k$. By linearity of trace, the equality above reduces to

$$(2.18) \qquad \sum_{\mathbf{q'}} \widehat{G}(\mathbf{p}, \mathbf{q'}) \left(d^{-k} \cdot Tr(\sigma_{q_1'} \otimes \cdots \otimes \sigma_{q_k'} T(\mathbf{q})) \right) = Tr(\sigma_{p_1} \otimes \cdots \otimes \sigma_{p_k} T(\mathbf{q}))$$

$$(2.19) \qquad \qquad \qquad \qquad = \widehat{G}(\mathbf{p}, \mathbf{q})$$

$\square$

Corollary 2.13.1. *$\widehat{G}$ has eigenvalues 0 and 1*

2.2.1. *Explicit calculation of $T(\boldsymbol{p})$.* The calculation is divided into the two pertinent cases $k = 1, 2$.

(1) *(For $k = 1$)* For this case, a random unitary sampled from the uniform distribution will randomize the initial state. Thus:

$$T(p) = \begin{cases} \sigma_0 & p = 0 \\ 0 & p \neq 0 \end{cases}$$

(2) *(For $k = 2$)* From Lemma (2.12) and Theorem (2.13), $\widehat{G}$ must fix the only two permutation operators $I, \mathcal{F}$ where $\mathcal{F}$ is the swap operator on two qubits. We note that we will normalize the swap operator

$$(2.20) \qquad \qquad \mathcal{F} = \frac{1}{d^2 - 1} \sum_{1 \leq q \leq d^2 - 1} \sigma_q \otimes \sigma_q$$

We then deduce the following about $\widehat{G}$:

(a) $\widehat{G}(p_1, p_2; q_1, q_2) = 0, \quad p_1 \neq p_2$ or $q_1 \neq q_2$

Proof. To see this, observe that both $I, \mathcal{F}$ have vanishing $\sigma_{q_1} \otimes \sigma_{q_2}$ components for $q_1 \neq q_2$. Thus, multiplying $\widehat{G}$ with either one should also result in vanishing $\sigma_{q_1} \otimes \sigma_{q_2}$ components. This with the fact that all other operators are sent to zero by $\widehat{G}$ yields the result. $\square$

(b) $\widehat{G}(\mathbf{p}; \mathbf{0}) = \delta_{p0}$ (Here we denote $\widehat{G}(\mathbf{p}; \mathbf{0}) := \widehat{G}(p, p; 0, 0)$.)

Proof. This just derives from $\widehat{G}$ acting on I $\square$

(c) $\widehat{G}(\mathbf{p}, \mathbf{a}) = \frac{1}{d^2-1}, \quad \mathbf{p}, \mathbf{a} \neq 0$

Proof. As above, this results from $\widehat{G}$ acting on $\mathcal{F}$ along with the condition that it must send all other operators to zero. $\qquad\square$

We conclude from the definition of $T(p, q)$ for $p, q \in \{0, \cdots, d^2 - 1\}$ that

$$(2.21) \qquad T(p, q) = \begin{cases} 0 & p \neq q \\ \sigma_0 \otimes \sigma_0 & p = q = 0 \\ \frac{1}{d^2-1} \sum_{\ell \neq 0} \sigma_\ell \otimes \sigma_\ell & p = q \neq 0 \end{cases}$$

Recall that $T(p, q)$ refers to the expected state from applying unitary $U \sim \mathbb{U}(d)$ to the basis state $\sigma_p \otimes \sigma_q$. From the determination above, it is simple to see that U uniformly samples from all diagonal basis states when acting on $\sigma_p \otimes \sigma_p$, $p \neq 0$. Strikingly enough, applying $U \sim \mathbb{U}(d)$ conserves the total sum of the coefficients of the diagonal basis states.

Lemma 2.14. *The total sum of the coefficients $\sum_p \gamma_0(p, p)$ is conserved by $U \sim \mathbb{U}(d)$ for state $\rho = \frac{1}{d} \sum_{p_1, p_2} \gamma_0(p_1, p_2) \sigma_{p_1} \otimes \sigma_{p_2}$.*

Normalizing this sum allows us to view it as a *probability distribution*. Furthermore, the evolution of our random circuit scheme is amenable to Markov chain analysis. As an illustrative example, suppose we have an n-qubit system and so far have applied t steps of our scheme. As per usual, we pick our 2-qubit unitary $U \sim \mathbb{U}(4)$. Now if we were to pick qubits $1, 2$, our random 2-qubit unitary would only act on the first two qubits non-trivially whereas every other qubit would just transform via the identity map. By the explicit calculation shown in Equation 2.21 and linearity, the expected *coefficients* for step $t + 1$ will work out to be

$$(2.22)$$

$$\gamma_{t+1}(p_1, \cdots, p_n, q_1, \cdots, q_n) = \begin{cases} 0 & (p_1, p_2) \neq (q_1, q_2) \\ \gamma_t(0, 0, \cdots, p_n, 0, 0, \cdots, q_n) & (p_1, p_2) = (q_1, q_2) = (0, 0) \\ \frac{1}{15} \sum_{\substack{r_1, r_2 \\ r_1 r_2 \neq 00}} \gamma_t(r_1, r_2, \cdots p_n, r_1, r_2, \cdots, q_n) & (p_1, p_2) = (q_1, q_2) \neq (0, 0) \end{cases}$$

Oliveira et.al [6] had the insight to map the evolution of the $\gamma_t(p, p)$ coeffcients to a Markov chain as follows:

Theorem 2.15. *Let M be a Markov chain governing the evolution of states under $\widehat{G}$ where $\widehat{G}$ samples from the Haar distribution over $\mathbb{U}(2^n)$. Then M exhibits the following transitions:*

(1) *The states of the Markov chain are $\Omega = \{0, 1, 2, 3\}^n$*
(2) *The transitions adhere to the guidelines below: Suppose we start with a state indexed by string $\alpha = \alpha_1 \cdots \alpha_n$*
  (a) *Pick two distinct sites $i < j$ uniformly at random.*

(b) *If $i = j = 0$, remain at the same state. Otherwise pick $\beta_i \beta_j = \{0, 1, 2, 3\}^2 / \{00\}$ uniformly at random. Transition to the string*

$$\alpha_1 \alpha_2, \cdots \alpha_{i-1} \beta_i \alpha_{i+1} \cdots \alpha_{j-1} \beta_j \alpha_{j+1} \cdots \alpha_n$$

In the context of the example above, if we start with the coeffcient $\gamma_t(\mathbf{q}, \mathbf{q})$ for $q \in \{0, 1, 2, 3\}^n$, the only two possibilities are $(\mathbf{q}, \mathbf{q}) = (0, 0)$ and $(\mathbf{q}, \mathbf{q}) \neq (0, 0)$ since it is impossible for the coefficient to vanish by the first case in Equation 2.22. Thus, the transitions presented in Theorem 2.15 indeed lead to the correct expected coefficient value $\gamma_{t+1}(\mathbf{p}, \mathbf{p})$. Note that this is not the complete value for γ_{t+1} as we still have to take the expectation over all the remaining qubit pairs.

2.3. Convergence of the Random Circuit.

Before embarking towards Markov chain techniques, we must prove that our random circuit construction indeed converges to Haar distribution. Specifcally, we will prove that any universal distribution (Definition 2.6) on $\mathbb{U}(d)$ will be k-copy gapped (Definition 2.5). In turn, this leads to convergence in sense of $\lim_{t \to \infty} (\widehat{G}_\mu)^t = \widehat{G}_{\mathbb{U}(d)}$. To see this, observe that another formulation of a k-copy gapped distributon over $\mathbb{U}(d)$ is

$$(2.23) \qquad\qquad\qquad ||G_\mu - G_{\mathbb{U}(d)}||_\infty < 1$$

where $G_\mu = \mathbb{E}_{U \sim \mu} \left[U^{\otimes k} \otimes (U^\dagger)^{\otimes k} \right]$. To justify this, first recall that $\widehat{G}$ and G are unitarily-related by Definition 2.11, so in proving 2.23, we can argue on the operators $\widehat{G}$ instead. In view of this, $\widehat{G}_{\mathbb{U}(d)}$ is a projector and thus has eigenvalues 0 and 1. If we can prove that any eigenvector of $\widehat{G}_{\mathbb{U}(d)}$ of eigenvalue one is also an eigenvector of $\widehat{G}_\mu$ of eigenvalue one i.e the k-symmetric subspace of $(\mathbb{C}^d)^{\otimes k}$ is also fixed by $\widehat{G}_\mu$. Thus, Equation 2.23 shows that indeed we can converge to $G_{\mathbb{U}(d)}$ as all eigenvectors shrink by iterating our construction. Note that $\widehat{G}_\mu$ may not even be Hermitian, so we cannot directly argue on its spectrum (as the spectrum may not completely contain real eigenvalues).

Theorem 2.16. *Let μ be a universal distribution over $\mathbb{U}(d)$. Then all eigenvectors of $G_{\mathbb{U}(d)}$ of eigenvalue one will also be eigenvectors of G_μ of eigenvalue one. Furthermore, μ will be k-copy gapped.*

For the sake of presentation, the proof of this will be delegated towards the appendix of this report.

2.4. Markov Chain Analysis.

As discussed in a previous section, the central idea of proving *rapid* convergence involves mapping the evolution of the diagonal states under our random circuit to that of a Markov chain's. Here, we refer to states of the form

$$\psi = \frac{1}{2^n} \sum_p \gamma(p, p) \sigma_p \otimes \sigma_p, \quad p \in \{0, 1, 2, 3\}^n, \ \gamma(p, p) \in \mathbb{C}$$

In fact, we can explicitly determine the operator P capturing our dynamics. Let $\widehat{G}^{(ij)}$ denote the unitary transformation resulting from sampling some unitary $U \sim \mathbb{U}(4)$ and applying U to qubits i and j while leaving all other qubits undisturbed.

$$(2.24) \qquad P = \frac{1}{n(n-1)} \sum_{i \neq j} \widehat{G}^{(ij)}$$

as evidenced by the equality

$$(2.25) \qquad \gamma_{t+1}(\mathbf{p}) = \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{\mathbf{q}} \widehat{G}^{(ij)}(\mathbf{p}, \mathbf{q}) \gamma_t(\mathbf{q})$$

where $\mathbf{p}, \mathbf{q} \in \{0, 1, 2, 3\}^2$ and $\gamma_t(\mathbf{p})$ represents the expected coeffcients at timestep t. A major caveat of interpreting P as a *transition matrix* for some suitable Markov chain on the diagonal states is the possibility that our initial coefficients of the diagonal states $\gamma_0(p, p)$ may not even sum to one or even be non-negative. It turns out that this doesn't bring too much trouble into the analysis as pointed out by Corollary 2.10.1. For now, let us tackle the case where the two conditions above are true and direct our focus on the evolution of the diagonal states. The ultimate goal of this line of thought will be to first prove Lemma 2.10 through a carefully-done investigation of the *mixing time* of P. Of course, we need to also show that the expected coeffcients of the non-diagonal states $\gamma_0(p, q)$ vanish rapidly as well. These matters will be addressed in a future section.

2.4.1. *Preliminaries on Markov Chains.* Let us introduce some definitions for properties of the Markov chains we are interested in. Although we assume familiarity with the definition of a Markov chain along with the its associated transition matrix P, we refer the reader to the reference [1] for greater detail. Recall that if $q^{(0)}$ is some initial distribution over the Ω then $q^{(t)} = q^{(0)} P^t$. The *stationary distribution* of a Markov chain π is a probability distribution over Ω such that $\pi P = \pi$.

Definition 2.17. A Markov chain is said to be *irreducible* if there exists a path $s \rightsquigarrow j$ for every pair of states $s, j \in \Omega$.

Definition 2.18. A Markov chain is said to be *periodic* at state $i \in \Omega$ if there exists a positive integer $T > 1$ such for that some initial distribution $q^{(}0)$, if $q_i^{(t)} > 0$ then $t \in \{a + Tm \mid m \in \mathbb{N}\}$. A Markov chain is deemed as *aperiodic* if none of its states are periodic.

Hence, a Markov chain containing self-loops i.e transitions to itself with positive probability, is aperiodic.

Definition 2.19. A Markov chain is *ergodic* if it is aperiodic and irreducible.

As the below theorem reveals, ergodicity shows that any initial distribution over Ω converges to a unique stationary distribution.

Theorem 2.20. *If a Markov chain is ergodic, then it converges to a unique stationary distributon π i.e P has a unique eigenvector with eigenvalue one.*

One can view this as special case of the *Perron-Frobenius Theorem for Regular Matrices.*

Theorem 2.21. *Suppose that P is transition matrix with positive entries and also that $P_{ij}^k > 0$ for some sufficiently large k and for all entries i, j (Regularity). Then the following must hold:*

(1) *The largest eigenvalue must be real and postive*
(2) *The largest left eigenvector π must be positive and unique*
(3) $\lim_{k \to \infty} q P^{(k)} = \pi$ *for all initial distributions q.*

Since our matrices are stochastic i.e $\sum_j P_{ij} = 1$ for all i, our largest eigenvalue will be one. Furthermore, the left eigenvector of eigenvalue one will be the stationary distribution π. Since our Markov chain is both irreducible and aperiodic, it must be regular as well.

Remark 2.22. It is not difficult to find counterexamples to Theorem 2.21 if the Markov chain is not ergodic. For instance, consider the chain
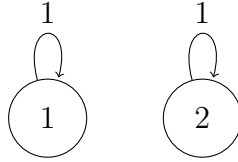

FIGURE 1. An aperiodic but reducible chain

The transition matrix will just be $P = I$. Although the largest eigenvalue is positive and real, its associated left eigenvector is not unique. Thus, it does not have an unique stationary distribution. Another simple chain is
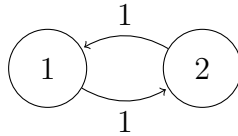

FIGURE 2. An irreducible but periodic chain

Once again, this chain does not have a unique left eigenvector of unit eigenvalue. Although the stationary distribution $\pi = [\frac{1}{2} \ \frac{1}{2}]$ is unique, no initial distribution converges to it besides π itself. Finally, consider the chain
The transition matrix will be

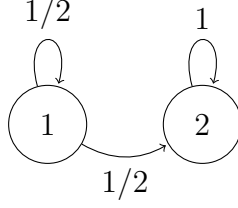$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$$

FIGURE 3. An aperiodic, reducible chain which converges

which has eigenvalues $\lambda_1 = 1, \lambda_2 = \frac{1}{2}$. Every initial distribution also converges to the unique stationary distribution $\pi = [0\ 1]$. Hence, this chain exhibits a counterexample to a converse of Theorem 2.21.

Definition 2.23. A Markov chain is *reversible* if the following equality holds for all states $i, j \in \Omega$

$$\pi(i)P_{ij} = \pi(j)P_{ji} \tag{2.26}$$

If the stationary distributon is uniform, then it is simple to see that reversability is equivalent to P being symmetric. Reversability is necessary to invoke certain mixing time lemmas which are introduced in the upcoming section.

2.4.2. *Mixing Times.* We introduce the definition of mixing times presented in [3]. For a deeper treatment of the properties below, see [5].

Definition 2.24. Let π be the stationary distribution of ergodic chain P. The 1-norm mixing time $\tau : [0, 1] \to \mathbb{R}$ is real-valued function defined as

$$\tau(\epsilon) = \max_q \min_{t \geq 0} ||qP^t - \pi|| \leq \epsilon \tag{2.27}$$

where $||p - q|| = \frac{1}{2} \sum_{i \in \Omega} |p_i - q_i|$

There is also an associated 2-norm version as well:

Definition 2.25. Let π be the stationary distribution of ergodic chain P. The 2-norm mixing time $\tau : [0, 1] \to \mathbb{R}$ is real-valued function defined as

$$\tau(\epsilon) = \max_q \min_{t \geq 0} ||qP^t - \pi||_2 \leq \epsilon \tag{2.28}$$

These definitions allow us to measure the length of time required for a random walk on an ergodic Markov chain to be within ϵ from the stationary distribution. Consquently, we can express a deep integral inequality bounding the mixing time of an ergodic, reversible chain using its *spectral gap*.

Theorem 2.26 (Refer to [5], Corollary 1.15).

$$(2.29) \qquad \tau(\epsilon) \leq \frac{1}{\Delta} \log \frac{1}{\pi_* \epsilon}$$

where $\pi_* = \min_{x \in \Omega} \pi(x)$ and $\Delta = \min 1 - \lambda_2, 1 + \lambda_{min}$ Here λ_2 refers to the second largest eigenvalue of P and λ_{min} refers to the smallest. This Δ is known as the spectral gap.

For irreversible chains, the same holds for PP^*

Theorem 2.27 ((Refer to [5]), Corollary 1.14).

$$(2.30) \qquad \tau(\epsilon) \leq \frac{2}{\Delta_{PP^*}} \log \frac{1}{\pi_* \epsilon}$$

The inequality can be converted for the 2-norm as well:

$$(2.31) \qquad \tau_2(\epsilon) \leq \frac{1}{\Delta} \log(1/\epsilon)$$

Finally, one more integral tool is required to lower bound the spectral gap of our matrix P in Equation 2.24. It is worth stressing that the Markov chain associated to P_μ which we will define as

$$(2.32) \qquad P_\mu - \frac{1}{n(n-1)} \sum_{i \neq j} \widehat{G_\mu}^{(ij)}$$

For this chain, we sample from a 2-copy gapped distribution μ over $\mathbb{U}(4)$ rather than over Haar distribution. It is unclear exactly how this chain evolves as we cannot explicitly compute P_μ as we did for $P_{\mathbb{U}(4)}$. However, with the *Comparision Theorem*, we can at least lower bound the spectral gap associated with P_μ, say Δ_μ, with $\Delta_{\mathbb{U}(4)}$. We first state a more general version of the theorem

Theorem 2.28 (Refer to [5], Theorem 2.3). *Let P and $\widehat{P}$ be two Markov chains with the same state space Ω and the same stationary distribution π. Then, for ever $x \neq y \in \Omega$ with $\widehat{P}_{xy} > 0$, construct a set of directed paths $\mathcal{P}_{xy}$ from x to y along edges in P. Let $\mathcal{P} = \bigsqcup_{x \neq y} \mathcal{P}_{xy}$. We define the flow function $f : \mathcal{P} \to [0,1]$ such that*

$$(2.33) \qquad \sum_{\gamma_{xy} \in \mathcal{P}_{xy}} f(\gamma_{xy}) = \widehat{P}_{xy}$$

then

$$\Delta \geq \widehat{\Delta}/A$$

where

$$A = \max_{a \neq b, P_{ab} > 0} \frac{1}{\pi_a P_{ab}} \sum_{x \neq y} \sum_{\substack{\gamma_{xy} \in \mathcal{P}_{xy} \\ (a,b) \in \gamma_{xy}}} \pi(x) f(\gamma_{xy}) |\gamma_{xy}|$$

where $|\gamma_{xy}|$ refers to the length of the path γ_{xy}.

Note that the bounds yielded will depend on the paths chosen for all $x \neq y \in \Omega$ such that $\widehat{P}_{xy} > 0$ and the values assigned to those paths by the constructed flow function f. Figure 2.4.2 shows an example of creating paths as specified in the theorem.
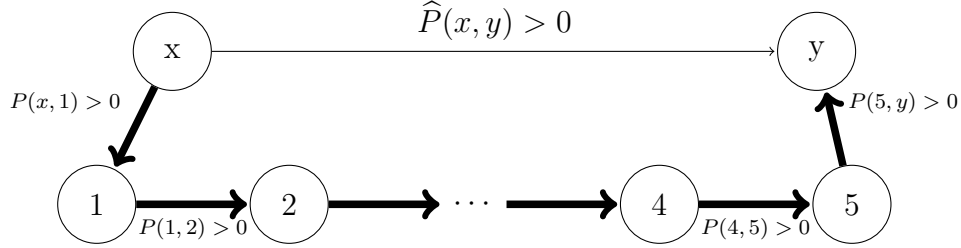

FIGURE 4. Creating paths in spirit of Theorem 2.28. Thick lines delineate the constructed path.

By Theorem 2.16, both P_μ and $P_{\mathbb{U}(4)}$ converge to same stationary distribution and both share the same state space. Using Theorem 2.28, we can bound the spectral gap of P_μ, Δ_μ with the with spectral gap of the $P_{\mathbb{U}(4)}$ chain, $\Delta_{\mathbb{U}(4)}$. However, we must first ensure that the Markov chain associated P_μ is ergodic. To achieve this, we simply delete the state $\sigma_0 \otimes \sigma_0$ from the state space. This will turn P into an irreducible chain as $\sigma_0 \otimes \sigma_0$ the only isolated component in the chain. This chain has self-loops by Theorem 2.15 and thus, it must be ergodic. This puts us in a good position to state the lower bound

Theorem 2.29. *Let μ be any 2-copy gapped distribution of gates. If μ is invariant under Hermitian conjugation then let Δ_μ be the spectral gap of resulting Markov chain P_μ. Then*

$$(2.34) \qquad\qquad \Delta_\mu = \Omega(\Delta_{\mathbb{U}(4)})$$

If μ is not invariant under Hermitian conjugation, then let Δ_{PP^*} be the spectral gap of the chain PP^*. Then

$$(2.35) \qquad\qquad \Delta_{PP^*} = \Omega(\Delta_{\mathbb{U}(4)})$$

Proof. We only show the symmetric case for now.

First, let us start with an simple example on a 4-qubit system to demonstrate the construction of the relevant paths and flow function. Consider the valid transiton in $P_{\mathbb{U}(4)}$, $1200 \rightsquigarrow 1300$ accomplished by picking sites $1, 2$ and transitioning to $1, 3$. This shows that $P_{\mathbb{U}(4)}(1200, 1300) > 0$. An immediate obstacle arises by observing that this may not a valid transition in the chain P_μ. However, from Section 2.2.1, we know that $\hat{G}_\mu$ acts transitively on this subspace of two qubits $\mathbb{C}^4 \otimes \mathbb{C}^4$. Hence, no cycles can exist during the evolution i.e there exists no path $12 \rightsquigarrow \cdots \rightsquigarrow 12$ with length < 14. Consequently, only a constant number of transitions have to occur from $12 \rightsquigarrow 13$ and the length of the path cannot be more than 14. Also, there exists two more possbile transition paths from $1200 \rightsquigarrow 1300$, namely the ones

which pick sites $2, 3$ and sites $2, 4$. The Markov chain just has to transition from $20 \rightsquigarrow 30$ for both cases. This gives us *three* paths from $1200 \rightsquigarrow 1300$ which motivates us to construct our flow function to assign $\frac{1}{3}$ to each path. The procedure for the general case follows similarly:

We begin by noting that there two cases to consider. Let $h(x, y)$ denote the *Hamming distance* between strings $x, y \in \{0, 1, 2, 3\}^n$ defined to be $h(x, y) = |\{i \in [n] \mid x_i \neq y_i\}|$. There only two cases where $P_{\mathbb{U}(4)}(x, y) > 0$:

(1) ($h(x,y) = 2$). Let i, j be the unique indices such that $x_i \neq y_j$. By the discussion above, there exists a constant length path in P_μ, $x \rightsquigarrow y$ acting on sites i, j.
(2) ($h(x,y) = 1$). Let i be the index where $x_i \neq y_i$. There are two other subcases in this instance. If the inital state $x_i \neq 0$, then the chain choose any other of the $n - 1$ sites to transiton to y. To see this, view the desired transition as

$$x_1 \cdots x_i \cdots x_n \rightsquigarrow y_1 \cdots y_i \cdots y_n$$

For any $x_j, j < i$ (w.l.o.g), we can transition by $x_j x_i \rightsquigarrow x_j y_i = y_j y_i$. Finally, there is the case when $x_i = 0$. This precludes us from choosing any site containing the identity state ($x_j = 0$) as $00 \rightsquigarrow 00$ only. This shows us that we can use $n - z(x)$ sites where $z(x)$ denotes the number of zero sites in string x.

We can now define the flow function $f : \mathcal{P} \to [0, 1]$ through the observations above:

(1) ($h(x,y) = 2$) There is only one path σ_{xy} in this case. So set $f(\sigma_{xy}) = P_{\mathbb{U}(4)}(x, y) = \Theta(1/n^2)$ by Theorem 2.15.
(2) ($h(x,y) = 1$) For the case of non-zero initial value, there are $n - 1$ constant-length paths. Equally distributing the weight across all paths shows that $f(\gamma_{xy}) = \frac{P_{\mathbb{U}(4)}(x,y)}{n-1} = \Theta(1/n^2)$ for $\gamma_{xy} \in \mathcal{P}_{xy}$ as $P_{\mathbb{U}(4)}(x, y) = \Theta(1/n)$. For the case of zero initial value: $f(\gamma_{xy}) = \frac{P_{\mathbb{U}(4)}(x,y)}{n-z(x)} = \Theta(1/n^2)$ since $P_{\mathbb{U}}(4)(x, y) = \Theta(\frac{n-z(x)}{n^2})$.

All that remains is to calculate the number of edges (a, b) such that $P_\mu(a, b) > 0$ and $(a, b) \in \gamma_{xy}$ for some path $\gamma_{xy} \in \mathcal{P}_{xy}$. First, we reduce the expression of the constant A found in Theorem 2.28:

$$(2.36) \qquad A = \max_{a \neq b, \ P_\mu(a,b) > 0} A(a, b)$$

$$(2.37) \qquad A(a, b) = \frac{1}{P(a, b)} \sum_{\substack{\gamma_{xy} \in \sum_{x \neq y} \mathcal{P}_{xy} \\ (a,b) \in \gamma_{xy}}} f(\gamma_{xy})$$

This arises as π is just the uniform distribution and the path lengths $|\gamma_{xy}|$ will always be constant.

(1) ($h(a,b) = 2$) Let i, j be the two differing sites. Recall that our paths pick two sites and evolve accordingly. In other words, no path picks two sites and changes sites in

the middle of the path (no path evolves like $1200 \rightsquigarrow 1010 \rightsquigarrow 1300$). Furthermore, any such x, y must have $h(x, y) = 1, 2$ in order for it's path γ_{xy} to use edge (a, b). There are only a constant number of such pairs x, y which fit the above criterion. Additionally, by the $h(x, y) = 2$ case, there is a unique path which *only uses* sites i, j to transition from $x \rightsquigarrow y$. To see this, the case where $h(x, y) = 2$ is simple as there is only one path by construction above. If $h(x, y) = 1$ and (w.l.o.g) only differ at site i, then there is only one path from x which picks the exact pair i, j and transitions to y. Hence, there are is overall a constant number of paths which can use (a, b). The sum in 2.36 will reduce to $\frac{1}{\Theta(1/n^2)} \cdot \Theta(1/n^2) = \Theta(1)$ as $P_\mu(a, b) = \Theta(1/n^2)$.

(2) *($h(a,b) = 1$)* Let r be the number of allowed transitions between a, b (not all pairs may yield direct transitions between a, b). We can reduce to the case above for every pair, showing that the sum cumulates to $\Theta(r/n^2)$. As $P_\mu(a, b) = \Theta(r/n^2)$, this once again shows that $A = \Theta(1)$.

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

A significant portion of [3] is dedicated to proving that $\Delta_{\mathbb{U}(4)} = \Theta(1/n)$. Invoking Lemma 2.26 and 2.31 along with this tight bound on the spectral gap of the $\mathbb{U}(4)$ chain shows that

Corollary 2.29.1. *The chain P_μ has 1-norm mixing time $\mathcal{O}(n(n + \log 1/\epsilon))$ and 2-norm $\mathcal{O}(n \log \frac{n}{\epsilon})$*

In fact, by Theorem 2.29 and Lemma 2.26, the mixing time of the $P_{\mathbb{U}(4)}$ chain upper-bounds the mixing time of the P_μ chain. This mixing time bound is required to finally prove Lemma 2.10 as intended.

2.5. **Tight Analysis of $\mathbb{U}(4)$ gap.** In the last section, we used the Comparision Theorem (Theorem 2.28) to bound the gap of the P_μ chain where μ is some 2-copy gapped distribution. The rest of this section will be devoted to proving tight bounds on the spectral gap of the $\mathbb{U}(4)$ chain, $\Delta_{\mathbb{U}(4)}$. We reiterate that the Markov chain transitions will be governed as described in Theorem 2.15.

For this case, we will just sketch the overarching ideas of the section and leave some details to the appendix of this report and [3]. Once again, we restrict our attention to the coefficients of diagonal states and delete the identity $\sigma_0 \otimes \sigma_0$ in order for the chain to become irreducible. The analysis starts with the consideration of an auxiliary chain called the *zero chain*. The zero chain describes the number of zeros present in a string. The idea is that the number of zeroes slow the mixing process since picking two zeroes effectively changes nothing. If we can show that distribution of zeros mixes well-enough, it can be used to further infer about the mixing of the total chain. The transition probabilities are worked out in a lemma

Lemma 2.30. *The zero chain has the following transition matrix P on the state space $\Omega = \{1, 2, \cdots, n\}$. This chain actually counts the number of non-zeros of state.*

(2.38)
$$P(x, y) = \begin{cases} 1 - \frac{2x(3n-2x-1)}{5n(n-1)} & y = x \\ \frac{2x(x-1)}{5n(n-1)} & y = x - 1 \\ \frac{6x(n-x)}{5n(n-1)} & y = x + 1 \\ 0 & otherwise \end{cases}$$

for $1 \leq x, y \leq n$

Proof. This can be derived through a simple counting argument. For example, for the number of zeros to decrease (x increases), two sites with a non-zero value and a zero value must be chosen and the chain must transition to pair of non-zero values. 9 out of the 15 possible transition values consist of two non-zero values. There the probabilty of picking a zero and a non-zero value is $\frac{2x(n-x)}{n(n-1)}$, bringing the total value to $\frac{9}{15} \frac{2x(n-x)}{n(n-1)} = \frac{6x(n-x)}{5n(n-1)}$. $\qquad\square$

It is also simple to show that the stationary distribution π_0 of the zero chain will be

(2.39)
$$\pi_0(x) = \frac{3^x \binom{n}{x}}{4^n - 1}$$

The integral property about this zero chain is that it mixes rapidly

Theorem 2.31. *The zero chain mixes in $\Theta(n \log \frac{n}{\epsilon})$ time.*

which leads to the rapid mixing of the total chain

Theorem 2.32. *The total chain mixes in time $\Theta(n \log \frac{n}{\epsilon})$*

Remark 2.33. Diniz and Johnathan point out in [2] a faulty assumption made by the authors of [3]. In the proof of Corollary 2.32, Harrow and Low make the argument that once every site has been paired with another site with non-zero value, the total chain will be uniformly mixed. This would reduce the mixing time to a coupon collectors problem. However, the authors of [2] show a simple counterexample to this statement. Let us focus our attention to a 3-qubit system with initial state $(0, 0, 1)$. We want to see the resulting distribution after hitting all three sites with two transitions. First, observe that we have the following transitions $(0, 0, 1) \rightsquigarrow (1, 0, 0)$

$$001 \rightsquigarrow 100 \rightsquigarrow 100$$
$$001 \rightsquigarrow 010 \rightsquigarrow 100$$
$$001 \rightsquigarrow 001 \rightsquigarrow 100$$

whereas we have the transitions $(0,0,1) \rightsquigarrow (0,0,1)$

$$001 \rightsquigarrow 001 \rightsquigarrow 001$$
$$001 \rightsquigarrow 001 \rightsquigarrow 001$$

by first hitting the sites $2, 3$ and $1, 3$. The second transition hits the sites in the reverse order. This reveals that the number of ways to get to states $(1,0,0), (0,0,1)$ will be in ratio $3 : 2$; the final state will not be uniformly distributed.

The proofs of these two theorems will also be assigned a portion of the appendix. By the discussion of Corollary 2.29.1, this proves Lemma 2.10.

2.6. **Tying Some Loose Ends.** We dedicate this section to addressing some details glossed over in the previous two sections. In particular, the vanishing rate of the non-diagonal terms must be handled before considering the proof for Theorem 2.8. First, the proof from Lemma 2.10 to the more general statement of Corollary 2.10.1 must be given. The proof will demonstrate how the more restricted set of initial states used in Lemma 2.10 can easily generalize to all initial states, even those with negative diagonal coefficients.

(Proof of Corollary 2.10.1). Begin by first noting that Lemma 2.10 immediately implies the corollary for coeffcients of non-diagonal terms $\gamma(p, q)$, $p \neq q$. Thus, it suffices to only consider the diagonal terms $\gamma(p, p)$. Now from the observations made in Theorem 2.15 and linearity, we can view the expected coeffcients $\mathbb{E}_W \gamma_W(p, p)$ over all t-length random circuits W to follow that of a random walk over the diagonal states. Let $g_t(p; q)$ be the probability of arriving at state p, p in t-steps starting from state q, q via the random walk defined in Theorem 2.15. This means that $g_0(p, q) = \delta_{p,q}$. We can calculate the expected coeffcients as

$$(2.40) \qquad \gamma_t(p, p) = \sum_{q \neq 0} g_t(p, q) \gamma_0(q, q)$$

$\square$

By performing an effective normalization, we can generalize Lemma 2.10 to the corollary as follows: Consider first the inequality for the 2-norm

$$\sum_{p \neq 0} \left(\gamma_t(p,p) - \frac{\sum_{q \neq 0} \gamma_0(q,q)}{4^n - 1} \right)^2 = \sum_{p \neq 0} \left(\sum_{q \neq 0} \gamma_0(q,q) \left(g_t(p,q) - \frac{1}{4^n - 1} \right) \right)^2$$

$$\leq \left(\sum_{q \neq 0} |\gamma_0(q,q)|^2 \right) \sum_{q' \neq 0} \sum_{p \neq 0} \left(g_t(p,q') - \frac{1}{4^n - 1} \right)^2$$

$$\leq \left(\sum_{q \neq 0} |\gamma_0(q,q)|^2 \right) (4^n - 1)\epsilon$$

$$= 4^n \epsilon \cdot Tr(\rho^2)$$

$$\leq 4^n \epsilon$$

The first inequality arises from the Cauchy-Schwartz inequality and rearranging terms. The second inequality is Lemma 2.10, and the final inequality is just $Tr(\rho^2) \leq 1$ for any state ρ. Invoking Lemma 2.10 shows that the above inequality holds for $t \geq n(n + \log 4^n/\epsilon) = Cn((\log 4 + 1)n + \log 1/\epsilon)$ for sufficiently large C chosen in the Lemma. Now choose a larger constant C' dominating the righthand expression to prove the desired result. The 1-norm case proceeds similarly:

$$\sum_{p \neq 0} \left(\gamma_t(p,p) - \frac{\sum_{q \neq 0} \gamma_0(q,q)}{4^n - 1} \right) = \sum_{p \neq 0} \left| \sum_{q \neq 0} \gamma_0(q,q) \left(g_t(p,q) - \frac{1}{4^n - 1} \right) \right|$$

$$\leq \sum_{q \neq 0} |\gamma_0(q,q)| \sum_{p \neq 0} \left| g_t(p,q) - \frac{1}{4^n - 1} \right|$$

$$\leq \sum_{q \neq 0} |\gamma_0(q,q)| 2^n \epsilon$$

$$\leq 2^n \epsilon$$

2.6.1. *First and Second Moments Convergence for 1-designs.* We show that our scheme acts an approximate 1-design in the sense below:

Lemma 2.34.

(2.41) $$\sum_{p \neq 0} (\mathbb{E}_{W_t} |\gamma_{W_t}(p)|)^2 \leq \epsilon$$

after $t \geq n \log 1/\epsilon$
    and

(2.42) $$\sum_{p \neq 0} \mathbb{E}_{W_t} |\gamma_{W_t}(p)| \leq \epsilon$$

after $t \geq n \log n / \epsilon$

Note that since we are considering 1-designs here, we index through only one n-string p. From subsection 2.2.1, recall that applying a unitary $U \sim \mathbb{U}(4)$ completely randomizes the qubit pair if at least one of them are non-zero i.e take basis operators $\sigma_1, \sigma_2, \sigma_3$ in the Pauli basis. This means that if sites i, j are picked, then coefficients of basis states $\gamma(p)$ where $p_i \neq 0$ or $p_j \neq 0$ vanish under U. In this case, we say that sites i, j have been *hit*. For an unitary sampled over a universal distribution $U \sim \mu$, the coefficients shrink by a factor $1 - \Delta$ where Δ is deemed as a gap for distribution μ. Hence, sites which have been hit r times shrink by a factor $(1 - \Delta)^r$. The case for Haar distribution follows the analysis of the standard coupon collectors problem (see [7] for an introduction). However, the case for any universal distribution follows from a similar line of thinking as demonstrated in the proof of Lemma 2.6.1.

(Proof of Lemma). Let us first focus on the case for the coefficient of a specific basis state $\mathbb{E}_{W_t} |\gamma_{W_t}(p)|$. From the discussion above, the coefficient shrinks by a factor of $1 - \Delta$ if a pair with at least one non-zero site is picked. The probablity of this event occuring is easily calculated to be

$$1 - \frac{(n-d)(n-d-1)}{n(n-1)} \geq 1 - \frac{n-d}{n} = \frac{d}{n}$$

if d denotes the number of non-zeros contained in n-string p. This gives us the recurrence inequality:

$$\mathbb{E}_{W_t} |\gamma_{W_t}(p)| \leq ((1-\Delta)(d/n) + (1 - d/n)) \cdot |\gamma_{W_{t-1}}(p)| = \left(2 - \frac{\Delta d}{n}\right) \cdot |\gamma_{W_{t-1}}(p)|$$

Iterating this t times gives the inequality:

$$\mathbb{E}_{W_t} |\gamma_{W_t}(p)| \leq \left(2 - \frac{\Delta d}{n}\right)^t |\gamma_0(p)| = \left(2 - \frac{\Delta d}{n}\right)^{n \cdot \frac{t}{n}} |\gamma_0(p)| \leq e^{-\Delta d t / n} |\gamma_0(p)|$$

Summing over all p then yields the total inequality

$$\sum_{p \neq 0} \mathbb{E}_{W_t} |\gamma_{W_t}(p)| \leq \sum_{d=1}^{n} e^{-\Delta d t / n} \sum_{d(p)=d} |\gamma_0(p)|$$

where $d(p)$ is the number of non-zeros in string p. With th– obvious bound $|\gamma_0(p)| \leq 1$, we give the total bound:

$$\sum_{d(p)=d} |\gamma_0(p)| \leq \binom{n}{d} 3^d$$

This is easily seen as the number of basis states with exactly d non-zero states since each non-zero site would take values in $\{\sigma_1, \sigma_2, \sigma_3\}$. Substituting this bound into the inequality

above:

$$\sum_{d=1}^{n} e^{-\Delta dt/n} \sum_{d(p)=d} |\gamma_0(p)| \leq \binom{n}{d}(3e^{-\Delta t/n})^d \leq (1 + 3e^{-\Delta t/n})^n - 1$$

where the last inequality is an invocation of the Binomial theorem. Setting $t \geq \frac{n}{\Delta} \log \frac{3n}{\epsilon}$ yields

$$\sum_{p \neq 0} \mathbb{E}_{W_t} |\gamma_{W_t}(p)| \leq \left(1 + \frac{\epsilon}{n}\right)^n - 1 = \mathcal{O}(\epsilon)$$

The bound for the 2-norm is dervied in a similar fashion. See [3], Lemma 4.1 for the corresponding 2-norm bound. □

The proof for Lemma 2.6.1 can be repurposed to prove rapid vanishing of the non-diagonal terms $\sigma_{p_1} \otimes \sigma_{p_2}$ as presented in the lemma below:

Lemma 2.35.

(2.43) $$\sum_{p \neq q} \left(\mathbb{E}_{W_t} |\gamma_{W_t}(p,q)|\right)^2 \leq \epsilon$$

for $t \geq n \log \frac{1}{\epsilon}$ and

(2.44) $$\sum_{p \neq q} \mathbb{E}_{W_t} |\gamma_{W_t}(p,q)| \leq \epsilon$$

Proof. From the analysis for the second moments given in Section 2.2.1, the cofficients decay by a factor of $1 - \Delta$ if $\sigma_{p_i} \otimes \sigma_{p_j}$ is different from the state $\sigma_{q_i} \otimes \sigma_{q_j}$. Thus, picking at least one differing site will trigger a shrinking of the coefficient $\gamma(p,q)$. This motivates the setting of the parameter d in the previous proof to be the number of *non-differing sites*. This results in the same inequality

$$\mathbb{E}_{W_t} |\gamma_{W_t}(p,q)| \leq \left(2 - \frac{\Delta d}{n}\right) \cdot |\gamma_{W_{t-1}}(p,q)|$$

which yields the identical inequality:

$$\sum_{\substack{p \neq q \\ pq \neq 00}} \mathbb{E}_{W_t} |\gamma_{W_t}(p,q)| \leq \sum_{d=1}^{n} e^{-\Delta dt/n} \sum_{d(p,q)=d} |\gamma_0(p,q)|$$

where $d(p,q)$ here would count the number of non-differing sites instead. We bound $\sum_{d(p,q)=d} |\gamma_0(p,q)| \leq \binom{n}{d} 4^d 12^{n-d}$ to arrive at

$$\sum_{\substack{p \neq q \\ pq \neq 00}} \mathbb{E}_{W_t} |\gamma_{W_t}(p,q)| \leq \sum_{d=1}^{n} e^{-\Delta dt/n} \binom{n}{d} 4^d 12^{n-d} = 4^n \sum_{d=1}^{n} \binom{n}{d} 3^{n-d} = 4^n(1 + 3e^{-\Delta t/n})^n - 1)$$

Set $t \geq \frac{n}{\Delta}(n \log 4 + \log 1/\epsilon)$ to finish the proof. $\hfill\square$

(Proof of Theorem 2.8). We are now in the position to prove the main result in respect to the *diamond norm.* The following string of inequalities hold:

$$(2.45) \quad ||\mathcal{G}_W - \mathcal{G}_H||_\diamond^2 = \sup_d ||\mathcal{G}_W \otimes I_d - \mathcal{G}_W \otimes I_d||_\infty = \sup_\rho ||(\mathcal{G}_W \otimes I_n)(\rho) - (\mathcal{G}_H \otimes I_n)(\rho)||_1^2$$

$$(2.46) \qquad\qquad \leq (2^{2(2n)}) \sup_\rho ||(\mathcal{G}_W \otimes I_n)(\rho) - (\mathcal{G}_H \otimes I_n)(\rho)||_2^2$$

$$(2.47) \qquad\qquad = 2^{4n} \sup_\rho || \sum_{p_1,p_2,p_3,p_4} \gamma_\rho(p_1,p_2,p_3,p_4)(\mathcal{G}_W(\sigma_{p_1} \otimes \sigma_{p_2}) \otimes \sigma_{p_3} \otimes \sigma_{p_4}$$

$$(2.48) \qquad\qquad - \mathcal{G}_H(\sigma_{p_1} \otimes \sigma_{p_2}) \otimes \sigma_{p_3} \otimes \sigma_{p_4})||_2^2$$

The first inequality is just the standard inequality $||\rho||_1 \leq d||\rho||_2$ where d is the dimension of the system $\rho \in \mathbb{C}^d$. The last equality arises from linearity. Recall that

$$\mathcal{G}_{W_t}(\sigma_{p_1} \otimes \sigma_{p_2}) = \sum_{\substack{q_1,q_2 \\ q_1 q_2 \neq 00}} g_t(p_1,p_2;q_1,q_2)\sigma_{q_1} \otimes \sigma_{q_2}$$

over all random circuits of length t. Here g_t represents the evolution of the coefficients as described by the combination of the analyses for the diagonal coefficients $g_t(p,p;q,q)$ and the non-diagonal ones presented above, culminating in Corollary 2.10.1. This yields

$$2^{4n} \sup_\rho ||\gamma_\rho(p_1,p_2,p_3,p_4)(\mathcal{G}_W(\sigma_{p_1} \otimes \sigma_{p_2}) \otimes \sigma_{p_3} \otimes \sigma_{p_4} - \mathcal{G}_H(\sigma_{p_1} \otimes \sigma_{p_2}) \otimes \sigma_{p_3} \otimes \sigma_{p_4})||_2^2$$

$$= 2^{4n} \sup_\rho || \sum_{p_3,p_4} \sum_{\substack{p_1,p_2,q_1,q_2 \\ q_1,q_2 \neq 00}} \gamma_\rho(p_1,p_2,p_3,p_4) \left(g_t(p_1,p_2;q_1,q_2) - \frac{\delta_{p_1,p_2}\delta_{q_1,q_2}}{2^n(2^n+1)} \right) \sigma_{q_1} \otimes \sigma_{q_2} \otimes \sigma_{p_3} \otimes \sigma_{p_4}||_2^2$$

Taking the Hilbert-Schmidt norm and squaring the result kills the non-aligning terms to output

$$2^{4n} \sup_\rho \sum_{\substack{p_1,p_2,p_3,p_4,q_1,q_2 \\ q_1,q_2 \neq 0}} \gamma_\rho(p_1,p_2,p_3,p_4)^2 \left(g_t(p_1,p_2;q_1,q_2) - \frac{\delta_{p_1,p_2}\delta_{q_1,q_2}}{2^n(2^n+1)} \right)^2$$

$$\leq 2^{4n} \sup_\rho \sum_{\substack{p_1,p_2,p_3,p_4,q_1,q_2 \\ q_1,q_2 \neq 0}} \gamma_\rho(p_1,p_2,p_3,p_4)^2 \epsilon^2$$

$$\leq 2^{4n} \epsilon^2$$

The last equality is from Lemma 2.10. This finally shows that

$$(2.49) \qquad\qquad\qquad ||\mathcal{G}_W - \mathcal{G}_H||_\diamond \leq 2^{2n}\epsilon$$

when $t \geq Cn(n + \log 1/\epsilon)$. Take the final constant C' to be sufficiently large such that $C'n(n + \log 1/\epsilon) \geq Cn(n + \log 2^{2n}/\epsilon)$ to finish the proof. $\hfill\square$

References

1. Thomas M Cover, *Elements of information theory*, John Wiley & Sons, 1999.
2. Igor Tuche Diniz and Daniel Jonathan, *Comment on "random quantum circuits are approximate 2-designs" by aw harrow and ra low (commun. math. phys. 291, 257–302 (2009))*, Communications in Mathematical Physics **304** (2011), no. 1, 281–293.
3. Aram W Harrow and Richard A Low, *Random quantum circuits are approximate 2-designs*, Communications in Mathematical Physics **291** (2009), no. 1, 257–302.
4. Emanuel Knill, *Approximation by quantum circuits*, arXiv preprint quant-ph/9508006 (1995).
5. Ravi R Montenegro and Prasad Tetali, *Mathematical aspects of mixing times in markov chains*, Now Publishers Inc, 2006.
6. Roberto Oliveira, Oscar CO Dahlsten, and Martin B Plenio, *Efficient generation of generic entanglement*, arXiv preprint quant-ph/0605126 (2006).
7. Prabhakar Raghavan and Rajeev Motwani, *Randomized algorithms*, Cambridge University Press Cambridge, 1995.