Notes on Unitary Designs

Yuzhen Zhang

July 29, 2023

Contents

1 Introduction

Random unitaries plays an important role in quantum information and holography. To quantum information theorists, it is a strong tool for analytical computation. To study the properties of "generic" states, one first compute the Haar averged quantities, and then show that the deviation from the average is small. An example is Page's entropy curve for generic random states. In information theoretic studies of black holes, people have extensively used random unitaries as toy models, due to their strong scrambling power.

In experiments, it is extremly expensive to prepare Haar ensembles[1]. So it would be very nice if we can find ensemble of finite number of states/unitaries that can captures the Haar randomness to a desirable extent. This can be achieved by k-designs—ensmebles of states or unitaries that reproduce the k'th moment of the Haar ensemble. The first moment describes thermalization. The higher moments are responsible for observables that depend on more usages of the density matrix: Renyi entropies, OTOC, etc. Experimentally they cannot be accessed by a single copy of the quantum state. Designs can also help us understand holography—to what extent is it ligitimit to replace the black hole evolution with random unitaries? It can also be related to complexity, which is extremely hard to analytically compute.

2 Random states and unitaries

Definition 1 (random pure states)**.** *Given a basis* $\{|i\rangle\}$*, define unormalized vectors*

$$|\phi\rangle = \sum_i c_i |i\rangle \tag{1}$$

whose coeffecients are drawn form the standard normal distribution

$$p(\{c_i\}) = \frac{1}{(2\pi)^d} e^{-\frac{1}{2}\sum_i |c_i|^2} \tag{2}$$

The probablity measure η *for random states is defined as*

$$\eta(\mathcal{A}) = \text{Prob}(|\psi\rangle = \alpha |\phi\rangle \ \textit{ for some } \alpha \in \mathbb{C},\ |phi\rangle \in \mathcal{H}) \tag{3}$$

The measure is invariant under the action of any unitary:

$$\eta(U\mathcal{A}) = \eta(\mathcal{A}) \tag{4}$$

It follows from the fact that the action of unitaries preserve the norm of vectors, and that the Gaussian probablity distribution is fixed by the norm.

[1]If we have n qubits, there are $\sim e^{e^n}$ number of ϵ balls in the unitary group. Not only are there too many unitaries, but they are also very hard to prepare.

Definition 2 (Haar)**.** *Haar measure is the unique measure in $U(d)$ that is invariant under the left/right action of the group. For $\forall \mathcal{S} \in \text{Borel}(U(d))$,*

$$\mu(\mathcal{S}) = \mu(U\mathcal{S}) = \mu(\mathcal{S}U), \quad \forall U \in U(d) \tag{5}$$

One concrete way to construct Haar random unitary matrices is to apply QR or Gram-Schimidt to the Ginibre ensemble [1], whose matrix elements are independenly Gaussain random.

We can generate random pure states by choosing a reference state and acting Haar random unitaries on it. This is gauranteed by the following theorem.

Theorem 1 (Random unitaries generate random states)**.** *Let μ denote the Haar measure, and let η denote the measure for random pure states in $\mathcal{H}$. For any $\mathcal{A} \in \text{Borel}(\mathcal{H})$ and $|\psi\rangle \in \mathcal{H}$, it holds that*

$$\eta(\mathcal{A}) = \mu(\{U \in U(d) : U|\psi\rangle \in \mathcal{A}\}) \tag{6}$$

Proof. Define a step funcion for $\mathcal{A}$:

$$f(|\phi\rangle) = \begin{cases} 1, & |\phi\rangle \in \mathcal{A} \\ 0, & |\phi\rangle \notin \mathcal{A} \end{cases} \tag{7}$$

$$\eta(\mathcal{A}) = \int f(|\phi\rangle) d\eta(|\phi\rangle) = \int f(U|\phi\rangle) d\eta(|\phi\rangle) = \iint f(U|\phi\rangle) d\eta(|\phi\rangle) d\mu(U) \tag{8}$$

In the second equality we used the unitary invariance of the measure for random states. We can first integrate $\int f(U|\phi\rangle) d\mu(U)$ over U. In this integration we can change $|\phi\rangle$ into $|\psi\rangle$ because they can be related by some unitary V:

$$\int f(U|\phi\rangle) d\mu(U) = \int f(UV|\phi\rangle) d\mu(U) = \int f(U|\psi\rangle) d\mu(U) \tag{9}$$

where we have used the right invariance of the Haar measure. So we have

$$\eta(\mathcal{A}) = \iint f(U|\psi\rangle) d\eta(|\phi\rangle) d\mu(U) = \mu(\{U \in U(d) : U|\psi\rangle \in \mathcal{A}\}) \tag{10}$$

□

3 Moments of Haar

Before explicitly constructing the k-th moment of Haar random states, let's look at the permutation symmetric subspace. Given the Hilbert space $\mathcal{H}^{\otimes k}$, we define the permutation symmetric subspace to be

$$\mathcal{H}^{\vee \otimes k} := \{|\psi\rangle \in \mathcal{H}^{\otimes k} : W_\pi |\psi\rangle = |\psi\rangle, \forall \pi \in S_k\} \tag{11}$$

The projection operator on this subspace is just the equal weight sum over all permutations.

Proposition 1.

$$P_{sym} = \frac{1}{k!} \sum_{\pi \in S_n} W_\pi, \quad W_\pi = |i_{\pi(1)}, \cdots i_{\pi(k)}\rangle \langle i_1, \cdots i_k| \tag{12}$$

Proof. For $\forall |\psi\rangle \in \mathcal{H}^{\vee\otimes k}$, $P_{sym}|\psi\rangle = \frac{1}{k!}\sum_{\pi\in S_n} W_\pi |\psi\rangle = |\psi\rangle$ and $P_{sym}^2|\psi\rangle = |\psi\rangle$. For $\forall |\psi\rangle \in \mathcal{H}^{\otimes k}$ and $\forall \pi \in S_n$, $W_\pi P_{sym}|\psi\rangle = P_{sym}|\psi\rangle \Rightarrow P_{sym}|\psi\rangle \in \mathcal{H}^{\vee\otimes k}$ □

With an orthomormal basis in each replica $\{|i\rangle_{i=1}^d\}$, we can get an orthogonal basis of $\mathcal{H}^{\vee\otimes k}$ by doing the symmetrized tensor product:

$$P_{sym}|i_1\rangle \otimes \cdots \otimes |i_k\rangle = \frac{1}{k!}\sum_\pi W_\pi |i_{\pi(1)}\rangle \otimes \cdots \otimes |i_{\pi(k)}\rangle \tag{13}$$

It forms a complete basis because $\mathcal{H}^{\otimes k} = \text{span}\{|i_1\rangle \otimes \cdots \otimes |i_k\rangle\}$ and $P_{sym}\mathcal{H}^{\otimes k} = \mathcal{H}^{\vee\otimes k}$. To count the number of the base vectors, we can imagine putting k identical balls into d bags labbeld by $i = 1, \cdots, d$. Empty bags are allowed. Each way of doing this correspond to a base vector. This is equivalent to lining up the k balls and putting $d-1$ partitions between balls. In other words, each base vector correspond to an arrangement of k balls and $d-1$ partitions. There are $\binom{d+k-1}{k}$ of them. Thus the dimension of $\mathcal{H}^{\vee\otimes k}$ is

$$d_{sym} = \binom{d+k-1}{k} \tag{14}$$

Theorem 2 ([2]).

$$\mathcal{H}^{\vee\otimes k} = \text{span}\{|\psi\rangle^{\otimes n} : |\psi\rangle \in \mathcal{H}\} \tag{15}$$

Proof. It is straightforwrad that $\text{span}\{|\psi\rangle^{\otimes n} : |\psi\rangle \in \mathcal{H}\} \subseteq \mathcal{H}^{\vee\otimes k}$. We will show $\text{span}\{|\psi\rangle^{\otimes n} : |\psi\rangle \in \mathcal{H}\} \supseteq \mathcal{H}^{\vee\otimes k}$. We make use of polynomials in a vector space: $|p(x)\rangle = |v_0\rangle + x|v_1\rangle + \cdots + x^d|v_d\rangle$, $|v_0\rangle, \cdots, |v_d\rangle \in \mathcal{H}$. If $|p(x)\rangle \in \mathcal{H}' \subset \mathcal{H}$ for all x, then we can take derivatives with respect to x and find all of them to be in $\mathcal{H}'$. Thus we conclude $|v_0\rangle, \cdots, |v_d\rangle \in \mathcal{H}'$. Building on this, we consider the polynomial $|p(x_1, \cdots, x_d)\rangle = (\sum_i x_i |i\rangle)^{\otimes k}$ that spans the right hand side of (15). We can put the terms that have the coefficients $x_1^{t_1}\cdots x_2^{t_2}$ together: $|p(x_1, \cdots, x_d)\rangle = \sum_{t_1+\cdots+t_d=k} x_1^{t_1}\cdots x_2^{t_2}|v_{t_1\cdots t_d}\rangle$. We immediately notice that each $|v_{t_1\cdots t_d}\rangle$ is proportional to a base vector in the form of (13). Since all $|p(x_1, \cdots, x_d)\rangle \in \text{span}\{|\psi\rangle^{\otimes n} : |\psi\rangle \in \mathcal{H}\}$ and all $|v_{t_1\cdots t_d}\rangle \in \mathcal{H}^{\vee\otimes k}$, we conclude $\mathcal{H}^{\vee\otimes k} \subseteq \text{span}\{|\psi\rangle^{\otimes n} : |\psi\rangle \in \mathcal{H}\}$ □

In fact, one can span it with a smaller set, using Schwartz-Zippel.

We can also define the symmetric subspace for operators

$$\mathcal{L}(\mathcal{H})^{\vee\otimes k} := \{A \in \mathcal{L}(\mathcal{H})^{\otimes k} : W_\pi A W_\pi^\dagger = A, \forall \pi \in S_k\} \tag{16}$$

Theorem 3 ([3]).

$$\mathcal{L}(\mathcal{H})^{\vee\otimes k} = \text{span}\{(|\psi\rangle\langle\phi|)^{\otimes k} : |\psi\rangle, \langle\phi| \in \mathcal{H}\} = \text{span}\{U^{\otimes k} : U \in U(d)\} \tag{17}$$

Proof. Let W_π^{12} denote the permutation operator for $\mathcal{H}_{12} \equiv \mathcal{H}_1 \otimes \mathcal{H}_2$. A nice thing is that is factorizes as $W_\pi^{12} = W_\pi^1 \otimes W_\pi^2$. Then it follow that $(\mathcal{H}_1 \otimes \mathcal{H}_2)^{\vee\otimes k} = \mathcal{H}_1^{\vee\otimes k} \otimes \mathcal{H}_2^{\vee\otimes k}$. The same reasoning suggests $\mathcal{L}(\mathcal{H})^{\vee\otimes k} = (\mathcal{H} \otimes \mathcal{H}^*)^{\vee\otimes k} = \mathcal{H}^{\vee\otimes k} \otimes \mathcal{H}^{*\vee\otimes k}$. The first identity follows from (15). proof of second identity

□

We will use the k-th moment of Haar random states defined as $\rho_H^{(k)} = \mathbb{E}_{|\psi\rangle\in H}\left[(|\psi\rangle\langle\psi|)^{\otimes k}\right]$ where H stands for the Haar average. The left invariance of the Haar measure implies

$$\left[\rho_H^{(k)}, U^{\otimes k}\right] = 0, \quad \forall U \in U(d) \tag{18}$$

To continue, we use the following theorem [3]:

Theorem 4. *For $A \in \mathcal{L}(\mathcal{H})^{\otimes k}$,*

$$\left[A, U^{\otimes k}\right] = 0, \quad \forall U \in U(d) \quad \Leftrightarrow \quad A = \sum_\pi c_\pi W_\pi \text{ for some choice of } c_\pi \tag{19}$$

Proof. By Theorem 3, the left hand side is equivalent to saying $A \in \mathrm{comm}(\mathcal{L}(\mathcal{H})^{\vee\otimes k})$. By definition, $\mathcal{L}(\mathcal{H})^{\vee\otimes k}$ is the set of operators that commutes with all W_π. Using the bilinearity of the commutator, it follows that $\mathcal{L}(\mathcal{H})^{\vee\otimes k}$ is the commutant of $\mathcal{W} = \{\sum_\pi c_\pi W_\pi : c_\pi \in \mathbb{C}\}$. Then we apply von Neumann's double commutant theorem: the double commutant of the self-adjoint, unital subalgebra $\mathcal{W}$ is itself. Thus $\mathrm{comm}(\mathcal{L}(\mathcal{H})^{\vee\otimes k}) = \mathcal{W}$. □

This fixes $\rho_H^{(k)}$ to be a summation over permutations between replicas. Further imposing $W_\pi \rho_H^{(k)} = \rho_H^{(k)}$ for any permutation, we conclude that the weight for the permutations are equal [4]. Thus $\rho_H^{(k)}$ is proportional to the projection on the replica permutation symmetric subspace:

$$\rho_H^{(k)} = \frac{1}{d_{sym}} P_{sym} = \frac{1}{d_{sym}} \frac{1}{k!} \sum_\pi X_\pi, \qquad d_{sym} = \binom{d+k-1}{k} = \frac{1}{k!}\sum_\pi d^{|\pi|} \tag{20}$$

Acting P_{sym} or any W_π on the symmetric product state $|00\cdots0\rangle$ does nothing, hence the $1/k!$ in the second equality. This formula can be viewed as a resolution of identity in the permutation symmetric subspace.

Take $k = 2$ for example:

$$\rho_H^{(k)} = \frac{1}{d(d+1)}(I + X) \tag{21}$$

where X is the swap operator.

We study the moments of the Haar ensemble with the k-fold channel $\Phi_{\mathcal{E}}^{(k)}(A) = \mathbb{E}_{U\in\mathcal{E}}\left[U^{\otimes k} A U^{\dagger\otimes k}\right]$.[2] Due to the invariance of the Haar measure, we have $[\Phi_H^{(k)}(A), V^{\otimes k}] = 0$ for all $V \in U(d)$. Then by Theorem 4 we have

$$\Phi_{\mathcal{E}}^{(k)}(A) = \sum_\pi W_\pi u_\pi(A) \tag{22}$$

Since $\Phi_H^{(k)}$ is a linear funcion, so should u_π. It can be written as $u_\pi(A) = \mathrm{Tr}(C_\pi A)$ for some operator C_π. We now assume $d \geq k$. In this case the W_πs are linearly independent. Using $\Phi_{\mathcal{E}}^{(k)}(V^{\otimes k} A V^{\dagger\otimes k}) = \Phi_H^{(k)}(A)$, we have $\left[C_\pi, V^{\otimes k}\right] = 0$ for all $V \in U(d)$. Again it can be written as a summation over permutations:

$$\Phi_H^{(k)}(A) = \sum_{\pi,\sigma} \mathcal{W} g(\pi,\sigma) W_\pi \,\mathrm{Tr}(W_\sigma A) \tag{23}$$

[2] The derivation in this paragraph follows [4].

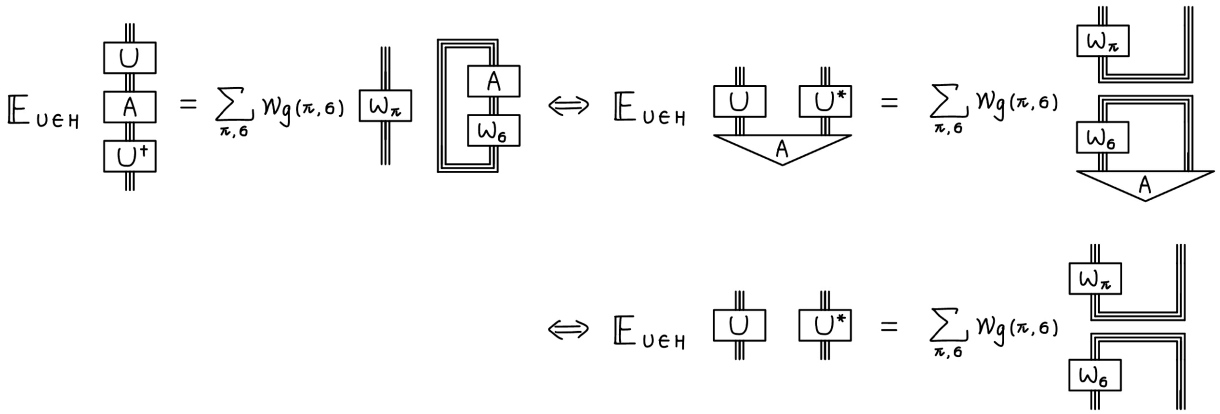


Figure 1: Computing $\widehat{\Phi}_{\mathcal{E}}^{(k)} = \mathbb{E}_{U\in\mathcal{E}}\left[U^{\otimes k}\otimes U^{*\otimes k}\right]$. In the drawing U means $U^{\otimes k}$.

The coefficients $\mathcal{W}g(\pi,\sigma)$ are known as Weingarten functions. Since $\Phi_H^{(k)}(W_\lambda) = W_\lambda$, we have

$$\delta_{\pi,\lambda} = \sum_\sigma \mathcal{W}g(\pi,\sigma)\,\mathrm{Tr}(W_\sigma W_\lambda) = \sum_\sigma \mathcal{W}g(\pi,\sigma) d^{|\sigma\lambda|} \tag{24}$$

where $|\pi|$ is the number of cycles in π. So the Weingarten functions can be evalutated by taking the inverse of $d^{|\sigma\lambda|}$. We can also compute the k-fold operator

$$\widehat{\Phi}_{\mathcal{E}}^{(k)} \equiv \mathbb{E}_{U\in\mathcal{E}}\left[U^{\otimes k}\otimes U^{*\otimes k}\right] = \sum_{\pi,\sigma}\mathcal{W}g(\pi,\sigma) W_\pi\big(\,|\widetilde{\mathcal{I}}\rangle\,\langle\widetilde{\mathcal{I}}|\,\big)^{\otimes k} W_\sigma \tag{25}$$

where $|\widetilde{\mathcal{I}}\rangle = \sum_{i=1}^d |i\rangle\,|i\rangle$ is the unnormalized EPR state. The pictorial representation and the derivation is shown in figure 1.

Example: $k = 2$

$$Q_{\sigma\pi} = \begin{pmatrix} d^2 & d \\ d & d^2 \end{pmatrix} \qquad Q^{-1}_{\sigma\pi} = \frac{1}{d(d^2-1)}\begin{pmatrix} d & -1 \\ -1 & d \end{pmatrix} \tag{26}$$

4 Unitary design

Suppose we want to integrate a funciton $f(x)$ on $x \in [a,b]$. We would like to replace this integral by a set of sample points and weights $\{x_i, p_i\}$:

$$\sum_i p_i f(x_i) = \int_a^b f(x)dx \tag{27}$$

The complexity of $\{x_i, p_i\}$ required might vary drastically between different types of functions. The idea of k-design is similar.

4.1 Exact designs

Definition 3 (state design)**.** *2 equivalent definitions:*

1. *(Polynomial) A balanced polynomial of degree k is a function $f(|\psi\rangle)$ whose constituents are*

$$c_{i_1}\cdots c_{i_t}c^*_{j_1}\cdots c^*_{j_t}, \qquad \max_t = k \tag{28}$$

The ensemble is a k-design if, for all balanced polynomial of degree at most k,

$$\mathbb{E}_{|\psi\rangle\in\mathcal{E}}[f(|\psi\rangle)] = \mathbb{E}_{|\psi\rangle\in H}[f(|\psi\rangle)] \tag{29}$$

2. *(Moments) Define the k-th moment of a state ensemble $\mathcal{E}$ as $\rho^{(k)}_{\mathcal{E}} = \mathbb{E}_{|\psi\rangle}(|\psi\rangle\langle\psi|)^{\otimes k}$. The ensemble is a k-design if*

$$\rho^{(k)}_{\mathcal{E}} = \rho^{(k)}_{H} \tag{30}$$

where H stands for the Haar ensemble.

They are equivalent because $c_{i_1}\cdots c_{i_k}c^*_{j_1}\cdots c^*_{j_k} = \langle i_1,\cdots,i_k|\left(|\psi\rangle\langle\psi|\right)^{\otimes k}|j_1,\cdots,j_k\rangle$.

Definition 4 (unitary design)**.** *There are 3 equivalent definitions:*

1. *(Polynomial) A balanced polynomial of degree k is a function $f(U)$ whose constituents are*

$$U_{i_1j_1}\cdots U_{i_tj_t}U^*_{k_1lj_1}\cdots U^*_{k_tl_t}, \qquad \max_t = k \tag{31}$$

The ensemble is a k-design if, for all balanced polynomial of degree at most k,

$$\mathbb{E}_{U\in\mathcal{E}}[f(U)] = \mathbb{E}_{U\in H}[f(U)] \tag{32}$$

2. *(Operator) Define operators $\widehat{\Phi}^{(k)}_{\mathcal{E}} = \mathbb{E}_{U\in\mathcal{E}}\left[U^{\otimes k}\otimes U^{*\otimes k}\right]$. The ensemble is a k-design if*

$$\widehat{\Phi}^{(k)}_{\mathcal{E}} = \widehat{\Phi}^{(k)}_{H} \tag{33}$$

3. *(Channel) Define the k-fold channel $\Phi^{(k)}_{\mathcal{E}}(O) = \mathbb{E}_{U\in\mathcal{E}}\left[U^{\otimes k}OU^{\dagger\otimes k}\right]$. The ensemble is a k-design if*

$$\Phi^{(k)}_{\mathcal{E}} = \Phi^{(k)}_{H} \tag{34}$$

To see the equivalence, take $k=1$ for example:

$$U_{ij}U^*_{kl} = \langle i,k|\,U\otimes U^*\,|j,l\rangle = \langle i|\left[U(|j\rangle\langle l|)U^\dagger\right]|k\rangle \tag{35}$$

The polynomial definition tells us that a k-design is automatically a $k-1$ design. This reflects that higher design capture finer properties of the ensemble.

Existence see aaronson2006 p4

Pauli operators form a 1-design

The Pauli operators $\{P_i\}_{i=1}^{d^2}$ are tensor products of single-qubit pauli matrices and identities. We start from 1 qubit. Using the anti-commutation relation of pauli operators, we have

$$\frac{1}{4}\sum_{\sigma_i=I,X,Y,Z}\sigma_i\sigma_j\sigma_i = \frac{1}{2}\begin{cases}\sigma_j, & \sigma_j = I\\ 0, & else\end{cases} \tag{36}$$

The Pauli operators forms a complete basis in the operator space, so for every one qubit operator A, we have

$$\frac{1}{4}\sum_{\sigma_i=I,X,Y,Z}\sigma_i A\sigma_i = \frac{1}{2}\operatorname{Tr}(A)I \tag{37}$$

This can be generalized to multi-qubits by taking the partial trace.

$$\frac{1}{d^2}\sum_{P} PAP = \frac{1}{d}\operatorname{Tr}(A)I \tag{38}$$

We conclude that the Pauli group is a 1-design. In fact the Pauli group saturates the lower bound for the cardinality of a 1-design, because they are orthogonal.

This can be easily seen in the bloch ball representation, where the Pauli operators generate points whose center of mass is at the origin.

$$\begin{aligned} \rho &= \frac{1}{2}(I + xX + yY + zZ) \qquad X\rho X = \frac{1}{2}(I + xX - yY - zZ) \\ Y\rho Y &= \frac{1}{2}(I - xX + yY - zZ) \quad Z\rho Z = \frac{1}{2}(I - xX - yY + zZ) \end{aligned} \tag{39}$$

Clifford group is a 3-design

The Clifford group is generated by Hadamard, Phase and CNOT gates[3]:

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \qquad CNOT = \begin{pmatrix} I & 0 \\ 0 & X \end{pmatrix} \tag{40}$$

The Clifford group is the normalizer of the Pauli group—under conjugation, it takes one Pauli operator to another. Pauli operators are automatically included. For example: $Z = S^2$, $X = HS^2H$.

The proof that the Clifford group is a 3-design involves a calculation of the frame potential [5], which we define below. The design property gives many practical merits, since they are efficient: any Clifford gate can be generated by $O(n^2)$ steps.explain

Definition 5 (frame potential of an ensemble of unitaries)**.**

$$F_{\mathcal{E}}^{(k)} = \mathbb{E}_{U,V\in\mathcal{E}}\left[\left|\operatorname{Tr}(U^\dagger V)\right|^{2k}\right] \tag{41}$$

Subtlety for $k > d$ The eigenvalues of a unitary are complex numbers with unit 1, thus $\operatorname{Tr}[U^\dagger V]$ reaches its maximum d when $U = V$. When U are V are distinct, $\operatorname{Tr}[U^\dagger V]$ will quickly become small due to dephasing. The more "dispersed" the unitaries are, the smaller the frame potential is. We will see that the Haar ensemble has the smallest frame potential, for they are maximally dispersed. The Haar value is

$$F_H^{(k)} = \mathbb{E}_{U\in H}\operatorname{Tr}[U^{\otimes k}\otimes U^{*\otimes k}] = \sum_{\pi,\sigma}\mathcal{W}g(\pi,\sigma)\operatorname{Tr}\left[W_\pi\left(|\widetilde{\mathcal{I}}\rangle\langle\widetilde{\mathcal{I}}|\right)^{\otimes k}W_\sigma\right] = \sum_{\pi,\sigma}\mathcal{W}g(\pi,\sigma)d^{|\pi\sigma|} = k! \tag{42}$$

[3]Adding one more gate outside the Clifford group such as $\frac{\pi}{8}$ makes the set universal.

For states there's no problem for $k > d$ but for unitary there isIn the first line we used the invariance of the Haar measure. The last equality is because $d^{|\pi\sigma|}$ is a $k! \times k!$ matrix. We expect a non-Haar ensemble to have a bigger frame potential. The difference in frame potential is related to the 2-norm distance of the k-fold operators as:

$$\begin{aligned}\|\widehat{\Phi}_{\mathcal{E}}^{(k)} - \widehat{\Phi}_{H}^{(k)}\|_2^2 &= \mathrm{Tr}\left(\widehat{\Phi}_{\mathcal{E}}^{(k)\dagger} - \widehat{\Phi}_{H}^{(k)^\dagger}\right)\left(\widehat{\Phi}_{\mathcal{E}}^{(k)} - \widehat{\Phi}_{H}^{(k)}\right) \\ &= \mathbb{E}_{U,V\in\mathcal{E}}\left[\left|\mathrm{Tr}(U^\dagger V)\right|^{2k}\right] - 2\mathbb{E}_{U\in\mathcal{E},V\in H}\left[\left|\mathrm{Tr}(U^\dagger V)\right|^{2k}\right] + \mathbb{E}_{U,V\in H}\left[\left|\mathrm{Tr}(U^\dagger V)\right|^{2k}\right] \\ &= F_{\mathcal{E}}^{(k)} - 2F_{H}^{(k)} + F_{H}^{(k)} = F_{\mathcal{E}}^{(k)} - F_{H}^{(k)}\end{aligned} \tag{43}$$

In the third line we used the left invariance of the Haar measure. The difference in frame potential also bounds the diamond distance between k-fold channels [6].

Theorem 5.

$$\|\Phi_{\mathcal{E}}^{(k)} - \Phi_{H}^{(k)}\|_\diamond^2 \leq d^{2k}\left[F_{\mathcal{E}}^{(k)} - F_{H}^{(k)}\right] \tag{44}$$

Proof. First, we establish a relation between the channel distance and the operator distance:

$$\|\Phi_{\mathcal{E}}^{(k)} - \Phi_{H}^{(k)}\|_{2\to 2} = \|\widehat{\Phi}_{\mathcal{E}}^{(k)} - \widehat{\Phi}_{H}^{(k)}\|_\infty \leq \|\widehat{\Phi}_{\mathcal{E}}^{(k)} - \widehat{\Phi}_{H}^{(k)}\|_2 \tag{45}$$

To prove the first equality, we map operators to states using the Choi–Jamiolkowski isomorphism.

$$A = \sum_{ij} A_{ij}\left|i\right\rangle\left\langle j\right| \mapsto \left|A\right\rangle = \sum_{ij} A_{ij}\left|i\right\rangle\left|j\right\rangle \tag{46}$$

The operator 2-norm is mapped to the usual norm in Hilbert space: $\|A\|_2 = \||A\rangle\|$. Define $\widehat{T} = \widehat{\Phi}_{\mathcal{E}}^{(k)} - \widehat{\Phi}_{H}^{(k)}$

$$\left|T(\left|i\right\rangle\left\langle j\right|)\right\rangle = V_{ki}V_{lj}^*\left|k\right\rangle\left|l\right\rangle = V\otimes V^*\left|i\right\rangle\left|j\right\rangle \quad \Rightarrow \quad \left|T(A)\right\rangle = \widehat{T}\left|A\right\rangle \tag{47}$$

(45) directly follows from the definition of the channel 2-norm:

$$\|T\|_{2\to 2} = \max_{A\neq 0}\frac{\|T(A)\|_2}{\|A\|_2} = \max_{|A\rangle\neq 0}\frac{\|\widehat{T}\left|A\right\rangle\|}{\||A\rangle\|} = \|\widehat{T}\|_\infty \tag{48}$$

Then, we notice that the channel 2-norm bounds the diamond norm (see (160)): $\|\Phi_{\mathcal{E}}^{(k)} - \Phi_{H}^{(k)}\|_\diamond \leq d^k\|\Phi_{\mathcal{E}}^{(k)} - \Phi_{H}^{(k)}\|_{2\to 2}$ □

It follows immediately that

$$F_{\mathcal{E}}^{(k)} \geq F_{H}^{(k)} \tag{49}$$

and $\mathcal{E}$ is a k-design if and only if $F_{\mathcal{E}}^{(k)} = F_{H}^{(k)}$.

4.2 Approximate designs

Definition 6 (approximate state design)**.**

$$\Delta_p^{(k)} \equiv \frac{\|\rho^{(k)} - \rho_H^{(k)}\|_p}{\|\rho_H^{(k)}\|_p} \leq \epsilon \qquad \|\rho_H^{(k)}\|_p = \frac{1}{d_{sym}^{1-\frac{1}{p}}} = \begin{cases} 1, & p = 1 \\ \frac{1}{d}, & p = \infty \end{cases} \tag{50}$$

Theorem 6 ([7])**.** *An approximate state k-design is automatically an approximate state k'-design for any $k' \leq k$. In other words $\Delta_p^{(k)} \leq \Delta_p^{(k+1)}$, $\forall p > 1$*

Proof. The proof is based on the fact that tracing out a replica in $\rho^{(k)}$ gives $\rho^{(k-1)}$. Let $|i_k\rangle$ denote the eigenvectors of $\rho^{(k)} - \rho_H^{(k)}$. See fig(2) for the proof.

For p=1 we can give a simpler proof using the Holevo-Helstrom theorem. There exist a measuremen M such that $\frac{1}{2}\Delta_1^{(k)} = \mathrm{Tr}[M(\rho^{(k)} - \rho_H^{(k)})]$. So

$$\frac{1}{2}\Delta_1^{(k)} = \mathrm{Tr}[M(\rho^{(k)} - \rho_H^{(k)})] = \mathrm{Tr}[(M \otimes I)(\rho^{(k+1)} - \rho_H^{(k+1)})] \leq \frac{1}{2}\Delta_1^{(k+1)} \tag{51}$$

□

Among all norms, the 2-norm is the most technically computable:

$$\Delta_2^{(k)} = \left(\frac{F_{\mathcal{E}}^{(k)}}{F_H^{(k)}} - 1\right)^{\frac{1}{2}}, \qquad F_{\mathcal{E}}^{(k)} = \mathrm{Tr}[(\rho^{(k)})^2] = \frac{1}{d_{sym}} \tag{52}$$

It immediately follows that $F_{\mathcal{E}}^{(k)} \geq F_H^{(k)}$.

Definition 7 (approximate unitary design)**.** *The ensemble is an ϵ-approximate k-design if*

1. *(Polynomial) For all balanced polynomial of degree at most k,*

$$\left|\mathbb{E}_{U\in\mathcal{E}}[f(U)] - \mathbb{E}_{U\in H}[f(U)]\right| \leq \epsilon \tag{53}$$

2. *(Operator)*

$$\|\widehat{\Phi}_{\mathcal{E}}^{(k)} - \widehat{\Phi}_H^{(k)}\|_1 \leq \epsilon \qquad \widehat{\Phi}_{\mathcal{E}}^{(k)} \equiv \mathbb{E}_{U\in\mathcal{E}}\left[U^{\otimes k} \otimes U^{*\otimes k}\right] \tag{54}$$

3. *(Channel)*

$$\|\Phi_{\mathcal{E}}^{(k)} - \Phi_H^{(k)}\|_\diamond \leq \epsilon \qquad \Phi_{\mathcal{E}}^{(k)}(O) \equiv \mathbb{E}_{U\in\mathcal{E}}\left[U^{\otimes k} O U^{\dagger\otimes k}\right] \tag{55}$$

These 3 definitions are equivalent up to $\epsilon \to \mathrm{poly}(d^k)\epsilon$ [8].

Theorem 7. *Under the (Polynomial) and (Channel) definitions, an ϵ-approximate k-design is automatically an ϵ-approximate k'-design for any $k' \leq k$.*

Proof. 1. (Polynomial) Straightforward from definition.

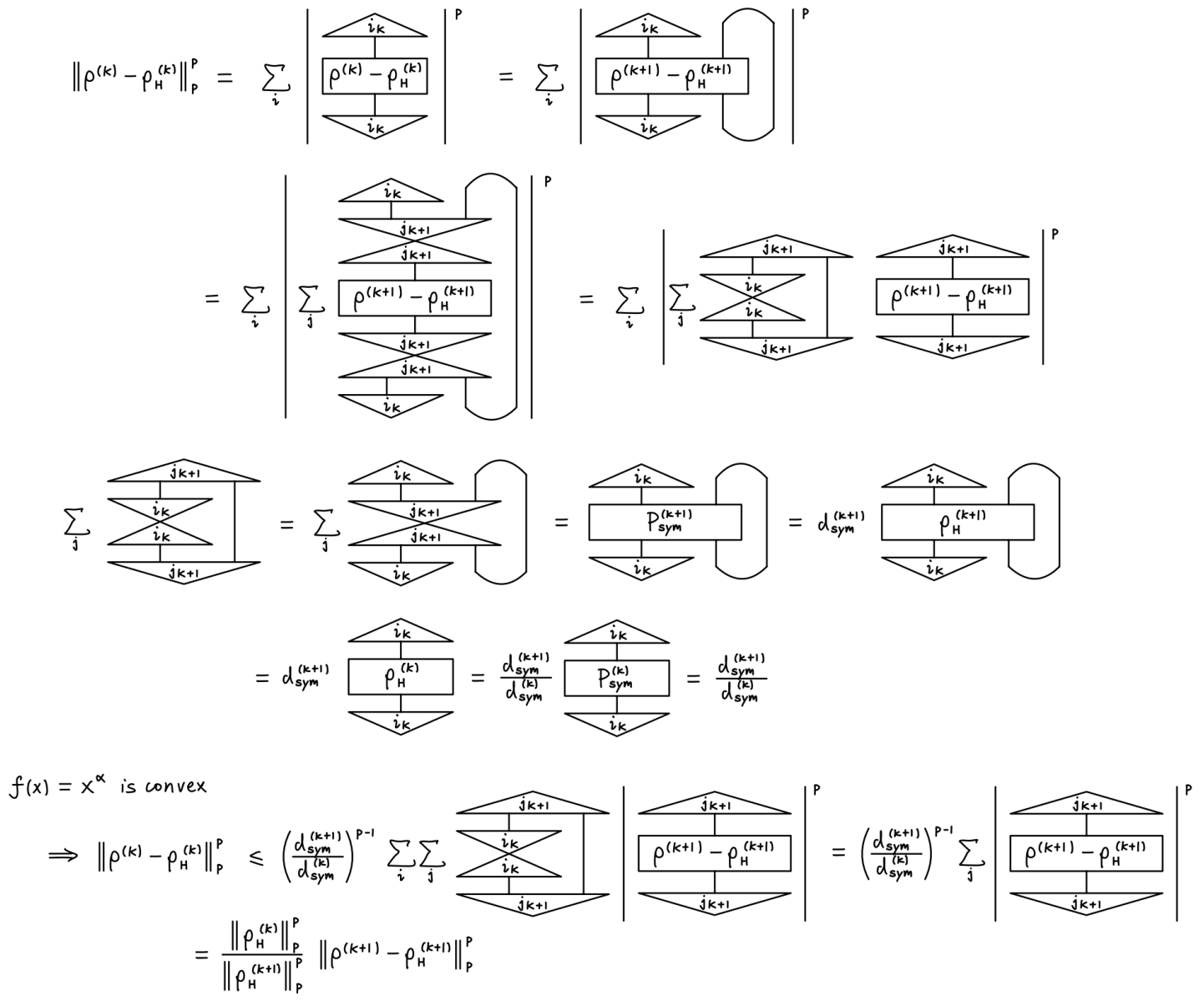


Figure 2: Proof of Theorem 6

2. (Channel) Define $\Delta^{(k)} = \Phi_{\mathcal{E}}^{(k)} - \Phi_H^{(k)}$ The operational meaning of the diamond norm (Theorem 14) tells us that there exist some measurement M and state ρ such that the distinguishability is saturated:

$$\begin{aligned}\frac{1}{2}\|\Delta^{(k)}\|_\diamond &= \mathrm{Tr}\left[M(\Delta^{(k)} \otimes I^{\otimes(k+1)})(\rho)\right] = \mathrm{Tr}\left[(I\otimes M)(\Delta^{(k+1)}\otimes I^{\otimes(k+1)})\left(\frac{I}{d}\otimes\rho\right)\right] \\ &\leq \frac{1}{2}\|\Delta^{(k+1)}\|_\diamond\end{aligned} \tag{56}$$

□

Unitary designs generate state designs, but the converse is not true

By Theorem 13

$$\left\|\mathbb{E}_{U\in\mathcal{E}}\left[\left(U|0\rangle\langle 0|U^\dagger\right)^{\otimes k}\right] - \mathbb{E}_{U\in H}\left[\left(U|0\rangle\langle 0|U^\dagger\right)^{\otimes k}\right]\right\|_1 \leq \|\Phi_{\mathcal{E}}^{(k)} - \Phi_H^{(k)}\|_\diamond \tag{57}$$

Thus the orbit of a unitary design is a state design.[4]

The converse is not true in general. Given a state k-design, one cannot obtain a unitary k-design by choosing a reference state, and gathering the unitaries whose orbit forms our state ensemble. For example $\{|0\rangle, |1\rangle\}$ each with probability $\frac{1}{2}$ forms a state 1-design. If we choose $|0\rangle$ as the reference, then $\{I, X\}$ is enough to generate the ensemble. But applying $\{I, X\}$ on $|+\rangle$ only result in $|+\rangle$. This is an example that the symmetry of the orbit of an ensemble can differ with the reference state. While the Haar ensemble does not discriminate different reference states, a general ensemble does.

5 Properties of a k-design ensemble

For simplicity of exposition, we scale the allowed error with a factor $\epsilon \to \epsilon\frac{k!}{d^{2k}}$ following [9].

5.1 Size and weights

The intuition is that to form a k-design, the probability distribution in the ensemble should not be too spicky and the cardinality should be large. We start from discrete ensembles.[5]

Theorem 8 (Lemma 5 of [9]). *Let $\mathcal{E} = \{p_i, U_i\}$ be an ϵ approximate k-design, then*

$$p_i \leq (1+\epsilon)\frac{k!}{d^{2k}}, \qquad |\mathcal{E}| \geq \frac{1}{1+\epsilon}\frac{d^{2k}}{k!} \tag{58}$$

[4]The state ensemble generated by acting unitaries on a reference state is called the orbit of the unitary ensemble.

[5]To talk about continuous ensembles, we should do coarse graining by dividing the state/operator space into ϵ balls.

Proof. Fix $U \in \mathcal{E}$.

$$\sum_i p_i \left|\mathrm{Tr}(U^\dagger V_i)\right|^{2k} = \mathbb{E}_{V\in\mathcal{E}}\left[\left|\mathrm{Tr}(U^\dagger V)\right|^{2k}\right] = k! + \underbrace{\mathbb{E}_{V\in\mathcal{E}}\left[\left|\mathrm{Tr}(U^\dagger V)\right|^{2k}\right] - \mathbb{E}_{V\in H}\left[\left|\mathrm{Tr}(U^\dagger V)\right|^{2k}\right]}_{\Delta} \tag{59}$$

where we have used $\mathbb{E}_{V\in H}\left[\left|\mathrm{Tr}(U^\dagger V)\right|^{2k}\right] = \mathbb{E}_{V\in H}\left[\left|\mathrm{Tr}\, V\right|^{2k}\right] = k!$. Δ is bounded by ϵ. Set $|\mathcal{I}\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^d |i\rangle\,|i\rangle$.

$$\begin{aligned}\Delta &= d^{2k}\langle\mathcal{I}|^{\otimes k}\left\{\mathbb{E}_{V\in\mathcal{E}}\left[\left((U^\dagger V\otimes I)|\mathcal{I}\rangle\langle\mathcal{I}|(V^\dagger U\otimes I)\right)^{\otimes k}\right] - \mathbb{E}_{V\in H}\left[\left((U^\dagger V\otimes I)|\mathcal{I}\rangle\langle\mathcal{I}|(V^\dagger U\otimes I)\right)^{\otimes k}\right]\right\}|\mathcal{I}\rangle^{\otimes k}\\ &\le d^{2k}\|(\Phi_{\mathcal{E}}^{(k)}\otimes I - \Phi_H^{(k)}\otimes I)(|\mathcal{I}\rangle\langle\mathcal{I}|)^{\otimes k}\|_\infty \le d^{2k}\|(\Phi_{\mathcal{E}}^{(k)}\otimes I - \Phi_H^{(k)}\otimes I)(|\mathcal{I}\rangle\langle\mathcal{I}|)^{\otimes k}\|_1\\ &\le d^{2k}\|\Phi_{\mathcal{E}}^{(k)} - \Phi_H^{(k)}\|_\diamond \le \epsilon k!\end{aligned} \tag{60}$$

In the second line U means $U\otimes I$. Hence $\sum_i p_i\left|\mathrm{Tr}(U^\dagger V_i)\right|^{2k} \le (1+\epsilon)k!$. $\sum_i p_i\left|\mathrm{Tr}(U^\dagger V_i)\right|^{2k}$ is lower bounded by the $U = V_i$ contributions[6]

$$\sum_i p_i\left|\mathrm{Tr}(U^\dagger V_i)\right|^{2k} \ge p_i\left|\mathrm{Tr}(V_i^\dagger V_i)\right|^{2k} = p_i d^{2k} \tag{61}$$

We now have $p_i \le (1+\epsilon)\frac{k!}{d^{2k}}$ and $|\mathcal{E}| \ge \frac{1}{1+\epsilon}\frac{d^{2k}}{k!}$. □

The important thing to remember is that the cardinality should be exponential in k.

Theorem 9 (Lemma 6 of [9])**.** *Let $\mathcal{E} = \{p_i, |\psi_i\rangle\}$ be an ϵ approximate state k-design, then*

$$p_i \le (1+\epsilon)\frac{1}{d_{sym}}, \qquad |\mathcal{E}| \ge \frac{1}{1+\epsilon}d_{sym} \tag{62}$$

When d is much greater than k, $d_{sym} \sim \frac{d^k}{k!}$. Replacing d with d^2 gives the results for approximate unitary designs.

Straightforward that for 1-design $|\mathcal{E}| \ge d$

5.2 Fooling power

How difficult is it to distinguish a state/unitary design from the maximally entangled state/completely depolarizing channel? At first sight they are drastically different. But here is an extreme example: we are only allowed to measure few-qubit observables, while we have a random pure state in a very large Hilbert space. Page tells us that the reduced density matrix of the qubits that we measure is very close to the maximally mixed state. This will fool us to think that the global state is also maximally mixed. Similarly, it is hard to distinguish between random unitaries and the maximally depolarizing channel by looking at a few qubits. When random unitaries are acted on a large number of qubits, a small subsystem may be fooled to think that it has undergone a maximally depolarizing channel.

[6] We expect this to be a fairly tight bound, because as we noted before, dephasing will make $\left|\mathrm{Tr}(U^\dagger V_i)\right|$ very small as U and V_i differ by a small amount

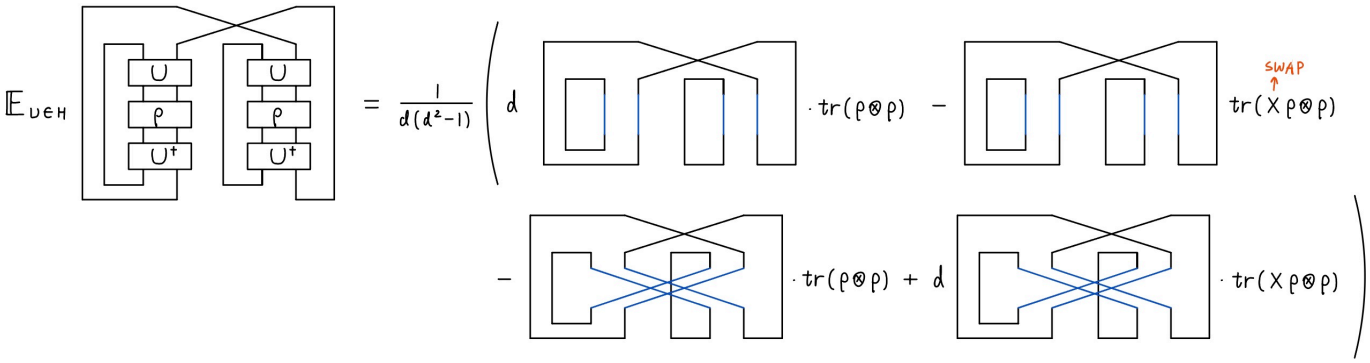


Figure 3: $\operatorname{Tr}\rho_{AB}^2$

To be more quantitative, we divide our system into A and B. After applying a random unitary on the whole system, it is a simple exercise to show that

$$\begin{aligned}
\rho_A(U) &\equiv \operatorname{Tr}_B[U\rho_{AB}U^\dagger] \\
\mathbb{E}_{U\in H}\left\|\rho_A(U)-\frac{I_A}{d_A}\right\|_2^2 &= \operatorname{Tr}[\rho_A(U)^2]-\frac{1}{d_A} = \frac{1}{d(d^2-1)}\left[d_B^2 d_A(d-\operatorname{Tr}(\rho_{AB}^2))+d_B d_A^2(d\operatorname{Tr}\rho_{AB}^2-1)\right]-\frac{1}{d_A} \\
&\le \frac{d_B+d_A}{d+1}-\frac{1}{d_A}\le\frac{1}{d_B} \\
\Rightarrow\ \mathbb{E}_{U\in H}\left\|\rho_A(U)-\frac{I_A}{d_A}\right\|_1 &\le \sqrt{\frac{d_A}{d_B}}
\end{aligned} \tag{63}$$

The calculation of $\operatorname{Tr}\rho_{AB}^2$ is shown in figure 3. In the second line we used $\operatorname{Tr}\rho_{AB}^2\le 1$. Notice that the random unitaries can be replaced by a unitary 2-design.

We see that measurements restricted to A can be fooled. Since random unitaries does not care about spacial locality, we naturally suspect that the fooling power is robust when we allow "simple" measurements that may not be restricted to a few qubits. In addition, we ask what happens when we have designs instead of Haar random states. We will make this more rigorous below.

"easy" measurements

First a brief review of POVM. When doing a measurement, we sometimes donnot care about the post-measurement state and only care about the probabilities assigned by the measurement. Suppose we are measuring an observable on our system. The probabilities are

$$p_i = \operatorname{Tr}(P_i\rho), \qquad P_i = \text{projectors to the eigenspaces of the observable} \tag{64}$$

Now let's consider a generalized scenario where we measure observables that has support on not only our system, but also an ancilla. The state is set to be a tensor product: $\rho_{SA}=\rho_S\otimes\rho_A$. Then the probabilities can still be cast in the form of (64):

$$p_i = \operatorname{Tr}(E_i\rho_S), \qquad E_i = \operatorname{Tr}_A(P_i\rho_A) \tag{65}$$

This motivates the definition of POVM.

Definition 8. *A POVM is an assignment of probabilities to the density matrix: $p_i = \mathrm{Tr}(E_i \rho_S)$. The operators E_i satisfy*

$$E_i > 0, \qquad \sum_i E_i = I \tag{66}$$

The physical implementability of POVM is gauranteed by the Naimark's theorem:

Theorem 10 (Naimark)**.** *Any POVM can be physically implemented by introducing an ancilla, and then doing a projective measurement on the system together with the ancilla. Suppose there are m POVM operators $\{E_1, \cdots, E_m\}$ with rank $l_1, \cdots, l_m$. After introducing an m-dimensional ancilla initiallized in some state $|0\rangle_A$, we can find a set of projection operators $P_1, \cdots, P_m$ such that*

$$E_i = \langle 0|_A P_i |0\rangle_A \tag{67}$$

So the P_i projective measurements gives the desired POVM results.

Proof. The proof is mainly based on [10]. Suppose E_i is a rank l_i operator. We diagonalize it into a sum of projectors:

$$E_i = \sum_{n=1}^{l_i} |\phi_{n,i}\rangle \langle \phi_{n,i}|, \qquad |\phi_{n,i}\rangle \text{ is unnormalized.} \tag{68}$$

$$\widetilde{d} \equiv \sum_{i=1}^{m} \mathrm{rank}(E_i) \geq \mathrm{rank}\left(\sum_{i=1}^{m} E_i\right) = d, \qquad \widetilde{d} \leq m \cdot d \tag{69}$$

Let α denote n, i. There are $\widetilde{d}$ different $|\phi_\alpha\rangle$ in total. They are not gauranteed to be orthogonal.

Now we claim that we are able to "extend" them into the $m \cdot d$ dimensional space such that they become orthonormal. That is, we can find an othonormal set $\{|\widetilde{\phi_\alpha}\rangle\}$ such that $\langle 0_A| \widetilde{\phi_\alpha}\rangle = |\phi_\alpha\rangle$. This is done in the following way. Write down a $d \times \widetilde{d}$ matrix whose columns are all the $|\phi_\alpha\rangle$s.

$$|\phi_\alpha\rangle = \begin{pmatrix} \phi_{1\alpha} \\ \vdots \\ \phi_{d\alpha} \end{pmatrix} \qquad E = \begin{pmatrix} \phi_{11} & \cdots & \phi_{1\widetilde{d}} \\ \vdots & \ddots & \vdots \\ \phi_{d1} & \cdots & \phi_{d\widetilde{d}} \end{pmatrix} \tag{70}$$

The condition $\sum_i E_i = I$ translates into $\sum_\alpha \phi_{\beta\alpha}\phi^*_{\gamma\alpha} = \delta_{\beta\gamma}$. This is the statement that the rows of the matrix

$$\begin{pmatrix} \phi_{\beta 1} & \cdots & \phi_{\beta\widetilde{d}} \end{pmatrix}, \qquad \beta = 1, \cdots, d \tag{71}$$

are orthonormal, when they are viewed as vectors in a $\mathbb{C}^{\widetilde{d}}$. Hence we can add more rows to get a complete orthonormal basis for $\mathbb{C}^{\widetilde{d}}$.

$$\widetilde{E} = \begin{pmatrix} \phi_{11} & \cdots & \phi_{1\widetilde{d}} \\ \vdots & \ddots & \vdots \\ \phi_{d1} & \cdots & \phi_{d\widetilde{d}} \\ \Phi_{d+1,1} & \cdots & \Phi_{d+1,\widetilde{d}} \\ \vdots & \ddots & \vdots \\ \Phi_{\widetilde{d}1} & \cdots & \Phi_{\widetilde{d}\widetilde{d}} \end{pmatrix}, \qquad |\widetilde{\phi}_\alpha\rangle = \begin{pmatrix} \phi_{1\alpha} \\ \vdots \\ \phi_{d\alpha} \\ \Phi_{d+1,\alpha} \\ \cdots \\ \Phi_{\widetilde{d}\alpha} \end{pmatrix} \tag{72}$$

$\widetilde{E}$ is a square matrix whose rows are orthonormal. So its columns $|\widetilde{\phi}_\alpha\rangle$ are also orthonormal. Now we extend these vectors to the $m \cdot d$ dimensional space $\mathcal{H}_S \otimes \mathcal{H}_A$ by setting

$$|\widetilde{\phi}_\alpha\rangle = \begin{pmatrix} \phi_{1\alpha} \\ \vdots \\ \phi_{d\alpha} \\ \Phi_{d+1,\alpha} \\ \vdots \\ \Phi_{\widetilde{d}\alpha} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{73}$$

They are still orthonormal. We arrange the basis of $\mathcal{H}_S \otimes \mathcal{H}_A$ in such a way that the first d rows of the vectors correspond to the basis $\{|\alpha\rangle_S |0\rangle_A\}_{\alpha=1}^d$. Hence

$$\langle 0_A| \widetilde{\phi}_\alpha\rangle = |\phi_\alpha\rangle, \qquad E_i = \sum_{n=1}^{l_i} \langle 0_A|\widetilde{\phi}_{n,i}\rangle\langle\widetilde{\phi}_{n,i}|0_A\rangle = \langle 0|_A P_i |0\rangle_A \tag{74}$$

where P_i is a rank l_i projection operator acting on $\mathcal{H}_S \otimes \mathcal{H}_A$. □

We define the hardness of a POVM operator M by looking at its corresponding projection operator in the larger space with the ancilla [9]

$$M = \langle 0|_A P_l |0\rangle_A \tag{75}$$

where P_l is a rank l projection. If the projection is on the computational basis $|b\rangle\langle b|$ (it does not need to have support on all the qubits), we say that M has length 0. A general P_l can be obtained by rotating the basis with some unitary: $P_l = V|b\rangle\langle b|V^\dagger$. We define the set of measurements $\mathcal{M}_r$ as those whose V can be realized by r implemenations of gates from a two qubit gate set G. Let $|G_r| \leq (n^2|G|)^r$ be the number of gates we can achieve with r implementations, then

$$|\mathcal{M}_r| \leq |P|(n^2|G|)^r \tag{76}$$

where $|P|$ is the number of $|b\rangle\langle b|$ that we could choose.

Now we are in the place to discuss the fooling power of k-designs. Roughly speaking, they have strong fooling power because the likelihood for every single measurement to not be fooled is tiny, and that the number of possible short measurements is small.

First, we bound the ability of a single aubitruary POVM to distinguish between the states in our ensemble and the maximally mixed state. Given a pure state $|\psi\rangle$, define the distinguishing power of a measurement $0 \leq M \leq I$

$$D(M, |\psi\rangle) = \mathrm{Tr}\left[M\left(|\psi\rangle\langle\psi| - \frac{I}{d}\right)\right] = \mathrm{Tr}(\overline{M}|\psi\rangle\langle\psi|), \qquad \overline{M} = M - \frac{\mathrm{Tr}\, M}{d} I \tag{77}$$

Given an ensemble of states $\psi \in \mathcal{E}$, Markov's inequality says the probability of D reaching some value is bounded by its expectation:

$$\Pr_{\mathcal{E}}[D(M, |\psi\rangle) \geq \tau] \leq \tau^{-1}\mathbb{E}_{|\psi\rangle\in\mathcal{E}} D(M, |\psi\rangle) \tag{78}$$

If our ensemble forms an approximate k-design, $D(M, |\psi\rangle)$ should be small for most of the time. Thus we expect $\mathbb{E}_{|\psi\rangle\in\mathcal{E}}\big[D(M, |\psi\rangle)^k\big]$ to decrease drastically with k. Using this, we can get a tight bound:

$$\Pr_{\mathcal{E}}[D(M, |\psi\rangle) \geq \tau] = \Pr_{\mathcal{E}}[D(M, |\psi\rangle)^k \geq \tau^k] \leq \tau^{-k}\mathbb{E}_{|\psi\rangle\in\mathcal{E}}\big[D(M, |\psi\rangle)^k\big] \tag{79}$$

The following lemma gives a bound to $\mathbb{E}_{|\psi\rangle\in\mathcal{E}}\big[D(M, |\psi\rangle)^k\big]$

Lemma 1. *Let the states $|\psi\rangle$ form an approximate k-design in the one norm: $\|\rho^{(k)} - \rho_H^{(k)}\|_1 \leq \frac{k!}{d^{2k}}\epsilon$. Then*

$$\mathbb{E}_{|\psi\rangle\in\mathcal{E}}\big[D(M, |\psi\rangle)^k\big] \leq \binom{d+k-1}{k}^{-1}(d^{k/2} + \epsilon) \leq (1+\epsilon)\left(\frac{k^2}{d}\right)^{k/2} \tag{80}$$

Proof.

$$\mathbb{E}_{|\psi\rangle\in\mathcal{E}}\big[D(M, |\psi\rangle)^k\big] = \mathbb{E}_{|\psi\rangle\in H}\big[D(M, |\psi\rangle)^k\big] + \underbrace{\mathbb{E}_{|\psi\rangle\in\mathcal{E}}\big[D(M, |\psi\rangle)^k\big] - \mathbb{E}_{|\psi\rangle\in H}\big[D(M, |\psi\rangle)^k\big]}_{\Delta} \tag{81}$$

$$\mathbb{E}_{|\psi\rangle\in H}\big[D(M, |\psi\rangle)^k\big] = \frac{1}{d_{sym}}\operatorname{Tr}[P_{sym}\overline{M}^{\otimes k}] \leq \frac{1}{d_{sym}}[\operatorname{Tr}(\overline{M}^2)]^{\frac{k}{2}} = \frac{1}{d_{sym}}\|\overline{M}\|_2^k \leq \frac{d^{k/2}}{d_{sym}}\|M\|_\infty^k \leq \frac{d^{k/2}}{d_{sym}} \tag{82}$$

In the first inequality we used $\operatorname{Tr}\overline{M} = 0$ and $\operatorname{Tr}(\overline{M}^l) \leq [\operatorname{Tr}(\overline{M}^2)]^{\frac{l}{2}}$ for $l \geq 2$ (can be proved using induction).

$$\Delta = \operatorname{Tr}\left[\overline{M}^{\otimes k}(\rho_{\mathcal{E}}^{(k)} - \rho_H^{(k)})\right] \leq \|\overline{M}\|_\infty^k \cdot \|\rho^{(k)} - \rho_H^{(k)}\|_1 \leq \frac{k!}{d^{2k}}\epsilon \leq \frac{\epsilon}{d_{sym}} \tag{83}$$

□

Plugging in (79), we get

$$\Pr_{\mathcal{E}}[D(M, |\psi\rangle) \geq \tau] \leq (1+\epsilon)\left(\frac{k}{\tau\sqrt{d}}\right)^k \tag{84}$$

This decreases exponentially in k. But this is a statement about a single measurement. We would like to allow some freedom to choose a measurement from a given set. We focus on easy measurement that have complexity r, denoted by $\mathcal{M}_r$. The equation above says that the portion in $\mathcal{E}$ that can be distinguished up to τ by a single M is smaller than $(1+\epsilon)\left(\frac{k}{\tau\sqrt{d}}\right)^k$. When we are allowed to choose in $\mathcal{M}$ for each state, the total proportion that can be distinguished up to τ is then

$$\Pr_{\mathcal{E}}\left[\max_{M\in\mathcal{M}_r} D(M, |\psi\rangle) \geq \tau\right] \leq (1+\epsilon)|\mathcal{M}_r|\left(\frac{k}{\tau\sqrt{d}}\right)^k \leq (1+\epsilon)|P|(n^2|G|)^r\left(\frac{k}{\tau\sqrt{d}}\right)^k \tag{85}$$

This is a small number for short r.

5.3 Entropy

5.4 Complexity

Suppose we choose a universal gate set G that consists $|G|$ different two-qubit gates. We wish to prepare our ensemble of unitaries in the following way: assign probabilities to the applicable gates, and then apply them probabilistically for r steps. [4] defined the complexity of an ensemble $C_{\mathcal{E}}$ as the minimum number of steps required to generate the ensemble.

We can lower bound this by a counting argument. Let $|G_r|$ be the number of different unitaries we can create by r applications of gates. Note that it is possible that two different application of gates create the same unitary, although this is rare. Taking these "collisions" into account, we have

$$|G_r| \leq (n^2|G|)^r \tag{86}$$

If $C_{\mathcal{E}}$ steps can generate the ensemble, then $|G_{C_{\mathcal{E}}}|$ must be greater than the cardinality $|\mathcal{E}|$ of our ensemble. Hence

$$C_{\mathcal{E}} \geq \frac{\log|\mathcal{E}|}{\log(n^2|G|)} \tag{87}$$

This is roughly the mean complexity of the ensemble, provided that the complexity does not fluctuate too drastically between members in the ensemble. Plugging in the cardinality bound (62) that says the cardinality should be exponential in k, we get a (almost) linear k dependence:

$$C_{\mathcal{E}} \geq \frac{k(2\log d - \log k) - \log(1+\epsilon)}{\log(n^2|G|)} \tag{88}$$

The error is related to the difference in frame potential as:

$$F_{\mathcal{E}}^{(k)} - F_H^{(k)} \leq \Delta_F \quad \Rightarrow \quad \epsilon \leq \frac{d^{3k}}{k^2}\sqrt{\Delta_F} \quad C_{\mathcal{E}} \geq \frac{k((2+3\epsilon)\log d - \log k) - \frac{d^{3k}}{k^2}\sqrt{\Delta_F}}{\log(n^2|G|)} \tag{89}$$

The traditional state complexity is defined as

Definition 9 (State complexity)**.** *A state has complexity at most r if a size r unitary is able to prepare the state starting from $|0\cdots 0\rangle$ up to some error δ. That is:*

$$C_\delta(|\psi\rangle) \leq r \quad \Leftrightarrow \quad \frac{1}{2}\min_{U\in G_r} \| U|0\rangle\langle 0|U^\dagger - |\psi\rangle\langle\psi| \|_1 \leq \delta \tag{90}$$

Another way of obtaining a lower bound is by using Markov's inequality to bound the preparing ability of a single unitary:

$$\begin{aligned}\Pr\left[|\langle\psi|U|0\rangle|^2 \geq 1-\delta^2\right] = \Pr\left[|\langle\psi|U|0\rangle|^{2k} \geq (1-\delta^2)^k\right] &\leq \frac{1}{(1-\delta^2)^k}\mathbb{E}_{|\psi\rangle\in\mathcal{E}}[|\langle\psi|U|0\rangle|^{2k}] \\ &\leq \frac{1+\epsilon}{(1-\delta^2)^k}\frac{1}{d_{sym}}\end{aligned} \tag{91}$$

$$\Pr_{\mathcal{E}}[C_\delta(|\psi\rangle) \leq r] \leq |G_r|\Pr\left[|\langle\psi|U|0\rangle|^2 \geq 1-\delta^2\right] \leq \frac{1+\epsilon}{(1-\delta^2)^k}\frac{1}{d_{sym}}(n^2|G|)^r \leq \frac{1+\epsilon}{(1-\delta^2)^k}\frac{k!}{d^k}(n^2|G|)^r \tag{92}$$

This remains small until

$$r \gtrsim \frac{k(\log d - \log k)}{\log(n^2|G|)} \tag{93}$$

To draw more quantitive conclusions, [9] refined these results by using the notion of "strong complexity". They also gave lower bounds to the number of high comlexity memberes and the distances between them.

5.4.1 Strong complexity

The idea is that the maximal distinguishability between a pure state and the maximally mixed state is a fixed number

$$\frac{1}{2}\left\| |\psi\rangle\langle\psi| - \frac{I}{d} \right\|_1 = 1 - \frac{1}{d} \tag{94}$$

and only measurements that know enough about the state are able to achieve this.

Definition 10 (Strong complexity)**.** *A state has strong comlexity at most r if a size r measurement is able to achieve maximal distinguishablity between the state and the maximally mixed state up to some error δ. That is:*

$$C_{\delta,\mathrm{str}}(|\psi\rangle) \leq r \quad \Leftrightarrow \quad \max_{M\in\mathcal{M}_r} \mathrm{Tr}\left[M\left(|\psi\rangle\langle\psi| - \frac{I}{d}\right)\right] \geq 1 - \frac{1}{d} - \delta \tag{95}$$

It is stronger than circuit complexity because it is easier to distinguish than to prepare.

Lemma 2 (Strong complexity is stronger)**.**

$$C_{\delta,\mathrm{str}}(|\psi\rangle) > r \quad \Rightarrow \quad C_{\delta}(|\psi\rangle) > r \tag{96}$$

Proof. By contraposition. If $C_\delta(|\psi\rangle) \leq r$, then there exist some $U \in G_r$ such that $\frac{1}{2}\max_{U\in\mathcal{U}_r}\|U|0\rangle\langle 0|U^\dagger - |\psi\rangle\langle\psi|\|_1 \leq \sqrt{\delta}$. We can use U to build the size r measurement $M = U|0\rangle\langle 0|U^\dagger$.

$$\mathrm{Tr}\left[M\left(|\psi\rangle\langle\psi| - \frac{I}{d}\right)\right] = |\langle 0|U^\dagger|\psi\rangle|^2 - \frac{1}{d} \geq 1 - \frac{1}{d} - \delta \quad \Rightarrow C_{\delta,\mathrm{str}}(|\psi\rangle) \leq r \tag{97}$$

where we used $\frac{1}{2}\||\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\| = \sqrt{1 - |\langle\psi|\phi\rangle|^2}$. □

The converse is not true because highly complex states can be locally simple. For example, set $|\psi\rangle = |0\rangle \otimes |\phi\rangle$ where $|\phi\rangle$ is very complex. We can simply measure the first qubit $M = |0\rangle\langle 0| \otimes I$ to achieve high distinguishability.

If we have an approximate k-design ensemble, we can use its fooling power (85) to say things about strong complexity:

$$\Pr_{\mathcal{E}}[C_{\delta,\mathrm{str}}(|\psi\rangle) \leq r] = \Pr_{\mathcal{E}}\left[\max_{M\in\mathcal{M}_r} D(M, |\psi\rangle) \geq 1 - \frac{1}{d} - \delta\right] \leq (1+\epsilon)|P|(n^2|G|)^r \left(\frac{k}{(1-d^{-1}-\delta)\sqrt{d}}\right)^k \tag{98}$$

This probability remains small until

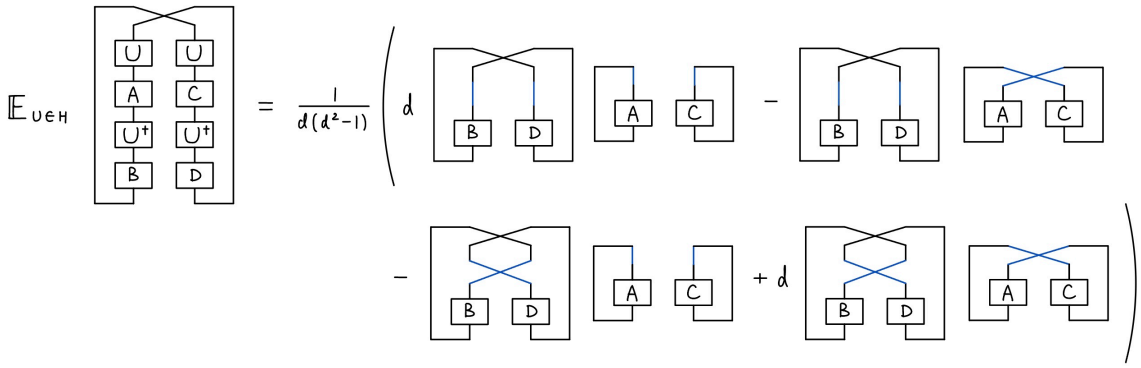
$$r \gtrsim \frac{k(\frac{1}{2}\log d - \log k) - \log|P|}{\log(n^2|G|)} \tag{99}$$

5.5 Scrambling

The folklore is that in chaotic systems, the OTOC will asymptote to a value given by Haar random Unitaries. The Haar value is [4]

$$\begin{aligned}\frac{1}{d}\mathbb{E}_{U\in H}\operatorname{Tr}\left[AB(t)CD(t)\right] =& \frac{1}{d(d^2-1)}\left[\operatorname{Tr}A\operatorname{Tr}C\operatorname{Tr}(BD)+\operatorname{Tr}(AC)\operatorname{Tr}B\operatorname{Tr}D\right]\\ &-\frac{1}{d^2(d^2-1)}\left[\operatorname{Tr}(AC)\operatorname{Tr}(BD)-\operatorname{Tr}A\operatorname{Tr}B\operatorname{Tr}C\operatorname{Tr}D\right]\end{aligned} \tag{100}$$

The diagrams are given below.



For one-qubit pauli operators,

$$\frac{1}{d}\mathbb{E}_{U\in H}\operatorname{Tr}\left[PQ(t)PQ(t)\right]=\begin{cases}-\frac{1}{d^2-1}, & P,Q\neq I\\ 1, & else\end{cases} \tag{101}$$

$$\left\langle\frac{1}{d}\mathbb{E}_{U\in H}\operatorname{Tr}\left[PQ(t)PQ(t)\right]\right\rangle_{P,Q}=\frac{7}{16}-\frac{9}{16}\frac{1}{d^2-1} \tag{102}$$

The averaged OTOC for some ensemble is given by

$$\mathbb{E}_{U\in\mathcal{E}}\operatorname{Tr}\left[AB(t)CD(t)\right]=\operatorname{Tr}\left[X\cdot\Phi^{(k)}_{\mathcal{E}(t)}(A\otimes C)B\otimes D\right] \tag{103}$$

As the 2-fold channel get closer and closer to the 2-fold channel of Haar, the averged OTOC will approach its equilibrium value given by Haar. It is often said that in strongly interacting systems, scrambling happens later than dissipation or the relaxation of two point functions. This is simply the statement that the ensemble of time evolutions appoach a 2-design later than 1-design.

At finite temperature, the out-of-time ordered correlators that people typically consider takes the form of

$$\operatorname{Tr}\left[\rho_\beta^{\frac{1}{4}}A\rho_\beta^{\frac{1}{4}}B(t)\rho_\beta^{\frac{1}{4}}C\rho_\beta^{\frac{1}{4}}D(t)\right]\qquad \rho_\beta=\frac{e^{-\beta H}}{Z} \tag{104}$$

To evaluate the averaged OTOC, one needs a "thermal k-th moment channel"

$$\Phi^{(k)}_{\mathcal{E}(t),\beta}(O)=\int_{\mathcal{E}}dJP(J)U\left(t-i\frac{\beta}{4}\right)^{\otimes k}OU\left(t-i\frac{\beta}{4}\right)^{\dagger\otimes k} \tag{105}$$

$$\overline{\mathrm{Tr}\left[\rho_\beta^{\frac{1}{4}} A \rho_\beta^{\frac{1}{4}} B(t) \rho_\beta^{\frac{1}{4}} C \rho_\beta^{\frac{1}{4}} D(t)\right]}^{\mathcal{E}} = \mathrm{Tr}\left[Swap \cdot A \otimes C \Phi^{(k)\dagger}_{\mathcal{E}\left(t - i\frac{\beta}{4}\right)}(B \otimes D)\right] \tag{106}$$

It is not really a quantum channel because it doen't preserve trace.

6 How to prepare designs

6.1 Time-dependent evolution

Random circuit model

$t = O(n^2 k^{11})$. Linear growth $t = O(n^2 k)$ in the large bond dimension limit (perhaps not tight).

Brownian spins/fermions

Effective Hamiltonian has $k!$ ground states. Gap gives linear growth.

$$F_{\mathcal{E}}^{(k)} = \mathbb{E}_{\mathcal{E}} \mathrm{Tr}[U(2t)^{\otimes k} \otimes U^*(2t)^{\otimes k}] \equiv \mathrm{Tr}[e^{-2tH_{\text{eff.}}}] \tag{107}$$

At long times, the first excited states gives the leading order beahavior:

$$F_{\mathcal{E}}^{(k)} = k! + Ce^{-2\Delta t} \tag{108}$$

$$d^{2k}\left[F_{\mathcal{E}}^{(k)} - F_H^{(k)}\right] \geq \epsilon \quad \Rightarrow \quad t \geq k \log d + \frac{1}{2}\log(C/\epsilon) \tag{109}$$

Nearly time independent

6.2 Time independent Hamiltonians

Given an ensemble of Hamiltonians $\{H\}$, we can form an emsemble of time evolutions $\mathcal{E}(t) = \{e^{-iHt}\}$. These ensembles cannot asymptote to k-designs because their eigenvalue statistics fails to asymptote to Haar random unitaries(CUE) [4]. More explicitly, we evaluate the long-time-averaged frame potential

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T F_{\mathcal{E}(t)}^{(k)} = \lim_{T\to\infty}\frac{1}{T}\int_0^T \left|\sum_i e^{-iE_i T}\right|^{2k} = \lim_{T\to\infty}\frac{1}{T}\int_0^T \sum_{i_1\cdots i_k}\sum_{j_1\cdots j_k}\exp\left[-iT\sum_{m=1}^k E_{i_m} + iT\sum_{m=1}^k E_{j_m}\right] \tag{110}$$

The terms that survive are those with $-\sum_{m=1}^k E_{i_m} + \sum_{m=1}^k E_{j_m} = 0$. Assuming all energy levels are incommensurate, this only happens when the is and js are paired up ($i_m = j_n$). There are $k!$ ways of paring $m = 1, \cdots, k$ and $n = 1, \cdots, k$. For each pair, $i_m = j_n$ can take d values. We have overcounted a bit. When different paris take the same value, different paring schemes can produce the same results. But this is negligible for $d \gg k$.

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T F_{\mathcal{E}(t)}^{(k)} \sim k! d^k \tag{111}$$

which is far greater than $F_H^{(k)} = k!$. For general $d > k$, it is still true that $\lim_{T\to\infty} \frac{1}{T}\int_0^T F_{\mathcal{E}(t)}^{(k)}$ is a lot greater than $k!$. We conclude that generic or chaotic time-independent evolution deos not converge to k-designs.

However, similar to spectral form factos, there is a "dip" at some intermediate time before the "plateau". The spectral form factor is

$$\mathbb{E}_H \left[Z(\beta,t)Z^*(\beta,t)\right] = \mathbb{E}_H \left[\operatorname{Tr} e^{-\beta H - iHt} \operatorname{Tr} e^{-\beta H + iHt}\right] \tag{112}$$

At infinite temperature, it computes the second frame potential of the corresponding unitary ensemble. This suggest that the behavior of frame potentials are somewhat similar. Indeed, the frame potential can get very close to $k!$ before rising up again [11].explain

SYK is different from RMT in that it is few-body (or k-local) randomness.

6.3 Projected ensemble

Divide the qubits in to subsystem A and system B, with N_A and N_B qubits. Given a single quantum state $|\psi\rangle$, we measure the computational basis on B. The probability for obtaining some result $|z_B\rangle$ and the resulting state on A is given by

$$\begin{aligned} p(z_B) &= \langle\psi|z_B\rangle\langle z_B|\psi\rangle \\ |\psi_A(z_B)\rangle &= \frac{1}{\sqrt{p(z_B)}}\langle z_B|\psi\rangle \end{aligned} \tag{113}$$

The ensemble $\mathcal{E}(\psi) = \{p(z_B), |\psi_A(z_B)\rangle\}$ is called the projected ensemble.

The nice thing about these ensembles is that the wave function $|\psi\rangle$ doesn't have to be the result of a time-dependent evolution [12]. Before getting to that, let's look at generic $|\psi\rangle$ first. Namely, we study k-th moment of the projected ensemble for some $|\psi\rangle$, average over random $|\psi\rangle$s, and bound the deviations from the averge. As usual, the k-th moment of the ensemble is defined as $\rho_{\mathcal{E}(|\psi\rangle)}^{(k)} = \sum_z p(z) \left(|\psi_A\rangle \langle\psi_A|\right)^{\otimes k}$

Lemma 3 (Averaged k-th moment)**.**

$$\mathbb{E}_{|\psi\rangle\in H} \rho_{\mathcal{E}(|\psi\rangle)}^{(k)} = \rho_H^{(k)} \tag{114}$$

Proof. We first prove that for a single z, the Haar average factorizes into

$$\mathbb{E}_{|\psi\rangle\in H}\left[p(z)\left(|\psi_A\rangle\langle\psi_A|\right)^{\otimes k}\right] = \mathbb{E}_{|\psi\rangle\in H}[p(z)] \cdot \mathbb{E}_{|\psi\rangle\in H}\left[\left(|\psi_A(z)\rangle\langle\psi_A(z)|\right)^{\otimes k}\right] \tag{115}$$

To prove this, we use the fact that $p(z)$ is invariant under unitaries on A. That is,

$$p(z,|\psi\rangle) = \langle\psi|z_B\rangle\langle z_B|\psi\rangle = \langle\psi|U_A^\dagger|z_B\rangle\langle z_B|U_A|\psi\rangle = p(z, U_A|\psi\rangle) \tag{116}$$

Where we added the second label in $p(z,|\psi\rangle)$ to label the state before projection. Using the invariance of Haar measure, we have

$$\begin{aligned} &\mathbb{E}_{|\psi\rangle\in H}\left[p(z,|\psi\rangle)\left(|\psi_A\rangle\langle\psi_A|\right)^{\otimes k}\right] = \mathbb{E}_{|\psi\rangle\in H}\mathbb{E}_{U_A\in H}\left[[p(z,U_A|\psi\rangle)]\left(U_A|\psi_A\rangle\langle\psi_A|U_A^\dagger\right)^{\otimes k}\right] \\ &= \mathbb{E}_{|\psi\rangle\in H}\left[p(z,|\psi\rangle)\mathbb{E}_{U_A\in H}\left(U_A|\psi_A\rangle\langle\psi_A|U_A^\dagger\right)^{\otimes k}\right] \\ &= \mathbb{E}_{|\psi\rangle\in H}[p(z,|\psi\rangle)] \cdot \rho_{H,A}^{(k)} = \operatorname{Tr}\left(\frac{I_A}{d_{AB}}\otimes|z\rangle\langle z|\right)\cdot\rho_{H,A}^{(k)} = \frac{1}{d_B}\rho_{H,A}^{(k)} \end{aligned} \tag{117}$$

□

We can also bound the deviation from the average using concentration of measure. Thus, roughly speaking, projected ensemble formed from generic states are pretty close to designs. To get generic states, just run some chaotic evolution that does not need to break energy conservation.

Another interesting thing to notice is that $|\psi_A(z)\rangle$ is exacly Haar distributed for fixed z. This follows from our definition of random state. Suppose

$$|\psi\rangle = \begin{pmatrix} c_1 & \cdots & c_{d_A} & c_{d_A+1} & \cdots & c_{d_A d_B} \end{pmatrix}^T \tag{118}$$

is a random state on AB, then the entries are (independently) Gaussian distributed before normalization. The first d_A entries are coefficients before $|i_A\rangle |z\rangle$. Projecting on $|z\rangle$ gives

$$\langle z|\psi\rangle = \begin{pmatrix} c_1 & \cdots & c_{d_A} \end{pmatrix}^T \tag{119}$$

before normalization. Now $c_1, \cdots, c_{d_A}$ are still Gaussian variables, thus defines random states on d_A after normalization.

7 Experimental applications

7.1 Quantum state tomography

Rank-one informationally complete POVM

We would like to do state tomography with a rank-1 POVM $\{E_i\}$. We write the operators as projectors:

$$E_i = \tau_i P_i = \tau_i |\phi_i\rangle \langle\phi_i|, \qquad w_i \equiv \tau_i/d, \qquad \sum_i w_i = 1 \tag{120}$$

Suppose $|\phi_i\rangle$ is the post measurement state corresponding to the outcome i.[7] Then we classically process the experimental data by superposing the post-measurement states weighted by the observed probablities

$$\mathcal{M}(\rho) = \sum_i \mathrm{Tr}(E_i\rho) |\phi_i\rangle \langle\phi_i| = d \sum_i w_i \mathrm{Tr}(|\phi_i\rangle \langle\phi_i| \rho) |\phi_i\rangle \langle\phi_i| \tag{121}$$

Note that w_i forms a probablity distribution. If our POVM is so nice such that $\{w_i, |\phi_i\rangle\}$ forms a 2-design, then using $\sum_i w_i (|\phi_i\rangle \langle\phi_i|^{\otimes 2} = \frac{1}{d(d+1)}(I + X)$, we have

$$\mathcal{M}(\rho) = \frac{1}{d+1}(\rho + I) \tag{122}$$

We have succeded in recovering the density matrix with experimental data only. The reconstuction fumula is linear and minimize some sort of mean-squared error [13].

Single observable tomography

[7] This can always be achieved from the construction in our proof of Naimark's theorem.

Shadow tomography

We do randomized measurements by implementing a unitary that is chosen from some ensemble, then measure $Z \otimes \cdots \otimes Z$ [14]. After obtaining a particular result, we act on the inverse of the unitary we just used. As before, we classically average over the outcomes, weighted by the probabilies of choosing the particular unitary and of obtaining the particular outcome. This defines a linear map

$$\mathcal{M}(\rho) = \mathbb{E}_{U \in \mathcal{E}} \sum_b \langle b| U \rho U^\dagger |b\rangle U^\dagger |b\rangle \langle b| U \tag{123}$$

For some choice of $\mathcal{E}$, $\mathcal{M}$ is invertible. In this case we say $\mathcal{E}$ is tomographically complete. The simplest example is when we choose random Clifford unitaries, which forms a three design. It is straightforward to show $\mathcal{M}(\rho) = \frac{1}{d+1}(\rho + I)$.

We can in principle revert $\mathcal{M}$ to reconstruct our density matrix, if we have access to the exact probabilities. However we can only approximate this by increasing the number of measurements we make. The idea of shadow tomography is for some classes of observables and by cleverly choosing an estimation scheme, the number of measurements required to reach some precision can be greatly reduced. This is natural because we don't need full knowledge of ρ to compute the expectation of some observables.

7.2 Randomized benchmarking of Clifford gates

We wish to measure how well our unitary gates are behaving in the precense of noise. In other words, we wish to measure some sort of distance between the noisy gate and the perfect noiseless gate. One choice is the state-dependent fidelity that measures the fidelity of the output of two channels $\mathcal{E}_1$ and $\mathcal{E}_2$ when we input the same state.

$$F_{\mathcal{E}_1,\mathcal{E}_2}(\rho) = \left(\operatorname{Tr} \sqrt{\sqrt{\mathcal{E}_1(\rho)} \mathcal{E}_2(\rho) \sqrt{\mathcal{E}_1(\rho)}} \right)^2 \tag{124}$$

Due to the concavity of the fidelity, its minimum occurs when we input pure states. This motivates us to define the average gate fidelity to be its averge over random pure states. Let $\mathcal{U}$ and $\mathcal{E}$ denote the (noiseless) unitary channel and the noisy channel. In this case the fidelity becomes the trace norm because $\mathcal{U}$ output pure states.

$$\overline{F_{\mathcal{E},\mathcal{U}}} = \mathbb{E}_{|\psi\rangle \in H} \left[\operatorname{Tr} \left[\mathcal{E}(|\psi\rangle \langle\psi|) \mathcal{U}(|\psi\rangle \langle\psi|) \right] \right] = \mathbb{E}_{|\psi\rangle \in H} \left[\langle\psi| \left(\mathcal{U}^\dagger \circ \mathcal{E} \right)(|\psi\rangle \langle\psi|) |\psi\rangle \right] \tag{125}$$

Since $|\psi\rangle \langle\psi|$ appeared 2 times, a 2-design ensemble (such as pure states generated by the Clifford group) is enough for the random average. However, this averaged fidelity is not practical because we cannot get an ideal noiseless $\mathcal{U}$. The use of design here is not necessary, but may help us better estimate global quantities [14].

We modify it a little bit [15, 16]. We want to benchmark the Clifford gates $\{C\}$[8]. First, randomly apply a series of group elements $C_{i_1}, \cdots, C_{i_m}$ to the initial state. Then apply $C_{inv} = (C_{i_m} \cdots C_{i_1})^{-1}$. Finally, we measure the overlap with the initial pure state. If the channels

[8] This is almost enough in the sense that adding $\frac{\pi}{8}$ gates to the Clifford gernerators makes the gates universal.

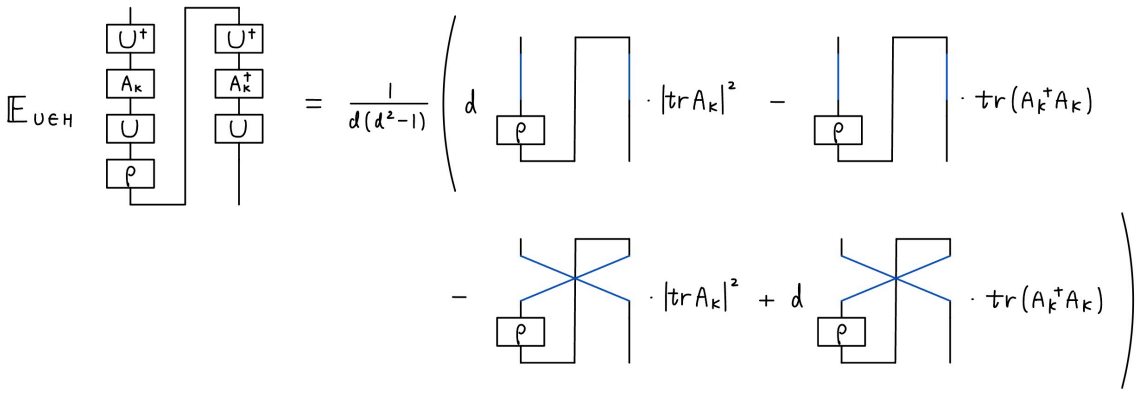


Figure 4: $\mathbb{E}_{U\in H}[U^\dagger A_k^i U \rho U^\dagger A_k^{i\dagger} U]$

are noiseless, then we would get 1. Due to the noise, we actually applied some noisy channel $\widetilde{C}_{inv} \circ \widetilde{C}_m \circ \cdots \circ \widetilde{C}_1$. The idea is the noise would accumulate with with m in a tractable manner. A simple model for the noise is $\widetilde{C}_i = \lambda \circ C_i$, $\Lambda(\rho) = \sum_k A_k \rho A_k^\dagger$. So we are measuring

$$\begin{aligned} F_m = \mathbb{E}_{C_1,\cdots,C_m\in C} \operatorname{Tr}\left[|\psi\rangle\langle\psi| \left(\mathcal{C}_1^{-1} \circ \cdots \circ \mathcal{C}_m^{-1} \circ \Lambda_m \circ \mathcal{C}_m \circ \cdots \circ \Lambda_1 \circ \mathcal{C}_1\right)(|\psi\rangle\langle\psi|)\right] \\ \mathbb{E}_{U_1,\cdots,U_m\in H} \operatorname{Tr}\left[|\psi\rangle\langle\psi| \left(\mathcal{U}_1^{-1} \circ \cdots \circ \mathcal{U}_m^{-1} \circ \Lambda_m \circ \mathcal{U}_m \circ \cdots \circ \Lambda_1 \circ \mathcal{U}_1\right)(|\psi\rangle\langle\psi|)\right] \end{aligned} \tag{126}$$

The second equality is because the Clifford group is a 2-design. Using the invariance of Haar measure

$$\mathbb{E}_{U_1,\cdots,U_i\in H}\left[\mathcal{U}_1^{-1} \circ \cdots \circ \mathcal{U}_i^{-1} \circ \Lambda_i \circ \mathcal{U}_i \circ \cdots \circ \mathcal{U}_1\right] = \mathbb{E}_{U\in H}\left[\mathcal{U}_i^{-1} \circ \Lambda_i \circ \mathcal{U}\right] \equiv \widehat{\Lambda}_i \tag{127}$$

$$F_m = \operatorname{Tr}\left[|\psi\rangle\langle\psi| \left(\widehat{\Lambda}_m \circ \cdots \circ \widehat{\Lambda}_1 |\psi\rangle\langle\psi|\right)\right] \tag{128}$$

Here comes the catch: the averaged error channel is simply a depolarizing channel!

$$\widehat{\Lambda}_i(\rho) = \mathbb{E}_{U\in H}[U^\dagger \sum_k A_k^i U \rho U^\dagger A_k^{i\dagger} U] = p_i \rho + (1-p_i)\frac{I}{d} \qquad p_i = \frac{\sum_k |\operatorname{Tr} A_k^i|^2 - 1}{d^2-1} \tag{129}$$

See figure 4 for the diagrams of the Weingarten functions. We further assume the errors are gate independent: $\Lambda_i = \Lambda$. Then

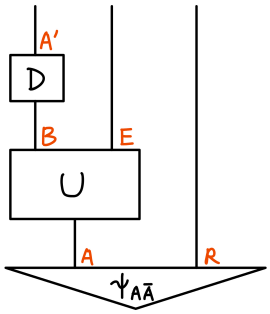
$$F_m = \operatorname{Tr}\left[|\psi\rangle\langle\psi| \left(p^m |\psi\rangle\langle\psi| + (1-p^m)\frac{I}{d}\right)\right] = \frac{1}{d} + \left(1-\frac{1}{d}\right) p^m \tag{130}$$

This is an exponential decay.

7.3 Error correction

Two designs can achieve very good deocupling—something that is required for a good error correction code. Suppose we have a state ρ_A to be encoded. We purify ρ_A with a reference R. Let $|\psi_{AR}\rangle$ denote the global state. We act a random unitary on A, as an encoding. Then we divide A into B and E(environment) and erase the environment. The errasure error is

correctible if we can find a decoding channel $\mathcal{D}$ acting on B such that it recovers the initial ρ_A. Let A' denote the output of $\mathcal{D}$.



Initially $I(A,R) = 2S_A$. If we are to success, the mutual information must retained: $I(A,R) = I(A',R)$. But in the procedure we acted many channels, which cannot increase mutual information: $I(A,R) \geq I(B,R) \geq I(A',R)$. So these inequalites should be saturated. In particular

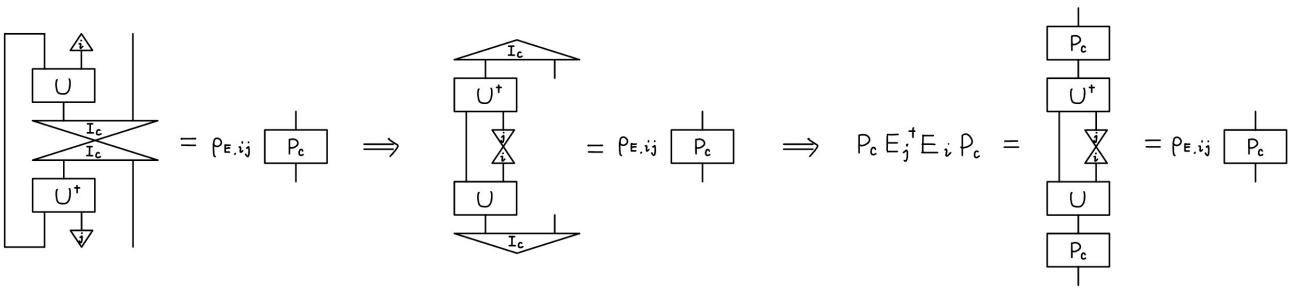
$$I(BE,\overline{A}) = I(B,R) \quad \Leftrightarrow \quad I(E,R) = 0 \quad \Leftrightarrow \quad \rho_{ER} = \rho_E \otimes \rho_R \tag{131}$$

E and R are decoupled. This is a necessary condition for QEC.

To prove that it's sufficient refxlq, we decompose the encoding channel (from A to B) into an operator sum and use the Knill-Laflamme quantaum error correction criterion. The Kraus operators are $E_i = \langle i_E| U$. Set the initial state to be the maximally mixed state in the code space $\rho_A = \frac{I_C}{d_C}$ with $|I_C\rangle = \sum_{j=1}^{d_C} |j_A\rangle |j_R\rangle$ being its purification. The matrix elements of ρ_{ER} can be related to Kraus operators:

$$\langle i_E| \rho_{ER} |j_E\rangle = \rho_{E,ij} \otimes \frac{I_C}{d_C} = \mathrm{Tr}_B \left[E_i |I_C\rangle \langle I_C| E_j^\dagger\right] \tag{132}$$

The first equality is due to decoupling. After rearranging the legs



We get

$$P_C E_j^\dagger E_i P_C = \rho_{E,ij} P_C \tag{133}$$

Since $\rho_{E,ij}$ is a Hermitian matrix, the error correction condition is satisfied.

The decoupling equality [17] says that random unitaries can achieve good decoupling, hence good codes.

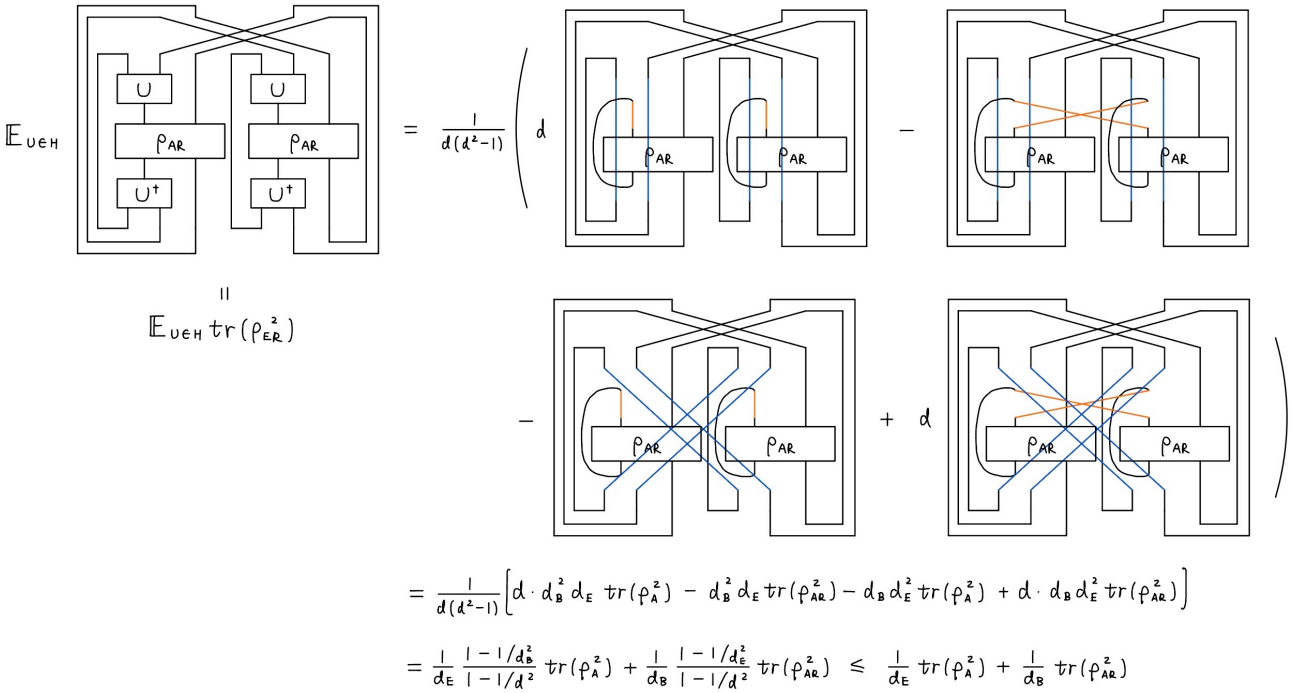


Figure 5: $\mathbb{E}_{U\in H}\left[\operatorname{Tr}\rho_{ER}^2(U)\right]$

Theorem 11 (Decoupling)**.** *For a general entangled state ρ_{AR},*

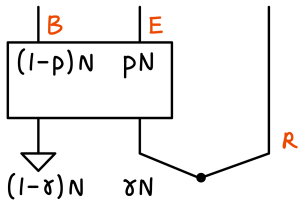
$$\mathbb{E}_{U\in H}\left\|\rho_{ER}(U) - \frac{I_E}{d_E}\otimes\rho_R\right\|_1 \le \sqrt{\frac{d_E d_R}{d_B}\operatorname{Tr}(\rho_{AR}^2)} \qquad \rho_{ER}(U) = \operatorname{Tr}_B[U_A\rho_{AR}U_A^\dagger] \tag{134}$$

Proof.

$$\begin{aligned}\mathbb{E}_{U\in H}\left\|\rho_{ER}(U) - \frac{I_E}{d_E}\otimes\rho_R\right\|_2^2 &= \mathbb{E}_{U\in H}\left[\operatorname{Tr}\rho_{ER}^2(U)\right] - \frac{1}{d_E}\operatorname{Tr}(\rho_R^2)\\ &= \frac{1}{d_E}\frac{1-1/d_B^2}{1-1/d_A^2}\operatorname{Tr}(\rho_R^2) + \frac{1}{d_B}\frac{1-1/d_E^2}{1-1/d_A^2}\operatorname{Tr}(\rho_{AR}^2) - \frac{1}{d_E}\operatorname{Tr}(\rho_R^2) \le \frac{1}{d_B}\operatorname{Tr}(\rho_{AR}^2)\end{aligned} \tag{135}$$

The second equality is derived using Weingarten functions in figure (5). □

In the proof we used $U\otimes U^*$ two times before averaging, thus a two design is enough to do the job. When $d_B \gg d_E$, E is well decoupled from R. Suppose we have N qubits in A and pN qubits in E. Initially, γN of them are maximally entangled with R, and the other $(1-\gamma)N$ are in $|0\rangle$[9]. So γ quantifies the entanglement of A.



[9]This is the setup that appeared in [18]. The question that they asked is whether measurements in E can destroy entanglement between A and R. The scrambling unitary protects entanglement $S(R)$ due to decoupling.

The right hand side of the decoupling inequality is $2^{-(1-2p-\gamma)N/2}$. In the $N \to \infty$ limit, there is a phase transition at $p = \frac{1}{2}(1-\gamma)$.

8 Open questions

More efficient tomography for Von-Neuman entropy.

Systems with energy conservation, and more symmetries.

Field theory analogs. Random projections in field theory.

How to measure the randomness in the quantum mechanical ensemble that is dual to gravity. Maybe it is not Haar, but it should preserve some features of Haar.

A Operator norms

Definition 11 (Schatten p-norm)**.**

$$\|A\|_p = (\operatorname{Tr}|A|^p)^{\frac{1}{p}} \tag{136}$$

If $\{\lambda_i\}$ are the eigenvalues of A, then

$$\begin{aligned} \|A\|_p &= \left(\sum_i |\lambda_i|^p\right)^{\frac{1}{p}} \\ \|A\|_\infty &= \max_i |\lambda_i| = \max_{|\psi\rangle \neq 0} \frac{\||A|\psi\rangle\|}{\||\psi\rangle\|} \end{aligned} \tag{137}$$

The p norm is monotonic:

$$\|A\|_1 \geq \|A\|_2 \geq \cdots \geq \|A\|_\infty \tag{138}$$

Proof. For $q > p$, assume without loss of generality $\|A\|_q = 1$, then $|\lambda_i| \leq 1$, $|\lambda_i|^p \leq |\lambda_i|^q$, $(\sum_i |\lambda_i|^p)^{\frac{1}{p}} \leq 1 = \|A\|_q = 1$. □

Also, the lower norm is upper bounded by the higher norms as

$$\|A\|_1 \leq \sqrt{d} \cdot \|A\|_2 \qquad \|A\|_1 \leq d \cdot \|A\|_\infty \tag{139}$$

where d is the dimension of the Hilbert space that A acts on. The first inequality follows from Cauchy-Schwartz $\sum_i |\lambda_i| \leq \sqrt{d} \cdot \sqrt{\sum_i |\lambda_i|^2}$. The second inequality is just $\sum_i |\lambda_i| \leq d \cdot \max |\lambda_i|$. A bound for general p and q(Wastrous p32):

$$\|A\|_p \leq \operatorname{rank}(A)^{\frac{1}{p}-\frac{1}{q}} \|A\|_q \leq d^{\frac{1}{p}-\frac{1}{q}} \|A\|_q, \quad 1 \leq p \leq q \tag{140}$$

In particular,

$$\|A\|_1 \leq \sqrt{\operatorname{rank}(A)} \|A\|_2, \qquad \|A\|_2 \leq \sqrt{\operatorname{rank}(A)} \|A\|_\infty \tag{141}$$

Hölder's inequality generalizeds Chauch-Schwartz. It says that for $\forall p, q \in [1, +\infty)$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$\operatorname{Tr}(A^\dagger B) \leq \|A\|_p \|B\|_q \tag{142}$$

Operational significance of the 1-norm

We would like to distinguish two states ρ and σ through a two-outcome POVM described by Kraus operators, the POVM is described by Kraus operators $\{M_0, M_1\}$. Denote $M = M_0^\dagger M$. The two states correspond to two probability distributions associated with the POVM.

$$p_0(\rho) = \mathrm{Tr}[M\rho], \quad p_1(\rho) = \mathrm{Tr}[(1-M)\rho], \quad p_0(\sigma) = \mathrm{Tr}[M\sigma], \quad p_1(\sigma) = \mathrm{Tr}[(1-M)\sigma] \tag{143}$$

The bias is given by

$$\Delta p = \mathrm{Tr}[M(\rho - \sigma)] \tag{144}$$

We can think of it as the success rate of distinguishing them. The Holevo-Helstrom theorem says that the bias is bounded by the trace distance:

Theorem 12 (Holevo-Helstrom)**.** *For any POVM $\{M, I - M\}$,*

$$\mathrm{Tr}[M(\rho - \sigma)] \leq \frac{1}{2}||\rho - \sigma||_1 \tag{145}$$

Equality is achieved when $M = P$, where P is the projector on the positive eigenspace of $\rho - \sigma$.

Proof. Let P and Q denote the projectors on the positive and negative eigenspace of $\rho - \sigma$.

$$\rho - \sigma = P - Q \qquad |\rho - \sigma| = P + Q \qquad \mathrm{Tr}(\rho - \sigma) = 0 \;\Rightarrow\; \mathrm{Tr}\, P = \mathrm{Tr}\, Q \tag{146}$$

$$\|\rho - \sigma\|_1 = \mathrm{Tr}\, P + \mathrm{Tr}\, Q = 2\,\mathrm{Tr}\, P \tag{147}$$

$$\mathrm{Tr}[M(\rho - \sigma)] = \mathrm{Tr}[M(P - Q)] \leq \mathrm{Tr}[MP] \leq \mathrm{Tr}\, P = \frac{1}{2}\|\rho - \sigma\|_1 \tag{148}$$

where we used $M \leq I$. Setting $M = P$, the inequalites in the equation obove become equalities. □

B Diamond norm

A natural norm for quantum channels is

Definition 12 (p-norm for maps)**.**

$$\|T\|_{p\to p} = \max_{A\neq 0} \frac{\|T(A)\|_p}{\|A\|_p} = \max_{\|A\|_p \leq 1} \|T(A)\|_p \tag{149}$$

However, it is not stable under tensoring with identity. Consider the transpose operation

$$T: \; |i\rangle\langle j| \mapsto |j\rangle\langle i|, \qquad i, j = 1, \cdots, d \tag{150}$$

Since transposition don't change singular values, $\|T\|_{1\to 1} = 1$. The norm of the operator $X = \sum_{i,j} |i,i\rangle\langle j,j|$ is $\|X\|_1 = d$. Acting $T \otimes I$ on X gives the swap operator whose norm is $\|(T \otimes I)X\|_1 = \mathrm{Tr}\sqrt{SWAP^2} = \mathrm{Tr}[I \otimes I] = d^2$, thus $\|T \otimes I\|_{1\to 1} \geq d$.

Definition 13 (diamond norm [19])**.**

$$\|T\|_\diamond = \|T \otimes I_{\mathcal{H}}\|_{1\to 1} \tag{151}$$

The following theorem says that the diamond norm is "just enough" to be stable under tensoring with identity.

Theorem 13 (Theorem 3.46 of [3])**.**

$$\|T \otimes I_{\mathcal{H}'}\|_{1\to 1} \leq \|T\|_\diamond \tag{152}$$

Equality holds when $\dim \mathcal{H}' \geq \dim \mathcal{H}$*.*

Proof. Since $T \otimes I_{\mathcal{H}'}$ is a linear map, its output can be written as a sum over $(T \otimes I_{\mathcal{H}'})(|u\rangle \langle v|)$. Due to the convexity of the 1-norm (any norm is convex), there exist some $|u\rangle, |v\rangle \in \mathcal{H} \otimes \mathcal{H}'$, such that

$$\|T \otimes I_{\mathcal{H}'}\|_{1\to 1} = \max_{\|A\|_1 \leq 1} \|(T \otimes I_{\mathcal{H}'})(A)\|_1 = \|(T \otimes I_{\mathcal{H}'})(|u\rangle \langle v|)\|_1 \tag{153}$$

We now need a lemma: there exist some $|x\rangle, |y\rangle \in \mathcal{H} \otimes \mathcal{H}$ such that $\|(T \otimes I_{\mathcal{H}'})(|u\rangle \langle v|)\|_1 = \|(T \otimes I_{\mathcal{H}})(|x\rangle \langle y|)\|_1$. For $\dim \mathcal{H}' \leq \dim \mathcal{H}$, we can construct isometry U and define $I_{\mathcal{H}} \otimes U |u\rangle = |x\rangle$ and $I_{\mathcal{H}} \otimes U |v\rangle = |y\rangle$. Now

$$\begin{aligned}\|(T \otimes I_{\mathcal{H}'})(|u\rangle \langle v|)\|_1 &= \|I_{\mathcal{H}} \otimes U(T \otimes I_{\mathcal{H}'})(|u\rangle \langle v|) I_{\mathcal{H}} \otimes U^\dagger\|_1 = \|(T \otimes I_{\mathcal{H}})(I_{\mathcal{H}} \otimes U |u\rangle \langle v| I_{\mathcal{H}} \otimes U^\dagger)\|_1 \\ &= \|(T \otimes I_{\mathcal{H}})(|x\rangle \langle y|)\|_1\end{aligned} \tag{154}$$

For $\dim \mathcal{H}' \geq \dim \mathcal{H}$, we construct isometies $U, V : \mathcal{H} \mapsto \mathcal{H}'$ such that $|u\rangle = I_{\mathcal{H}} \otimes U |x\rangle$ and $|v\rangle = I_{\mathcal{H}} \otimes V |y\rangle$. Then

$$\begin{aligned}\|(T \otimes I_{\mathcal{H}'})(|u\rangle \langle v|)\|_1 &= \|(T \otimes I_{\mathcal{H}'})(I_{\mathcal{H}} \otimes U |x\rangle \langle v| I_{\mathcal{H}} \otimes V^\dagger)\|_1 = \|I_{\mathcal{H}} \otimes U(T \otimes I_{\mathcal{H}})(|x\rangle \langle v|) I_{\mathcal{H}} \otimes V^\dagger\|_1 \\ &= \|(T \otimes I_{\mathcal{H}})(|x\rangle \langle y|)\|_1\end{aligned} \tag{155}$$

Using this Lemma, we have

$$\|T \otimes I_{\mathcal{H}'}\|_{1\to 1} = \|(T \otimes I_{\mathcal{H}})(|x\rangle \langle y|)\|_1 \leq \|T\|_\diamond \tag{156}$$

When $\dim \mathcal{H}' \geq \dim \mathcal{H}$, there exist an isometry $V : \mathcal{H} \mapsto \mathcal{H}'$. For $\forall A$ with $\|A\|_1 \leq 1$, we have

$$\begin{aligned}\|(T \otimes I_{\mathcal{H}})(A)\|_1 &= \|I_{\mathcal{H}} \otimes V(T \otimes I_{\mathcal{H}})(A) I_{\mathcal{H}} \otimes V^\dagger\|_1 = \|(T \otimes I_{\mathcal{H}'})(I_{\mathcal{H}} \otimes V A I_{\mathcal{H}} \otimes V^\dagger)\|_1 \\ &\leq \|T \otimes I_{\mathcal{H}'}\|_{1\to 1} \|I_{\mathcal{H}} \otimes V A I_{\mathcal{H}} \otimes V^\dagger\|_1 = \|T \otimes I_{\mathcal{H}'}\|_{1\to 1} \|A\|_1 \leq \|T \otimes I_{\mathcal{H}'}\|_{1\to 1}\end{aligned} \tag{157}$$

Then $\|T\|_\diamond \leq \|T \otimes I_{\mathcal{H}'}\|_{1\to 1}$ □

The diamond norm has a nice operational meaning, analogous to the operator 1-norm.

Theorem 14. *For any* $\mathcal{H}'$ *and any choice of measurement* M *that acts on* $\mathcal{H} \otimes \mathcal{H}'$*,*

$$\mathrm{Tr}[M(\Phi_1 \otimes I_{\mathcal{H}'})(\rho) - M(\Phi_2 \otimes I_{\mathcal{H}'})(\rho)] \leq \frac{1}{2}\|\Phi_1 - \Phi_2\|_\diamond \tag{158}$$

For $\dim(\mathcal{H}') \geq \dim(\mathcal{H})$*, equality can be achieved for some choice of* ρ *and* M*.*

Proof. By the Holevo-Helstrom theorem,

$$\mathrm{Tr}[M(\Phi_1 \otimes I_{\mathcal{H}'})(\rho) - M(\Phi_2 \otimes I_{\mathcal{H}'})(\rho)] \leq \frac{1}{2}\|(\Phi_1 \otimes I_{\mathcal{H}'} - \Phi_2 \otimes I_{\mathcal{H}'})(\rho)\|_1 \leq \frac{1}{2}\|\Phi_1 \otimes I_{\mathcal{H}'} - \Phi_2 \otimes I_{\mathcal{H}'}\|_\diamond \tag{159}$$

When $\dim(\mathcal{H}') \geq \dim(\mathcal{H})$, the last inequality becomes equality. The Holevo-Helstrom theorem gaurantees that the first equality can be achieved. □

$$\|T\|_\diamond = \|T \otimes I_{\mathcal{H}}\|_{1\to 1} = \max_{A\neq 0} \frac{\|(T \otimes I_{\mathcal{H}})(A)\|_1}{\|A\|_1} \leq \max_{A\neq 0} \frac{d \cdot \|(T \otimes I_{\mathcal{H}})(A)\|_2}{\|A\|_2} = d \cdot \|T\|_2 \tag{160}$$

C Concentration of measure

Lemma 4 (Levy's lemma)**.** *Consider a unit sphere S^{2n-1} embedded in $\mathbb{R}^{2n}$. If a function $f : S^{2n-1} \to \mathbb{R}$ is Lipchitz with constant η, meaning*

$$|f(x) - f(y)| \leq \eta\|x - y\| \tag{161}$$

where $\|x - y\|$ is the usual distance in the embedding space $\mathbb{R}^{2n}$.

For a uniform measure on S^k, the function's value is concentrated near its expectation value:

$$\Pr\left[\,|f(x) - \mathbb{E}_x f(x)| > \alpha\right] \leq 2\exp\left(-\frac{n\alpha^2}{9\pi^3\eta^2}\right) \tag{162}$$

References

[1] F. Mezzadri, *How to generate random matrices from the classical compact groups*, *arXiv e-prints* (Sept., 2006) math–ph/0609050, [math-ph/0609050].

[2] A. W. Harrow, *The Church of the Symmetric Subspace*, *arXiv e-prints* (Aug., 2013) arXiv:1308.6595, [1308.6595].

[3] J. Watrous, *The Theory of Quantum Information*. Cambridge University Press, 2018, 10.1017/9781316848142.

[4] D. A. Roberts and B. Yoshida, *Chaos and complexity by design*, *Journal of High Energy Physics* **2017** (apr, 2017) .

[5] H. Zhu, *Multiqubit clifford groups are unitary 3-designs*, *Physical Review A* **96** (dec, 2017) .

[6] N. Hunter-Jones, *Unitary designs from statistical mechanics in random quantum circuits*, 1905.12053.

[7] M. Ippoliti and W. W. Ho, *Dynamical purification and the emergence of quantum state designs from the projected ensemble*, 2023.

[8] R. A. Low, *Pseudo-randomness and learning in quantum computation*, 2010.

[9] F. G. Brandão, W. Chemissany, N. Hunter-Jones, R. Kueng and J. Preskill, *Models of quantum complexity growth*, *PRX Quantum* **2** (Jul, 2021) 030316.

[10] http://theory.caltech.edu/~preskill/ph229/notes/chap3.pdf.

[11] J. S. Cotler, G. Gur-Ari, M. Hanada, J. Polchinski, P. Saad, S. H. Shenker et al., *Black holes and random matrices*, *Journal of High Energy Physics* **2017** (may, 2017) .

[12] P. W. Claeys and A. Lamacraft, *Emergent quantum state designs and biunitarity in dual-unitary circuit dynamics*, *Quantum* **6** (June, 2022) 738.

[13] A. Roy and A. J. Scott, *Weighted complex projective 2-designs from bases: Optimal state determination by orthogonal measurements*, *Journal of Mathematical Physics* **48** (jul, 2007) 072110.

[14] H.-Y. Huang, R. Kueng and J. Preskill, *Predicting many properties of a quantum system from very few measurements*, *Nature Physics* **16** (jun, 2020) 1050–1057.

[15] J. Emerson, R. Alicki and K. Ż yczkowski, *Scalable noise estimation with random unitary operators*, *Journal of Optics B: Quantum and Semiclassical Optics* **7** (sep, 2005) S347–S352.

[16] E. Magesan, J. M. Gambetta and J. Emerson, *Characterizing quantum gates via randomized benchmarking*, *Physical Review A* **85** (apr, 2012) .

[17] A. Abeyesinghe, I. Devetak, P. Hayden and A. Winter, *The mother of all protocols: restructuring quantum information's family tree*, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **465** (jun, 2009) 2537–2563.

[18] S. Choi, Y. Bao, X.-L. Qi and E. Altman, *Quantum error correction in scrambling dynamics and measurement-induced phase transition*, *Physical Review Letters* **125** (jul, 2020) .

[19] A. Y. Kitaev, A. H. Shen and M. N. Vyalyi, *Classical and quantum computation*, in *Graduate Studies in Mathematics*, 2002.