Exact and Approximate Unitary 2-Designs: Constructions and Applications

Christoph Dankert,[1] Richard Cleve,[1,2] Joseph Emerson,[3] and Etera Livine[2]

[1]*David R. Cheriton School of Computer Science and Institute for Quantum Computing, University of Waterloo*
[2]*Perimeter Institute for Theoretical Physics, Waterloo*
[3]*Department of Applied Mathematics and Institute for Quantum Computing, University of Waterloo*

(Dated: Dated: June 20, 2006)

We consider an extension of the concept of spherical t-designs to the unitary group in order to develop a unified framework for analyzing the resource requirements of randomized quantum algorithms. We show that certain protocols based on *twirling* require a unitary 2-design. We describe an efficient construction for an exact unitary 2-design based on the Clifford group, and then develop a method for generating an ϵ-approximate unitary 2-design that requires only $\mathcal{O}(n\log(1/\epsilon))$ gates, where n is the number of qubits and ϵ is an appropriate measure of precision. These results lead to a protocol with exponential resource savings over existing experimental methods for estimating the characteristic fidelities of physical quantum processes.

The importance of generating random states and random unitary operators in quantum information processors has become increasingly clear from the growing number of algorithms and protocols that presume such a resource [1, 2, 3, 4, 5, 6, 7]. In several of the above applications the Haar-measure on the unitary group is the relevant randomization measure [3, 5, 7]. It is well-known that generating Haar-random unitary operators on a quantum information processor is inefficient: the number of gates grows exponentially with the number of qubits n. Consequently it is useful to identify subsets of the unitary group that admit efficient gate decompositions and that can adequately simulate the Haar-measure for a given randomization task.

In this context we find it useful to classify finite subsets of the unitary group according to the highest degree polynomial for which averages over the subset are indistinguishable from averages with respect to the Haar measure. Following recent work that has extended the related concept of *spherical t-design* to classify finite sets of quantum states [8], we define a *unitary t-design* as a set $\{U_k\}_{k=1}^K \subset U(D)$ of unitary operators such that,

$$\frac{1}{K}\sum_{k=1}^{K} P_{(m,l)}(U_k) = \int_{U(D)} dU P_{(m,l)}(U) \qquad (1)$$

for all $(m,l) \leq (t,t)$ where $P_{(m,l)}(U)$ denotes a polynomial of degree (m,l) in the (complex) matrix elements of U [9]. In the above dU denotes the Haar measure on the unitary group $U(D)$.

It is easy to see from this definition that the Heisenberg-Weyl group and the generalized Pauli group (defined below) both generate unitary 1-designs. An example of a randomization task for which a *unitary 1-design* is necessary and sufficient is the private quantum channel [1]. This protocol requires sampling from a set of unitary operators such that an average over the set takes any input state to the completely mixed (identity) state: this condition corresponds to averaging a polynomial of degree $(1,1)$. As is known, exact randomization with respect to a unitary 1-design can be done efficiently: only n quantum gates acting in parallel are required.

Several quantities of experimental and theoretical interest can be expressed as Haar averages of polynomials of degree $(2,2)$. Important examples are the entanglement fidelity [10] and the average gate fidelity [11] of a physical quantum process, and the average entanglement between subsystems expressed in terms of subsystem purity [6]. In this Letter we show that the Clifford group, which has been considered previously as an appropriate Haar-substitute for certain protocols [2, 12], constitutes an exact unitary 2-design. In addition, we devise a circuit construction for generating an ϵ-approximate unitary 2-design that requires only $\mathcal{O}(n\log(1/\epsilon))$ quantum gates. We then describe how our circuit construction enables an efficient, scalable protocol for experimentally estimating the entanglement and average gate fidelities mentioned above.

We start by observing that for $t = 2$ the definition above is equivalent to:

$$\frac{1}{K}\sum_{k=1}^{K} U_k^\dagger A U_k X U_k^\dagger B U_k = \int_{U(D)} dU\, U^\dagger A U X U^\dagger B U \qquad (2)$$

for all linear operators A, X, B. This condition emphasizes that the close relationship between unitary 2-designs and the physical transformation known as *twirling* [13].

Definition 1. For a mapping $\Lambda : \mathbb{C}^{D\times D} \to \mathbb{C}^{D\times D}$, and a probability measure S on $U(D)$, an *S-twirl* applied to Λ consists of the mapping

$$X \mapsto \int_{U(D)} dS(U)\, U^\dagger \Lambda(U X U^\dagger) U. \qquad (3)$$

It follows that a unitary 2-design will adequately simulate the Haar-measure for any randomization protocol based on the expectation of a twirled superoperator [7, 13, 14]. A similar result holds for protocols based on

bilateral twirling of a bipartite quantum state [2] (see [15] for further details.)

We now specialize to dimensions $D = 2^n$ and prove that the Clifford group $\mathcal{C}_n$ forms a generic unitary 2-design. Our strategy is to show that a $\mathcal{C}_n$-twirl of any linear superoperator of the form $\Lambda(X) = AXB$ is equivalent to a Haar-twirl. $\mathcal{C}_n$ is defined as the normalizer of the generalized Pauli group $\mathcal{P}_n$, which consists of all n-fold tensor products of the one-qubit Pauli operators $\{\mathbb{I}, X, Y, Z\}$. We denote the elements of $\mathcal{P}_n$ as $\{P_j\}_{j=1}^{D^2}$, where P_1 is the n-fold tensor product of $\mathbb{I}$.

Lemma 2. *Applying a $\mathcal{P}_n$-twirl to the mapping $\Lambda(X) = AXB$, where $A, B \in \mathbb{C}^{D\times D}$, results in a mapping of the form*

$$X \mapsto \sum_{k=1}^{D^2} r_k P_k X P_k, \tag{4}$$

where $r_1 = \mathrm{Tr}(A)\mathrm{Tr}(B)/D^2$ and $\sum_{k=1}^{D^2} r_k = \mathrm{Tr}(AB)/D$.

Proof. Note that we can express $A = \sum_{a=1}^{D^2} \alpha_a P_a$ and $B = \sum_{b=1}^{D^2} \beta_b P_b$. The resulting operation maps X to

$$\begin{aligned}
&1/D^2 \sum_{k=1}^{D^2} P_k A P_k X P_k B P_k \\
&= 1/D^2 \sum_{k=1}^{D^2} P_k \left(\sum_{a=1}^{D^2} \alpha_a P_a\right) P_k X P_k \left(\sum_{b=1}^{D^2} \beta_b P_b\right) P_k \\
&= 1/D^2 \sum_{a=1}^{D^2}\sum_{b=1}^{D^2} \alpha_a\beta_b \left(\sum_{k=1}^{D^2} (-1)^{(k,a\oplus b)s_P}\right) P_a X P_b \\
&= \sum_{a=1}^{D^2} \alpha_a\beta_a P_a X P_a,
\end{aligned} \tag{5}$$

with the symplectic inner product s_P on the index space (see [15] for details). Therefore, setting $r_k = \alpha_k\beta_k$, satisfies the conditions of the Lemma. □

Using Lemma 2, we prove the following.

Theorem 3. *Let Λ be any mapping of the form $\Lambda(X) = AXB$, where $A, B \in \mathbb{C}^{D\times D}$. Then applying a Clifford-twirl to Λ is equivalent to applying a Haar-twirl to Λ. That is, for all X,*

$$\int_{U(D)} dU\; U^\dagger A U X U^\dagger B U = \frac{1}{|\mathcal{C}_n|}\sum_{U\in\mathcal{C}_n} U^\dagger A U X U^\dagger B U.$$

Proof. First, as shown in Ref. [7], we can interpret the arguments in the 2-design as the action of the linear superoperator Λ to obtain

$$\begin{aligned}
\int_{U(D)} dU\; U^\dagger A U X U^\dagger B U &= \frac{\mathrm{Tr}(AB)\mathrm{Tr}(X)}{D}\frac{\mathbb{I}}{D} \\
&+ \frac{D\mathrm{Tr}(A)\mathrm{Tr}(B) - \mathrm{Tr}(AB)}{D(D^2-1)}\left(X - \mathrm{Tr}(X)\frac{\mathbb{I}}{D}\right).
\end{aligned} \tag{6}$$

We can express each $U \in \mathcal{C}_n$ as $U = C_j P_k$, where $P_k \in \mathcal{P}_n$ and $C_j \in \mathcal{C}_n/\mathcal{P}_n$. First we twirl Λ by $\mathcal{P}_n$ and then by $\mathcal{C}_n/\mathcal{P}_n$.

The effect of $\mathcal{P}_n$-twirling is established in Lemma 2, and applying a $\mathcal{C}_n/\mathcal{P}_n$-twirl to the resulting operator yields

$$\frac{|\mathcal{P}_n|}{|\mathcal{C}_n|}\sum_{j=1}^{|\mathcal{C}_n|/|\mathcal{P}_n|}\sum_{k=1}^{D^2} r_k C_j^\dagger P_k C_j X C_j^\dagger P_k C_j. \tag{7}$$

Next we distinguish the identity element $P_1 = \mathbb{I}$ and make use of the fact that the Clifford group is the normalizer of the Pauli group and hence, under conjugation, maps each non-identity Pauli element to every other non-identity Pauli element with equal frequency. It follows that the final state is

$$\begin{aligned}
&r_1 X + \frac{|\mathcal{P}_n|}{|\mathcal{C}_n|}\sum_{k=2}^{D^2} r_k \sum_{j=1}^{|\mathcal{C}_n|/|\mathcal{P}_n|} C_j^\dagger P_k C_j X C_j^\dagger P_k C_j \\
&= r_1 X + \frac{1}{D^2-1}\left(\sum_{k=2}^{D^2} r_k\right)\sum_{l=2}^{D^2} P_l X P_l.
\end{aligned} \tag{8}$$

Using the fact that $\sum_{j=1}^{D^2} P_j X P_j = D\mathrm{Tr}(X)\mathbb{I}$, and the equations in Lemma 2, it is straightforward to show that the right sides of Eqs. (8) and (6) are equivalent. □

By linearity, we deduce the following.

Corollary 4. *For an arbitrary trace-preserving CP map $\Lambda(\rho) = \sum_k A_k \rho A_k^\dagger$ (with $\sum_k A_k^\dagger A_k = \mathbb{I}$), applying a Clifford-twirl to Λ is equivalent to applying a Haar-twirl to Λ. That is,*

$$\int dU \sum_k U A_k U^\dagger \rho U A_k^\dagger U^\dagger = p\rho + (1-p)\mathrm{Tr}(\rho)\mathbb{I}/D, \tag{9}$$

where

$$p = \frac{\sum_k |\mathrm{Tr}(A_k)|^2 - 1}{D^2 - 1}. \tag{10}$$

Since the size of the Clifford group is exponential in n^2, the implementation of a Clifford group element requires at least order $n^2/\log n$ quantum gates [16]. In light of the practical usefulness of Clifford-twirling in the fidelity-estimation protocol described below and other contexts (e.g., [2]) it is useful to analyze the resource savings that may be achieved by generating an approximate unitary 2-design with a subset of the Clifford group.

From Lemma 2, it can be deduced that applying a Pauli-twirl to any channel yields a Pauli channel, and the cost of this is $\mathcal{O}(n)$ gates. In order to convert an arbitrary Pauli channel into a good approximation of a depolarizing channel, we add slightly more than $\mathcal{O}(n)$

further twirling operations to approximately uniformize the probabilities associated with each P_a for all $a \neq 1$.

This process consists of a series of repetitions of the following basic procedure:

1. $\mathcal{C}_2/\mathcal{P}_2$-twirl qubit k for all $k \in \{1, \ldots, n\}$. (This operation is defined below.)

2. Conjugate the first qubit by a random XOR. (This operation is defined below.)

3. H-conjugate the first qubit, and $\mathcal{C}_2/\mathcal{P}_2$-twirl qubit k for all $k \in \{2, \ldots, n\}$.

4. Conjugate the first qubit by a random XOR.

5. H-conjugate the first qubit, and $\mathcal{C}_2/\mathcal{P}_2$-twirl qubit k for all $k \in \{2, \ldots, n\}$.

6. With probability $1/2$, S-conjugate the first qubit.

7. Conjugate the first qubit by a random XOR.

A $\mathcal{C}_2/\mathcal{P}_2$-twirl of a qubit is defined as follows. Let $R = SH$, where $S = |0\rangle\langle 0| + i|1\rangle\langle 1|$, and H is the Hadamard transform. Select $i \in \{0, 1, 2\}$ uniformly and conjugate the register by R^i. This operation has the property that, if it is applied to the identity channel $\mathbb{I}$, it has no net effect; however, for a Pauli channel of the form X, Y, or Z, this operation causes the channel to become a uniform mixture of X, Y, and Z.

Conjugating the first qubit by a random XOR is the following operation. For each $k \in \{2, \ldots, n\}$, with independent probability $3/4$, conjugate by a CNOT gate with the first qubit as target and qubit k as control.

Consider the effect of starting with a channel of the form $\rho \mapsto P_a \rho P_a$, for some fixed $a \neq 1$, and applying the above procedure. To analyze the result, we trace through the effect of each of the seven steps:

1. For each k, if component k of P_a is $\mathbb{I}$ then it remains $\mathbb{I}$, and if component k is X, Y, or Z then it becomes a uniform mixture of X, Y, and Z.

2. Call an execution of this procedure *good* if, after Step 2, the first component of the channel is X or Y. This happens with probability at least $1/2$.

   Case 1: For all $k \in \{2, \ldots, n\}$, component k of P_a is $\mathbb{I}$. In this case, the CNOT gates have no effect, but since $a \neq 1$, component 1 of P_a is *not* $\mathbb{I}$. Therefore, after the previous step, the first component of P_a is uniformly distributed over X, Y, and Z. Hence the first component of the channel is X or Y with probability $2/3$.

   Case 2: For some $k \in \{2, \ldots, n\}$, component k of P_a is not $\mathbb{I}$. With probability $(2/3)(3/4) = 1/2$, component k has X or Y *and* the CNOT gate is present. This causes the first component to evolve as follows. If it is X or $\mathbb{I}$ then it becomes an equal mixture of $\mathbb{I}$ and X. Also, if it is Y or Z then it becomes an equal mixture of Y and Z. In either case, the first component is X or Y with probability $1/2$.

3. If the execution is good then the first component is Y or Z. For each $k \in \{2, \ldots, n\}$, component k is either $\mathbb{I}$ or a uniform mixture of X, Y, and Z.

4. If the execution is good then for each $k \in \{2, \ldots, n\}$, component k is $\mathbb{I}$ with independent probability $1/4$, and some mixture of X, Y, Z with probability $3/4$. To see why this is so, for each k, consider the effect of the back-action of the CNOT gates in the following two cases separately.

   Case 1: After the previous step, component k is $\mathbb{I}$. In this case, it remains $\mathbb{I}$ with probability $1/4$, and it becomes Z with probability $3/4$.

   Case 2: After the previous step, component k is a uniform mixture of X, Y, and Z. In this case, with probability $3/4$, the channel becomes a uniform mixture of Y, X, and $\mathbb{I}$. Hence the component becomes $\mathbb{I}$ with probability $(3/4)(1/3) = 1/4$.

5. If the execution is good then, after this step, the first component of the channel is X or Y, and, for each $k \in \{2, \ldots, n\}$, component k is independently a uniform mixture of $\mathbb{I}$, X, Y, and Z.

6. If the execution is good then, after this step, the first component of the channel is a uniform mixture of X and Y.

7. Call an execution *very good* if it is good and, after Step 6, there is at least one component $k \in \{2, \ldots, n\}$ that is not $\mathbb{I}$. If the execution is very good, the first component of the channel is a uniform mixture of $\mathbb{I}$, X, Y, and Z (independent of the other components of the channel).

   To see why this is so, consider the effect of any non-$\mathbb{I}$ component $k \in \{2, \ldots, n\}$. Prior to the potential conjugation by CNOT, the first component is uniformly distributed among X and Y and component k is uniformly distributed among X, Y, and Z. Therefore, with probability $(2/3)(3/4) = 1/2$, the first component becomes a uniform mixture of I and Z.

Now, suppose that an execution of the basic procedure is very good (this occurs with probability $(1/2)(1 - (1/4)^{n-1})$). Then the resulting channel is an approximately uniform distribution of all P_a, for $a \neq 1$ in the following sense. All Paulis that are not $\mathbb{I}$ in all components 2 through n occur with probability $1/4^n$, yielding a total variation distance $\mathcal{O}(1/4^n)$ from the uniform distribution.

Repeating the procedure $O(\log(1/\varepsilon))$ times, we can increase the probability of a very good execution to $1-\varepsilon$. It follows that the channel after the repeated procedure has variation distance $\varepsilon + 1/4^n$ from a depolarizing channel.

Each execution of the procedure consists of $\mathcal{O}(n)$ gates, that can be implemented in $\mathcal{O}(\log n)$ depth (see [15] for details). Therefore, the repeated procedure can be implemented with $\mathcal{O}(n \log 1/\varepsilon)$ gates in $\mathcal{O}(\log n \log 1/\varepsilon)$ depth.

We now turn to a discussion of the fidelity estimation problem for which the above unitary 2-design constructions lead to an efficient, scalable protocol. Consider the Haar-averaged fidelity [7, 17]

$$\begin{aligned} \langle F \rangle &\equiv \int_{U(D)} dU \mathrm{Tr}[U|0\rangle\langle 0|U^\dagger \Lambda(U|0\rangle\langle 0|U^\dagger] \\ &= \sum_k \frac{|\mathrm{Tr}(A_k)|^2 + D}{D^2 + D}. \end{aligned} \tag{11}$$

of a quantum operation $\Lambda(\rho) = \sum_k A_k \rho A_k^\dagger$. The Haar-averaged fidelity is trivially related to two standard fidelity benchmarks: the entanglement-fidelity F_e, which has been proposed as means of characterizing the noise strength in a physical quantum channel Λ [10], and the gate-fidelity F_g, which has been used to characterize the quality of quantum memory [18] or of an implementation of a target unitary U_g on a noisy quantum processing device [11, 19]. In the latter scenario we imagine the implementation of a gate sequence U_g followed immediately by its inverse $U_g^\dagger$, and make the identification $\Lambda(\rho) = U_g^\dagger \mathcal{E}(U_g \rho U_g^\dagger) U_g$, where the map $\mathcal{E}(\rho)$ represents the noise accumulated over the course of implementing $U_g^\dagger U_g$. Then, using the results of Ref. [7, 10, 11, 17], we find the following simple relationship between the Haar-average fidelity and the previously proposed gate-fidelity and entanglement-fidelity,

$$\langle F \rangle = \frac{DF_g + 1}{D+1} = \frac{DF_e + 1}{D+1}. \tag{12}$$

Standard methods for estimating F_e and F_g are based on either state or process tomography and the best known methods require a number of experiments that grows exponentially with $n = \log_2 D$ [17, 20]. However, as described in Ref. [7], we can estimate $\langle F \rangle$ directly by the following protocol: apply a random unitary operator U to the initial state $|0\rangle$, followed by the quantum operation Λ, and then apply $U^\dagger$ to the output state. Then from Eq. (11) we see that $\langle F \rangle$ can be estimated by repeating this procedure with U sampled randomly from the Haar measure in each experiment. Given that F is a polynomial function of homogeneous degree (2,2), Theorem 3 implies that we can estimate $\langle F \rangle$ by sampling from any unitary 2-design. For an arbitrary, but fixed, average fidelity $0 \leq \langle F \rangle \leq 1$, the Chernoff bound guarantees that the number of experiments required to estimate $\langle F \rangle$ to precision $\delta > 1/4^n$ is independent of the dimension D. Finally, the ϵ-approximate unitary 2-design described above implies that each experiment requires only $\mathcal{O}(n \log(1/\epsilon))$ gates. Hence the fidelity $\langle F \rangle$, and equivalently F_g and F_e, may be estimated by a scalable, efficient experimental protocol.

Turning to dimensions other than powers of two, consider the case where $D = p$, for an odd prime p. We can replace the Pauli group by the Heisenberg-Weyl (HW) group generated by the two operators $X|k\rangle = |k+1\rangle$, $Z|k\rangle = \omega|k\rangle$ (where $\omega = e^{i2\pi/p}$). Following a similar approach to above, one can show that the HW normalizer forms a unitary 2-design. Furthermore, this normalizer is generated approximately by repeated conjugation under a unitary drawn uniformly at random from the set of MUB unitaries and their inverses [21]. Since the MUB unitaries have circuit decompositions of $O(\log^2 p)$ gates (see [15]), we have an efficient approximate 2-design for general odd prime dimensions. Using the MUB unitaries may also be a promising approach to efficiently constructing 2-designs for general odd prime power dimensions.

It remains an interesting open question whether an arbitrary quantum randomization algorithm can be reduced to a t-design condition, and hence classified within this framework. This would provide further motivation to generalize the methods of this paper to obtain unitary and state t-designs for $t > 2$. The alternate definition proposed in [9] might be a good starting point for research in this direction.

We thank Robin Blume-Kohout, David Cory, Daniel Gottesman, Debbie Leung, and Pranab Sen for helpful discussions. This work was supported in part by Canada's NSERC, MITACS, and CIAR, and the U.S. ARO.

[1] A. Ambainis, M. Mosca, A. Tapp, R. de Wolf, Proc. 41st FOCS, 547-553 (2000).
[2] D. DiVincenzo, D. Leung, B. Terhal, IEEE Trans. Inform. Theory, 48(3):580-599, (2002).
[3] J. Radishkran, M. Roetteler, P. Sen, LNCS 3580, 1399-1411 (2005); P. Sen, quant-ph/0512085 (2005).
[4] A. Ambainis, A. Smith, LNCS 3122, 249-260 (2004).
[5] P. Hayden, D. Leung, P. Shor, A. Winter, Commun. Math. Phys. 250(2):371-391, 2004; C. H. Bennett, P. Hayden, D. Leung, P. Shor, A. Winter, IEEE Trans. Inform. Theory 51(1):56-74, 2005; A. Harrow, P. Hayden, and D. Leung, Phys. Rev. Lett. 92, 187901 (2004).
[6] J. Emerson, Y. Weinstein, M. Saraceno, S. Lloyd, D. Cory, Science 302: 2098-2100 (Dec. 19, 2003).
[7] J. Emerson, R. Alicki, K. Zyczkowski, J. Opt. B: Quantum and Semiclassical Optics, **7** S347-S352 (2005).
[8] J. Renes, R. Blume-Kohout, A. Scott, C. Caves, J. Math. Phys. 45(6):2171-2180 (2004); A. Klappenecker and M. Roetteler, Proc. ITIS 2005, 1740-1744 (2005).

[9] An equivalent definition of a unitary t-design is such that $\sum_{k=1}^{K} D^{\mathcal{I}}(U_k) = 0$ for all non-trivial irreducible representations $D^{\mathcal{I}}$ contained in the tensor power $V^{\otimes t} \otimes \overline{V}^{\otimes t}$ of the fundamental representation V and its conjugate. This characterization is especially relevant when analyzing the efficiency of "random circuit" constructions for generating pseudo-random sets of unitaries [22].

[10] B. Schumacher, Phys. Rev. A 54, 2614-2628 (1996).

[11] E. M. Fortunato, M. A. Pravia, N. Boulant, G. Teklemariam, T. F. Havel, D. G. Cory, J. Chem. Phys. 116, 7599-7606 (2002).

[12] H. Chau, IEEE Trans. Inform. Theory, 51(4):1451-1468 (2005).

[13] C.H. Bennett, D.P. DiVincenzo, J.A. Smolin, W.K. Wootters, Phys. Rev. A 54(5):3824-3851 (1996).

[14] W. Dür, M. Hein, J.I. Cirac, H.-J. Briegel, Phys. Rev. A 72, 052326 (2005).

[15] C. Dankert, M.Math. Thesis, University of Waterloo, quant-ph/0512217 (2005).

[16] D. Gottesman, Ph.D. Thesis, quant-ph/9705052 (1997).

[17] M.A. Nielsen, Phys. Lett. A 303(4): 249-252 (2002).

[18] N. Boulant, T. F. Havel, M. A. Pravia, D. G. Cory, Phys. Rev. A 67, 042322 (2003).

[19] Y. Weinstein, T.F. Havel, J. Emerson, N. Boulant, M. Saraceno, D. Cory, J. Chem. Phys. 121, 6117 (2004).

[20] I. Chuang and M. Nielsen, J. Mod. Opt. 44, 2455 (1997); J. Altepeter et al., Phys. Rev. Lett. 90, 193601 (2003).

[21] The MUB unitaries are $U_b^{(a)} |k\rangle = \frac{1}{\sqrt{p}} \sum_l \omega^{al^2+bkl} |l\rangle$, with $a \in \{0, \dots p\}$ and $b \in \{0, \dots p-1\}$. The non-trivial abelian subgroups of the HW group are the lines $H = \{Z^\alpha : \alpha = 0, \dots, p-1\}$ and $G_i = \{(XZ^i)^\alpha\}$, which are diagonalized by the MUBs: H by the computational basis while G_i by the basis $|\psi_b^a\rangle = U_b^{(a)}|0\rangle$ with $i = 2a$. The $U_b^{(a)}$ allow to go from the computational basis to any of the $\{|\psi_b^a\rangle\}$, and thus generate all permutations between the abelian subgroups, H and G_i, of the HW group. The eigenvalues of the doubly stochastic matrix associated with random conjugation are $\{0, 1, (1/2 - 1/(2p)), (-1/2 - 1/(2p))\}$. Asymptotically for large p, the spectral gap governing the convergence to the uniform distribution on the normalizer is $1/2$.

[22] J. Emerson, E. Livine, and S. Lloyd, Phys. Rev. A 72, 060302(R) (2005).