Approximate Randomization of Quantum States With Fewer Bits of Key

Paul A. Dickinson *
University of Waterloo

Ashwin Nayak †
University of Waterloo, and
Perimeter Institute

June 1, 2006

Abstract

Randomization of quantum states is the quantum analogue of the classical one-time pad. We present an improved, efficient construction of an approximately randomizing map that uses $O(d/\epsilon^2)$ Pauli operators to map any d-dimensional state to a state that is within trace distance ϵ of the completely mixed state. Our bound is a $\log d$ factor smaller than that of Hayden, Leung, Shor, and Winter [7], and Ambainis and Smith [5].

Then, we show that a random sequence of essentially the same number of unitary operators, chosen from an appropriate set, with high probability form an approximately randomizing map for d-dimensional states. Finally, we discuss the optimality of these schemes via connections to different notions of pseudorandomness, and give a new lower bound for small ϵ.

1 Introduction

1.1 Encryption of quantum states

Randomization of quantum states is a procedure analogous to encryption of classical messages such as in the "one-time pad". Imagine that two parties wish to exchange sensitive data in the form of quantum states over an insecure quantum communication channel. They would like to encrypt the quantum data so that any eavesdropper with access to the channel will not gain any information about the data. The idea is to use a secret key, such as a uniformly random bit string, to transform a quantum state so that without access to the key, an adversary is unable to distinguish two different encrypted states, when averaged over the random choice of key. Equivalently, every state is mapped to the same mixed state by the encryption

*Department of Combinatorics and Optimization, and Institute for Quantum Computing, University of Waterloo. Address: 200 University Ave. W., Waterloo, Ontario N2L 3G1, Canada. Email: padickinson@iqc.ca. Research supported in part by an NSERC Canada Graduate Scholarship, and grants from NSERC Canada, CIAR, MITACS, CFI, and OIT (Canada).

†Department of Combinatorics and Optimization, and Institute for Quantum Computing, University of Waterloo, and Perimeter Institute for Theoretical Physics. Address: 200 University Ave. W., Waterloo, Ontario N2L 3G1, Canada. Email: anayak@math.uwaterloo.ca. Research supported in part by NSERC Canada, CIAR, MITACS, CFI, and OIT (Canada). Research at Perimeter Institute is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MEDT.

procedure. The encrypted state may then be sent over the insecure channel, and the receiver, who also knows the key may decrypt to recover the message.

It has been known for several years that applying an independently chosen random Pauli operator to each qubit of an n-qubit state maps it to the completely mixed state $\mathbb{I}/2^n$. This gives rise to a scheme for perfect encryption of n-qubit quantum states with $2n$ secret uniformly random classical bits [6, 4]. This was also shown to be optimal in terms of the number of bits of key required [6, 4, 8, 12].

The requirement of perfect encryption may be relaxed a little without compromising security, so that the encrypted states are all *close* to being completely mixed, rather than being exactly so. By using a probabilistic argument, Hayden, Leung, Shor, and Winter [7] showed that the number of bits of key required then drops by a factor of approximately 2: to approximately randomize n-qubit states to within ϵ of $\mathbb{I}/2^n$ (in trace norm), we need at most $n + \log n + 2\log\frac{1}{\epsilon} + \mathrm{O}(1)$ bits of key. Subsequently, Ambainis and Smith [5] gave an efficient (quadratic time) scheme for approximate state randomization with respect to the trace norm using

$$n + \min\left\{2\log n + 2\log\frac{1}{\epsilon}, \log n + 3\log\frac{1}{\epsilon}\right\} + \mathrm{O}(1)$$

bits of key. Their construction is based on *small-bias spaces* (see, e.g., Ref. [11]). They also showed how to reduce the key length to $n + 2\log\frac{1}{\epsilon}$ at the cost of increasing the length of the ciphertext by $2n$ bits.

The amount of key required for approximate encryption with respect to the Hilbert-Schmidt norm and the operator norm has been studied by Kerenidis and Nagaj [9]. They show that key length is quite sensitive to the norm chosen to specify the security requirement.

In this article, we revisit approximate randomization with respect to the trace norm, which reflects most closely our ability to physically distinguish quantum states. We first observe that an explicit scheme of Ambainis and Smith may be improved by using an optimal construction of small-bias spaces due to Alon, Goldreich, Håstad, and Peralta [3]. This reduces the key size to $n + 2\log\frac{1}{\epsilon} + 4$, and avoids the need for ciphertext that is longer than the original message. This construction avoids another rather subtle issue. The length-preserving schemes suggested in Ref. [5] require that the two communicating parties agree on a prime number of length $\Theta(n)$. Since there is no known polynomial-time deterministic procedure to generate a prime number of a specified length, additional communication is required to establish this shared prime number. (The prime may be generated locally by one party by a randomized procedure.) The encryption and decryption procedures we suggest require a common irreducible polynomial over GF(2) of degree $\Theta(n)$, which may be computed independently by the two parties using an efficient deterministic algorithm due to Shoup [15].

Next, we investigate collections of unitary operators that give rise to approximately randomizing maps. We show by a probabilistic argument that any sequence of

$$\mathrm{O}\left(\frac{d}{\epsilon^2}\log\frac{1}{\epsilon}\right)$$

unitary operators chosen independently from a perfectly randomizing set with high probability defines an approximately randomizing map for d-dimensional quantum states.

A simple rank argument shows that at least $d(1-\frac{\epsilon}{2})$ unitary operators are always needed, for approximate encryption in d dimensions. No better lower bound is known. Methods for showing lower bounds for perfect encryption all fail, since they crucially rely on the property of completely randomizing maps to destroy all quantum correlation between the encrypted state and any state previously entangled to it.

We take a different approach, and derive necessary conditions on distributions over Pauli operators that correspond to approximately randomizing maps. These conditions are similar to notions such as "almost n-wise independence" in the theory of pseudorandom distributions (see, e.g., Ref. [11]). As a corollary, we get a tighter lower bound on randomizing sets of Pauli operators in the regime of exponentially small ϵ.

We describe our results more formally in the following two subsections, and then prove them in the remaining sections.

1.2 Preliminaries

We refer the reader to the text [13] or the lecture notes [14] for definitions of basic concepts in quantum information.

Let $\mathrm{L}(\mathcal{H})$ denote the space of linear operators on the Hilbert space $\mathcal{H}$. This includes the cone of positive semi-definite operators (density operators) on $\mathcal{H}$. Let $\mathrm{U}(\mathcal{H})$ denote the set of unitary linear operators on the Hilbert space $\mathcal{H}$.

Definition 1.1 *Let $\epsilon \geq 0$. A completely positive, trace-preserving (CPTP) linear operator $R : \mathrm{L}(\mathbb{C}^d) \to \mathrm{L}(\mathbb{C}^d)$ is said to be ϵ-randomizing with respect to the norm $\|\cdot\|$ if, for all density operators (mixed states) $\rho \in \mathrm{L}(\mathbb{C}^d)$,*

$$\left\| R(\rho) - \frac{\mathbb{I}}{d} \right\| \quad \leq \quad \epsilon.$$

We say that R is completely randomizing if $\epsilon = 0$.

Remark 1.1 *Due to convexity, a map R that randomizes all pure states (rank 1 density operators) also randomizes all* mixed *states to the same extent.*

We will mainly discuss randomization with respect to the trace norm. For any linear operator $M \in \mathrm{L}(\mathbb{C}^d)$, the trace norm is defined by $\| M \|_{\mathrm{tr}} = \mathrm{Tr}\sqrt{M^\dagger M}$. Equivalently, it is the sum of the singular values of M, and therefore also referred to as the "1-norm". The trace norm is arguably a more appropriate measure of distinguishability in the context of eavesdropping, since it is directly related to information that measurements reveal about quantum states. We will also use the Frobenius (or Hilbert-Schmidt) norm in our proofs. This norm is defined as $\| M \|_F = \sqrt{\mathrm{Tr}(M^\dagger M)}$. Since this is the ℓ_2 norm of the vector of singular values of M, this is also referred to as the "2-norm" by some authors.

Randomizing maps are easy to construct. For example, the map $R : \rho \mapsto \mathrm{Tr}(\rho)\frac{\mathbb{I}}{d}$ is completely randomizing. However, these maps are most useful when they can be inverted by a quantum operation to recover the original state, as is required in the case of encryption.

The protocols for encryption we study involve two parties, labeled Alice and Bob, who share a secret, uniformly random bit-string, called the private key k. Alice wishes to send a d-dimensional quantum state ρ to Bob. She would like to apply an invertible quantum operation [12] to the state, and send it to Bob so that when averaged over k, the map is randomizing. This would ensure that no eavesdropper be able to distinguish two different messages with non-trivial probability. Such protocols have also been called "private quantum channels" by some authors (see, e.g., Ref. [4]). We have implicitly assumed that the quantum channel is noiseless unless an eavesdropper tampers with it. Therefore Bob, who also has the key k, can apply the inverse operation to decrypt the message ρ perfectly.

A natural way to create such an invertible randomizing operator is to select a sequence of unitary operators $U_1, \ldots, U_m$ and define

$$R(\rho) \quad = \quad \frac{1}{m} \sum_{i=1}^{m} U_i \rho U_i^\dagger. \tag{1}$$

Here, the index i corresponds to the shared secret key held by the communicating parties, and is unknown to any eavesdropper. With a suitable choice of unitary operators the map R would be ϵ-randomizing. In fact, any orthogonal set of unitary operations on $\mathbb{C}^d$, such as the set of d^2 Pauli operators, form a completely randomizing map [6].

The most general one-way encryption scheme may in addition involve an ancilla that depends upon the key [12]:

$$R(\rho) \quad = \quad \frac{1}{m} \sum_{i=1}^{m} U_i (\rho \otimes \sigma_i) U_i^\dagger.$$

This is slightly more general than the form claimed in Ref. [4]. However, the results in the latter article extend to the more general maps above (see also Ref. [8]). This general form of encryption uses more qubits in the ciphertext than originally present in the message, which is undesirable from an efficiency point of view. We will only study randomizing maps as in equation (1), which correspond to encryption without ancilla.

The randomizing maps we construct will involve the Pauli operators. We will denote the Pauli operators on a single qubit by $\mathrm{I}, \mathrm{X}, \mathrm{Y}, \mathrm{Z}$:

$$\mathrm{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathrm{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathrm{Z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathrm{Y} = \mathrm{iXZ} = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}.$$

These operators are unitary, Hermitian, and they square to the identity. The non-identity Pauli matrices anti-commute with other non-identity Pauli matrices. For example, $\mathrm{XY} = -\mathrm{YX}$. Where the overall phase of 'i' is irrelevant, we will substitute Y with the matrix XZ.

For two n-bit strings a, b, let $|a \wedge b| = \sum_{j=1}^{n} a_j b_j$. We will often represent a tensor product of n single qubit Pauli operators by a string of $2n$ bits $(a, b) \in \{0,1\}^{2n}$ using the correspondence

$$\begin{aligned} (a,b) \quad &\leftrightarrow \quad \mathrm{i}^{|a \wedge b|} \, \mathrm{X}^a \mathrm{Z}^b, \quad \text{where} \\ \mathrm{X}^a \quad &= \quad \mathrm{X}^{a_1} \otimes \mathrm{X}^{a_2} \otimes \cdots \otimes \mathrm{X}^{a_n}, \end{aligned} \tag{2}$$

and Z^b is defined similarly. Let P_n denote the set $\left\{ \mathrm{i}^{|a \wedge b|} \mathrm{X}^a \mathrm{Z}^b : (a,b) \in \{0,1\}^{2n} \right\}$ of all tensor products of n single qubit Pauli operators.

For two n-bit strings $a, b \in \{0,1\}^n$, considered as elements of $\mathrm{GF}(2)^n$, the standard scalar product is defined as $\langle u, x \rangle = \sum_i u_i x_i \pmod 2$. The symplectic inner product of a pair of $2n$-bit strings (a, b) and (c, d), considered as elements of $\mathrm{GF}(2)^{2n}$, is given by $\langle a, d \rangle + \langle b, c \rangle \pmod 2$. The symplectic inner product tells us when two Pauli operators commute: $\mathrm{X}^a \mathrm{Z}^b$ commutes with $\mathrm{X}^c \mathrm{Z}^d$ if and only if the symplectic inner product of (a, b) and (c, d) is 0.

A distribution p over $\{0,1\}^{2n}$ defines a CPTP map on n qubits via the above bijection:

$$R_p(\rho) \quad = \quad \sum_{(a,b) \in \{0,1\}^{2n}} p(a,b) \; \mathrm{X}^a \mathrm{Z}^b \rho \mathrm{Z}^b \mathrm{X}^a. \tag{3}$$

We will study randomizing maps of this form more closely. A special case is when p is uniformly distributed over a set $S \subset \{0,1\}^{2n}$. In this case, we will denote the associated CPTP map by R_S.

The single qubit Pauli operators P_1 form an orthogonal basis for $\mathrm{L}(\mathbb{C}^2)$ under the inner product $(A, B) = \mathrm{Tr}(B^\dagger A)$. The set P_n of 2^{2n} tensor products of n such Pauli operators similarly form an orthogonal basis for $\mathrm{L}(\mathbb{C}^{2^n})$. There are also bases of d^2 orthogonal unitary operators on $\mathrm{L}(\mathbb{C}^d)$ for general dimension d (that is not a power of 2).

We will also make use of the concept of a *stabilizer state* [13, Section 10.5.1, page 454]. A *stabilizer group* G is an abelian group generated by a subset $T \subset \mathrm{P}_n$ of the Pauli operators on n qubits. Each stabilizer group G defines a linear subspace C_G of $\mathbb{C}^{2^n}$ which is the common $+1$-eigenspace of all the Pauli operators in G. If G is generated by k independent Pauli operators, and does not contain $-\mathbb{I}$, then the linear subspace C_G has dimension 2^{n-k}. By a stabilizer state, we will mean a pure state which spans the one dimensional subspace C_G stabilized by a group G of order 2^n.

Every stabilizer group generated by k independent Pauli matrices may be specified by listing its generators row-wise in a $k \times 2n$ boolean matrix M via the bijection in equation (2). Since the generators all commute, different rows of the matrix have symplectic inner product 0 with each other. For a $2n$-bit vector $w = (u, v)$, let $M \cdot w$ denote the k-bit vector obtained by taking the symplectic inner product of the k rows of M with w.

1.3 Statement of Results

The problem we address in this paper is the construction of approximately randomizing maps which preserve the number of qubits in the message.

First, in Section 2, we describe an an explicit construction for a sequence of unitaries that approximately randomize. This construction combines the work of Refs. [5, 3, 15] to give an improvement over the explicit construction by Ambainis and Smith [5].

Theorem 1.2 *For any $\epsilon \in (0, 2]$, and dimension $d = 2^n$, there is a sequence of $m = \frac{16d}{\epsilon^2}$ unitary operations $\{U_i : 1 \leq i \leq m\}$, each a tensor product of Pauli operators, such that the map*

$$R(\rho) \quad = \quad \frac{1}{m} \sum_{i=1}^{m} U_i \rho U_i^\dagger$$

is ϵ-randomizing with respect to the trace norm. The sequence of Pauli operators defining U_i may be determined from the index i in time $\tilde{\mathrm{O}}((\log m)^4) = \tilde{\mathrm{O}}(n^4)$.

Remark 1.3 *The notation $\tilde{\mathrm{O}}(T)$ above suppresses factors poly-logarithmic in T. Since there is an linear-time completely randomizing map consisting of a sequence of d^2 unitary operators, the above theorem is only useful when $\epsilon > 4/\sqrt{d} = 4/2^{n/2}$. We were therefore able to assume that* $\log m = \log d + 2 \log \frac{1}{\epsilon} + \mathrm{O}(1) = \mathrm{O}(n)$.

Next, we study which sequences of unitary operations are suitable for approximate encryption. In Section 3, we prove that almost all sequences of $\mathrm{O}(\frac{d}{\epsilon^2} \ln \frac{1}{\epsilon})$ unitary operations form an ϵ approximately randomizing map for d dimensional states.

Theorem 1.4 *For all $\epsilon \in (0, 2]$, a random sequence of $m = \frac{37d}{\epsilon^2} \ln\left(\frac{15}{\epsilon}\right)$ unitary operations $\{U_i : 1 \leq i \leq m\}$ in $\mathrm{U}(\mathbb{C}^d)$ defines a map*

$$R(\rho) \quad = \quad \frac{1}{m} \sum_{i=1}^{m} U_i \rho U_i^\dagger$$

and R is ϵ-randomizing with respect to the trace norm, with probability at least $1 - \mathrm{e}^{-d/2}$. Each unitary operation U_i may be chosen independently from an arbitrary distribution over $\mathrm{U}(\mathbb{C}^d)$ that gives rise to a completely randomizing map.

In the above theorem, each unitary in the sequence may be chosen independently according to an arbitrary completely randomizing distribution of unitaries, not necessarily the same for each i. For instance, it may be chosen according to the Haar measure on $\mathrm{U}(\mathbb{C}^d)$, or the uniform distribution over any orthogonal unitary basis for $\mathrm{U}(\mathbb{C}^d)$. For the case that ρ is an n-qubit state, the unitary operators can be chosen from among the Pauli operators, which are particularly simple operators.

Theorem 1.4 is in general incomparable to Theorem II.2 of Hayden *et al.* [7]. Our theorem reduces by a factor of $\log d$ the number of unitaries required for approximate encryption in the trace norm. However, it does not imply the stronger bound of ϵ/d with respect to the spectral norm ("∞-norm") on the distance from the completely mixed state, even with a $\log d$ factor more unitaries.

We conjecture that the construction of the approximately randomizing map in Theorem 1.2 is optimal in the use of secret key bits, up to an additive constant. We are unable to establish this rigorously at present, but take some steps towards this.

We derive conditions on distributions over $\{0,1\}^{2n}$ that define randomizing maps. These conditions are similar in flavour to other notions of pseudo-randomness such as "almost k-wise independence". We believe these will help prove the optimality of our constructions.

Theorem 1.5 *Let R_p be a CPTP map on n qubits induced by a distribution p over $\{0,1\}^{2n}$, as in equation (3). Let V be the random variable corresponding to p. If R_p is an ϵ-randomizing map with respect to the trace norm, then the random variable $M \cdot V$ is ϵ-close to the uniform distribution over $\{0,1\}^n$ in ℓ_1 distance for every $n \times 2n$ matrix M over $\mathrm{GF}(2)$ that defines a stabilizer state.*

As a corollary, we prove that any distribution corresponding to an ϵ randomizing map (with respect to the trace norm) is necessarily ϵ-biased (cf. Definition 2.1 in Section 2). This implies a new lower bound on the number of bits of key in the regime of extremely small ϵ, when it is smaller than $2^{-n/2}$.

Corollary 1.6 *Let R_p be a CPTP map on n qubits induced by a distribution p over $\{0,1\}^{2n}$, as in equation (3). If R_p is an ϵ-randomizing map with respect to the trace norm, then the distribution p is ϵ-biased. Therefore, if p has support $S \subset \{0,1\}^{2n}$, then $|S|$ is at least a universal constant times*

$$\min\left\{2^{2n}, \quad \frac{n}{\epsilon^2 \log \frac{1}{\epsilon}}\right\}.$$

2 An explicit randomizing set

In this section, we prove Theorem 1.2. We describe an explicit sequence of unitary (Pauli) operators that are approximately randomizing. The i-th unitary in the sequence can be determined from the index i in polynomial time. To obtain this result, we use the connection made by Ambainis and Smith [5] between randomizing maps and *small-bias spaces*, together with a more efficient construction of such spaces due to Alon, Goldreich, Håstad, and Peralta [3].

Recall that the Pauli operators form an orthogonal basis for matrices, so we may express any density matrix over n qubits as

$$\begin{aligned} \rho &= \sum_{M \in \mathrm{P}_n} \frac{\mathrm{Tr}(M^\dagger \rho)}{\mathrm{Tr}(M^\dagger M)} \, M \\ &= \frac{1}{2^n} \sum_{M \in \mathrm{P}_n} \alpha_M \, M, \end{aligned}$$

where $\alpha = (\alpha_M)$ is a vector in $\mathbb{C}^{2n}$ with $\| \alpha \|_2^2 \leq 2^n$. The component $\alpha_{\mathbb{I}}/2^n$ of any quantum state along the identity operator is exactly $1/d = 1/2^n$. If a CPTP map E is completely randomizing, then

$$\begin{aligned} E(\rho) &= \frac{1}{2^n} \sum_{M \in \mathrm{P}_n} \alpha_M \, E(M) \\ &= \frac{1}{2^n} \mathrm{I}. \end{aligned}$$

Thus, the map annihilates all the non-identity components of the state. The idea behind the construction for approximate randomization is to construct a map that shrinks the non-identity components of a density matrix sufficiently, so that it becomes close to completely mixed. Such a map may be constructed from *small-bias* sets.

Definition 2.1 (Naor and Naor [11]) *The* bias *of a subset $S \subset \{0,1\}^k$ with respect to a string $u \in \{0,1\}^k$ is defined as*

$$\begin{aligned} \mathrm{bias}(S,u) &= \left| \mathrm{E}_{x \in S} \, (-1)^{\langle u,x \rangle} \right| \\ &= \left| 1 - 2 \, \mathrm{E}_{x \in S} \, \langle u,x \rangle \right|, \end{aligned}$$

where the expectation is taken over strings x chosen uniformly at random from S, and $\langle u,x \rangle = \sum_i u_i x_i \pmod 2$ is the standard scalar product over GF(2).

The subset $S \subset \{0,1\}^k$ is said to be δ-biased *if the bias with respect to every non-zero string is bounded by δ:* $\mathrm{bias}(S,u) \leq \delta$ *for all $u \in \{0,1\}^k - \{0^k\}$.*

This definition extends to arbitrary distributions p over $\{0,1\}^k$ in the natural way: $\mathrm{bias}(p,u) = \left| \mathrm{E} \, (-1)^{\langle u,X \rangle} \right|$, *where the random variable X is distributed according to p, and p is said to be δ-biased if its bias with respect to every non-zero string u is bounded by δ.*

The bias with respect to a string u is the bias of the XOR (exclusive OR) of the bits selected by the string u, i.e., the difference of the probabilities that this XOR is 0 or 1. The set of all strings has bias zero, and small-bias spaces are more efficient substitutes for this set.

Recall from equation (3) in Section 1.2 that a subset of strings $S \subset \{0,1\}^{2n}$ defines a CPTP map on n qubits as follows:

$$R_S(\rho) \quad = \quad \frac{1}{|S|} \sum_{(a,b)\in S} \mathrm{X}^a \mathrm{Z}^b \rho \mathrm{Z}^b \mathrm{X}^a. \tag{4}$$

If we choose $S = \{0,1\}^{2n}$, we get a completely randomizing map. Ambainis and Smith showed that if we choose S to be a δ-biased set, then the operator R_S scales every non-identity Pauli operator by a factor at most δ. We then get an ϵ-randomizing map by setting δ to be suitably small, namely, $\epsilon \cdot 2^{-n/2}$.

Proposition 2.1 (Ambainis and Smith [5]) *Let $S \subset \{0,1\}^{2n}$ be a set with bias at most $\epsilon/2^{n/2}$. Then the map R_S as defined in equation (4) is an ϵ-randomizing map with respect to the trace norm for n-qubit states.*

For completeness, we give a proof of this proposition in Appendix A.

We now use an optimal construction of δ-biased sets to get our randomizing map.

Proposition 2.2 (Alon, Goldreich, Håstad, Peralta [3]) *Let r, s be positive integers. There is a subset $S \subset \{0,1\}^{rs}$, of size 2^{2r}, with bias at most $\frac{s}{2^r}$. Given a monic irreducible polynomial of degree r over $\mathrm{GF}(2)$, and an index $1 \leq i \leq rs$, the i-th string in S may be constructed with $\mathrm{O}(rs)$ multiplications in $\mathrm{GF}(2^r)$, and a further $r^2 s$ bit operations.*

We describe this construction in Appendix B.

For our purposes, we need r, s such that the length of the strings is $2n$, and the bias of the set S is at most $\epsilon \cdot 2^{-n/2}$. In other words,

$$\begin{aligned} rs &= 2n, \\ \frac{s}{2^r} &\leq \epsilon \cdot 2^{-n/2}. \end{aligned}$$

Solving for the smallest such r, we get that the length of key $2r$ is at most

$$2r \quad \leq \quad \left\lceil n + 2\log\frac{1}{\epsilon} + 4 \right\rceil.$$

So a δ-biased set of size $m = 2^{2r} \leq 16 \cdot 2^n/\epsilon^2$ with $\delta \leq \epsilon \cdot 2^{-n/2}$ exists. This gives us an ϵ-randomizing map R_S with m unitary operations, corresponding to a key length of $2r$, as above.

Since a completely randomizing map exists with 2^{2n} unitaries, we may assume that $\epsilon \geq 2^{-n/2}$ in our construction. In other words, we may assume that $r \leq n$.

Given a key of length $2r$, and an irreducible polynomial of degree r over $\mathrm{GF}(2)$, the associated tensor product of single qubit Pauli operators may be computed with $\mathrm{O}(rs) = \mathrm{O}(n)$ multiplications in $\mathrm{GF}(2^r)$, and a further $\mathrm{O}(r^2 s) = \mathrm{O}(n^2)$ bit operations. Multiplication in $\mathrm{GF}(2^r)$ can be implemented with $\mathrm{O}(r \log r) = \mathrm{O}(n \log n)$ bit operations (see, e.g., Theorem 8.7 and its corollary on page 288, Chapter 8, in Ref. [1]). The bit-complexity of these computations is therefore $\mathrm{O}(n^2 \log n)$. Furthermore, a monic irreducible polynomial of degree r over $\mathrm{GF}(2)$ may be computed by a deterministic algorithm that takes $\tilde{\mathrm{O}}(r^4)$ bit operations [15, page 40, Theorem 3.6]. Thus, this part of the construction dominates the time complexity, which is in effect $\tilde{\mathrm{O}}(n^4)$. These observations conclude the proof of Theorem 1.2.

3 The abundance of randomizing maps

In this section, we prove Theorem 1.4, which states that there is a plethora of randomizing maps that use essentially the same number of bits of key as in the explicit construction. We use a probabilistic argument that is similar in structure to that of Hayden *et al.* [7]. To show that m unitaries suffice, we first show that a sequence of m random unitary operations approximately randomize any fixed state with high probability. To extend the approximate randomizing property to all states, we show that it suffices to randomize a set of *finitely many* pure states that in a certain precise sense approximately cover the unit sphere in $\mathbb{C}^d$. Finally, a "union bound" shows that with probability exponentially close to 1 *every* state is approximately randomized.

In our argument, each unitary operator is independently distributed according to the Haar measure, or any other distribution over unitary operations corresponding to a completely randomizing map. In particular, the operators could be chosen uniformly at random from an orthogonal basis for $\mathrm{L}(\mathbb{C}^{2^n})$, such as the Pauli basis P_n.

Proof: (of Theorem 1.4) Consider a sequence of m unitaries $\{U_i\}$ independently chosen from a measure μ_i on $\mathrm{U}(\mathbb{C}^d)$. We require that the measure μ_i give us a completely randomizing map. For any density matrix $\rho \in \mathrm{L}(\mathbb{C}^d)$, and U distributed according to μ_i,

$$\mathrm{E}_U \, U\rho U^\dagger \;=\; \int U\rho U^\dagger \, d\mu_i \;=\; \frac{\mathbb{I}}{d}. \tag{5}$$

The sequence $\{U_i\}$ define the map

$$R(\rho) \;=\; \frac{1}{m}\sum_{i=1}^{m} U_i \rho U_i^\dagger.$$

Fix a pure state $\rho \in L(\mathbb{C}^d)$. We first bound the expected distance of $R(\rho)$ from the completely mixed state $\mathbb{I}/d$. Define the random variable Y_ρ as follows

$$Y_\rho \;=\; \left\| R(\rho) - \frac{\mathbb{I}}{d} \right\|_{\mathrm{tr}}.$$

While we may carry out a similar analysis for a mixed state ρ, it is sufficient (and also simpler) to restrict ourselves to pure states; cf. Remark 1.1.

Proposition 3.1 $\mathrm{E}\, Y_\rho \;\leq\; \sqrt{d/m}.$

Proof: From Corollary A.2, we have

$$Y_\rho^2 \;\leq\; d \;\| R(\rho) \|_F^2 - 1. \tag{6}$$

By the definition of Frobenius norm,

$$\begin{aligned} \| R(\rho) \|_F^2 &= \mathrm{Tr}\; R(\rho)^2 \\ &= \frac{1}{m^2}\sum_i \mathrm{Tr}\left(U_i\rho U_i^\dagger\right)^2 + \frac{1}{m^2}\sum_{i\neq j} \mathrm{Tr}\left(U_i\rho U_i^\dagger U_j \rho U_j^\dagger\right) \\ &= \frac{1}{m} + \frac{1}{m^2}\sum_{i\neq j} \mathrm{Tr}\left(U_i\rho U_i^\dagger U_j \rho U_j^\dagger\right). \end{aligned} \tag{7}$$

Here, we have used the linearity of the trace function, and the fact that $\mathrm{Tr}(\sigma^2) = 1$ for any pure state density matrix σ.

Recall that the unitary operators U_i are chosen randomly according to a measure μ_i satisfying equation (5). Taking expectation over the random choice of unitaries, we get

$$\begin{aligned}
\mathrm{E}_{\{U_i\}}[\mathrm{Tr}\, R(\rho)^2] &= \frac{1}{m} + \frac{1}{m^2}\sum_{i\neq j} \mathrm{E}_{\{U_i\}} \mathrm{Tr}\left(U_i \rho U_i^\dagger U_j \rho U_j^\dagger\right) \\
&= \frac{1}{m} + \mathrm{Tr}\left[(\mathrm{E}_{U_i} U_i \rho U_i^\dagger)(\mathrm{E}_{U_j} U_j \rho U_j^\dagger)\right] \qquad (8) \\
&= \frac{1}{m} + \mathrm{Tr}\,\frac{\mathbb{I}}{d^2} \\
&= \frac{1}{m} + \frac{1}{d}. \qquad (9)
\end{aligned}$$

In equation (8), we used the fact that U_i and U_j are chosen independently according to measures μ_i, μ_j.

Putting equations (6) and (9) together gives us

$$\begin{aligned}
\mathrm{E}\, Y_\rho &\leq \sqrt{\mathrm{E}\, Y_\rho^2} \\
&\leq \sqrt{d \parallel R(\rho) \parallel_F^2 - 1} \\
&= \sqrt{d/m}.
\end{aligned}$$

the claimed bound on $\mathrm{E}\, Y_\rho$. ∎

Thus, the random sequence of unitary operators $\{U_i\}$ randomizes any *fixed* state ρ very well in expectation, provided m is chosen suitably larger than d.

We now note that the function $f_\rho(U_1, U_2, \ldots, U_m)$ defining the random variable Y_ρ has bounded difference. In other words, if we replace any one of the unitaries U_i by another unitary $\tilde{U}_i$, the function value changes by a small amount. Denote the randomizing map given by the modified sequence

$$(U_1, U_2, \ldots, U_{i-1}, \tilde{U}_i, U_{i+1}, \ldots, U_m)$$

by $\tilde{R}$. Then, we have

$$\begin{aligned}
&\left| f_\rho(U_1, U_2, \ldots, \tilde{U}_i, \ldots, U_m) - f_\rho(U_1, U_2, \ldots, U_m) \right| \qquad (10) \\
&= \left| \left\| R(\rho) - \frac{\mathbb{I}}{d} \right\|_{\mathrm{tr}} - \left\| \tilde{R}(\rho) - \frac{\mathbb{I}}{d} \right\|_{\mathrm{tr}} \right| \\
&\leq \left\| R(\rho) - \tilde{R}(\rho) \right\|_{\mathrm{tr}}, \qquad \text{By the triangle inequality} \\
&= \frac{1}{m} \left\| U_i \rho U_i^\dagger - \tilde{U}_i \rho \tilde{U}_i^\dagger \right\|_{\mathrm{tr}} \\
&\leq \frac{2}{m}. \qquad (11)
\end{aligned}$$

The McDiarmid bound from probability theory states that any random variable with such a bounded difference property is concentrated around its mean.

Theorem 3.2 (McDiarmid's Inequality [10]) *Let $X_1, X_2, \ldots, X_m$ be m independent random variables, with X_k taking values in a set A_k for each k. Suppose that the measurable function $f : \prod_{i=1}^m A_i \to \mathbb{R}$ satisfies*

$$\left|f(x) - f(x')\right| \quad \leq \quad c_k$$

whenever the vectors x and x' differ only in the k-th coordinate. Let $Y = f(X_1, X_2, \ldots, X_m)$ be the corresponding random variable. Then for any $t \geq 0$,

$$\Pr\left[Y - \mathrm{E}(Y) \geq t\right] \quad \leq \quad \exp\left(\frac{-2t^2}{\sum_{i=1}^m c_k^2}\right)$$

Theorem 3.2, along with equation (11) immediately implies that for any fixed pure state $\rho \in \mathbb{C}^d$,

$$\Pr[Y_\rho - \mathrm{E}\, Y_\rho \geq \delta] \quad \leq \quad \exp\left(\frac{-\delta^2 m}{2}\right).$$

This implies, using our bound from Proposition (3.1) on the expected value of Y_ρ,

$$\Pr[Y_\rho \geq \delta + \sqrt{d/m}] \quad \leq \quad \exp\left(\frac{-\delta^2 m}{2}\right). \tag{12}$$

The probability that $R(\rho)$ deviates from the completely mixed state decays exponentially in its distance, and the number of unitary operators m. We would like to extend this property to *all* pure states. For this, it suffices to randomize a suitably large, but finite, set of pure states (a "net") given by the following proposition (see, e.g. Ref. [7] for a proof).

Proposition 3.3 *For every $0 < \eta < 1$, there is a set $\mathcal{M}$ of pure states in $\mathbb{C}^d$ with $|\mathcal{M}| \leq (5/\eta)^{2d}$, such that for every pure state $|\phi\rangle \in \mathbb{C}^d$, there is a state $|\tilde{\phi}\rangle \in \mathcal{M}$ with $\left\| |\phi\rangle\langle\phi| - |\tilde{\phi}\rangle\langle\tilde{\phi}| \right\|_{\mathrm{tr}} \leq \eta$.*

From Proposition 3.3, we know that every pure state $\rho \in \mathbb{C}^d$ is η-close in trace norm to a pure state $\tilde{\rho}$ from a finite set $\mathcal{M}$ of size $|\mathcal{M}| \leq \left(\frac{5}{\eta}\right)^{2d}$. By the triangle inequality, and the unitary equivalence of the trace norm, it is straightforward to show that $|Y_\rho - Y_{\tilde{\rho}}| \leq \eta$. Therefore, if $Y_\rho \geq \epsilon$, then $Y_{\tilde{\rho}} \geq \epsilon - \eta$ for some $\tilde{\rho} \in \mathcal{M}$.

We can now bound the probability that the map R fails to randomize some pure state.

$$\begin{aligned}
&\Pr\left[\exists \rho : Y_\rho > \epsilon\right] \\
&\leq \quad \Pr\left[\exists \tilde{\rho} \in \mathcal{M} : Y_{\tilde{\rho}} > \epsilon - \eta\right] && \text{From the discussion above} \\
&\leq \quad |\mathcal{M}| \cdot \Pr\left[Y_{\tilde{\rho}} > \epsilon - \eta\right] && \text{By the union bound, for the worst case state } \tilde{\rho} \\
&\leq \quad \left(\frac{5}{\eta}\right)^{2d} \exp\left(\frac{-m}{2}(\epsilon - \eta - \sqrt{d/m}\,)^2\right) && \text{By equation (12)} \\
&\leq \quad \mathrm{e}^{-d/2},
\end{aligned}$$

if η is chosen to be at most $\epsilon/3$, and m at least

$$\frac{37d}{\epsilon^2} \ln\left(\frac{15}{\epsilon}\right).$$

Thus, there is an overwhelming majority of $m = \mathrm{O}(\frac{d}{\epsilon^2} \log \frac{1}{\epsilon})$ unitaries such that the corresponding map is randomizing to within ϵ, with respect to the trace norm. ■

4 Towards proving optimality

The best known lower bound for an ϵ-randomizing map R with respect to the trace norm, defined by a distribution over unitary maps on n qubits, is $n + \log(1 - \frac{\epsilon}{2})$. This follows directly from a rank argument: consider the image of a pure state. It has rank at most m, the number of unitary matrices defining R. The distance of any rank m density matrix from the completely mixed state $\mathbb{I}/2^n$ is at least $2(1 - m \cdot 2^{-n})$. Since R is ϵ-randomizing, this distance is at most ϵ, and the bound on the number of bits of key, which is $\log m$, follows.

The above lower bound does not reflect the amount of key required to achieve better security, as $\epsilon \to 0$. (At $\epsilon = 0$, the optimal number of key bits is $2n$.) To get stronger bounds, we focus on the simplest and perhaps most natural maps, those defined by distributions of Pauli operators, as in equation (3).

Recall that the n-qubit Pauli operators are in one-to-one correspondence with the set $\{0,1\}^{2n}$, and we may therefore study distributions on this set instead. We derive conditions on these distributions (stated in Theorem 1.5) which we believe will help prove the optimality of our constructions. As a corollary, we prove that any distribution corresponding to an ϵ randomizing map (with respect to the trace norm) is necessarily ϵ-biased (cf. Definition 2.1). This implies a new lower bound on the number of bits of key. The bound makes the strong dependence of key length on the parameter ϵ explicit, while sacrificing the strong dependence on message length n.

In Theorem 1.5, we stated constraints on distributions over Pauli matrices that are randomizing. We prove these constraints here by considering the action of randomizing maps on stabilizer states.

Proof: (of Theorem 1.5) Let $|\psi\rangle$ be an n-qubit (pure) stabilizer state, stabilized by a group whose n generators are given by the set T. We claim that for any Pauli operator P, the state $P|\psi\rangle$ is either parallel to $|\psi\rangle$ or orthogonal to it.

If P commutes with every Pauli operator in T, then $P|\psi\rangle$ is also stabilized by T: For $g \in T$, we have $gP|\psi\rangle = Pg|\psi\rangle = P|\psi\rangle$. Since the linear subspace stabilized by T is one-dimensional, $P|\psi\rangle$ belongs the linear span of $|\psi\rangle$. If P anticommutes with some $g \in T$, then $\langle\psi|P|\psi\rangle = \langle\psi|Pg|\psi\rangle = -\langle\psi|gP|\psi\rangle = -\langle\psi|P|\psi\rangle = 0$.

It follows that for any two Pauli operators P, Q, the states $P|\psi\rangle$ and $Q|\psi\rangle$ are either parallel or orthogonal—we use the matrix $PQ = P^\dagger Q$ in the above argument. In fact, we can say something stronger. Let M be the $n \times 2n$ matrix representation of the generator set T. The states $P|\psi\rangle$ and $Q|\psi\rangle$ are parallel iff $M \cdot w = M \cdot w'$, where w and w' are the $2n$-bit representations of the Pauli operators P and Q, respectively, and $M \cdot z$ is the vector of symplectic inner products of the rows of M with z. This is because

$$M \cdot w = M \cdot w' \quad \text{iff} \quad M \cdot (w + w') = 0,$$

which is equivalent to saying that PQ commutes with the stabilizer.

Let $|\psi_x\rangle$ be a canonical pure state in the linear span of $P|\psi\rangle$, where P is any Pauli matrix such that $M \cdot w = x$, and $w \in \{0,1\}^{2n}$ represents P. Since the n generators in T are independent, the matrix M has rank n. Therefore, the image of set $\{0,1\}^{2n}$ under M is all of $\{0,1\}^n$, the states $|\psi_x\rangle$ are well-defined as x ranges in $\{0,1\}^n$, and they form an orthonormal basis for $\mathbb{C}^{2^n}$.

Now consider a randomizing map R_p specified by a distribution p over $\{0,1\}^{2n}$, and its action on the

stabilizer state $|\psi\rangle$. We use ψ, ψ_x, etc. as shorthand for the density matrices $|\psi\rangle\langle\psi|, |\psi_x\rangle\langle\psi_x|$, etc.

$$\begin{aligned} R_p(\psi) &= \sum_{(a,b)\in\{0,1\}^{2n}} p(a,b)\ \mathrm{X}^a\mathrm{Z}^b\psi\mathrm{Z}^b\mathrm{X}^a \\ &= \sum_{(a,b)} p(a,b)\ \psi_{M\cdot(a,b)}. \end{aligned}$$

This mixed state is diagonal in the basis $\{|\psi_x\rangle\}$, and therefore its trace distance from the completely mixed state is

$$\left\| R_p(\psi) - \frac{\mathbb{I}}{2^n} \right\|_{\mathrm{tr}} = \sum_{x\in\{0,1\}^n} \left| \Pr[M\cdot V = x] - \frac{1}{2^n} \right|,$$

where V is the random variable corresponding to the distribution p. Since R_p is ϵ-randomizing, the above expression is bounded by ϵ. This is precisely the ℓ_1 distance of the random variable $M \cdot V$ from uniform on n-bits. ■

These conditions imposed by on distributions over Pauli matrices are similar to conditions such as "almost k-wise independence" (see, e.g., Ref. [11]), but are not equivalent to any of the standard notions of pseudo-randomness. As claimed in Corollary 1.6, it is however a stronger notion than that of having bias at most ϵ. We finish with a proof of this corollary, which also gives us a stronger lower bound for the key size for exponentially small ϵ.

Proof: (of Corollary 1.6) Consider any non-zero string $w \in \{0,1\}^{2n}$. Let $w = (u, v)$, where $u, v \in \{0,1\}^n$. We would like to show that the random bit $\langle w, V\rangle$ has bias at most ϵ, where V is the random variable corresponding to the distribution p.

We first prove this property for w such that for each $i = 1, \ldots, n$, at least one of u_i, v_i is 1. Consider n stabilizer generators, the i-th one g_i defined as $g_i = \bigotimes_{j=1}^n P_j$, where $P_j = \mathrm{I}$ for all $j \neq i$, and P_i is equal to

$$\begin{array}{ll} \mathrm{Z} & \text{if } u_i = 1 \neq v_i, \\ \mathrm{X} & \text{if } v_i = 1 \neq u_i, \text{ and} \\ \mathrm{Y} & \text{if } u_i = 1 = v_i. \end{array}$$

These n generators $\{g_i\}$ commute and are independent, and therefore specify a pure stabilizer state. This state is a tensor product of n single qubit Pauli eigenvectors,

$$|0\rangle, \quad \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad \text{or } \frac{1}{\sqrt{2}}(|0\rangle + \mathrm{i}|1\rangle),$$

depending upon whether the i-th generator g_i has Z, X, or Y, respectively, in its i-th tensor factor.

For $i = 1, \ldots, n$, let e_i be the n-bit string which is zero in all positions except the i-th. Then the $2n$-bit string representing the generator g_i is

$$g_i \quad \leftrightarrow \quad (\langle e_i, v\rangle\, e_i, \langle e_i, u\rangle\, e_i).$$

Consider the action of the map R_p on this stabilizer state. From Theorem 1.5, we get that the random variable $M \cdot V$ is ϵ close to uniform on n-bits, where M is the matrix representing the stabilizer $\{g_i\}$. Its rows are given by the equation (13). Note that $M \cdot V$ is the sequence of n bits $u_iV_i + v_iV_{n+i} \pmod 2$. Any

distribution that is ϵ close to uniform in ℓ_1-norm is also ϵ-biased. Therefore, the XOR of the bits in $M \cdot V$ has bias at most ϵ. The XOR is precisely the scalar product $\langle (u,v), V \rangle = \langle w, V \rangle$, so we have proven the first part of the claim for strings w of the type described above.

For an arbitrary non-zero string $w = (u,v)$, we consider a string $w' = (u',v)$ such that $u'_i = 1$ for all i such that $u_i = v_i = 0$, and $u'_i = u_i$ for the remaining i. From the argument above, we have that $M' \cdot V$ is close to uniform, where M' is defined by the string w'. The scalar product $\langle w, V \rangle$ is the XOR of a *subset* of the bits in $M' \cdot V$. Therefore its bias is also at most ϵ.

When p is uniform over a subset $S \subset \{0,1\}^{2n}$, we get that the set is ϵ-biased, and the stated lower bound on its size is given in Ref. [3, equation (3), page 13]. The same lower bound holds for possibly non-uniform distributions p with support on the subset S [2]. ∎

Acknowledgements

We thank Debbie Leung, Harold Ollivier, and David Woodruff for helpful discussions, Noga Alon for providing a copy of [2], and Alfred Menezes and Jeff Shallit for pointers to the literature.

References

[1] Alfred V. Aho, John E. Hopcroft, and Jeffrey D. Ullman. *The Design and Analysis of Computer Algorithms*. Addison-Wesley Series in Computer Science and Information. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1974.

[2] Noga Alon. On ϵ-biased random variables. Manuscript.

[3] Noga Alon, Oded Goldreich, Johan Håstad, and Réne Peralta. Simple construction of almost k-wise independent random variables. *Random Structures and Algorithms*, 3(3):289–304, 1992. Addendum in *Random Structures and Algorithms* 4 (1993), pages 119–120.

[4] Andris Ambainis, Michele Mosca, Alain Tapp, and Ronald de Wolf. Private quantum channels. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pages 547–553. IEEE Press, Los Alamitos, CA, USA, 2000.

[5] Andris Ambainis and Adam Smith. Small pseudo-random families of matrices: Derandomizing approximate quantum encryption. In *Proceedings of RANDOM*, pages 249–260, 2004.

[6] P. Oscar Boykin and Vwani Roychowdhury. Optimal encryption of quantum bits. *Physical Review A*, 67, 2003. Article no. 042317.

[7] Patrick Hayden, Debbie Leung, Peter W. Shor, and Andreas Winter. Randomizing quantum states: Constructions and applications. *Communications in Mathematical Physics*, 250(2):371–391, 2004.

[8] Rahul Jain. Resource requirements of private quantum channels. Pre-print quant-ph/0507075, ArXiv.org e-print Archive, http://www.arxiv.org/abs/quant-ph, July 2005.

[9] Iordanis Kerenidis and Daniel Nagaj. On the optimality of quantum encryption schemes. *Journal of Mathematical Physics*. To appear. Also Technical Report quant-ph/0509169, ArXiv.org Preprint Archive, `http://www.arxiv.org/abs/quant-ph/`, September 2005.

[10] Colin McDiarmid. On the method of bounded differences. *Surveys in Combinatorics*, 141:148–188, 1989. Invited papers at the Twelfth British Combinatorial Conference.

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

[12] Ashwin Nayak and Pranab Sen. Invertible quantum operations and perfect encryption of quantum states. *Quantum Information and Computation*. To appear. Also Technical Report quant-ph/0605041, ArXiv.org Preprint Archive, `http://www.arxiv.org/abs/quant-ph/`, May 2006.

[13] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, UK, 2000.

[14] John Preskill. Quantum computation. Lecture Notes, available at `http://www.theory.caltech.edu/people/preskill/ph229/`, California Institute of Technology, Pasadena, CA, 1998.

[15] Victor Shoup. *Removing Randomness from Computational Number Theory*. PhD thesis, University of Wisconsin, Madison, WI, 1989. Available at http://shoup.net/papers/.

A Proofs of some claims

In this section of the Appendix, we present proofs of some statements made in the article.

We use the following relation between trace norm and Frobenius norm, which is essentially an application of the Cauchy-Schwartz inequality.

Proposition A.1 *For any rank d matrix M, $\| M \|_{\mathrm{tr}} \leq \sqrt{d} \cdot \| M \|_F$.*

We use this relation in the following form.

Corollary A.2 *Let $M \in \mathrm{L}(\mathbb{C}^d)$ be a density matrix. Then, its trace distance from the completely mixed state $\mathbb{I}/d$ is bounded as*

$$\left\| M - \frac{\mathbb{I}}{d} \right\|_{\mathrm{tr}}^2 \quad \leq \quad d \| M \|_F^2 - 1.$$

Proof: By the definition of Frobenius norm in terms of the trace inner product,

$$\begin{aligned} & \left\| M - \frac{\mathbb{I}}{d} \right\|_F^2 \\ &= \| M \|_F^2 - 2 \operatorname{Tr} \frac{M}{d} + \left\| \frac{\mathbb{I}}{d} \right\|_F^2 \\ &= \| M \|_F^2 - \frac{2}{d} \operatorname{Tr} M + \operatorname{Tr} \frac{\mathbb{I}}{d^2} \\ &= \| M \|_F^2 - \frac{1}{d}. \end{aligned}$$

The corollary now follows from Proposition A.1. ∎

We can now prove Proposition 2.1.

Proof: (of Proposition 2.1) First, we express a state ρ in the Pauli basis:

$$\rho \quad = \quad \frac{1}{2^n} \sum_{(u,v)\in\{0,1\}^{2n}} \alpha_{uv}\, \mathrm{X}^u\mathrm{Z}^v,$$

where $\alpha = (\alpha_{uv}) \in \mathbb{C}^{2^{2n}}$ with $\| \alpha \|_2^2 \leq 2^n$.

Since X and Z anti-commute,

$$\begin{aligned} R_S(\mathrm{X}^u\mathrm{Z}^v) &= \frac{1}{|S|} \sum_{(a,b)\in S} \mathrm{X}^a\mathrm{Z}^b\, (\mathrm{X}^u\mathrm{Z}^v)\, \mathrm{Z}^b\mathrm{X}^a \\ &= \frac{1}{|S|} \sum_{(a,b)\in S} (-1)^{\langle u,b\rangle + \langle v,a\rangle}\, \mathrm{X}^u\mathrm{Z}^v \\ &= \delta_{v,u}\, \mathrm{X}^u\mathrm{Z}^v, \end{aligned}$$

where $\langle x, y\rangle$ is the standard scalar product of two strings over GF(2), and $\delta_{v,u} \in \mathbb{R}$ is given by the equation above. Note that $|\delta_{v,u}| = \mathrm{bias}(S,(v,u))$. Thus, if S is δ-biased, then each non-identity component of any density matrix will be scaled by a factor of δ:

$$\begin{aligned} R_S(\rho) &= \frac{1}{2^n} \sum_{(u,v)\in\{0,1\}^{2n}} \alpha_{uv}\, R_S(\mathrm{X}^u\mathrm{Z}^v) \\ &= \frac{1}{2^n} \sum_{(u,v)} \alpha_{uv}\, \delta_{v,u}\, \mathrm{X}^u\mathrm{Z}^v, \end{aligned}$$

where $|\delta_{v,u}| = \mathrm{bias}(S,(v,u)) \leq \delta = \epsilon/2^{n/2}$, for all $(v,u) \neq 0^{2n}$. The Frobenius norm of the randomized state is thus concentrated in the first term, the completely mixed state.

$$\begin{aligned} \| R_S(\rho) \|_F^2 &= \frac{1}{2^{2n}} \sum_{(u,v)} |\alpha_{uv}|^2 \cdot \delta_{v,u}^2 \cdot \| \mathrm{X}^u\mathrm{Z}^v \|_F^2 \\ &\leq \frac{1}{2^{2n}} \left(2^n + \sum_{(u,v)\neq 0^{2n}} |\alpha_{uv}|^2 \cdot \delta^2 \cdot 2^n \right) \\ &\leq \frac{1}{2^n}(1+\epsilon^2). \end{aligned}$$

Here, we used the bound of 2^n on $\| \alpha \|_2^2$.

From Corollary A.2,

$$\begin{aligned} \| R_S(\rho) \|_{\mathrm{tr}}^2 &\leq 2^n \| R_S(\rho) \|_F^2 - 1 \\ &\leq \epsilon^2, \end{aligned}$$

and Proposition 2.1 now follows. ■

B A construction of small bias sets

In this section, we present the construction described in Proposition 2.2 of small-bias spaces due to Alon, Goldreich, Håstad, and Peralta [3, Section 5] (see the remarks at the end of the Section 5 in the reference). This construction is optimal in the regime of extremely small biases that we are interested in.

Let r, s be positive integers. We would like to identify a set $S \subset \{0,1\}^{rs}$ of size 2^{2r} and with bias at most $s2^{-r}$. We construct S by describing a string s_{xy} for each pair of string $x, y \in \{0,1\}^r$. We identify both x and y with elements of the vector space $\mathrm{GF}(2^r)$ over the field $\mathrm{GF}(2)$ in the natural way. Let $\{e_i\}, i \in [r] = \{0, 1, \ldots, r-1\}$ be a basis for the vector space $\mathrm{GF}(2^r)$.

We define the string s_{xy} bit-by-bit. For $i \in [r], j \in [s]$, the (i,j)-th bit of s_{xy} is given by $\langle e_i x^j, y\rangle$. All multiplications in the expression $e_i x^j$ are in the field $GF(2^r)$, and $\langle \cdot, \cdot \rangle$ is the standard scalar product in $\mathrm{GF}(2)$. The string s_{xy} is thus given by the following array of bits:

$$\begin{array}{cccc} \langle e_0, y\rangle & \langle e_0 x, y\rangle & \cdots & \langle e_0 x^{s-1}, y\rangle \\ \langle e_1, y\rangle & \langle e_1 x, y\rangle & \cdots & \langle e_1 x^{s-1}, y\rangle \\ & & \vdots & \\ \langle e_{r-1}, y\rangle & \langle e_{r-1} x, y\rangle & \cdots & \langle e_{r-1} x^{s-1}, y\rangle \end{array}$$

Note that computing all the rs bits of s_{xy} takes $\mathrm{O}(rs)$ multiplications in $\mathrm{GF}(2^r)$, and a further $\mathrm{O}(r^2 s)$ bit operations to compute scalar products.

It only remains to argue that the bias of the set $S = \{s_{xy}\}$ so constructed has bias at most $s2^{-r}$.

Proposition B.1 *The set $S = \{s_{xy}\} \subset \{0,1\}^{rs}$ constructed as above has bias at most $\frac{s-1}{2^r}$.*

Proof: Let $u \in \{0,1\}^{rs}$ be any non-zero string. For any string $s_{xy} \in S$, we have

$$\begin{aligned} \langle u, s_{xy}\rangle &= \sum_{i\in[r], j\in[s]} u_{ij} \langle e_i x^j, y\rangle \\ &= \left\langle \sum_{ij} u_{ij}\, e_i x^j,\ y \right\rangle \\ &= \langle p_u(x), y\rangle, \end{aligned}$$

where

$$p_u(x) = \sum_{j\in[s]} \left(\sum_{i\in[r]} u_{ij} e_i \right) x^j$$

is a polynomial in x with coefficients in $\mathrm{GF}(2^r)$ and with degree at most $s-1$. Since u is non-zero, and $\{e_i\}$ are linearly independent, the polynomial p_u is not identically 0.

The bias of S with respect to u is then given by

$$\begin{aligned} \mathrm{bias}(S,u) &= |1 - 2\,\mathrm{E}_{xy}\langle u, s_{xy}\rangle| \\ &= \left|1 - 2 \Pr_{xy}[\langle u, s_{xy}\rangle = 1]\right| \\ &= \left|1 - 2 \Pr_{xy}[\langle p_u(x), y\rangle = 1]\right|. \end{aligned}$$

We may estimate the above probability as

$$\begin{aligned}\Pr_{xy}[\langle p_u(x), y\rangle = 1] &= \Pr_{xy}[\langle p_u(x), y\rangle = 1 \mid p_u(x) \neq 0] \cdot \Pr_x[p_u(x) \neq 0] \\ &= \frac{1}{2} \Pr_x[p_u(x) \neq 0],\end{aligned}$$

since the scalar product of any non-zero $p_u(x)$ with a uniformly random y has zero bias. Putting these together, we have

$$\begin{aligned}\mathrm{bias}(S, u) &= 1 - \Pr_x[p_u(x) \neq 0] \\ &= \Pr_x[p_u(x) = 0] \\ &\leq \frac{s-1}{2^r},\end{aligned}$$

since any non-zero polynomial of degree $s-1$ has at most $s-1$ roots in any field. ■

This finishes the description of the small bias set in Proposition 2.2.