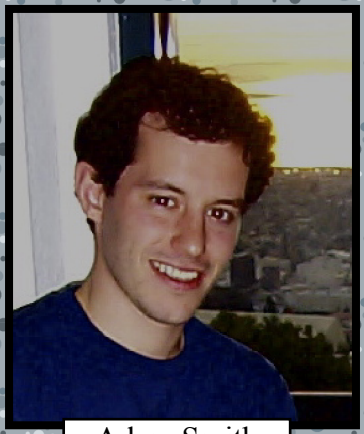
Andris Ambainis and Adam Smith

Small Pseudo-Random Families of Matrices: Derandomizing Approximate Quantum Encryption

Andris Ambainis



Adam Smith

1 Introduction

A *quantum encryption scheme* (or *private quantum channel*, or *state randomization protocol*) allows Alice, holding a *classical* key[3], to scramble a quantum state and send it to Bob (via a quantum channel) so that (1) Bob, given the key, can recover Alice's state exactly and (2) an adversary Eve who intercepts the ciphertext learns nothing about the message, as long as she doesn't know the key. We do not assume any shared quantum states between Alice and Bob, nor any back channels from Bob to Alice.[4]

There are two variants of this definition. An encryption scheme is called *perfect* if Eve learns zero information from the ciphertext, and *approximate* if Eve can learn some non-zero amount of information. A perfect encryption ensures that the distributions (density matrices) of ciphertexts corresponding to different messages are exactly identical, while an approximate scheme only requires that they be very close; we give formal definitions further below. In the classical case, both perfect and approximate encryption require keys of roughly the same length—n bits of key for n bits of message. In the quantum case, the situation is different.

For perfect encryption, Ambainis et al. [3] showed that $2n$ bits of key are necessary and sufficient to encrypt n qubits. The construction consists of applying two classical one-time pads—one in the "standard" basis $\{|0\rangle, |1\rangle\}$ and another in the "diagonal" basis $\{\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\}$.

Approximate encryption was studied by Hayden, Leung, Shor and Winters [8]. They introduced an additional, useful relaxation: they showed that if the plaintext is not entangled with Eve's system to begin with, then one can get *approximate* quantum encryption using only $n + o(n)$ bits of key—roughly half as many as are necessary for perfect encryption. The assumption that Eve's system is unentangled with the message is necessary for this result; otherwise roughly $2n$ bits are needed, even for approximate encryption. The assumption holds in the quantum counterpart of the one-time pad situation (one party prepares a quantum message and sends it to the second party, using the encryption scheme) as long as the message is not part of a larger cryptographic protocol.

Hayden et al. [8] showed that a *random* set of $2^{n+o(n)}$ unitary matrices leads to a good encryption scheme with high probability (to encrypt, Alice uses the key to choose one of the matrices from the set and applies the corresponding operator to her input). However, verifying that a particular set of matrices yields a good encryption scheme is not efficient; even writing down the list of matrices is prohibitive, since there are exponentially many of them.

The main tools in our construction are small-bias sets [10] of strings in $\{0,1\}^{2n}$. Such sets have proved useful in derandomizing algorithms [10], constructing short PCPs [6] and the encryption of high-entropy messages [12]. Thus, one of the contributions of this paper is a connection between classical combinatorial derandomization and constructions of pseudo-random matrix families in a continuous space. Specifically, we connect Fourier analysis over $\mathbb{C}^{\mathbb{Z}_2^{2n}}$ to Fourier analysis over the matrices $\mathbb{C}^{2^n \times 2^n}$. This parallels to some extent the connection between quantum error-correcting codes over n qubits and classical codes over $GF(4)^n$.

Definitions We assume that the reader is familiar with the basic notation of quantum computing (see [11] for an introduction). Syntactically, an approximate quantum encryption scheme is a set of 2^k invertible operators $\left\{E_\kappa | \kappa \in \{0,1\}^k\right\}$. The E_κ's may be unitary, but need not be: it is sufficient that one be able to recover the input ρ from the output $E_\kappa(\rho)$, which may live in a larger-dimensional space than ρ. Each E_κ takes n qubits as input and produces $n' \geq n$ qubits of output. If $n' = n$ then each operator E_κ corresponds to a unitary matrix U_κ, that is $E_\kappa(\rho) = U_\kappa \rho U_\kappa^\dagger$.

For an input density matrix[6] ρ, the density matrix of the ciphertext from the adversary's point of view is:

$$\mathcal{E}(\rho) = \mathbb{E}_\kappa\left[E_\kappa(\rho)\right] = \frac{1}{2^k} \sum_{\kappa \in \{0,1\}^k} E_\kappa(\rho)$$

When the scheme is length-preserving, this yields $\mathcal{E}(\rho) = \frac{1}{2^k} \sum_\kappa U_\kappa \rho U_\kappa^\dagger$.

Definition 1. *The set of operators* $\{E_\kappa\}$ *is an approximate quantum encryption scheme with leakage* ϵ *(also called "*ϵ*-randomizing scheme") for* n *qubits if*

$$\textit{for all density matrices } \rho \textit{ on } n \textit{ qubits:} \qquad D(\mathcal{E}(\rho), \tfrac{1}{2^{n'}}\mathbb{I}) = \left\| \mathcal{E}(\rho) - \tfrac{1}{2^{n'}}\mathbb{I} \right\|_{tr} \leq \epsilon. \tag{1}$$

Here $\mathbb{I}$ refers to the identity matrix in dimension $2^{n'}$, and $D(\cdot,\cdot)$ refers to the trace distance between density matrices. The trace norm of a matrix σ is the trace of the absolute value of σ (equivalently, the sum of the absolute values of the eigenvalues). The *trace distance* between two matrices ρ, σ is the trace norm of their difference:

$$D(\rho,\sigma) \triangleq \|\rho - \sigma\|_{tr} = \mathrm{Tr}(|\rho - \sigma|)$$

This distance plays the same role for quantum states that statistical difference plays for probability distributions: the maximum probability of distinguishing between two quantum states ρ, σ via a single measurement is $\frac{1}{2} + \frac{1}{4}D(\rho,\sigma)$. One can also measure leakage with respect to other norms; see below.

Remark 1. This definition of quantum encryption implicitly assumes that the message state ρ is not entangled with the adversary's system. Without that assumption the definition above is not sufficient, and it is *not* possible to get secure quantum encryption using $n(1+o(1))$ bits of key (roughly $2n$ bits are provably necessary[7]). Thus, this sort of construction is not universally applicable, and must be used with care.

Our Results We present three explicit, polynomial time constructions of approximate state randomization protocols for the trace norm. All are based on existing constructions of δ-biased sets [10, 2, 1], or on families of sets with small average bias. The three constructions are explained and proven secure in Sections 3.1, 3.2 and 3.3, respectively.

The first construction is length-preserving, and requires $n + 2\log n + 2\log(1/\epsilon) + O(1)$ bits of key, roughly matching the performance of the non-explicit construction. The second construction is length-increasing: it encodes n qubits into n qubits and $2n$ classical bits but uses a shorter key: only $n + 2\log(1/\epsilon)$ bits of key are required. Both of these constructions are quite simple, and are proven secure using the same Fourier-analytic technique.

The final construction has a more sophisticated proof, but allows for a length-preserving scheme with slightly better dependence on the number of qubits:

$$n + \min\{2\log n + 2\log(1/\epsilon), \log n + 3\log(1/\epsilon)\} + O(1)$$

bits of key. The right-hand term provides a better bound when $\epsilon > \frac{1}{n}$.

Randomization Schemes for Other Norms? Definition 1 measures leakage with respect to the trace norm on density matrices, $\|\cdot\|_{tr}$. This is good enough for encryption since the trace norm captures distinguishability of states. However, Hayden et al. [8] also considered randomization schemes which give guarantees with respect to a different norm, the operator norm.

A guarantee on the operator norm implies a guarantee for the trace norm, but schemes with the operator norm guarantee also have a host of less cryptographic applications, for example: constructing efficient quantum data hiding schemes in the LOCC (local operation and classical communication) model; exhibiting "locked" classical correlations in quantum states [8]; relaxed authentication of quantum states using few bits of key [9]; and transmitting quantum states over a classical channel using $n + o(n)$ bits of communication, rather than the usual $2n$ bits required for quantum teleportation [5].

More formally, for a $d \times d$ Hermitian matrix A with eigenvalues $\{\lambda_1, \lambda_2, ..., \lambda_d\}$, the *operator norm* (or ∞-norm) is the largest eigenvalue, $\|A\|_\infty = \max |\lambda_i|$, the Frobenius norm is the Euclidean length of the vector of eigenvalues, $\|A\|_2 = (\sum_i \lambda_i^2)^{1/2}$, and the trace norm is the sum of the absolute values of the eignvalues, $\|A\|_{tr} = \sum_i |\lambda_i|$. It is easy to see the chain of inequalities:

$$\|A\|_{tr} \leq \sqrt{d}\,\|A\|_2 \leq d\,\|A\|_\infty .$$

We can then state the condition for a map $\mathcal{E}$ to be ϵ-randomizing map for n qubits in three forms of increasing strength. For all input states ρ on n qubits:

$$\left\|\mathcal{E}(\rho) - \tfrac{1}{2^n}\mathbb{I}\right\|_{tr} \leq \epsilon; \qquad \left\|\mathcal{E}(\rho) - \tfrac{1}{2^n}\mathbb{I}\right\|_{2} \leq \epsilon/\sqrt{2^n}; \qquad \left\|\mathcal{E}(\rho) - \tfrac{1}{2^n}\mathbb{I}\right\|_{\infty} \leq \epsilon/2^n.$$

Our constructions satisfy the definition with respect to the Frobenius norm, but they are not known to satisfy the stronger operator-norm definition. This suggests two interesting questions. First, is it possible to prove that the other applications of state randomization schemes require only a guarantee on the Frobenius norm? Second, is it possible to design explicit (i.e. polynomial-time, deterministic) randomization schemes that give good guarantees with respect to the operator norm?

2 Preliminaries

Small-Bias Spaces The bias of a random variable A in $\{0,1\}^n$ with respect to a string $\alpha \in \{0,1\}^n$ is the distance from uniform of the bit $\alpha \odot A$, where $\odot$ refers to the standard dot product on $\mathbb{Z}_2^n$:

$$\hat{A}(\alpha) = \mathbb{E}_A\left[(-1)^{\alpha \odot A}\right] = 2\Pr[\alpha \odot A = 0] - 1.$$

The function $\hat{A}$ is the Fourier transform of the probability mass function of the distribution, taken over the group $\mathbb{Z}_2^n$.

The bias of a set $S \in \{0,1\}^n$ with respect to α is simply the bias of the uniform distribution over that set. A set S is called δ-biased if the absolute value of its bias is at most δ for all $\alpha \neq 0^n$.

Small-bias sets of size polynomial in n and $1/\delta$ were first constructed by Naor and Naor [10]. Alon, Bruck et al. (ABNNR , [1]) gave explicit (i.e. deterministic, polynomial-time) constructions of δ-biased sets in $\{0,1\}^n$ with size $O(n/\delta^3)$. Constructions with size $O(n^2/\delta^2)$ were provided by Alon, Goldreich, et al. (AGHP , [2]). The AGHP construction is better when $\delta = o(1/n)$. In both cases, the ith string in a set can be constructed in roughly n^2 time (regardless of δ).

One can sample a random point from a δ-biased space over $\{0,1\}^n$ using either $\log n + 3\log(1/\delta) + O(1)$ bits of randomness (usingABNNR) or using $2\log n + 2\log(1/\delta)$ bits (using AGHP).

Small-bias Set Families One can generalize small bias to *families* of sets (or random variables) by requiring that on average, the bias of a random set from the family with respect to every α is low [7]. Specifically, the expectation of the *squared* bias must be at most δ^2. Many results on δ-biased sets also hold for δ-biased families, which are easier to construct.

Definition 2. *A* family *of random variables (or sets)* $\{A_i\}_{i \in I}$ *is* δ*-biased if*

$$\mathbb{E}_{i \leftarrow I}\left[\hat{A}_i(\alpha)^2\right] \leq \delta^2 \textit{ for all } \alpha \neq 0^n.$$

Note that this is *not* equivalent, in general, to requiring that the expected bias be less than δ. There are two important special cases:

1. If S is a δ-biased set, then $\{S\}$ is a δ-biased set family with a single member;
2. A family of linear spaces $\{C_i\}_{i \in I}$ is δ-*biased* if no particular word is contained in the dual $C_i^{\perp}$ of a random space C_i from the family with high probability. Specifically:

$$\hat{C}_i(\alpha) = \begin{cases} 0 \text{ if } \alpha \notin C_i^{\perp} \\ 1 \text{ if } \alpha \in C_i^{\perp} \end{cases}$$

Hence a family of codes is δ-biased if and only if $\Pr_{i \leftarrow I}[\alpha \in C_i^{\perp}] \leq \delta^2$, for every $\alpha \neq 0^n$. Note that to meet the definition, for linear codes the expected bias must be at most δ^2, while for a single set the bias need only be δ.

One can get a good δ-biased family simply by taking $\{C_i\}$ to be the set of all linear spaces of dimension k. The probability that any fixed non-zero vector α lies in the dual of a random space is exactly $\delta^2 = \frac{2^{n-k}-1}{2^n-1}$, which is at most 2^{-k}.

One can save some randomness in the choice of the space using a standard pairwise independence construction. View $\{0,1\}^n$ as $GF(2^n)$, and let $K \subseteq GF(2^n)$ be an additive subgroup of size 2^k. For every non-zero string a, let the space C_a be given by all multiples $a\kappa$, where $\kappa \in K$, and the product is taken in $GF(2^n)$. The family $\{C_a \mid a \in GF(2^n), a \neq 0\}$ has the same bias as the set of all linear spaces ($\delta < 2^{-k/2}$). To see this, let $\{\kappa_1, ..., \kappa_k\}$ be a basis of K (over $GF(2)$). A string α is in $C_a^{\perp}$ if and only if $\alpha \odot (a\kappa_1) = \cdots = \alpha \odot (a\kappa_k) = 0$. This is a system of k linearly independent constraints on a, and so it is satisfied with probability $\delta^2 = 2^{-k}$ when $a \leftarrow GF(2^n)$, and even lower probability when we restrict a to be non-zero. Choosing a set C_a from the family requires n bits of randomness.

Entropy of Quantum States As with classical distributions, there are several ways to measure the entropy of a quantum density matrix. We'll use the analogue of collision entropy (a.k.a. Renyi entropy).

For a classical random variable A on $\{0,1\}^n$, the collision probability of two independent samples of X is $p_c = \sum_a \Pr[A = a]^2$. The Renyi entropy of A is $-\log p_c$.

For a quantum density matrixρ, the analogous quantity is$-\log \mathrm{Tr}(\rho^2)$. If the eigenvalues of ρ are $\{p_x\}$, then the eigenvalues of ρ^2 are $\{p_x^2\}$, and so $\mathrm{Tr}(\rho^2)$ is exactly the collision probability of the distribution obtained by measuring ρ in a basis of eigenvectors. $\sqrt{\mathrm{Tr}(\rho^2)}$ is called the Frobenius norm of ρ.

If ρ is the completely mixed state in d dimensions, $\rho = \frac{1}{d}\mathbb{I}$, then $\mathrm{Tr}(\rho^2)$ is $1/d$. The following fact states that any other density matrix for which this quantity is small must be very close to $\mathbb{I}$. The fact follows by applying the (standard) inequality $\mathrm{Tr}\ (|\Delta|)^2 \leq d\mathrm{Tr}(\Delta^2)$ to the Hermitian matrix $\Delta = \rho - \mathbb{I}/d$.

Fact 1. *If ρ is d-dimensional quantum state and Tr $(\rho^2) \leq \frac{1}{d}(1+\epsilon^2)$, then $D(\rho, \frac{1}{d}\mathbb{I}) \leq \epsilon$.*

Pauli matrices The 2×2 Pauli matrices are generated by the matrices:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The Pauli matrices are the four matrices $\{\mathbb{I}, X, Z, XZ\}$. These form a basis for the space of all 2×2 complex matrices. Since $XZ = -ZX$, and $Z^2 = X^2 = 1$, the set generated by X and Z is given by the Pauli matrices and their opposites: $\{\pm\mathbb{I}, \pm X, \pm Z, \pm XZ\}$.

If u and v are n-bit strings, we denote the corresponding tensor product of Pauli matrices by $X^u Z^v$. That is, if we write $u = (u1, ..., u_n)$ and $v = (v_1, ..., v_n)$, then

$$X^u Z^v = X^{u_1} Z^{v_1} \otimes \cdots \otimes X^{u_n} Z^{v_n}.$$

(The strings x and z indicate in which positions of the tensor product X and Z appear, respectively.) The set $\{X_u Z_v \mid u, v \in \{0,1\}^n\}$ forms a basis for the $2^n \times 2^n$ complex matrices. The main facts we will need are given below:

1. Products of Pauli matrices obey the group structure of $\{0,1\}^{2n}$ up to a minus sign. That is, $(X^u Z^v)(X^a Z^b) = (-1)^{a \odot v} X^{u \oplus a} Z^{v \oplus b}$.
2. Any pair of Pauli matrices either commutes or anti-commutes. Specifically, $(X^u Z^v)(X^a Z^b) = (-1)^{u \odot b + v \odot a} (X^a Z^b)(X^u Z^v)$.
3. The trace of $X^u Z^v$ is 0 if $(u, v) \neq 0^{2n}$ (and otherwise it is $\mathrm{Tr}(\mathbb{I}) = 2^n$).
4. $(X^u Z^v)^\dagger = Z^v X^u = (-1)^{u \odot v} X^u Z^v$

Pauli matrices and Fourier Analysis The Pauli matrices form a basis for the set of all $2^n \times 2^n$ matrices. Given a density matrix ρ, we can write $\rho = \sum_{u,v \in \{0,1\}^n} \alpha_{u,v} X^u Z^v$. This basis is orthonormal with respect to the inner product given by $\frac{1}{2^n}\mathrm{Tr}(A^\dagger B)$, where A, B are square matrices. That is, $\frac{1}{2^n}\mathrm{Tr}((X^u Z^v)^\dagger X^a Z^b) = \delta_{a,u}\delta_{b,v}$.

Thus, the usual arithmetic of orthogonal bases (and Fourier analysis) applies. One can immediately deduce certain properties of the coefficients $\alpha_{u,v}$ in the decomposition of a matrix ρ. First, we have the formula $\alpha_{u,v} = \frac{1}{2^n}\mathrm{Tr}(Z^v X^u \rho)$. Second, the squared norm of ρ is given by the squared norm of the coefficients, that is $\frac{1}{2^n}\mathrm{Tr}(\rho^\dagger \rho) = \sum_{u,v} |\alpha_{u,v}|^2$. Since ρ is a density matrix, it is Hermitian ($\rho^\dagger = \rho$). One can use this fact, and the formula for the coefficients $\alpha_{u,v}$, to get a compact formula for the Renyi entropy (or Frobenius norm) in terms of the decomposition in the Pauli basis:

$$\mathrm{Tr}(\rho^2) = \frac{1}{2^n} \sum_{u,v} |\mathrm{Tr}(X^u Z^v \rho)|^2.$$

3 State Randomization and Approximate Encryption

3.1 Encrypting with a Small-Bias Space

The ideal quantum one-time pad applies a random Pauli matrix to the input [3]. Consider instead a scheme which first chooses a $2n$-bit string from some set with small bias δ (we will set δ later to be $\epsilon 2^{-n/2}$). If the set of strings is B we have:

$$\mathcal{E}(\rho_0) = \frac{1}{|B|} \sum_{(a,b)\in B} X^a Z^b \rho_0 Z^b X^a = \mathbb{E}_{a,b}\left[X^a Z^b \rho_0 Z^b X^a\right]$$

That is, we choose the key from the set B, which consists of $2n$-bit strings. To encrypt, we view a $2n$-bit string as the concatenation (a, b) of two strings of n bits, and apply the corresponding Pauli matrix.

(The intuition comes from the proof that Cayley graphs based on ϵ-biased spaces are good expanders: applying a Pauli operator chosen from a δ-biased family of strings to ρ_0 will cause all the Fourier coefficients of ρ_0 to be reduced by a factor of δ, which implies that the "collision probability" (Frobenius norm) of ρ_0 also gets multiplied by δ. We expand on this intuition below.)

As a first step, we can try to see if a measurement given by a Pauli matrix $X^u Z^v$ can distinguish the resulting ciphertext from a totally mixed state. More explicitly, we perform a measurement which projects the ciphertext onto one of the two eigenspaces of the matrix $X^u Z^v$. We output the corresponding eigenvalue. (All Pauli matrices have two eigenvalues with eigenspaces of equal dimension. The eigenvalues are always either -1 and 1 or $-i$ and i.)

To see how well a particular Pauli matrix $X^u Z^v$ will do at distinguishing, it is sufficient to compute

$$|\mathrm{Tr}(X^u Z^v \mathcal{E}(\rho_0))|.$$

This is exactly the statistical difference between the Pauli measurement's outcome and a uniform random choice from the two eigenvalues. We can compute it explicitly:

$$\begin{aligned}
\mathrm{Tr}(X^u Z^v \mathcal{E}(\rho_0)) &= \mathrm{Tr}\left(X^u Z^v \mathbb{E}_{(a,b)\in B}\left[X^a Z^b \rho_0 Z^b X^a\right]\right) \\
&= \mathbb{E}_{a,b}\left[\mathrm{Tr}(X^u Z^v X^a Z^b \rho_0 Z^b X^a)\right] \\
&= \mathbb{E}_{a,b}\left[\mathrm{Tr}(Z^b X^a X^u Z^v X^a Z^b \rho_0)\right] \\
&= \mathbb{E}_{a,b}\left[(-1)^{a\odot v + b\odot u}\right] \mathrm{Tr}(X^u Z^v \rho_0)
\end{aligned}$$

Since $a \odot v + b \odot u$ is linear in the concatenated $2n$-bit vector (a, b), we can take advantage of the small bias of the set B to get a bound:

$$|\mathrm{Tr}(X^u Z^v \mathcal{E}(\rho_0))| \leq \delta |\mathrm{Tr}(X^u Z^v \rho_0)| \qquad \text{when } (u, v) \neq 0^{2n}$$

Equivalently: if we express ρ_0 in the basis of matrices $X^u Z^v$, then each coefficient shrinks by a factor of at least δ after encryption. We can now bound the distance from the identity by computing $\mathrm{Tr}(\mathcal{E}(\rho_0)^2)$:

$$\begin{aligned}\mathrm{Tr}(\mathcal{E}(\rho_0)^2) &= \frac{1}{2^n} \sum_{u,v} |\mathrm{Tr}(X^u Z^v \mathcal{E}(\rho_0))|^2 \\ &\leq \frac{1}{2^n} + \frac{\delta^2}{2^n} \sum_{(u,v) \neq 0^{2n}} |\mathrm{Tr}(X^u Z^v \rho_0)|^2 \leq \frac{1}{2^n}(1 + \delta^2 2^n \mathrm{Tr}(\rho_0^2))\end{aligned}$$

Setting $\delta = \epsilon 2^{-n/2}$, we get approximate encryption for all states (since $\mathrm{Tr}\ (\rho_0^2) \leq 1$). Using the constructions of AGHP [2] for small-bias spaces, we get a polynomial-time scheme that uses $n + 2\log n + 2\log(1/\epsilon)$ bits of key.

THE END