Concentration of measure effects in quantum information

Patrick Hayden

ABSTRACT. Most applications of quantum information require many qubits, which means that they must be described using state spaces of very high dimension. The geometry of such spaces is invariably simple but often surprising. Subspaces, in particular, can be interpreted as quantum error correcting codes and, when the dimension is high enough, random subspaces form remarkably good codes. This is because information stored in random subspaces gets encoded into highly entangled states. The entanglement properties of random subspaces also have other applications, such as making it possible to extend superdense coding from bits to qubits.

1. Introduction

In quantum information theory, we're fond of saying that Hilbert space is a big place, the implication being that there's room for the unexpected to occur. A number of results in quantum information theory derive from the initially counterintuitive geometry of high-dimensional vector spaces, where subspaces with nearly extremal properties are the norm rather than the exception. Randomly selected subspaces can be used, for example, to send quantum information through a noisy quantum channel at the highest known systematically achievable rate [**16, 24, 5**]. In another example, a randomly chosen subspace of a bipartite quantum system will likely contain nothing but nearly maximally entangled states, even if the subspace is nearly as large as the original system in qubit terms [**12**]. This observation makes it possible to invent a version of superdense coding in which each transmitted qubit somehow contains two qubits worth of quantum data [**9, 9**].

2. Quantum codes

That quantum computers could perform tasks like factoring large integers is surprising enough. That they can also in principle be made robust to noise is a small miracle. In a companion article, Daniel Gottesman provides an introduction to the mathematics of quantum error correction making this

robustness possible [8]. For example, the simultaneous $+1$ eigenspace of the four operators

$$
\begin{array}{ccccccccc}
X & \otimes & Z & \otimes & Z & \otimes & X & \otimes & I \\
I & \otimes & X & \otimes & Z & \otimes & Z & \otimes & X \\
X & \otimes & I & \otimes & X & \otimes & Z & \otimes & Z \\
Z & \otimes & X & \otimes & I & \otimes & X & \otimes & Z
\end{array}
$$

is two-dimensional and robust to arbitrary errors acting on any single qubit. This construction is obviously not haphazard: the four operators are selected using a procedure starting with a self-dual linear classical error-correcting code. While this mathematical structure is beautiful, it is also daunting. How might we go about designing optimal codes for specific error models?

Suppose that the noise is described by a quantum channel $\mathcal{N}$ taking $\mathcal{D}(A)$, the density operators of an input space A, to $\mathcal{D}(B)$, those of an output space B. (Mathematically, a quantum channel is a trace-preserving, completely positive linear map.) $\mathcal{N}$ could represent the effect of sending photons through an optical fiber, the relaxation of a quantum memory made of nuclear spins or any other corruption of quantum mechanical data. Now suppose that you can use this channel many times, either sequentially or in parallel. The resulting channel will have the form $\mathcal{N}^{\otimes k}$ for some large k, assuming that the uses of the channel are independent.

A quantum error correcting code for this situation will consist of an encoding channel $\mathcal{E} : \mathcal{D}(\mathbb{C}^d) \to \mathcal{D}(A^{\otimes k})$ and a recovery channel $\mathcal{R} : \mathcal{D}(B^{\otimes k}) \to \mathcal{D}(\mathbb{C}^d)$ such that $\mathcal{R} \circ \mathcal{N}^{\otimes k} \circ \mathcal{E}$ approximates the identity channel, meaning that it leaves the quantum data in $\mathbb{C}^d$ essentially invariant. (To design the best codes, it's essential to leave in that bit of wiggle room rather than requiring perfect reconstruction.) The rate of the code is defined to be $R = \frac{1}{k} \log_2 d$, meaning that each use of $\mathcal{N}$ allows R qubits to be transmitted. The quantum capacity $Q(\mathcal{N})$ of the channel $\mathcal{N}$ is then defined to be the supremum over rates R such that the quality of the approximation of the identity can be made arbitrarily good for sufficiently large k. The definition is quite insensitive to how one defines "quality of approximation" provided it is related to some meaningful measure of statistical distinguishability. The metric induced by the completely bounded trace norm is one well-motivated choice [14].

Evaluating $Q(\mathcal{N})$ for arbitrary channels is one of the biggest open problems in quantum information theory. Perhaps that isn't surprising given the daunting nature of the definition and the the combinatorial nature of the design the five qubit code introduced earlier. Nonetheless, an approach based on a combination of optimism and naivety proves to be remarkably successful, yielding the most efficient known codes for all but a very small number of channels. That approach is simply to encode $\mathbb{C}^d$ as a random subspace of $A^{\otimes k}$ and hope for the best. Of course, there is a bit of art involved in specifying the distribution over subspaces. A density operator $\rho \in \mathcal{D}(A)$ singles out a *typical subspace* of $A^{\otimes k}$ which contains nearly all the

support of $\rho^{\otimes k}$ [**22**]. For any given ρ, one can choose a code by choosing subspaces of the typical subspace at random according to the unitarily invariant measure. Provided the subspace isn't too big, it will be possible to design an appropriate recovery operation $\mathcal{R}$.

Taking many copies of $\mathcal{N}^{\otimes k}$ and using this random coding strategy washes away many of the detailed features of the channel $\mathcal{N}$ so that the rates achievable this way are just linear combinations of entropies. For a state σ, let $H(\sigma) = -\operatorname{Tr}\sigma\log_2\sigma$ be the von Neumann entropy of σ. This function is zero for pure states and maximal when σ is proportional to the identity or "maximally mixed". If σ^{AB} is a state in $\mathcal{D}(A\otimes B)$, write $H(AB)_\sigma$ for $H(\sigma^{AB})$ and $H(B)_\sigma$ for $H(\operatorname{Tr}_A\sigma^{AB})$.

THEOREM 2.1 (Lloyd-Shor-Devetak [**16, 24, 5**]). *Let $A' \cong A$ and $|\varphi\rangle \in A' \otimes A$. Then*

$$(1) \qquad\qquad Q(\mathcal{N}) \geq H(B)_\sigma - H(A'B)_\sigma$$

where $\sigma = (\mathrm{id}_{A'} \otimes \mathcal{N})(|\varphi\rangle\langle\varphi|)$.

(See [**11**] for a detailed proof along the lines described here.) Readers familiar with information theory will recognize $H(B)_\sigma - H(AB)_\sigma$ as the *negative* of conditional entropy, which is defined as the entropy of the whole minus the entropy of a part. For joint random variables, the conditional entropy is always nonnegative. For quantum states, however, a negative conditional entropy is a signature of the presence of entanglement. For example, consider the state $|\sigma\rangle^{A'B} = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle)$. Since the state is pure, $H(A'B)_\sigma = 0$. However, $H(B)_\sigma = 1$ because $\operatorname{Tr}_A\sigma^{AB}$ is maximally mixed. Therefore, $H(B)_\sigma - H(A'B)_\sigma = 1$, the maximum possible for a pair of qubits. The Lloyd-Shor-Devetak theorem states that the more negative conditional entropy a channel can produce, the more qubits can be sent through it using the random coding strategy.

So, random subspaces are pretty good quantum error correcting codes. Why should this stroke of good luck hold? In order for quantum information to be well-protected, the no-cloning theorem requires that none of it leak into the environmental degrees of freedom responsible for making the channel $\mathcal{N}$ noisy. From that perspective, it is necessary to design a code that hides the information from the environment. One good way to do this is to make sure that the states produced in the environment look random, a task for which it's hard to beat random subspaces. Another source of intuition is to notice that quantum error correcting codes succeed by delocalizing quantum information, encoding it in complicated multipartite correlations that are resistant to local perturbations. States in random subspaces generically contain nearly maximal amounts of nonlocal correlation as we'll see in the following sections.

3. Surprises in high dimension

Suppose, for the moment, that you are an astronaut orbiting Earth in a space shuttle. Imagine, slightly less plausibly, that you are also mathonaut, meaning that you observe not our Earth but instead a highly idealized version of it in which the population is evenly distributed over the whole surface of the planet. Bored with the daily routine of gyroscope failures and rebreather malfunctions, you decide to look out the window and count the number of people living within a hundred kilometer band of the equator. (You have both a good telescope and lots of time on your hands.) Give or take a few, you find fifty million people, with the rest of the population of six billion living elsewhere. Now, bold mathonaut that you are, you repeat your observations in higher and higher dimensions, first counting the inhabitants of a hundred kilometer thickening about the equator of a 3-sphere version of the Earth (in four dimensions), then of a 4-sphere and so on up. The number of equator dwellers increases steadily. Eventually, once the dimension gets large enough, you discover a great time saver: count the people *outside* of the band. There aren't any. Perplexed, you decide to check if your luck was bad by selecting other equators for the high-dimensional sphere, but each time you find that every single inhabitant of the planet lives within hundred kilometers of the equator.

What's going on? Nothing too sophisticated, it turns out. The calculation itself is elementary, an exercise in spherical coordinates, but the effect is an example of the broader "concentration of measure" phenomenon: naturally defined random variables on high-dimensional spaces tend to concentrate strongly around their average values [**15**]. The most familiar example of this is probably the case of the sum of n independent, bounded random variables. According to Chernoff's bound, the probability that the sum deviates more than ϵ from its mean value is less than $\exp(-Cn\epsilon^2)$ for some positive constant C. The analogous statement for functions on the k-sphere is known as Levy's lemma:

LEMMA 3.1 (Levy. See [**19**], Appendix IV, and [**15**]). *Let $f : \mathbb{S}^k \to \mathbb{R}$ be a function with Lipschitz constant η (with respect to the Euclidean norm) and a point $X \in \mathbb{S}^k$ be chosen uniformly at random. Then*

$$(2) \qquad \Pr\left\{ f(X) - \bar{f} \gtrless \pm \alpha \right\} \leq \exp\left(-C(k-1)\alpha^2/\eta^2\right)$$

for some constant $C > 0$.

Here $\bar{f}$ is used to denote either the mean value or a median for f; the median is actually a more natural quantity in the theory of concentration of measure. The function relevant to our mathonaut investigations is simply $f(x_1, \ldots, x_n) = x_1$, which obviously has Lipschitz constant one and both mean and median of zero.

4. Random states and random subspaces

Quantum states, of course, are represented as unit vectors, so Levy's lemma provides a ready-made tool for exploring the properties of random quantum states in high-dimensional systems. We need only choose the function f and plug in its mean value.

The example that will occupy us is the entanglement of a bipartite system. Earlier we saw that for a bipartite state $\sigma \in \mathcal{D}(A \otimes B)$, positivity of function $H(B)_\sigma - H(AB)_\sigma$ was a signature of the presence of entanglement in σ. If $\sigma = |\varphi\rangle\langle\varphi|$ is a pure state, however, $H(AB)_\varphi$ is zero and the function $H(B)_\varphi - H(AB)_\varphi$ reduces to $H(B)_\varphi = H(A)_\varphi$, a function known as the entropy of entanglement, which quantifies entanglement in units of *ebits*. In the asymptotic setting where one considers many copies of the pure state $|\varphi\rangle$, the entropy of entanglement is essentially the *unique* measure of entanglement in the sense that any pair of bipartite pure states can be interconverted using only local operations and classical communication, with the optimal interconversion rate given by the ratio of the states' entropy of entanglement [2].

Let $|\varphi\rangle$ be a random pure state in $A \otimes B$, chosen according to the unitarily invariant measure, which in turn corresponds to the uniform measure on the $(2d_A d_B - 1)$-sphere, where $\dim(X)$ is abbreviated as d_X. Assuming without loss of generality that $d_A \leq d_B$, the expected value of the entropy of entanglement is known [17, 20, 7, 21, 23] and satisfies

$$(3) \qquad \mathbb{E}H(A)_\varphi \geq \log_2 d_A - \frac{d_A}{2 \ln 2 d_B}.$$

Since the maximum value of $H(A)_\varphi$ is $\log_2 d_A$, any state of this bipartite system can have no more than $\log_2 d_A$ ebits of entanglement. (A state whose entropy of entanglement is maximal is said to be *maximally entangled*.) The inequality therefore asserts that on average the entanglement is within one ebit of being maximal. Levy's lemma allows us to quantify how likely it is that the entanglement of a random state will fall significantly below the mean. Define $\beta = \frac{1}{\ln 2}\frac{d_A}{d_B}$. Once all the calculations are done, we get the following bound:

$$(4) \qquad \Pr\left\{H(A)_\varphi < \log_2 d_A - \alpha - \beta\right\} \leq \exp\left(-\frac{(d_A d_B - 1)C\alpha^2}{(\log d_A)^2}\right),$$

for some $C > 0$ provided $d_B \geq d_A \geq 3$. Ignoring the small $(\log d_A)^2$ factor in the denominator of the exponent, this is the same type of exponential convergence to the mean that occurs for population evenly distributed on the k-sphere.

The convergence is so rapid, in fact, that it is possible to strengthen these results about random states into statements about random subspaces. The idea is to fix a subspace S_0 of dimension s and choose a very fine net of states $\mathcal{N}_0 \subset S_0$, so fine that given any state $|\varphi\rangle \in S_0$, there is an approximating $|\tilde{\varphi}\rangle \in \mathcal{N}_0$ such that $\|\varphi - \tilde{\varphi}\|_1 \leq \epsilon$. If we choose a random unitary U according

to the Haar measure, it takes S_0 to a random subspace US_0 and it takes the net $\mathcal{N}_0$ to a net $U\mathcal{N}_0$ for the new subspace. The probability that a given state in $U\mathcal{N}_0$ has entanglement less than $\log_2 d_A - \alpha - \beta$ is given by Equation (4) while the probability that any one of them has entanglement less than $\log_2 d_A - \alpha - \beta$ is then bounded above by

$$(5) \qquad\qquad |\mathcal{N}_0| \exp\left(-\frac{(d_A d_B - 1)C\alpha^2}{(\log d_A)^2}\right).$$

As a net on the unit ball of a subspace of real dimension $2s$, the size of $\mathcal{N}_0$ will scale as $(C/\epsilon)^{2s}$ for some constant $C > 0$. Proving the existence of a subspace in which all states are highly entangled then becomes a matter of tuning the resolution of the net $\mathcal{N}_0$ and the value of α. We find that when $d_B \geq d_A \geq 3$ and $0 < \alpha < 1$, there exists a subspace of $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$ of dimension

$$(6) \qquad\qquad \left\lfloor d_A d_B \frac{C\alpha^{2.5}}{(\log d_A)^3} \right\rfloor,$$

where $C > 0$ is, as always, a constant. From now on, a subspace having this property (for fixed α) will be referred to as a *maximally entangled subspace*. In qubit terms, in a bipartite system of n by $n + o(n)$ qubits, this is a subspace of size $2n - o(n)$ qubits in which *all* of the states of entanglement at least $n - o(1)$ ebits. The maximally entangled subspace is nearly as large as the whole system.

For the sake of unfair comparison, we could consider the subspace spanned by any two Bell states of a pair of qubits. Any such subspace will not only fail to contain only nearly maximally entangled states, it will always contain some product states!

5. Superdense coding of quantum states

Another way to place the existence of these maximally entangled subspaces in context is to study their applications to communication, which include s a surprising strengthening of a venerable result in quantum information theory is known as superdense coding [4].

A single qubit can carry at most one bit of information because storing k bits requires 2^k mutually orthogonal states and the state space of a qubit only contains room for two mutually orthogonal states. Suppose, however, that Bob would like to send Alice *two* bits and that they happen to share the entangled state $|\Phi_2\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. (Alice and Bob's roles are reversed from the usual convention in order to be consistent with the rest of the paper.) If Bob applies the unitary operators I, σ_x, σ_y and σ_z to his half of

$|\Phi_2\rangle$ then he produces, respectively, the states

$$
\begin{aligned}
|00\rangle &+ |11\rangle, \\
|01\rangle &+ |10\rangle, \\
|01\rangle &- |10\rangle \quad \text{and} \\
|00\rangle &- |11\rangle,
\end{aligned}
$$

ignoring the normalization and global phase. These states are all orthogonal, however, so if he were to send his half to Alice, she would be able to determine which of the four operators he had applied. Using this method, he can therefore send Alice two bits of classical information using one qubit of communication and one maximally entangled qubit pair (an ebit).

This observation can be summarized by the following schematic inequality:

$$1 \text{ qubit} + 1 \text{ ebit} \succeq 2 \text{ cbits}.$$

It is natural to ask whether it is possible, using the same resources, to communicate two *qubits* worth of quantum information rather than just two classical bits. A simple thought experiment reveals that this should not be the case. Indeed, if the schematic inequality

$$1 \text{ qubit} + 1 \text{ ebit} \succeq 2 \text{ qubits}$$

were true, then the two qubits communicated using just a qubit and an ebit could themselves be paired with two ebits, resulting in the communication of four qubits worth of quantum information. Repeating the process, an arbitrary amount of quantum information could be transmitted by sending just the single original qubit and a correspondingly large amount of entanglement. This is clearly impossible [**13**].

Suppose, however, that Bob has in mind a specific state $|\psi\rangle$ on a quantum system S that he would like to send to Alice. If S were a bipartite system $A \otimes B$ and $|\psi\rangle$ was promised to be maximally entangled, then Bob *could* take advantage of superdense coding: Alice and Bob would pre-share a fixed maximally entangled state and in order to send $|\psi\rangle$, Bob would apply a local unitary transformation V_ψ before sending his half of the system to Alice. Because all maximally entangled states are related by local unitary transformations by either Alice or Bob, such a V_ψ is always guaranteed to exist. To see this, consider a bipartite pure state $|\psi\rangle = \sum_{ij} c_{ij} |i\rangle|j\rangle$ and the associated matrix $C = (c_{ij})$ of coefficients. If $|\psi\rangle$ is transformed according to $(U \otimes V)|\psi\rangle$ then C gets mapped to UCV^T. By the singular value decomposition, there exist unitaries U and V such that UCV^T is diagonal and equal to $\mathrm{diag}(s_1, \ldots, s_{d_A})$ for some nonnegative s_j. These nonnegative numbers, sometimes known as the Schmidt coefficients of $|\psi\rangle$, are therefore the only invariants of a bipartite state under local unitary transformations; for maximally entangled states they are all equal to $d_A^{-1/2}$. On the other hand, the identity $(U \otimes V)|\psi\rangle = (I \otimes VU^T)|\psi\rangle$ holds for the particular maximally entangled state $|\psi\rangle = d_A^{-1/2} \sum_{j=1}^{d_A} |j\rangle|j\rangle$, which ensures that local unitary

transformations of the form $U \otimes V$ can always be applied to maximally entangled states by acting with a unitary transformation on one subsystem alone.

That's fine, of course, but the promise that $|\psi\rangle$ be maximally entangled would seem to make this a very special case, especially since Alice ends up with both halves of the bipartite system. Actually, thanks to the existence of a maximally entangled subspace, this is essentially the *general* case. If Alice and Bob pre-share a fixed maximally entangled state $|\Phi_d\rangle$ and agree on an embedding $U : S \hookrightarrow \mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$ of a maximally entangled subspace, then Bob can send to Alice any state $|\omega\rangle := U|\psi\rangle$ for $|\psi\rangle \in S$ using the simple protocol, up to small errors, since they are all nearly maximally entangled. To decode, Bob simply applies $U^\dagger$.

(7)
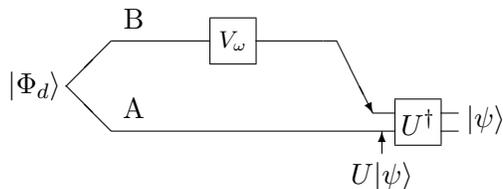
The qubit accounting then works as follows: Bob can send Alice an arbitrary $2n - o(n)$ qubit state by consuming n ebits of entanglement and sending $n + o(n)$ qubits, achieving the two-for-one savings normally associated with sending only classical information [**12, 9**].

The superdense coding idea can be pushed even further, to the case where the state to be prepared in Alice's lab is entangled with Bob's system and, therefore, no longer pure. A quick check of the extremal situation suggests that this should be easier: if the goal is prepare a fixed maximally entangled state between Alice and Bob's labs, then, provided Alice's system is no larger than Bob's, no communication is required at all; Bob need just perform an appropriate local unitary on his own system. The interpolation between the two-for-one of pure states and the no communication of maximally entangled states is analyzed in [**1**] using techniques similar to those discussed here, with the result that the largest Schmidt coefficient $\lambda_{\max}$ of all the states to be prepared controls the trade-off. To leading order in the asymptotics, $\frac{1}{2} \log_2 s + \frac{1}{2} \log_2 \lambda_{\max}$ qubits and $\frac{1}{2} \log_2 s - \frac{1}{2} \log_2 \lambda_{\max}$ ebits are required. s is defined as before to be the dimension of the system being prepared on Alice's side alone.

6. Consequences for mixed state entanglement measures

Mixed states in high dimensional state spaces have their own peculiarities. Unlike the pure state case, there is no privileged, unique measure of entanglement for mixed states. Instead, there are many different measures with the appropriate choice depending on the choice at hand. One of the

simplest is known as the entanglement of formation [**3**]:

$$(8) \qquad\qquad E_f(\rho) = \min \sum_j p_j H(A)_{\varphi_j}$$

where the minimization if over pure state decompositions $\rho = \sum_j p_j |\varphi_j\rangle\langle\varphi_j|$ of the bipartite mixed state ρ. This quantity is related to the number of ebits required to produce many copies of ρ, which is itself known as the entanglement cost [**10**].

Consider the maximally mixed state ρ on one of the maximally entangled subspaces. Because the range of ρ consists only of these states, any convex decomposition of ρ into pure states will again be into these nearly maximally entangled states. In an n by $n + o(n)$ qubit system, ρ will therefore have entanglement of formation

$$(9) \qquad\qquad E_f(\rho) = n - o(1),$$

which is nearly maximal. On the other hand, as the maximally mixed state on a subspace of $2n - o(n)$ qubits, ρ will have entropy at least $H(\rho) = 2n - o(n)$. In fact, the parameters can be tuned such that the quantum mutual information satisfies

$$(10) \qquad\qquad H(A)_\rho + H(B)_\rho - H(AB)_\rho = O(\log n).$$

This is an upper bound on essentially any sensible measure of the usable correlation present in ρ. In particular, the distillable entanglement of the ρ, the rate at which ebits can be extracted from ρ using local operations and classical communication, is therefore also $O(\log n)$ [**25**]. This leaves a huge gap between the entanglement of formation and the entanglement of distillation, the first being almost as large as it can be with the second simultaneously nearly as small as it can be. Ignoring potential discrepancies between the entanglement of formation and the entanglement cost, the state ρ provides an example of a state that is nearly as hard to make as a maximally entangled state and yet is nearly useless as a resource. In other words, this ρ would be an example of a state exhibiting near-maximal irreversibility under local operations and classical communication.

7. Multipartite entanglement

The results on bipartite entanglement extend easily to the multipartite realm. For convenience, consider a random state of n qudits, so that $|\varphi\rangle \in (\mathbb{C}^d)^{\otimes n}$ and assume that n is held fixed while d is allowed to increase. The following conclusions about random states are essentially corollaries of what we've already seen:

- The pure state entanglement across *every* bipartite cut is likely to be near maximal simultaneously.
- If $k > n/2$ then the reduced state of any k qudits will likely have near-maximal entanglement of formation. Meanwhile, if $k < n/2$

then it is likely that the entanglement of formation becomes less than any positive constant.

- With the participation of the remaining $n - 2$ parties, any pair of parties can distill a nearly maximally entangled pure state.

The last item is at first glance probably the most surprising but no harder to prove than the others. The distillation protocol consists of the remaining $n - 2$ parties each measuring in a random local basis. The state shared by the other two conditioned on the outcome of this measurement is essentially random and, therefore, nearly maximally entangled.

8. Conclusion

In retrospect it is no surprise that techniques for dealing with random subspaces should prove useful in quantum information theory. Random subspace techniques have been a mainstay of the "local theory of Banach spaces" ever since Milman [**18**] gave a proof of Dvoretsky's Theorem [**6**] using concentration of measure ideas. It is amusing and perhaps instructive to note that the title of a classic book by Milman and Schechtman on the subject, "Asymptotic theory of finite dimensional normed spaces," concisely sums up in mathematical terms one of the main goals of quantum information theory.

Acknowledgments

References

[1] A. Abeyesinghe, P. Hayden, G. Smith, and A. Winter. Optimal superdense coding of entangled states. arXiv:quant-ph/0407061.

[2] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher. Concentrating partial entanglement by local operations. *Phys. Rev. A*, 53:2046–2052, April 1996.

[3] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed state entanglement and quantum error correction. *Phys. Rev. A*, 54:3824–3851, 1996. arXiv:quant-ph/9604024.

[4] C. H. Bennett and S. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Phys. Rev. Lett.*, 69(20):2881–2884, 1992.

[5] Igor Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Trans. Inform. Theory*, 51(1):44–55, 2005. arXiv.org:quant-ph/0304127.

[6] A. Dvoretzky. Some results on convex bodies and Banach spaces. In *Proc. Symp. on Linear Spaces*, pages 123–160, Jerusalem, 1961.

[7] S. K. Foong and S. Kanno. Proof of Page's conjecture on the average entropy of a subsystem. *Phys. Rev. Lett.*, 72:1148–1151, 1994.

[8] D. Gottesman. Fault tolerant quantum computation. AMS Short Course Proceedings, 2009.

[9] A. Harrow, P. Hayden, and D. W. Leung. Superdense coding of quantum states. *Phys. Rev. Lett.*, 92:187901, 2004. arXiv:quant-ph/0307221.

[10] P. Hayden, M. Horodecki, and B. M. Terhal. The asymptotic entanglement cost of preparing a quantum state. *J. Phys. A*, 34(35):6891–6898, 2001. arXiv:quant-ph/0008134.

[11] P. Hayden, M. Horodecki, A. J. Winter, and J. T. Yard. A decoupling approach to the quantum capacity. *Information dynamics and open systems*, 15:7–19, 2008.

[12] P. Hayden, D. W. Leung, and A. Winter. Aspects of generic entanglement. arXiv:quant-ph/0407049.

[13] A. S. Holevo. Bounds for the quantity of information transmittable by a quantum communications channel. *Problemy peredači Informacii*, 9(3):3–11, 1973. English translation: A. S. Holevo, *Probl. Inf. Transm.* 9:177-183, 1973.

[14] D. Kretschmann and R. F. Werner. Tema con variazioni: quantum channel capacity. *New Journal of Physics*, 6:26–+, February 2004.

[15] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2001.

[16] S. Lloyd. Capacity of the noisy quantum channel. *Phys. Rev. A*, 55:1613, 1996.

[17] S. Lloyd and H. Pagels. Complexity as thermodynamic depth. *Ann. Phys.*, 188(1):186–213, 1988.

[18] V. D. Milman. A new proof of the theorem of a. dvoretzky on sections of convex bodies. *Funct. Anal. Appl.*, 5:28–37, 1971. (translated from Russian).

[19] V. D. Milman and G. Schechtman. *Asymptotic theory of finite dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, 1986.

[20] D. N. Page. Average entropy of a subsystem. *Phys. Rev. Lett.*, 71:1291, 1993.

[21] J. Sanchez-Ruiz. Simple proof of Page's conjecture on the average entropy of a subsystem. *Phys. Rev. E*, 52:5653, 1995.

[22] B. Schumacher. Quantum coding. *Phys. Rev. A*, 51:2738–2747, 1995.

[23] S. Sen. Average entropy of a quantum subsystem. *Phys. Rev. Lett.*, 77(1):1–3, 1996.

[24] P. W. Shor. Quantum error correction. http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1/.

[25] V. Vedral and M. B. Plenio. Entanglement measures and purification procedures. *Phys. Rev. A*, 57:1619–1633, 1998. arXiv:quant-ph/9702035.

SCHOOL OF COMPUTER SCIENCE, MCGILL UNIVERSITY, 3480 UNIVERSITY ST., MC-CONNELL ENGINEERING BUILDING, RM. 318, MONTREAL, QUEBEC, CANADA H3A 2A7