Classical capacities of a qubit and simulation of a qubit by a classical bit

Aidan Klobuchar
Budapest Semesters in Mathematics

February, 2011

1 Introduction

In Classical Information theory the smallest piece of information is the classical bit, which can take the values of either 0 or 1. That is, a classical bit has only two pure states. Two bits have 4 pure states, 3 bits have $2^3 = 8$ pure states and so on.

When talking about a quantum bit or **qubit**, one considers a two-level quantum system: a quantum system which is described by a Hilbert space $\mathcal{H}$ of dimension two. The state space of a quantum system with Hilbert space $\mathcal{H}$ can be identified with the space of **density operators** $S_1^+(\mathcal{H})$; that is, operators ρ acting on $\mathcal{H}$ having the properties

$$\rho \geq 0, \quad \mathrm{Tr}(\rho) = 1. \tag{1}$$

When $\dim(\mathcal{H}) = 2$, the convex body $S_1^+(\mathcal{H})$ is precisely a 3-dimensional ball and thus each of its border points is extremal. (In general — when $\dim(\mathcal{H}) > 2$ — the shape of $S_1^+(\mathcal{H})$ is much less understood, and cannot be so simply described as in the 2-dimensional case. It will not be simply a ball and not all of its border points will be extremal. Nevertheless, it will still have continuously many extremal points.) So in contrast to a classical bit which has only 2 pure states, a qubit has infinitely many. However, this does not neccessarily mean that we can store more (classical) information in a qubit than in classical one. The point is that though our qubit has infinitely many different pure states, it is impossible to distinguish these states *with*

certanity. This is a fundamental fact, and cannot be circumvented by some better measuring device.

Let us recapture the material disclussed in class in regard of this issue. Suppose we have some kind of device for distinguishing between two states. That is, we have a device such that

- whenever we put in our quantum system S, it either gives a "yes" or a "no" signal,
- the device gives a "yes" signal with prob. 1 if S was in the state described by ρ_1,
- the device gives a "yes" signal with prob. 0 if S was in the state described by ρ_2.

In what follows let us denote by $r(\rho)$ the probability of a "yes" signal when the system is in state ρ. Suppose we have 3 copies of the same system: one in state ρ and two in state $\tilde{\rho}$. We draw one and put it in our device. It seems clear that the device should signal "yes" with probability

$$\frac{1}{3}r(\rho) + \frac{2}{3}r(\tilde{\rho}). \tag{2}$$

In other words, r should **preserve the convex structure** of states:

$$r(t\rho + (1-t)\tilde{\rho}) = tr(\rho) + (1-t)r(\tilde{\rho}) \qquad (t \in [0,1]). \tag{3}$$

So when can we have a device which distinguishes between two states with certanity? We can now give a mathematical answer to this question.

Theorem 1.1. *For two density operators $\rho_1, \rho_2 \in S_1^+(\mathcal{H})$ there exists a real function r on $S_1^+(\mathcal{H})$ such that*

- *$r(\rho) \in [0,1]$ for all density operator ρ,*
- *r preserves the convex structure of $S_1^+(\mathcal{H})$,*
- *and $r(\rho_1) = 1$ while $r(\rho_2) = 0$*

if and only if $\rho_1\rho_2 = 0$; that is, if and only if the images of ρ_1 and ρ_2 are orthogonal.

For the moment we postpone the proof — in later sections, after introducing the concept of *positive operator valued measures*, it will be much easier to show the above theorem —- and note only that as a consequence, if $\dim(\mathcal{H}) = n$ then a collection of states in which any two can be distinguished with certainity can at most consist of n states. (A density operator's image cannot be the zero subspace, and in an n-dimensional space one can fit at most n pairwise orthogonal nonzero subspaces.) In particular, in case of a qubit — although we have infinitely many different pure states — we can distinguish with certanity between at most 2 states. So in this respect a single qubit performs very like a classical one.

Of course, it is well known that when our qubit is not *independent* from the rest of the world but is for example *entangled* to another qubit then the situation changes. In the process of *dense coding*, as was discussed in class, by sending a single qubit, Alice maneges to transmit 2 classical bits of information to Bob. However, in this case apart from the qubit sent by Alice, Bob also had another qubit which was *preveously entangled to the qubit sent by Alice*. So actually here we are dealing with a 2 qubit system and one cannot restrict its attention to the qubit sent by Alice, only. The "surprise" (the classically unexpected element) is not that Bob manages to read out 2 classical bits of information from the 2 qubit in his pocession, but the fact that Alice was in contact with *only one* of these qubits and yet Alice can code these 2 classical bits (later decoded by Bob) into the system.

So let us repeat: if our qubit is independent from anything which is within reach of Bob (or in general, within reach of the one who tries to read out information from this qubit), then Bob can distinguish with certanity between at most 2 states of the qubit in question. However, this does not make a single qubit and a single classical bit necessarily equivalent. Perhaps it is possible to distinguish between $n > 2$ states of the qubit not with certainty, but in a way that is — in some sense — "closer" to certanity than what we can have with a classical bit.

How to define 'closer to certanity' is an issue the following sections will describe. In general, we may view our qubit as a memory — or as it is often called in the literature: a channel — in which there is a ingoing and an outgoing information (someone chooses a certain state from a preveously fixed set of states and puts the qubit into the selected state, then passes it to another person who will have to try to determine: in which state the electron is). So one may investigate the issue from the point of view some kind of a *channel capacity*.

As we shall see, various capacity-like quantities of a qubit coincides with that of a classical bit. However, we shall then pose a (seemingly) more general question. Namely, what is the set of possible (classical, discrete) channels that one can realize with a single qubit: is it just the same as the ones realizable with a single classical bit? Or perhaps there exist channels realizible by a single qubit, which — although they have no more capacity then a single bit has — cannot be "simulated" by a classical bit?

We shall give the precize formulation of this question in a later section. Our conjecture is that in this respect a channel realized by a single quantum bit can always be simulated by (a mixture of) channels realizable by a classical single bit. Here we shall only prove this conjecture for channels whose output alphabet contains at most 3 letters. The statement will be generalized for the case when the transfer of an n-level system is used to realize a channel with $n+1$ possible outputs. However, at the moment this equivalence of quantum and classical realizability of a channel remains a conjecture when the number of outputs is more than $n+1$ (where n is the number of levels of the underlying system used to realize the channel).

2 "Money game" and a capacity concept

To model the amount of information a single classical or quantum bit can carry, we consider a channel which is realized by passing a single bit from a sender, Alice, to a receiver, Bob. We assume that the bit, when passed to Bob, is found in a state which is independent from the state of the rest of the world in reach of Bob.

One can then introduce various different quantities all trying to reflect the capacity of this channel to send useful information. Here, instead of the usually considered *Shannon capacity*, in order to familiarize with the concepts, we shall start with a somewhat simpler quantity. We shall introduce this quantity by considering a game involving a single bit.

In what follows, we shall refer to our game as the "Money game". In this game, a \$1 bill is put randomly and with equal probability into one of n boxes. Bob must pick one of the boxes and he gets what is inside that box. Now, to Alice it is revealed under which box the \$1 bill is. However, Alice cannot directly tell this to Bob (in which case Bob could always get the \$1 bill with certanity). Instead, she is allowed to pass to Bob a classical or a quantum bit whose state she can manipulate as she wishes. (That is, she is

allowed to pass a classical or a quantum bit of information.)

They may agree on some scheme before hand. For example, played with a classical bit, Alice and Bob can agree that the bit-value 0 will mean that the money is in box number 1 and the bit-value 1 will mean that the money is not in that box. Alternatively, they may agree that the bit-value 0 will mean that the money is in boxes 1 - $\lfloor \frac{n}{2} \rfloor$ and the bit-value 1 will mean that the money is in boxes $\lfloor \frac{n}{2} + 1 \rfloor$ - n.

The question then becomes, what is the expected value of the money Bob may win? This expected value (after a certain normalization) will be our measure of capacity.

In the classical case, the answer is straightforward. It is not too difficult to see that the ideal strategy is to have the measured value 0 correspond to half of the boxes and the measured value of 1 correspond to the other half of the boxes. Then with n defined as the number of boxes and $E[\$]$ as the expected value of the money won, we obtain

$$E[\$] = \text{ prob. of finding the bill} = \frac{1}{n/2} = \frac{2}{n}. \tag{4}$$

In the above, we have argued by using "common sense". To make things more rigorous and also to be able to proceed to more complex arguments, let us try now to formalize the basic concepts.

The chosen encoding scheme followed by Alice can be described by a $n \times 2$ matrix A. This matrix contains the probability values of Alice putting her bit into a particular state upon seeing that the money is in a particular box. That is, $A_{i,j}$ is the probability that upon seeing the money in the i-th box, Alice will pass the bit to Bob in state $j \in \{0, 1\}$. Naturally, all entry values need to be between 0 and 1 (since they are probability values) and the sum of the elements in each row must be 1 (given, that the money is in a particular box, Alice will *either* put her bit into state 0 *or* state 1: the sum of the respective probabilities must be 1).

The chosen decoding scheme followed by Bob can be described by a $2 \times n$ matrix B. This matrix contains the probability values of Bob picking a particular box upon receiving the bit from Alice in a particular state. That is, $B_{j,k}$ is the probability that upon receiving the bit in state $j \in \{0, 1\}$, Bob will pick box number k. Likewise to Alice's encoding matrix, also B must have its entry values between 0 and 1 and must have its rows sum to 1.

For the $n = 3$ case (there are 3 boxes) these encoding and decoding "tables" (matrices) would like this:

Table 1: Alice's encoding

	bit = 0	bit = 1
box nr 1	$p^A_{1\to 0}$	$p^A_{1\to 1}$
box nr 2	$p^A_{2\to 0}$	$p^A_{2\to 1}$
box nr 3	$p^A_{3\to 0}$	$p^A_{3\to 1}$

Table 2: Bob's decoding

	box nr 1	box nr 2	box nr 3
bit = 0	$p^B_{0\to 1}$	$p^B_{0\to 2}$	$p^B_{0\to 3}$
bit = 1	$p^B_{1\to 1}$	$p^B_{1\to 2}$	$p^B_{1\to 3}$

Let us now talk about the probability $p_{i\to k}$; that is, the probability of Bob picking box k given that the \$1 bill is under box k. Of course, to obtain its value, we must take account of both the encoding and the decoding probabilities:

$$p_{i\to k} = (p^A_{i\to 0})(p^B_{0\to k}) + (p^A_{i\to 1})(p^B_{1\to k}). \tag{5}$$

That is, the values $\{p_{i\to k}\}$ are obtained by multiplying Alice's encoding matrix with Bob's decoding matrix through standard matrix multiplication. We shall refer to the obtained matrix T of transitional probabilities as the **channel table** of the certain encoding-decoding scheme. This is an $n \times n$ matrix with nonnegative entries in which every row adds to 1.

Now, back on the topic of the maximum amount of money that can be won in the money game, it is clear that the expected value of the money won is given by

$$E(\$) = \frac{1}{n}\sum_j p_{j\to j} = \frac{1}{n}\mathrm{Tr}(T) = \frac{1}{n}\mathrm{Tr}(AB), \tag{6}$$

where T is the channel table, A is the encoding and B is the decoding matrix. Now, as every entry of A is less or equal than one and every entry of B is nonnegative, we have that

$$Tr(AB) = \sum_{k,l} A_{k,l}B_{l,k} \leq \sum_{k,l} B_{l,k} = 2, \tag{7}$$

as the sum of each row of B must be 1 and there are 2 rows in B. Thus

$$E[\$] = \frac{1}{n} Tr(AB) \leq \frac{2}{n}. \tag{8}$$

So classically, the money we can expect to obtain when a single \$1 bill is placed randomly into one of n boxes is at most $\$\frac{2}{n}$. The question then becomes, can this value be improved quantumly? We will answer this question in the following section.

3 The "money bound" on the qubit channel

Let us discuss the problem in a quantum framework using the Hilbert space $\mathcal{H}$ of the qubit. After learning the location of the money, Alice will put her qubit into a state given by a density operator. If the money is in box j, Alice will put her qubit into state $\rho_j \in S_1^+(\mathcal{H})$. That is, encoding is nothing else than the map $\{1, 2, \ldots, m\} \to S_1^+(\mathcal{H})$. Note that in the quantum description it seems we have avoided of considering probabilities. However, it only seems so: we did not assume ρ_j to be *pure*.

The qubit will be then passed to Bob's measuring device. Informally, this is a device with n lights, where the incoming qubit will trigger one of the lights to go off. Bob assumes that if light k goes off, then the state of the incoming qubit was ρ_k and that the money is in box k.

Formally, such a device is described by a function

$$\Phi : S_1^+(\mathcal{H}) \mapsto \{p_1, p_2, \ldots, p_n | p_k \geq 0 \ \forall k, \sum p_k = 1\}.$$

giving for each density operator (i.e. incoming state of our qubit) the probability values that that incoming state will trigger a specific light on the measuring device. As was already discussed in the introduction, for a function Φ to be realizable by some physical device, Φ must preserve the convex structure. That is, we must have

$$\Phi(t\rho + (1-t)\tilde{\rho}) = t\Phi(\rho) + (1-t)\Phi(\tilde{\rho}) \quad (t \in [0,1]). \tag{9}$$

Theorem 3.1. *$\rho \mapsto \Phi = (p_1(\rho), p_2(\rho), \ldots, p_n(\rho))$ preserves the convex structure if and only if it is of the form*

$$p_k(\rho) = \mathrm{Tr}(\rho E_k)$$

where $E_1, \ldots E_m$ are positive operators such that $\sum_k E_k = I$.

Proof. It is rather trivial to check the "if" part of the proof. Indeed, by linearity of trace, the given map preservex convex combinations, and since both ρ and E_k are positive operators, $\mathrm{Tr}(\rho E_k) \geq 0$ and moreover we have

$$\sum_k \mathrm{Tr}(\rho E_k) = \mathrm{Tr}(\rho \sum_k E_k) = \mathrm{Tr}(\rho I) = \mathrm{Tr}(\rho) = 1. \tag{10}$$

The somewhat more difficult is the "only if" part. Recall that for self-adjoint operators X, Y the formula

$$\langle X, Y \rangle = \mathrm{Tr}(XY) \tag{11}$$

defines a (real) scalar product. As p_k is assumed to be convex combination preserving, it extends to a (real) linear map from the full Euclidean space of self-adjoints $S(\mathcal{H})$ to $\mathbb{R}$. It follows that there exists a self-adjoint E_k such that

$$p_k(\cdot) = \langle \cdot, E_k \rangle = \mathrm{Tr}(\cdot E_k). \tag{12}$$

Since $\sum_k p_k = 1$, we further have that $\mathrm{Tr}(\rho \sum_k E_k) = 1$ for every density operator ρ. So

$$\langle \rho, (\sum_k E_k - I) \rangle = \mathrm{Tr}(\rho(\sum_k E_k - I)) = \mathrm{Tr}(\rho \sum_k E_k) - 1 = 0; \tag{13}$$

that is, $\sum_k E_k - I$ is orthogonal to every density operator. But the density operators span the full space, so $\sum_k E_k - I = 0$. All what remains then is to show the positivity of E_k. If $X \geq 0$ is a nonzero operator, then $\mathrm{Tr}(X) > 0$ and $\rho := X/\mathrm{Tr}(X)$ is a density operator. Hence

$$\mathrm{Tr}(XE_k) = \mathrm{Tr}(X)\mathrm{Tr}(\rho E_k) = \mathrm{Tr}(X)p_k(\rho) \geq 0 \tag{14}$$

and the positivity of E_k follows as the cone of positive operators is *self-dual*. □

The listed two properties regarding the operators $E_1, \ldots E_n$ describe what is known as a *Positive Operator Valued Measure* (POVM) and which may also be know as a *partition* or *resolution of identity*. The theorem mentioned (and not proved) in the introduction can be regarded as a corollary of the above theorem. Indeed, to show the statement made in the introduction, by what we have now, all we have to verify is that there exists a POVM E_1, E_2 such that

$$\mathrm{Tr}(\rho_1 E_1) = 0 \quad \text{and} \quad \mathrm{Tr}(\rho_2 E_1) = 0 \tag{15}$$

if and only if $\rho_2\rho_1 = 0$. Now for the only if part, since all operators involved are positive, from (15) it follows that $\rho_2 E_1 = 0$. Similarly, we have that $\rho_1 E_2 = 0$ since

$$\mathrm{Tr}(\rho_1 E_2) = \mathrm{Tr}(\rho_1(I - E_1)) = 1 - 1 = 0. \tag{16}$$

Then $\rho_2 = \rho_2 I = \rho_2(E_1 + E_2) = \rho_2 E_2$ and so

$$\rho_2\rho_1 = \rho_2 E_2 \rho_1 = \rho_2(\rho_1 E_2)^* = \rho_2 0 = 0. \tag{17}$$

Let us move on to the question of channel tables. As we have seen, encoding is a map $j \mapsto \rho_j \in S_1^+(\mathcal{H})$, whereas decoding is given by a POVM $(E_1, \ldots E_n)$. The channel table containing the transitional probabilities is nothing else than the matrix $(\mathrm{Tr}(\rho_j E_k))_{\{j,k\}}$. What can we say about the amount of money that can be won with a qubit in our money game?

The spectrum of a density operator ρ is always contained in the interval $[0, 1]$. Hence $\rho \leq I$ and $I - \rho$ is a positive operator and so if E is another positive operator then

$$\mathrm{Tr}((I - \rho)E) \geq 0 \;\Leftrightarrow\; \mathrm{Tr}(\rho E) \leq \mathrm{Tr}(E). \tag{18}$$

Thus all elements in the k-th column of the channel table are smaller or equal than $\mathrm{Tr}(E_k)$ and so in particular the expected value of the money won is smaller or equal than

$$\frac{1}{n}(\mathrm{Tr}(E_1) + \mathrm{Tr}(E_2) + \ldots + Tr(E_n)) = \frac{1}{n}\mathrm{Tr}(I) = \frac{1}{n}\dim(\mathcal{H}) = \frac{2}{n}, \tag{19}$$

since the dimension of the Hilbert space of a qubit is 2. Thus, a single quantum bit can win no more money in our little game than a classical bit. This amount of money that can be won is a form of channel capacity, as it gives an indication of the amount of information a bit may hold.

4 Shannon Channel Capacity

It would be interesting to see if a single classical and quantum bit share the same maximum capacity in the sense of the *Shannon Channel Capacity*. To look at a quantum system as classical channel, we need to fix an *encoding*; that is we need to fix a map $i \mapsto \rho_i$ from letter of the input alphabet

$\{a, b, ...\}$ to the set of density operators $S_1^+(\mathcal{H})$ (i.e. to the set of states of our quantum system). Decoding, from the mathematical point, is a convex structure preserving map from $S_1^+(\mathcal{H})$ to the output alphabet $\{\alpha, \beta, \ldots\}$ and as was discussed, is given by a POVM $\{E\}$. (From the physical point of view decoding is the actual device chosen by Bob, which picks up the sent quantum system and after examining it produces an output letter. To take account of a certain device, one then needs to specify how do the probabilities of the outcoming letters depend on the incoming state of the system; this is why we are considering decoding as the discussed map.) In our money game example,the input and output alphabets were identical, though this does not need to be the case.

The **Shannon channel capacity** is simply the maximum[1] amount of *mutual information* $I(\pi : \tilde{\pi})$ between the *coding probability distribution* $\{\pi\}$ and the probability distribution $\tilde{\pi}$ of the outcoming letter. Here

- the **coding probability distribution** is the list of probabilities that Alice will code a particular letter of the input alphabet (for example, the probability of Alice coding a is given by the value π_a — i.e. π_a describes how often a appears in Alice's messages)

- the **outcome probability distribution** $\tilde{\pi}$ is the list of probabilities that Bob will decode a particular leter of the output alphabet (for example, the probability of Bob will finally decode α is given by the value π_α.

The outcome probability distribution is determined by the *transitional probabilities* and the coding probability distribution π. The **transitional probability** $p_{i\to j}$ is the probability that if the input is set to i the output will be j. With a fixed coding $\{\rho\}$ and decoding $\{E\}$, as was discussed

$$p_{i\to j} = Tr(\rho_i E_j). \tag{20}$$

That is, what we called a channel table is merely the collection of these values. Knowing π and the values $\{p_{i\to j}|i, j\}$ the outcome distribution can be calculated as

$$\tilde{\pi}_j = \sum_i \eta_{i,j} \tag{21}$$

[1]In the finite case the existence of a maximum can be easily shown. In general however, one should be more careful and consider a *supremum* instead of a maximum.

Here $\{\eta_{i,j}|i,j\}$ is the *joint distribution* of the income and outcome:

$$\eta_{i,j} = \pi_i p_{i\to j} \tag{22}$$

is the probability that Alice will encode i and Bob will receive j. The mutual information $I(\pi : \tilde{\pi})$ is then

$$I(\pi : \tilde{\pi}) = H(\pi) + H(\tilde{\pi}) - H(\eta) \tag{23}$$

where $H(X)$ is the **entropy** of a probability distribution $X = (x_1, \ldots, x_n)$:

$$H(X) = -\sum_k x_k \log(x_k) \tag{24}$$

where the logarithms are traditionally taken base 2 (so that a single classical bit would turn out to have a channel capacity of 1 unit). Using that a probability distribution always adds to 1, and using the properties of the log function, by substitution one arrives to the following well-known formula:

$$I(\pi : \tilde{\pi}) = \sum_{i,j} \pi_i p_{i\to j} \log\left(\frac{p_{i\to j}}{\sum_k \pi_k p_{k\to j}}\right). \tag{25}$$

(Here i runs over the input alphabet, that is, the 'letters' which Alice can code in, and j runs over the output alphabet, or the different 'letters' which Bob's measuring device can read out.)

Now, suppose our channel relies on an n-level quantum system (that is, our density operators $\rho_1, \rho_2, \ldots$ and POVM are given on an n-dimensional Hilbert space). In this case then, what is the maximum value that the (classical) Shannon capacity C of the channel may be? By [1, Thm. 2.1] we have that

$$C \leq \sup_{\pi}\{H(\sum_k \pi_k \rho_k) - \sum_k \pi_k H(\rho_k)\} \tag{26}$$

where the supremum is taken over all probability distributions $\{\pi\}$ and $H(X) = \mathrm{Tr}(h(X))$ is the **von Neumann entropy** of a density operator X. Here h is the entropy function

$$h(x) = \begin{cases} -x\log(x), & \text{if } \mathrm{x} > 0 \\ 0, & \text{if } x = 0 \end{cases} \tag{27}$$

and $h(X)$ is defined *via* the spectral calculus. In other words, $H(X)$ is the (classical) entropy of the distribution of eigenvalues (taken with multiplicities) of the density operator X.

Since the von Neumann entropy of a density operator is nonnegative, we further have that

$$C \leq \sup_{\pi} H(\sum_{k} \pi_k \rho_k). \tag{28}$$

For any probability distribution $\{\pi\}$, the convex combination $\sum_k \pi_k \rho_k$ is a density operator. So we can further estimate the capacity by taking a supremum over the set of *all* density operators and hence

$$C \leq \sup_{\rho} H(\rho) = H((\frac{1}{n})I) = \log(n). \tag{29}$$

(It is well known that the entropy of a probability distribution is maximal if the distribution is uniform. That is, the highest von Neumann entropy is achieved when all eigenvalues of the density operator coincide; that is, when the density operator is a multiple of the identity.)

This upper bound indicates that the maximum channel capacity of a quantum channel is no greater than the maximum value of a classical channel, which is $\log(n)$. Note that the upper bound of $\log(n)$, on the other hand, is achievable. Indeed, let $\rho_1, \ldots, \rho_n$ be n 1-dimensional orthogonal projections summing to the identity. Then setting $E_j := \rho_j$ $(j = 1, \ldots, n)$ we have that $\{E_j\}_j$ is a POVM (actually it is more specifically a PVM: a projection valued measure). Using our choice of density operators and POVM, the channel table we obtain is simply the $n \times n$ identity matrix, since we have

$$\mathrm{Tr}(\rho_i E_j) = \mathrm{Tr}(\rho_i \rho_j) = \mathrm{Tr}(\delta_{i,j} \rho_j) = \delta_{i,j}. \tag{30}$$

Then further setting π to be the uniform distribution $(1/n, 1/n, \ldots, 1/n)$ we get that with our choices $I(\pi, I) = \log(n)$. Since the capacity C is obtained as a supremum, this shows that $C \geq \log(n)$. Together with the upper bound (29) this shows that in this case C is precisely $\log(n)$.

Note that the Shannon Channel Capacity and the "Money Capacity" reflect different ideas and it is easy to find cases where two schemes (channel tables) can have an equivalent Money / Shannon capacity and have a differing Shannon / Money capacity, respectively. Regardless, they are both criteria by which a single qubit and a single classical bit perform equivalently.

5 The set of channel tables

A channel table obtained by a classical encoding-decoding scheme (based on the transition of a classical bit) is the matrix product of an $n \times 2$ matrix

with a $2 \times n$ matrix. In particular, it must have a rank ≤ 2. Thus, the channel table which is a matrix of rank 3, cannot be obtained classically,

	Bob picks box nr 1	Bob picks box nr 2	Bob picks box nr 3
money is in box nr 1	2/3	1/6	1/6
money is in box nr 2	1/6	2/3	1/6
money is in box nr 3	1/6	1/6	2/3

even allowing for individual (but not common) sources of randomness (i.e. eventhough both encoding and decoding may contain random choices). On the other hand, it is easy to see that the above table can be obtained by passing a qubit (see an explanation of the example in the next section).

However, the situation changes, if we allow for a *common* source of randomness. Say for example that instead of individually tossing coins, both Alice and Bob take note of the actual whether. Their strategy may be something like "if it rains, we will do encoding A and decoding scheme B, if it doesn't rain, we will follow $\tilde{A}$ and $\tilde{B}$". If it then rains with a probability of 1/2, then this mixed strategy which uses a common source of randomness results the channel table

$$\frac{1}{2}AB + \frac{1}{2}\tilde{A}\tilde{B} = \frac{1}{2}T + \frac{1}{2}\tilde{T}. \tag{31}$$

That is, allowing for a common source of randomness means that we may consider convex combinations; hence the obtained set of channel tables is the *convex hull* of all those that can be obtained without a common source of randomness. For example, the previous table is simply the equal-weighted convex combination of the following three tables (each of which is realizable by a classical pure strategy). Naturally, the "money content" cannot be

1	0	0
0	1	0
0	1	0

0	1	0
0	1	0
1	0	0

1	0	0
1	0	0
0	0	1

increased by adding a common source of randomness. Let $m_j(T)$ be the maximum element of each column of a channel table realizable by a encoding-decoding scheme based on the transition of a quibit or a classical bit. Then, as was seen in general for the quantum case (note that the classical case can be considered as a subcase of the quantum one), we have that

$$c_M(T) := \sum_j m_j(T) \leq \text{ nr. of levels of the system sent from A to B} = 2 \tag{32}$$

and the expected value of the money won is

$$E(\$) = (1/n) \sum_j T_{j,j} \leq (1/n) c_M(T) \leq 2/n. \tag{33}$$

Now it is clear that m_j, and hence also c_M is a convex function:

$$m_j(tT + (1-t)\tilde{T}) \leq t m_j(T) + (1-t) m_j(\tilde{T}). \tag{34}$$

Thus a convex combination of channel tables with $c_M \leq 2$ can only result a channel table with $c_M \leq 2$. (In particular, we cannot win more money even if a common source of randomness is allowed.)

We have seen that a qubit cannot perform better in the money game than a classical one. However, one may ask:

I is it true that actually *all* channel tables realizable by a qubit can be realized by a mixture of classically realizable ones?

As was noted, if we did not allow a common source of randomness (i.e. if we considered the above question but without taking mixtures), the answer would be clearly a "no". One may also ask the more general question:

II is every channel table with $c_M \leq 2$ is realizable by a mixture of classically realizable ones?

By what was explained, if the answer to question II is yes, then so it is to question I. If on the other hand, there is a channel table with $c_M \leq 2$ which *cannot* be obtained by taking convex combination of classically realizable ones, then it is still possible that this table is realized by a quantum scheme. So in any case it seems essential to understand the answer to the second question, since even if the answer is no, at least we would give us an idea where to look for if we wanted to find counter-examples to question I.

Finally, let us pose one more question. We have seen, that it may happen that a channel table which can only be realized classically if we allow for a common source of randomness, may be realized by a quantum scheme *with no* common source of randomness. So here is another natural question:

III is allowing for a common random source increases the set of channel tables in the quantum case (or is the set of channel tables realizable by a qubit already convex)?

In the following sections we shall fully answer this third question and give a partial answer to the first two.

6 Geometric representation of channel tables

One open question raised by this investigation was whether or not, allowing for a common source of random variance, a classical bit and qubit had the same set of states that they could produce. To begin, we note that any channel table can be seen as a point in space with each table value as a single coordinate. In this way, each three by three channel table, for example, can be described as a single point in a nine dimensional space. As the channel tables represent transitional probability values, the entries must meet the following conditions

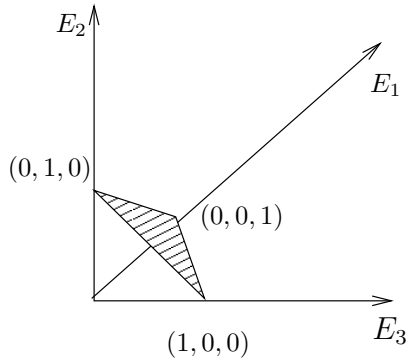
$$p_{e_k \to e_l} \in [0, 1] \tag{35}$$

$$\sum_l p_{e_k \to e_l} = 1, \tag{36}$$

which just describe the entries as probability values. In the three by three case, the first equation gives inequalities of the form $p_{e_1 \to e_2} \geq 0$ and the second equation gives a set of equalities. Continuing the geometric description of the channel tables, it should be noted that due to the encoding/decoding process the rows of the channel table, that is the complete set of transitional probabilities for some ρ_a can each be described in a specific geometrical manner. Because the decoding process must presere convex combinations, there must exist an affine map from the state space of the input system to the interior of the simplex created by the POVM representing the probability values $\{p_k\}$. With the input and output alphabets each fixed at three letters, for example, there are three output states and thus the simplex is a

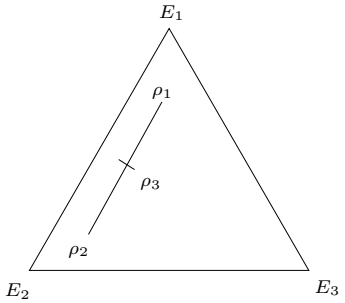
triangle. If we model this triangle (simplex) in three dimensional space and set the points $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$ as the vertices of the triangle then the coordinates of every point in the triangle represent its convex structure in relation to the output states. The value of each coordinate represents the probability that the represented incoming encoded state will trigger the output state represented by that particular coordinate. So if ρ_1 has the coordinates $(1,0,0)$, it means that when read by Bob's machine it will be read as E_1 100% of the time and as E_2 and E_3 0% of the time.

Figure 1: The Simplex in Three Dimensional Space



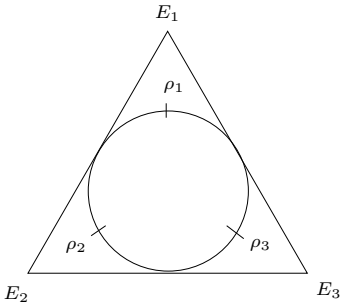
This means, when excluding a common random source, for a classical system we must fit a line into the triangle. As such, for three states to be classically realizable they must be represented by three collinear points in the triangle. Thus if only two states need to distinguished with certainty, they can be, by placing the state space (line) along one of the triangle's edges. However, with the addition of a third incoming state certainty can no longer be reached as the the third state must trigger the two states signaled by the first two states with some probability, and will not signal the remaining state as it must be on the triangle's edge.

Figure 2: The Graphical Representation of the Classical Case



In the case of the two level quantum system, the three dimensional state space is inserted inside the two dimensional simplex by an affine map, which by definition maps the sphere to an ellipse. This ellipse must be wholly contained by the simplex and the three chosen states must be wholly contained by the ellipse. As the ellipse can be affinely mapped into a one dimensional line, the two level quantum system can fully emulate the classical system. With the addition of a common source of random variance, convex combinations of the allowed states become possible, a fact which is difficult to describe geometrically using our model.

Figure 3: The Graphical Representation of the Quantum Case



This geometrical model can generalize to higher dimensions by changing the state space of the input system to match the appropiate dimension of the input alphabet and taking the simplex on the appropiate dimension representing the output alphabet.

Using these facts, it becomes easy to answer one of the questions posed in the introduction: whether or not adding a common source of random variance allows for more states or channel tables than is allowed with a quantum

system without a source of random variance. Of course, as the quantum system can model the classical system, then the quantum system allowing for a common random source can model the classical allowing for a common random source and if the classical system allowing for a common random source can model the quantum system fully then it can also model the quantum system allowing for a common random source, as the classical system could just model the quantum states making up the convex combination and then just emulate the convex combination. Back to the question, it is easy to see that the quantum system cannot consist of three states such that two are on the same edge of the simplex and the third is off the edge, as an ellipse can only be tangent to a line at one point, unless that ellipse is collapsed into a line. Thus if such a state can be described as a convex combination of classical states, then it shows that adding a common random source to a quantum system does allow for additional channel tables. One such table that a quantum system cannot acheive is:

	A	B	C
A	2/3	1/3	0
B	1/3	2/3	0
C	1/3	1/3	1/3

Represented geomtrically, this table becomes one point in the center of the simplex and two points evenly spaced along the E_1 - E_2 edge. Now this table can be modeled as a convex combination of classical states, specifically the states described by these three tables:

	A	B	C
A	1	0	0
B	0	1	0
C	0	1	0

	A	B	C
A	0	1	0
B	0	1	0
C	1	0	0

	A	B	C
A	1	0	0
B	1	0	0
C	0	0	1

This shows that adding a common source of random variance can in fact increase the number of possible channel tables for a two level quantum system.

7 A classical bit's ability to simulate the qubit

To begin to tackle the main problem it's best to review the previous results and use them as best as possible to create equations to describe the situation. Given the that the resultant equation regarding the Shannon Channel Capacity is a supremum of an equation involving logarithms, it is not very useful to use as a bound. Also, as discussed later in the paper, the Shannon Channel cpacity bound is easily shown to not clearly define the set of allowed channel tables. The result from the Money Channel capacity is useful, however. As described, the result was that the maximum expected value for the amount of money that could be won, in both the classical and the quantum case was $\$\frac{t}{n}$ where t was the level of the system and n was the number of boxes that had to be chosen from; essentially the size of the channel table. This result can be generalized by introducing the value m_k which is the maximum channel table value in the kth column. Then the money channel condition is given by

$$C_M = \sum_k m_k \leq t, \tag{37}$$

where we designated the sum of the maximum column elements as C_M. This value, C_M is our new money game channel capacity. This equation imposes a strict condition on the entries of the channel table. It requires that any sum of the table elements containing exactly one member from each column must be less than t, the level of the system. The set of channel tables meeting this condition is a convex compact set and is described geometrically as a polytope, as all of the constraining inequalities are linear and the set is clearly bounded, as every coordinate at the very least must be between one and zero.

The question then becomes, does this condition completely describe the body of allowed channel tables or are there channel tables which fit this condition but which are not realizable classically or quauntumly? The easiest way to look at this problem is to compare the polytope described by the bound of C_M, to the body of classically allowed channel tables. If the body described by the equations is equivalent to the classical body, it is a proof that the classical bit and quantum bit produce the same set of channel tables.

Unlike the body of quantumly allowed tables, the body of classical tables is easy to define based on its extremal states. As you will recall, the extremal states are those which cannot be formed as a convex combination of other states, thus a full description of the extremal states describes the body, as

the body is just the convex hull of the extremal states. To be more specific, the body of classically allowed tables is also a polytope and the extremal states are its vertices. So it is possible for the body described by the C_M bound and the body of classical tables to coincide. Now the extremal states of the classical body are easy to describe and are merely all of the channel tables composed solely of ones and zeroes with a single one in each row. So if we can show that the extremal states of each body exactly coincide, that is, that all of extremal states described by C_M bound are composed solely of ones and zeroes, we have a proof that the classical bit can simulate the quantum bit.

The proof that the money game condition fully describes the body of allowed classical and quantum channel tables is not currently complete in full generality. It is however, complete for the case of a t level system and a $t+1 \times t+1$ channel table. The case of the two level system and the 3×3 channel table described earlier is an example of such a case. We want to show that there does not exist any extremal table satisfying the C_M bound that is not composed solely of zeroes and one. We will call an entry that is a zero or one a *border* entry, as, in a geometrical sense, the described coordinate places the channel table along the edge of the polytope. We assume that in all further theorems that

$$\#\{e_k \to e_l | p_{e_k \to e_l}\} \neq 0, 1 > 0, \tag{38}$$

which just says that the table contains at least one non border point. Of course, as the sum of the elements of a row must add to one, there must be at least one additional non-border element in the same row. So we have that

$$\#\{e_k \to e_l | p_{e_k \to e_l}\} \neq 0, 1 \geq 2. \tag{39}$$

With this condition in mind, we begin with the following theorem:

Theorem 4.1: If $C_M(T) < t$, and condition (39) holds, then T is not extremal.

Proof: To show that T is note extremal, we will create a T' and a T'' such that

$$(\frac{1}{2})T' + (\frac{1}{2})T'' = T. \tag{40}$$

We merely have to take the two non-border entries, which we'll designate as x and y for brevity and respectively add and subtract some ϵ from them. That is, all the elements of T, T', and T'' are equivalent except for x and y. T' contains $x + \epsilon$ and $y - \epsilon$ while T'' contains $x - \epsilon$ and $y + \epsilon$. Thus the sum of the affected row is still 1 in every case. Also, C_M changes by at most $+\epsilon$, and is thus still strictly below t. This shows that T cannot be extremal, as it is a convex combination of T' and T'', both realizable tables. Thus an extremal table must meet the condition

$$C_M = t. \tag{41}$$

We will call this process of iterating the channel table up and down by some ϵ 'roaming', as in a geometrical sense, this is just a check to see if the point represented by the channel table can be moved in some direction and in the opposite of that direction. Only vertices, or extremal points, would not be able to move in such a manner. We continue with another theorem;

Theorem 4.2: If conditions (39) and (41) hold and there exists some row l in table T such that there exists no m_k in l, then T is not extremal.

Proof: If l has no m_k in it, then none of its values can be 1. Thus, at least two of the values of the row must be non-border entries. Thus a T' and a T'' can be created by 'roaming' on these two non-border entries, which will not effect the value of C_M. The rest follows as it does in the previous proof. This gives the condition

$$\forall \text{ rows } l \ \#\{m_k \mid m_k \in l\} \geq l. \tag{42}$$

We continue with the following:

Theorem 4.3: If conditions (39), (41), and (42) are met and there exists some row l in table T such that there exists > 1 m_k in l, then T is not extremal.

Proof: From (39) we have that $m_1 + m_2 + \ldots m_{t+1} = t$. Let m_1 and m_2 be the maximum values which share a row. Then $m_1 + m_2 \leq 1$. Equality must hold, as the maximum sum of the remaining $t - 1$ values is $t - 1$. This also means that the remaining $t - 1$ values must all be one. Thus we are guaranteed at least $t - 1$ rows composed solely of ones and zeroes. If either

m_1 and m_2 are one or zero then all but one row is composed solely of zeroes and ones. The remaining row has either a one in it, in which case we are done, or a zero in it. If the row has a zero in it and no ones (or else we are done), then two of the non-border entries can be safely 'roamed' in order to create a T' and a T'' without changing C_M.

So, we now assume that $m_1, m_2 \neq 0, 1$. In the two rows not composed solely of ones and zeros, one contains both m_1 and m_2, with the remaining entries being zero and the other row contains either m_1 or m_2 (by (42)) and some other non-border entry. These two rows look like this,

	A	B	...
A	m_1	m_2	...
B	x	m_2	...

where x is some non-border entry, which is shown in the same column as m_1 for neatness, as x may not actually be in said column, and m_2 is designated as the maximum value which appears in a second row. Now a T' and a T'' can be created by adding a $\pm\epsilon$ to the m_1 and $\mp\epsilon$ to the m_2 which share a row. In the other row, depending on which maximum value it contains, add or subtract ϵ in the opposite manner and then add/subtract some ϵ to some other non-border entry (if this second row contains also contains both m_1 and m_2, then these must be the two effected entries). This is shown below;

	A	B	...
A	$m_1 \pm \epsilon$	$m_2 \mp \epsilon$	...
B	$x \mp \epsilon$	$m_2 \pm \epsilon$	...

This proves that an extremal table cannot contain a row with two maximum values in it, giving the condition

$$\forall \text{ rows } l \ \#\{m_k \mid m_k \in l\} = 1. \tag{43}$$

This leaves us with a very specific criteria for an extremal table. Either all of the table values must be zeroes and ones or every row must have exactly one maximum column value in it and the sum of these values must be exactly t. This leaves us with one last theorem;

Theorem 4.4: If conditions (39), (41), and (43) are met by a table T and

T contains some value $x \neq 0, 1$, then T is not an extremal table.

Proof: Given the conditions (39), (41), and (43) our table T must be of the form

	A	B	C	...
A	m_1	a	b	...
B	c	m_2	d	...
C	e f	m_3	...	

and so on, where $(a, b, \ldots)$ are just some undefined table values. As $C_M = t$, an integer, if there exists some $m_k \neq 0, 1$, there must exist at least one other $m_l \neq 0, 1$. Using this fact it is easy to construct a suitable T' and T''. One must merely take $m_k \pm \epsilon$ from one row (taking $\mp\epsilon$ to some other non-border entry in that row which will not be a maximum column value) and $m_l \mp \epsilon$ from the other row (with $\pm\epsilon$ to some other non-border, non-maximum value in that row). This is shown below, showing the two effected rows,

	A	B	...
A	$m_k \pm \epsilon$	$x \mp \epsilon$	...
B	$y \mp \epsilon$	$m_l \pm \epsilon$	...

where x and y are some non-border table entry that is not a maximum column entry, by (43). This will leave C_M unchanged and thus both T' and T'' are realizable tables. Thus every m_k must be either zero or one which means t of the maximum values will be one and one of the maximum values will be zero. As described in the proof of Theorem 4.3, if the row which contains the maximum value of zero does not have a one in it, then the table is not extremal. Thus we have proved all the extremal states described by the condition that $C_M \leq t$ for the case of a t level system and a $t + 1 \times t + 1$ channel tables are the same as the extremal states of the classical body.

Of course, this proof was only for the special case where the size of the channel table is one larger than the level of the quantum system. Generalizing the proof for larger table sizes does not appear to be trivial. For example, in the case of the 4×4 table for the 2-level quantum system, the simplex has four vertices and is thus a three dimensional tetrahedron. This means that

the state space for the quantum system can remain as a sphere and does not need to be affinely mapped onto a lower dimensional body. This increase in the allowed dimension of the quantum state space generalizes for all channel tables of dimension less than $t^2 \times t^2$. Once the table is of dimension $t^2 \times t^2$ any increase in size doesn't change the allowed form of the quantum state space. This is because the simplex of this table will be of dimension $t^2 - 1$, the same as the quantum state space. Thus increasing the dimension of the simplex futher beyond this point is meaningless. This is a good rationale for why the generalized case does not follow directly from the proof of the special case. Now, the method used for the proof of the $t + 1 \times t + 1$ seems that it would eventually work for higher cases, but it becomes very involed and there does not seem to be an easy way to generalize it using such a method.

If our conjecture holds and the condition that $C_M \leq t$ fully describes the body of allowed classical and quantum channel tables, regardless of table size, then our created capacity,C_M, is very useful and interesting. For one, it is easy to calculate, much easier than the Shannon capacity for example. Also, its property of fully defining the body of allowed tables or probability schemes, is very useful and something which, for example, the Shannon Channel capacity does not do. For instance, consider the matrices

	A	B	C
A	2/3	1/6	1/6
B	1/6	2/3	1/6
C	1/6	1/6	2/3

&

	A	B	C
A	2/3	1/3	0
B	0	2/3	1/3
C	1/3	0	2/3

It is clear that the Shannon capacity of the second table is higher than the first, as the first table gives one no information if the incorrect box is signaled, while the second table eliminates a box from consideration if there is an error. This fact actually ties into the nature of our capacity and the game it was based on. The C_M capacity, in a sense, only measures correct answers. It does not care about the details of a mistake, as all mistakes are the same to it. This can be easily contrasted with the creation of a second game. This second game is the same as the first, except that if Bob's initial guess is incorrect he gets to take a second guess for half of the money. In such a case, the first table has an expected win value of \$0.75 and the second table an expected value of \$0.83 while both tables have an expected win amount of \$0.66 with the original game. Regardless, it is clear to see that the Shannon

Capacity of the first table is strictly less than one as it must be less than the capacity of the second table. Then the table

	A	B	C
A	$2/3 + \epsilon$	$1/6 - \epsilon$	$1/6$
B	$1/6$	$2/3$	$1/6$
C	$1/6$	$1/6$	$2/3$

must then have a Shannon Capacity that is also less than one and thus does not violate the upper bound set by the Shannon Capacity. However, the table has $C_M = 2 + \epsilon$ and thus is not realizable. This shows that the bound set by the Shannon Capacity is not enough to fully characterize the set of allowed channel tables, while the bound set by the money game capacity is enough, if our conjecture holds.

References

[1] A. Holevo: The Capacity of Quantum Channel with General Signal States. *IEEE Transactions on Information Theory* **44** (1998), pg. 269–273.