Introduction to the Holevo Additivity Violation Search

Sam Bader

August 14, 2012

This report aims to provide enough detail and references on the Holevo additivity violation search that anyone with a background in quantum mechanics could quickly learn all of the quantum information necessary to get up to speed on the project. The early sections explain Holevo capacity, why we're interested in its additivity properties, and how we do computations with it. The middle sections explain three- and four- state channels and the progress with this project over the summer of 2012, while the last section suggests future directions.

1 Background on Holevo Capacity

1.1 Motivation and Definition

Suppose Alice wants to send classical information to Bob via a channel whose operation is described quantum mechanically. That is, she encodes some classical information onto a quantum system, sends it through a potentially imperfect channel, Bob receives the quantum state on the other end, and makes a measurement to decode the classical information. The channel can be represented as some transformation from the quantum states which Alice inputs to the quantum states which Bob receives.

The classical messages Alice sends are composed of independently selected letters x_i with *a priori* distribution p_i. So Alice's input can be viewed as the random variable X. To begin, Alice puts in some quantum state, or mixture, ρ_i^A for each of the classical values x_i which she wishes to send. We can write Alice's expected quantum input as the mixed state

$$\rho^A = \sum_i p_i \rho_i^A \tag{1}$$

Then the channel Λ, which can be represented as a CPTP map, acts upon the state, giving an output to Bob:

$$\rho^B = \Lambda(\rho^A) = \sum_i p_i \Lambda(\rho_i^A) = \sum_i p_i \rho_i^B \tag{2}$$

Bob then performs some measurement to determine which ρ_i^B he has, and his result can be viewed as a random variable Y, correlated to X. According to classical information theory, the amount of classical information which can be transmitted per channel use (via a suitable asymptotically long block encoding scheme) is the "mutual information" between Bob's Y and Alice's X.

$$I(X:Y) = H(X) - H(X|Y)$$

Where $H(X) = \sum_x p_x \log_2(p_x)$ is the Shannon entropy of X and $H(X|Y) = \sum_{x,y} p_{x,y} \log_2(p_{(x|y)})$ is the conditional Shannon entropy of X given Y. This has the nice intuitive interpretation that the information transmitted is the entropy of the source (Alice) given no knowledge about Bob's measurement minus the entropy of the source given the knowledge of the Bob's measurement; heuristically, it is how much entropy Bob has eliminated with his indirect measurement of X.

One distinction of the quantum channel from a classical one is that the mutual information depends on the choice of Bob's decoding observable (ie what measurement he applies). For a given input encoding, one can place an upper bound, χ, on the greatest mutual information, $I(X:Y)$, which could be achieved with the best possible measurement scheme.

$$\chi = H(\rho^B) - \sum_i p_i H(\rho_i^B)$$

where here $H(\rho) = \mathrm{Tr}(\rho \log_2 \rho)$ is the Von Neuman entropy. Heuristically, one could possibly view this again as an entropy of a source minus an expectation of the entropy reduced in deciding which message the signal represents. This quantity is called the Holevo information of the ensemble, and Holevo's bound is asymptotically achievable if Bob's measurement may act on on entire strings of characters [1]. [1]

This quantity still depends on the choice of ensemble. Its maximum over all possible ensembles we shall call χ^*.

$$\chi^* = \max_{\{\rho_i^A, p_i\}} \left(H(\rho^B) - \sum_i p_i H(\rho_i^B) \right) \tag{3}$$

χ^* is then the best possible information capacity achievable with one copy of the channel. It is known as the *Holevo capacity* of the channel.

For the optimal ensemble, we will refer to the ρ_i^A as optimal inputs and ρ^A as the optimal input ensemble average (1). Likewise, the ρ_i^B are optimal outputs and ρ^B is the optimal output ensemble average (2).

1.2 Additivity

Classically, the mutual information of a channel is additive. That is, if one channel allows for a mutual information I_1 and another channel allows for I_2, then the channel formed by

[1] That is, the encoding/decoding scheme might operate on many characters in the message at once (as in classical information and compression theory), and Bob is not restricted to product measurements. The only restriction is that the input states of the multiple uses of the channel are not entangled with each other.

using both allows for $I_1 + I_2$. For example, having two copies of a given channel allows for twice the mutual information.

This begs the question of whether the Holevo capacity of a quantum channel is additive. Given two channels Λ_1 and Λ_2, with capacities $\chi^*(\Lambda_1)$ and $\chi^*(\Lambda_2)$, is it ever the case that

$$\chi^*(\Lambda_1 \otimes \Lambda_2) \neq \chi^*(\Lambda_1) + \chi^*(\Lambda_2)?$$

It is easy to show that the left side is never less than the right side, because one could just supply the tensor products of optimal inputs to the joint channel to achieve a capacity equal to the sum of individual capacities, so the question is really whether one can find *superadditivity*

$$\chi^*(\Lambda_1 \otimes \Lambda_2) > \chi^*(\Lambda_1) + \chi^*(\Lambda_2).$$

What makes this imaginable is that, unlike the classical joint channel, the tensor product quantum channel could have inputs which are more rich than simply tensor products of individual inputs. Its inputs could be entangled states of the two spaces.

Given that motivation, it comes as no surprise that channels known to break entanglement are also known to be Holevo additive. In fact, there is no explicit example of a channel which satisfies the above inequality. However, Hastings [3] proved indirectly that there exist superadditive channels. Unfortunately, the proof made use of large dimensional (ie 10^{32} [4]) spaces and channels, actually shows violations of an equivalent conjecture (additivity of minimum output entropy) and does not provide any concrete example. Searching spaces of that size to find a specific example of a superadditive channel (or even studying such an enormous matrix) is unthinkable. We would like to find a more approachable example, hopefully on two-dimensional spaces (qubit channels).

What does superadditivity mean for the Holevo capacity? Having two copies of a channel to add could be simulated by simply using the same channel twice, allowing for the two inputs to be entangled, and the measurement at the end to be an entangled measurement. So, if such a procedure could allow for more than twice the information to pass through the channel, it implies that the Holevo capacity does not fully capture the asymptotic ability of the channel to transmit information. One would have to fall back upon the regularized Holevo capacity:

$$\chi^*_\infty = \lim_{n\to\infty} \frac{\chi^*(\Lambda^{\otimes n})}{n}$$

which is a dreadful measure to attempt to evaluate.

It would be incredibly interesting to have an explicit superadditive channel, with dimensions small enough to study. The aim of this project is to search for superadditive qubit channels.

2 Computations with χ^*

2.1 χ^* as a geometric property of the channel

The definition (3) given for χ^* is not easy to evaluate because of the number of parameters in the maximization. For computational work, we use a different expression. Schumacher and Westmoreland [2] derived many properties of χ^* and the optimal ensemble from a clever geometric perspective.

It is straightforward to see that χ can be written in terms of the *relative entropy* of the output ensemble states to the ensemble average.

$$\chi = \sum_i p_i D(\rho_i^B || \rho^B)$$

where the relative entropy of ρ with respect to σ is given by

$$D(\rho||\sigma) = \mathrm{Tr}\, \rho \log \rho - \mathrm{Tr}\, \rho \log \sigma$$

Despite not satisfying the properties of a mathematical *metric* (eg, it is not symmetric), the relative entropy does have the feel of a kind of "directed distance" from σ to ρ. And this expression for χ is intuitive in that it reads that the information transmissable is the expected entropy of the signals relative to the average signal ensemble.

This approach easily yields several nice properties of the optimal ensemble. First, one can show that all optimal output states ρ_i^B are maxima of the entropy relative to the optimal average output ensemble ρ^B. And in fact, all optimal output states are equidistant (in terms of relative entropy) from the ensemble average output, so, for any ρ_i^B,

$$\chi^* = D(\rho_i^B || \rho^B) = \max_{\sigma^B} D(\sigma^B || \rho^B)$$

where the optimization is over the image of Λ.

In the above expression, if ρ^B were replaced with a other state γ^B, then γ^B would be nearer to some outputs and further from others, so the value of the optimum would increase. Thus

$$\chi^* = \max_{\sigma^B} D(\sigma^B || \rho^B) = \min_{\gamma} \left(\max_{\sigma^B} D(\sigma^B || \gamma^B) \right)$$

where, again, the optimizations are over the image of Λ. We could rewrite this as an optimization over the inputs like so:

$$\chi^* = \min_{\rho} \left(\max_{\rho_0} D(\Lambda(\rho_0) || \Lambda(\rho)) \right)$$

Here the maximization is unconstrained over the possible states Alice could send. And, in fact, we can reduce the effort even further, based on the observation that only pure states are optimal inputs.

$$\chi^* = \min_{\rho} \left(\max_{\psi} (\Lambda(|\psi\rangle\langle\psi|) || \Lambda(\rho)) \right) \tag{4}$$

This is the min-max expression we use for χ^*. The more detailed derivation can be found in [2].

2.2 Optimizations

Testing for additivity violations, then, involves evaluating this min-max problem for single and for multiple channels and comparing the results. There are some nice tricks which reduce the amount of work necessary.

2.2.1 Ignoring the minimization

Suppose you have done this full optimization (4) for a single channel (obtaining capacity C), so that an optimal average input ensemble, ρ, for the single channel, is known. Consider checking for additivity violations in a two-channel setup. Apply *just the inner maximization* from (4), using, as the two-channel average input, the tensor product of single-channel average $\rho \otimes \rho$.

$$\max_{\psi} D(\Lambda(|\psi\rangle\langle\psi|)||\Lambda(\rho \otimes \rho)) \tag{5}$$

As noted in [5], if this equals $2C$, then we know immediately that the two channel setup does not violate additivity. If it is greater than $2C$, we know the maxima must be attained by some state which is entangled between the two channels (since the maxima over non-entangled states would just be $2C$). That channel would violate additivity. *So, once the optimal average ensemble has been found for a single copy of the channel, only the inner optimization is necessary for checking violations.*

2.2.2 Shor's Optimization

A neat way of doing the inner optimization was noted by Shor and detailed in [6]. The expression to be maximized inside (5), in its full glory, is given by

$$\mathrm{Tr}\left[\Lambda(\gamma)(\log\Lambda(\gamma) - \log\Lambda(\rho))\right]$$

where $\gamma = |\psi\rangle\langle\psi|$. If $\Lambda^\dagger$ denotes the adjoint of Λ with respect to the norm $\langle A, B\rangle = \mathrm{Tr}\{A^\dagger B\}$, then we can rewrite the above as

$$\mathrm{Tr}\left[\gamma \cdot \Lambda^\dagger(\log\Lambda(\gamma) - \log\Lambda(\rho))\right]$$

View γ as a projector onto the 1-D subspace of $|\psi\rangle$. Intuitively, to maximize the trace, you want the $|\psi\rangle$ to be an eigenvector of the crazy matrix to its right with a large eigenvalue. In fact, you can quickly prove [6] that the iterative process of determining the largest eigenvalue of

$$\Lambda^\dagger(\log\Lambda(\gamma) - \log\Lambda(\rho))$$

and choosing this as the new value of $|\psi\rangle$ to plug in will find the local maxima. In the code, we try to search the entire space for all maxima by sampling many random starting points for Shor's algorithm.

2.2.3 Symmetries

Symmetries can reduce the space we need to search for maxima. For instance, if a qubit channel (discussed in the next section) is symmetric about the z-axis, then we need check only one arbitrary vertical slice of the Bloch sphere for optimal $|\psi\rangle$. Choosing the x-z axis, for example, allows us to work with only real values.

3 Additional background on qubit channels

As mentioned, the target of our search is a superadditive single-qubit channel. For qubit channels, there exists a clean, geometric picture of the channel's action, and a bit more we can say about the possible ensembles.

3.1 Geometric Description

A general CPTP map, Λ acts upon the Bloch sphere by shrinking it to an ellipsoid within the Bloch sphere [7]. Further, when representing quantum states by their Bloch vectors, Λ can be written as an affine map on the Bloch ball:

$$\Lambda(x, y, z) = (\lambda_1 x + t_1, \lambda_2 y + t_2, \lambda_3 z + t_3)$$

There are constraints on the λ shrinks and t shifts which ensure that the resulting ellipsoid is completely within the Bloch ball (and thus all outputs are valid states), but it turns out that not all possible ellipsoids in the Bloch ball are valid CPTP maps. The CPTP constraint is actually quite strong.

For simplicity, let us consider channels with $t_1 = t_2 = 0$, so the only shifts are in the z direction. For these channels, the CPTP condition can be expressed by the requirement that both of the following equations are satisfied:

$$(\lambda_1 \pm \lambda_2)^2 \leq (1 \pm \lambda_3)^2 - t_3^2 \tag{6}$$

From now on, when discussing qubit channels, we will freely use λ's and t's.

3.2 Input ensembles

One might expect that, encoding classical information onto a two-dimensional quantum space, the optimal method would always involve simply choosing two orthogonal (maximally distinguishable) quantum states—perhaps along whichever axis is shrunk least by the channel. One of those states represents "0," one represents "1," and you send a bit at a time.

However, Fuchs [8] showed that for some channels, the optimal input states are non-orthogonal. And, furthermore, some channels can require more than two inputs to achieve their maximal capacity [9], which is managed with long block encodings. However, one can show that no qubit channel requires more than four inputs. We will discuss three- and four-state channels further, as the search has focused on these examples thus far.

4 Four-state Search

At the start of the summer, the project began its search with the following four-state channel (similar to that reported in [10], but with x and y switched)

$$\Lambda(x, y, z) = (.601x, .6y + .21, .5z + .495)$$

It is known that, for this channel, at least one relative entropy optimum of the output lies on the $x - z$ plane. And since all the maxima should be equidistant in terms of relative entropy from the optimal average, we only need to find one to determine the capacity. So, when performing the inner maximization as discussed before, we only have to consider states with real coefficients (along the "real circle").

Still, the channel proved troublesome for two reasons. The first was what we initially suspected were local maxima (of relative entropy) distracting the optimization algorithm, because the code frequently converged upon sub-optimal capacities. This problem only worsened when running code on multiple copies of the channel, since the tensor products of local maxima became exponentially many local maxima of the tensor product channels.

As it turns out, the local maximum was an artifact of our restriction to the real circle. The maximum was merely a saddle point, a constrained optimum, and would only possibly appear if the initial point for Shor's algorithm happened to be a state with pure real coefficients. So we expanded our distribution of initial points to the entire Bloch sphere.

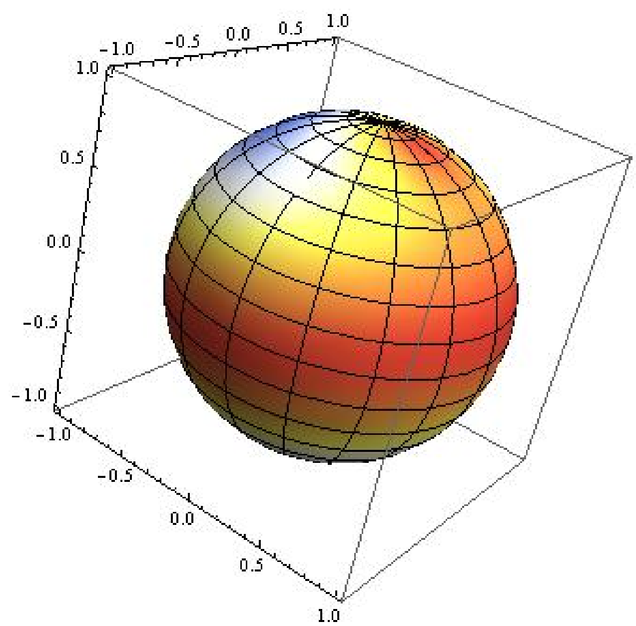


Figure 1: A temperature plot of the output relative entropy from the optimal average output for the four-state channel.

However, we found that, given the entire space of the Bloch sphere, Shor's algorithm became painfully slow to converge. The reason is evident from a visual examination (see

Figure 1). One optimum is the red spot near the north pole, and the other three lie on the dark red circle about the equator. The algorithm quickly finds its way to either the maxima at the north pole or to the equator. But once on the equator, the output relative entropy is extremely flat as a function of azimuthal angle, and thus it is quite difficult to find any of the three maxima to high precision.

We hoped that three-state channels might provide a computationally less obnoxious route to high-precision additivity tests.

5 Three-state Channels

While four-state channels are delicate creatures and difficult to come by, three-state channels arise from a wider range of parameters, so we have a good deal of freedom to tune our target. Since much of the work has gone into choosing a good set of maps to study, some more background on the construction of three-state channels will be necessary.

5.1 Restricted Capacities: C_V, C_H, C_2

One way [9] to generate a range of three-state channels is by taking convex combinations of certain two-state channels. Two common two-state channels are the shifted depolarizing channel:

$$\Phi_D(\rho) = (\mu x, \mu y, (1-\mu) + \mu z)$$

and the amplitude damping channel:

$$\Phi_{\text{amp}}(\rho) = (\sqrt{\mu} x, \sqrt{\mu} y, (1-\mu) + \mu z)$$

Let us choose $\mu = .5$ for specific examples of both.

The optimal input ensemble to the first channel includes just $|0\rangle$ and $|1\rangle$ (ie it is entirely along the classical z-axis). So for Φ_D, the unrestricted Holevo capacity is the same as the capacity restricted to ensembles which lie on a single vertical line. From here in, we notate Holevo capacity by C. We will call the capacity of any channel for ensembles restricted to two inputs C_2. We will call the capacity of a channel restricted to ensembles on a vertical line C_V. For Φ_D, $C = C_2 = C_V$.

For the second channel, the optimal ensemble is any pair of pure states along a horizontal line at $z = .596$ (note the azimuthal symmetry implies any horizontal line will do). The capacity restricted to horizontal lines we will call C_H. So for Φ_{amp}, $C = C_2 = C_H$.

We could take a combination of these two channels (which [9] calls Φ_{str}) and vary the relative weighting until, for the new channel, $C_V = C_H$. Now [9] argues that the channel generated should be three-state. The details of the proof are not important[2]—the take-away

[2]Roughly the argument proceeds by (1) arguing that the equal C_V and C_H capacities are the best that can be done with two states (2) constructing a convex combination of the optimal horizontal ensemble and optimal vertical ensemble to produce a new four-state ensemble which (by strict convexity of the capacity function) must have a higher capacity. (3) This four-state ensemble is actually reducible to a three-state ensemble. And thus the three-state capacity is greater than the two-state capacity.

from this section is the meanings of C_V, and C_H, and the understanding that their equality is one way to ensure a three-state channel.

5.2 Computations with Three-state Channels

When dealing with single instances of three-state channels, the optimization procedure described earlier for general channels can be simplified. Assuming $t_1 = t_2 = 0$ as before, $\lambda_1 \geq \lambda_2$, $t_3 > 0$, and $\lambda_3 > 0$, we can take the following to be true about the optimal ensemble. One state is at the north pole, and the other two are on the real circle, both at the same declination, and equally probable. So the solution can be described by just two parameters: the probability p_0 for the north pole state, and the declination angle θ of the other two input states.

So the optimization can be managed by numerically solving two equations in two variables. Thus when we use three-state channels, we can get the details of a single channel instance efficiently to high precision. This is important, because, as mentioned previously, we can use the single-channel average input ensemble for the many-channel optimizations. (By comparison, getting an ensemble average for the four-state channel is a min-max optimization and incredibly time-consuming to do with near as much precision).

5.3 The "Most Three-state" Channel

We began our three-state hunt by trying to find the channel which was the most unmistakably three-state, that is, the channel for which the ability to give three inputs offered the greatest advantage. So we attempted to maximize $C - C_2$ on channels of the form

$$\Lambda(\rho) = (\lambda_1 x, \lambda_2 y, \lambda_3 z + t_3), \quad \lambda_2 \leq \lambda_1$$

with the constraint $C_V = C_H$.

What we found was that, for the optimal channel, $\lambda_3 + t_3 = 1$. Unfortunately, plugging this into the CPTP conditions (6) given above requires that $\lambda_1 = \lambda_2$, which forces azimuthal symmetry upon the channel, as can be seen in the relative entropy plots below (Figure 2 and 3).

We would prefer a channel with well-defined optimal inputs, that is, peaked relative entropy in both the declination and azimuthal degrees. So for our next hunt, we will step away a little from the "most three-state" channel and see how much we can increase λ_1/λ_2.

5.4 Trade-offs

As mentioned, the difference between λ_1 and λ_2 is constrained to zero by the CPTP condition and the optimal situation $\lambda_3 + t_3 = 1$. So we will take $k = \lambda_3 + t_3 < 1$. For each channel of the form

$$\Lambda(\rho) = (\lambda_1 x, \lambda_2 y, \lambda_3 z + (k - \lambda_3)), \quad \lambda_1 > \lambda_2$$

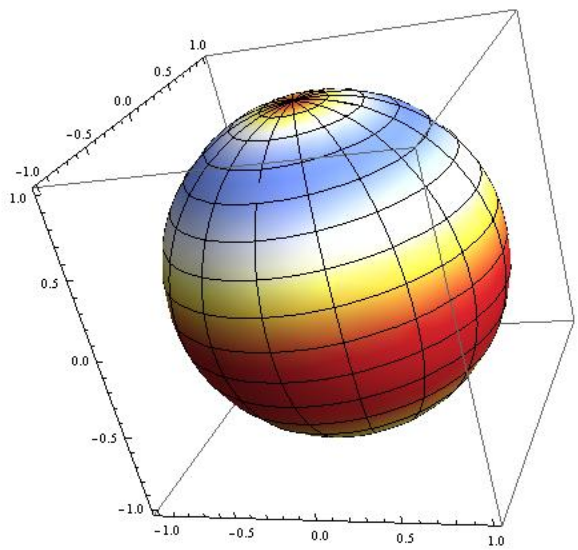


Figure 2: Output relative entropy (relative to the optimal output average) on the Bloch sphere for the most three-state channel

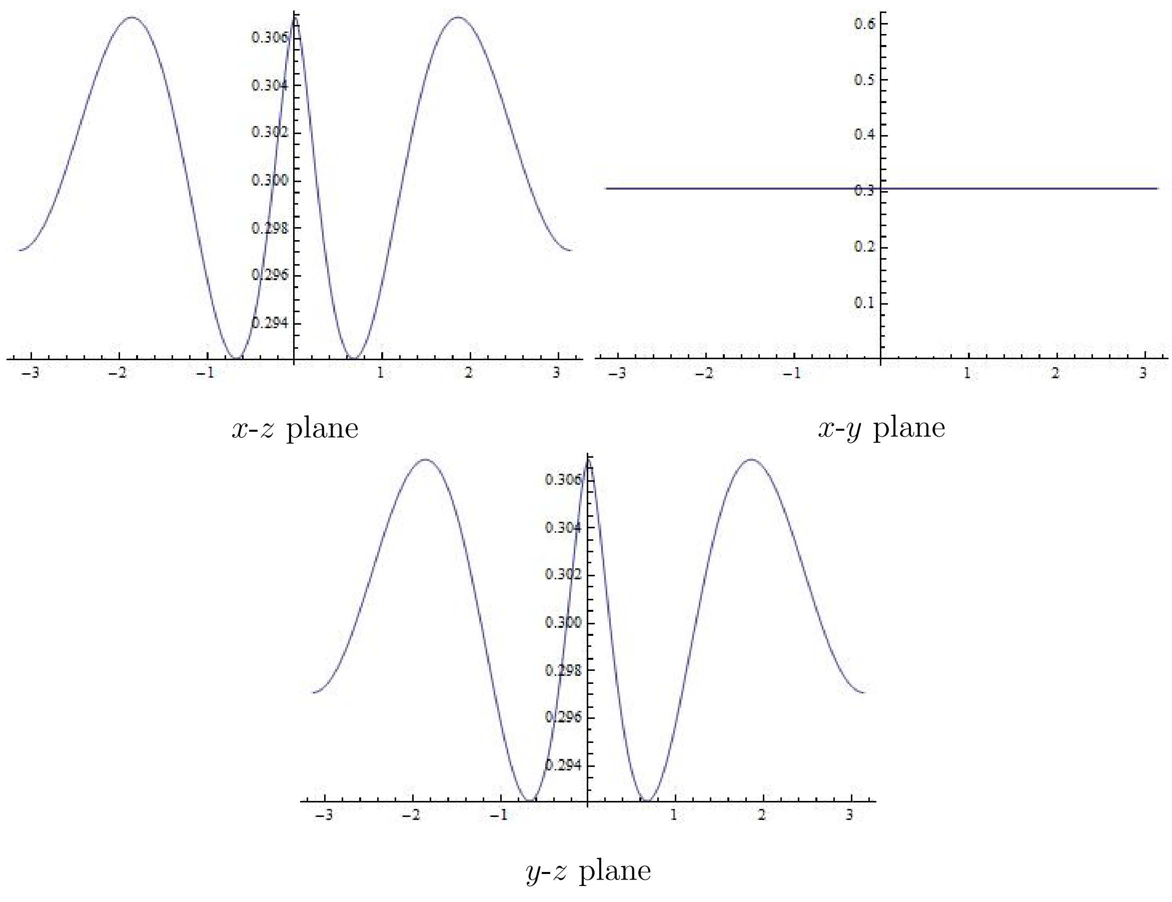


Figure 3: Relative entropy plots for specific slices of the Bloch sphere (clockwise). The horizontal axis of each plot gives the angle in radians from the $+z$, $+x$, and $+z$ direction respectively, and the vertical axis gives the output entropy of the pure state at that location relative to the optimal output average.

λ_2 has no effect on the capacities since, with $\lambda_1 > \lambda_2$, the maxima are along the real circle. So we can, for each k, find the optimal λ_1, λ_3, subject to the constraint $C_V = C_H$. And then we plug into the CPTP condition (6) to find how big a difference we can create between λ_2 and λ_1:

$$|\lambda_1 - \lambda_2| \leq \sqrt{(1+\lambda_3)^2 - (k-\lambda_3)^2}$$

So for each k parameter, we have a $C - C_2$ value and a λ_1/λ_2 value. We may as well just ignore the k parameter and plot one value against the other (Figure 4).

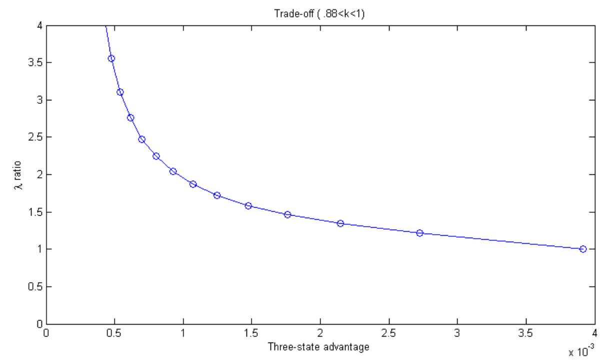


Figure 4: The trade-off of λ ratio versus three-state advantage

From this we see that we are actually able to get computationally workable λ ratios (i.e. not unity) without losing very much in terms of the three-state advantage. For example, consider the channel with λ ratio of 1.5. See in the below plots (Figure 5 and 6) that the relative entropy is now nicely peaked in all directions:

So now we have taken seven channels from along this trade-off curve to run high-precision Holevo additivity tests on, which will probably consume the rest of the week.

6 Where to now?

As we test three-state channels, we also keep in mind other ideas for the testing. For instance, Professor Ruskai suggested that one sign of non-additivity would be a shift in the location of maxima of tensor product channels away from the tensor products of single-channel maxima. One possible test might then involve initializing Shor's algorithm with all the tensor products of single-channel optima on tensor-product channels and seeing whether the first iteration produces any significant change.

Another possibility is that we may have to expand our search to channels of dimension larger than two. Larger channels do not have the simple Bloch sphere geometry, and may hide more secrets. However, less is known about larger channels, and figuring out where to start would be quite a challenge.

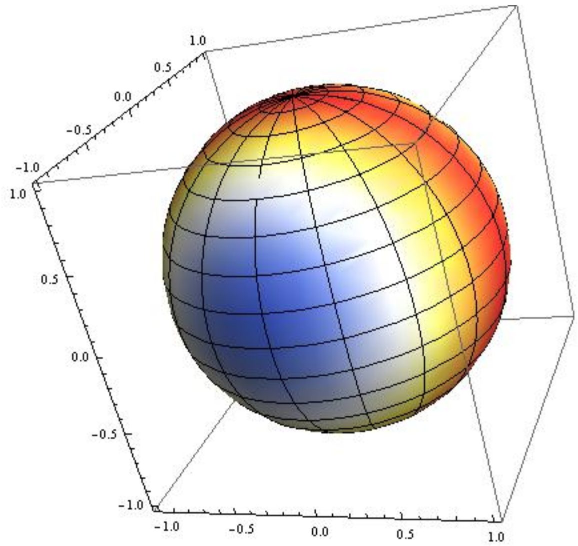
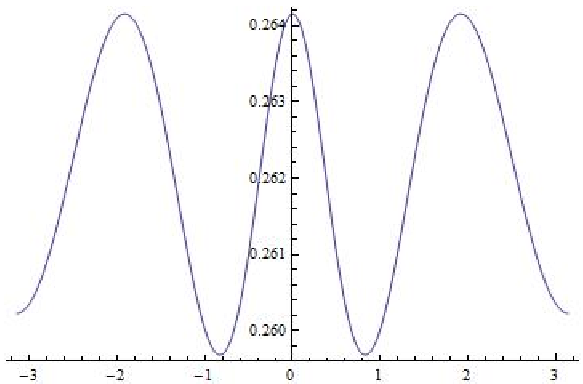
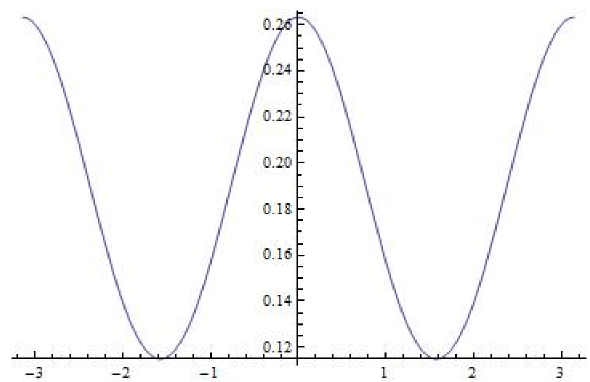


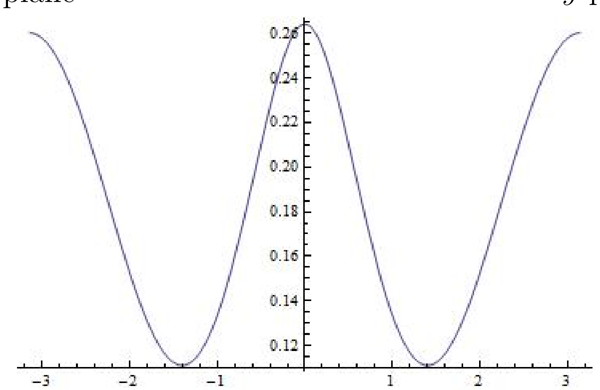
Figure 5: Output relative entropy (relative to the optimal output average) on the Bloch sphere for a sample trade-off channel



x-z plane



x-y plane



y-z plane

Figure 6: Relative entropy plots for specific slices of the Bloch sphere (clockwise). The horizontal axis of each plot gives the angle in radians from the $+z$, $+x$, and $+z$ direction respectively, and the vertical axis gives the output entropy of the pure state at that location relative to the optimal output average.

References

[1] Hausladen et al, Phys. Rev. A 55, 1869 (1996).

[2] B. Schumacher and M. Westmoreland, Phys. Rev. A 63, 022308 (2001).

[3] Hastings, M.B. Nature Physics 5, 255 - 257 (2009)

[4] http://arxiv.org/abs/0905.3697v1

[5] Hayashi, Imai, Matsumoto, Ruskai, Shimono, Quantum Inf. Comput. 5, 13–31 (2005)

[6] Datta and Ruskai, J. Phys. A: Math. Gen. 3, 9785-9802 (2005)

[7] Ruskai et al. Linear Algebra and its Applications, 347, 159-187.

[8] Fuchs, Phys. Rev. Lett., 79 (1997), pp. 11621165

[9] King, Nathanson, and Ruskai, Phys. Rev. Lett. 88, 057901 (2002)

[10] Hayashi et al. Quantum Inf. Comput. 5, 13–31 (2005)