Available online at www.sciencedirect.com
ScienceDirect
Electronic Notes in Theoretical Computer Science
ELSEVIER
Electronic Notes in Theoretical Computer Science 270 (1) (2011) 81–97
www.elsevier.com/locate/entcs

How to Randomly Flip a Quantum Bit

Keye Martin[1]

Center for High Assurance Computer Systems (Code 5540)
Naval Research Laboratory, Washington, DC 20375.

Abstract

We show that an important class of quantum channels, the convex closure of the spin channels, can be algebraically represented by *classical* channels that have four inputs and four outputs. This result is used to develop an experimentally realizable scheme for randomly flipping qubits whose basis of preparation is unknown. This scheme can be used to interrupt, but not necessarily eradicate, any form of hidden communication based on quantum information. It can also be used to remove steganographic information embedded in quantum data.

Keywords: Quantum channel, spin channel, classical channel, steganography

1 Introduction

The ability to randomly flip classical bits is a fundamental operation with many uses. It can be used to interrupt undesirable communication [2], for instance, or to remove steganographic information hidden in data [3]. Each of these help to ensure the security of systems. The scheme for randomly flipping classical bits is as follows: given a bit, we toss a coin, and then based on the result of this coin toss, either flip the bit or leave it alone. This scheme eliminates any and all correlation between the information sent and the information received. As a result, transmitting information with bits that have been randomly flipped is impossible.

In the quantum case, interrupting communication is more difficult. A common misconception is that one can randomly flip quantum bits by applying the "bit flipping" operator $\varepsilon(\rho) = p \cdot \rho + (1-p) \cdot \sigma_x \rho \sigma_x$ with $p = 1/2$. However, if the $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$ basis is being used to represent information, then this has no effect at all, since $\varepsilon(|\pm\rangle\langle\pm|) = |\pm\rangle\langle\pm|$. The states randomly flipped by ε when $p = 1/2$ are $|0\rangle$ and $|1\rangle$. Put another way, quantum bit flipping with $p = 1/2$ only has the effect we intend for it to have if we *know* the basis being used to represent

[1] Email: keye.martin@nrl.navy.mil

1571-0661 Published by Elsevier B.V.
doi:10.1016/j.entcs.2011.01.008

information. Naturally, it is not realistic to assume that we can know the basis being used to represent information: infinitely many representations are possible. For instance, suppose that two parties share a private key obtained from QKD and that they then use this private key to randomly alternate between representations in the $X = \{|+\rangle, |-\rangle\}$ and $Z = \{|0\rangle, |1\rangle\}$ bases. Since each bit of information is equally likely to have been coded in either the X or Z basis, applying ε with $p = 1/2$ only results in $1/4$ of the bits being flipped, when we intend to flip $1/2$ of them.

So the question is: how do we randomly flip a quantum bit, assuming that we do *not* know the basis being used to represent information? Is this even possible? And if it is possible, we are not just looking for a mathematical solution, such as some operator that gives rise to the correct probability of error. Fundamentally, we are trying to identify a simple and inexpensive procedure that can be easily performed in a laboratory.

In this paper we show that it *is* possible to randomly flip quantum bits. The procedure we give is based on experimental operations that are routinely performed in laboratories today. Our path to the solution though is also of interest. The way we solve this problem is by establishing an isomorphism between a class of classical channels and a class of quantum channels. The existence of this isomorphism means that we can reason about certain quantum channels *as though they were* classical channels. Once we become aware of this, the ability to randomly flip quantum bits *is a consequence* of our ability to randomly flip classical bits. Moreover, the class of such quantum channels covered by our representation theorem includes most of the models currently used to describe noise: bit flipping, phase flipping, bit-phase flipping, depolarization and projective measurements. Thus, channels like these can be reasoned about as though they were classical channels having four inputs and four outputs. This provides additional evidence of the importance of *algebraic structure* [2] in information theory, whether it be classical or quantum.

2 The classical paradigm

A binary channel has two inputs ("0" and "1") and two outputs ("0" and "1"). An input is sent through the channel to a receiver. Because of noise in the channel, what arrives may not necessarily be what the sender intended. The effect of noise on input data is modelled by a noise matrix u. If data is sent through the channel according to the distribution x, then the output is distributed as $y = x \cdot u$. The noise matrix u is given by

$$u = \begin{pmatrix} a & \bar{a} \\ b & \bar{b} \end{pmatrix}$$

where $a = P(0|0)$ is the probability of receiving 0 when 0 is sent and $b = P(0|1)$ is the probability of receiving 0 when 1 is sent and $\bar{x} := 1 - x$ for $x \in [0, 1]$. Thus, the noise matrix of a binary channel can be represented by a point (a, b) in the unit square $[0, 1]^2$ and all points in the unit square represent the noise matrix of some binary channel.

We begin with the classical case to gain some intuition about properties of random bit flipping that might aid us in the quantum case. Two parties are communicating, sending classical bits across a channel whose noise matrix is unknown to them. For a couple of a good reasons, it is reasonable to assume that they will send bits with equal probability. These reasons include:

- Creating the appearance that their transmission of data is random i.e. that it is simply noise will help serve as a deterrent to those who might seek to interrupt (i.e., who would care about interrupting noise?)
- Given that the matrix is unknown, the excellent result of Majani-Rumsey says that sending bits with equal probability will get them to within ninety-four percent of capacity.

Now when we say interrupt, we are talking about inserting a second channel with noise matrix (a, b) between the parties whose effect is to introduce a certain probability of error. When data is transmitted according to $(x, \bar{x})$, the probability of error achieved is $x\bar{a} + \bar{x}b$. Since we are assuming the parties send bits with equal frequency, the probability of error we introduce[2] is

$$e(a,b) = \frac{1}{2} \cdot \bar{a} + \frac{1}{2} \cdot b = \frac{1}{2} - \frac{1}{2}(a-b)$$

Now assume that we would like to introduce a probability of error $p = e(a, b)$. Then in order to interrupt communication as much as possible, we seek to *minimize* capacity

$$\mathrm{C}(a,b) = \log_2 \left(2^{\frac{\bar{a}H(b) - \bar{b}H(a)}{a-b}} + 2^{\frac{bH(a) - aH(b)}{a-b}} \right)$$

where $\mathrm{C}(a, a) := 0$ and $H(x) = -x \log_2(x) - (1-x)\log_2(1-x)$ is the base two entropy, over the set of channels $S = \{(a, b) : e(a, b) = p\}$ which cause a probability of error equal to p. Because S is convex (being a line) and capacity is a convex function of the noise matrix, C assumes its minimum value at the midpoint (a, b) of S. Such a midpoint (a, b) satisfies $a + b = 1$. That is, we must use a *binary symmetric channel* whose capacity is then $1 - H(b) = 1 - H(a)$. Let us point out two elementary but important facts:

(i) Within the class of binary channels, the binary symmetric channels are exactly those which *increase* entropy i.e. $H(x \cdot u) \geq H(x)$ for all x,

(ii) Within the class of binary *symmetric* channels, the only channel with capacity zero is $\perp := \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$

Thus, on the assumption that all inputs are equally likely, the best way to interrupt communication is to inject $\perp$ into the communication line: the unique entropy increasing channel of capacity zero, which corresponds to randomly flipping classical

[2] Notice that the actual probability of error may be larger since we are only interested in what *we* introduce and ignore effects from the environment

bits. Ideally, one hopes to see these properties also reflected in the case of qubits, since each choice of basis defines a classical channel, as we will see in the next section. Though it seems a bit naive at first glance, we nevertheless now restrict our attention to *entropy increasing* channels on qubits.

3 Quantum channels

We begin with basic intuitions that help us relate qubit channels to classical binary channels. Let Δ^2 denote the set of finite probability distributions over a two element set:

$$\Delta^2 := \{(x, y) \in [0, 1]^2 : x + y = 1\}$$

The noise matrix u of a binary channel defines a function $f : \Delta^2 \to \Delta^2$, given by $f(x) = x \cdot u$, which maps an input distribution $x \in \Delta^2$ to an output distribution $f(x) \in \Delta^2$.

The fact that a binary channel operates on Δ^2 is indicative of the fact that only two symbols are being sent and that we have chosen a particular and fixed way of representing these two symbols. By contrast, in the case of a quantum channel, there are an infinite number of ways to represent bits: each basis of the state space $\mathcal{H}^2$, a two dimensional complex Hilbert space, offers a different possible representation. Let us suppose that we choose a *particular* quantum representation for the classical bits '0' and '1', denoted by orthogonal unit vectors $|0\rangle$ and $|1\rangle$ in $\mathcal{H}^2$. In doing so, we are implicitly saying that we will use a quantum system to represent a classical bit. When the system is in state $|0\rangle$, it represents the classical bit '0'; when in state $|1\rangle$, it represents the classical bit '1'. There is a subtle but relevant caveat here though.

Physically, states are equal "to within a phase factor." So for example, the states $|0\rangle, -|0\rangle, i|0\rangle, -i|0\rangle, e^{i\theta}|0\rangle$ are all equivalent in the sense that quantum mechanics makes the same predictions about a system in any one of these states. Mathematically, though, we know that we cannot go around writing things like "$|0\rangle = -|0\rangle$," for the simple reason that in a vector space the only such element is the zero vector and the zero vector is not a unit vector. One way around this difficulty is to say that a 'state', specified by a unit vector $|\psi\rangle \in \mathcal{H}^2$, is mathematically represented by the operator $f : \mathcal{H}^2 \to \mathcal{H}^2$ given by

$$f(u) = \langle\psi|u\rangle \cdot |\psi\rangle$$

The operator f takes as input a vector u and returns as output the vector $|\psi\rangle$ multiplied by the complex number $\langle\psi|u\rangle$, which is the inner product of the vector u and the vector $|\psi\rangle$. For this reason, the operator f is traditionally denoted $f = |\psi\rangle\langle\psi|$. Such an operator is called a *pure state* since it refers to a state that the system can be in; pure states are the quantum analogues of $e_0 = (1, 0)$ and $e_1 = (0, 1)$ in Δ^2, the latter of which we think of as the classical representation of the bits '0' and '1'.

A classical binary channel $f : \Delta^2 \to \Delta^2$ takes an input distribution to an output distribution. In a similar way, a qubit channel will map input distributions to output

distributions. But what is the quantum analogue of a distribution? Let us return to the classical case. Each distribution $x \in \Delta^2$ may be written

$$x = x_0 \cdot e_0 + x_1 \cdot e_1$$

i.e., as a convex sum of classical 'pure' states. The meaning of such an expression is that the system is in state e_0 with probability x_0 and in state e_1 with probability x_1. Thus, if a quantum system is in state $|\psi_i\rangle\langle\psi_i|$ with probability x_i, a natural way to represent this 'distribution' is given by the operator

$$\rho = \sum_{i=1}^{n} x_i \cdot |\psi_i\rangle\langle\psi_i|$$

Such an operator is called a *density operator*. A density operator is also called a *mixed state*. The set of all density operators on $\mathcal{H}^2$ is denoted by Ω^2. Thus, in analogy with the classical case, a qubit channel will be a function of the form $\varepsilon : \Omega^2 \to \Omega^2$. Specifically,

Definition 3.1 *A* qubit channel *is a function $\varepsilon : \Omega^2 \to \Omega^2$ that is convex linear and completely positive.*

To say that ε is convex linear means that ε preserves convex sums i.e. sums of the form $x \cdot \rho + (1 - x) \cdot \sigma$. Complete positivity, defined in [4], is a condition which ensures that the definition of a qubit channel is compatible with natural intuitions about joint systems. For our purposes, there is no need to get lost in too many details of the Hilbert space formulation – very soon, we will represent qubit channels as mappings on the unit ball in Euclidean space.

3.1 Unitality

In the case of classical binary channels, the binary symmetric channels are the entropy increasing channels: they are exactly the channels f which preserve the uniform distribution $f(1/2, 1/2) = (1/2, 1/2)$. The *von Neumann entropy* of a density operator ρ is given by $S(\rho) = -\mathrm{tr}(\rho \log(\rho))$. In a similar way, the entropy increasing qubit channels are precisely those ε for which $\varepsilon(I/2) = I/2$. Such channels are called *unital* [4].

Definition 3.2 *A qubit channel ε is* unital *if $\varepsilon(I/2) = I/2$.*

Let us now consider several important examples of unital channels.

Example 3.3 *Unitary channels.* If U is a unitary operator on $\mathcal{H}^2$, then $\varepsilon(\rho) = U\rho U^\dagger$ is unital since $UU^\dagger = I$.

Example 3.4 *Projective measurements.* If $\{P_0, P_1\}$ are projections with $P_0 + P_1 = I$, then

$$\varepsilon(\rho) = P_0\rho P_0 + P_1\rho P_1$$

is a unital channel since $P_0^2 = P_0$ and $P_1^2 = P_1$.

Example 3.5 *Convex sum and composition.* If ε_1 and ε_2 are unital channels, then $\varepsilon_1 \circ \varepsilon_2$ and $p \cdot \varepsilon_1 + (1-p) \cdot \varepsilon_2$ are unital for $p \in [0,1]$.

Using these three examples, we can construct many examples of unital channels. For instance, to construct the "bit-flipping" channel, we use the spin operator $\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, which is unitary, to define a unital channel ε_x. Next, the identity channel $\varepsilon_I(\rho) = \rho$ is unital, so their convex sum

$$\varepsilon(\rho) = (1-p)\varepsilon_I(\rho) + p \cdot \varepsilon_x(\rho)$$

is a unital channel. In a similar way, phase flipping, bit-phase flipping, the two-Pauli channel, phase damping ("decoherence") and depolarization are also seen to be unital. Not all qubit channels are unital of course, amplitude damping is one well-known example.

3.2 The Bloch representation of unital channels

Every unital qubit channel can be represented by a linear mapping that takes the unit ball in three space into itself. Using this representation, the *Bloch representation*, one is able to avoid many of the complications of the Hilbert space formulation, so we consider it now.

There is a 1-1 correspondence between density operators on a two dimensional state space and points on the unit ball $\mathbb{B}^3 = \{x \in \mathbb{R}^3 : |x| \leq 1\}$: each density operator $\rho : \mathcal{H}^2 \to \mathcal{H}^2$ can be written uniquely as

$$\rho = \frac{1}{2} \begin{pmatrix} 1 + r_z & r_x - ir_y \\ r_x + ir_y & 1 - r_z \end{pmatrix}$$

where $r = (r_x, r_y, r_z) \in \mathbb{R}^3$ satisfies $|r| = \sqrt{r_x^2 + r_y^2 + r_z^2} \leq 1$. The vector $r \in \mathbb{B}^3$ is called the *Bloch vector* associated to ρ. Bloch vectors have a number of aesthetically pleasing properties.

If ρ and σ are density operators with respective Bloch vectors r and s, then (i) the eigenvalues of ρ are $(1 \pm |r|)/2$, (ii) the von Neumann entropy of ρ is $S\rho = H((1+|r|)/2) = H((1-|r|)/2)$, where $H : [0,1] \to [0,1]$ is the base two Shannon entropy, (iii) if ρ and σ are pure states and $r + s = 0$, then ρ and σ are orthogonal, and thus form a basis for the state space; conversely, the Bloch vectors associated to a pair of orthogonal pure states form antipodal points on the sphere, (iv) the Bloch vector for a convex sum of mixed states is the convex sum of the Bloch vectors, (v) the Bloch vector for the completely mixed state $I/2$ is $0 = (0,0,0)$ and (vi) $\mathrm{tr}(\rho \cdot \sigma) = (1 + (r,s))/2$, where (r,s) is the Euclidean inner product on $\mathbb{R}^3$.

Definition 3.6 *For a qubit channel $\varepsilon : \Omega^2 \to \Omega^2$, the mapping it induces on the Bloch sphere $f_\varepsilon : \mathbb{B}^3 \to \mathbb{B}^3$ is called the* Bloch representation *of ε.*

If ε is a qubit channel and f_ε is its Bloch representation, then (i) ε is unital iff $f_\varepsilon(0) = 0$, (ii) the function f_ε is convex linear, (iii) composition of quantum channels corresponds to composition of Bloch representations: for channels $\varepsilon_1, \varepsilon_2$, we have $f_{\varepsilon_1 \circ \varepsilon_2} = f_{\varepsilon_1} \circ f_{\varepsilon_2}$, (iv) convex sum of quantum channels corresponds to convex sum of Bloch representations: for channels $\varepsilon_1, \varepsilon_2$ and $x \in [0,1]$, we have $f_{x\varepsilon_1 + \bar{x}\varepsilon_2} = x f_{\varepsilon_1} + \bar{x} f_{\varepsilon_2}$.

To illustrate how these properties make it simple to calculate the Bloch representation of a qubit channel, let us return to the example of "bit flipping" $\varepsilon(\rho) = (1-p)\varepsilon_I(\rho) + p \cdot \varepsilon_x(\rho)$. The Bloch representation of ε_I is $f_{\varepsilon_I}(r) = r$. Using the correspondence between density operators and Bloch vectors, we calculate directly that the Bloch representation of ε_x is $f_{\varepsilon_x}(r_x, r_y, r_z) = (r_x, -r_y, -r_z)$. Thus, by property (iv) of Bloch representations,

$$f_\varepsilon(r_x, r_y, r_z) = (1-p)(r_x, r_y, r_z) + p(r_x, -r_y, -r_z) = (r_x, (1-2p)r_y, (1-2p)r_z)$$

If we set $p = 1/2$ in an attempt to randomly flip qubits, we see that states of the form $(r_x, 0, 0)$ are unchanged by this form of noise, which explains why "bit flipping" only removes all correlations when the basis used to represent information is *known*. Let us now turn to the case in which the representation is unknown.

3.3 Zero

In section 2, we saw that randomly flipping classical bits amounted to being able to construct the unique entropy increasing channel of capacity zero, $\perp$. Specifically, the way we 'construct' $\perp$ is by basing the decision on whether to flip a bit on the result of a coin toss. We now use this observation as a guide in approaching the more difficult problem of how to randomly flip a quantum bit.

Suppose Alice sends a qubit represented by ρ, that ρ suffers the effect of noise described by the quantum channel ε. Bob then receives $\varepsilon(\rho)$ and performs a measurement in *some* basis $\{|0\rangle, |1\rangle\}$. The measurement operators in this case are the projections $P_0 = |0\rangle\langle 0|$ and $P_1 = |1\rangle\langle 1|$ and form a complete set since $P_0 + P_1 = I$. Thus, by standard quantum mechanics, the probability that Bob obtains the result 0 is

$$p_0 = \mathrm{tr}(P_0^\dagger P_0 \cdot \varepsilon(\rho))) = \mathrm{tr}(P_0 P_0 \cdot \varepsilon(\rho)) = \mathrm{tr}(P_0 \cdot \varepsilon(\rho))$$

while the probability that Bob obtains the result 1 is

$$p_1 = \mathrm{tr}(P_1^\dagger P_1 \cdot \varepsilon(\rho))) = \mathrm{tr}(P_1 P_1 \cdot \varepsilon(\rho)) = \mathrm{tr}(P_1 \cdot \varepsilon(\rho))$$

where we recall that projections satisfy $P_i^2 = P_i$. Now both projections P_0 and P_1, being density operators, also have a Bloch vector associated with them, given by s and t, respectively. Thus, if r is the Bloch vector for ρ and f_ε is the Bloch representation of ε, then the probabilities p_0 and p_1 can be succinctly written as

$$p_0 = \frac{1 + (s, f(r))}{2} \quad \& \quad p_1 = \frac{1 + (t, f(r))}{2}$$

Further, since $|0\rangle$ and $|1\rangle$ form a basis for the state space, $s + t = 0$, which helps us see that $p_0 + p_1 = 1$. Thus, choosing any two density operators as inputs, we obtain a classical binary channel.

In the particular case that the two density operators chosen by Alice form a basis for the state space (perhaps different from the one that Bob measures in), the channel we get is *binary symmetric* and thus we see that shutting down all communication will require that all our probabilities be $1/2$. That is, the Bloch representation f_ε will have to be zero, meaning that the channel should be $\varepsilon(\rho) = I/2$. From an experimental point of view, this information is useless – how do we force a system into the completely mixed state $I/2$? Put another way, how do we factor zero into a product of nontrivial, experimentally realizable operations? Surprisingly, the answer to the question can be found by examining the algebraic structure of *classical channels*.

4 The involution group in classical information theory

For a classical channel with n inputs and n outputs, the noise matrix u has n rows and n columns, each entry is a probability, and as with binary channels, each row sums to one. Thus, each (n, n) classical channel induces a function $f : \Delta^n \to \Delta^n$ given by $f(x) = x \cdot u$ where

$$\Delta^n := \left\{ x \in [0,1]^n : \sum_{i=1}^{n} x_i = 1 \right\}$$

In particular, the (n, n) channel $\perp$ is the matrix all of whose entries are $1/n$,

$$\perp := \begin{pmatrix} 1/n & \dots & 1/n \\ \vdots & & \vdots \\ 1/n & \dots & 1/n \end{pmatrix}$$

It is the unique entropy increasing (n, n) channel with capacity zero.

With both classical and quantum channels, there is structure in common that we can formulate more abstractly – doing so not only helps us to think more clearly about channels, this more abstract stance is what ultimately enables us to solve the problem of randomly flipping quantum bits.

A *monoid* $(M, \cdot, 1)$ is a set M with an associative binary operation $\cdot : M^2 \to M$ such that $1 \in M$ is an *identity*: $x \cdot 1 = 1 \cdot x = x$ for all $x \in M$. A *group* $(G, \cdot, 1)$ is a monoid $(G, \cdot, 1)$ in which every element has an *inverse*: $(\forall x \in G)(\exists y \in G)\ xy = yx = 1$. An *algebra* over the reals, also called a *real algebra*, is a vector space $(A, +, 0)$ over the reals which is also a monoid $(A, \cdot, 1)$ in such a way that multiplication $\cdot$ and addition $+$ satisfy the kind of properties one would expect. To prevent getting lost in too many mathematical definitions, let us just give the main example of a real algebra that we are interested in: the algebra $M_n(\mathbb{R})$ of $n \times n$ real matrices.

Our interest in these mathematical structures is that they can be used to identify previously unknown similarities between classes of channels, both classical and quantum. A good starting point for a mathematical model of a "class of channels" is a *convex monoid*: a convex subset of a real algebra that is also a monoid. Noise matrices for (n, n) channels are closed under finite convex sums and composition, so they form a convex monoid. The same is true of the doubly stochastic classical channels, unital qubit channels, qubit channels and quantum channels in general.

Definition 4.1 *An* involution group *is a monoid* $(M, \cdot, 1)$ *in which* $x^2 = 1$ *for all* $x \in M$.

Notice that an involution group, while defined only to be a monoid, is in fact a group, since every element is its own inverse. In addition, such a group is *commutative*: since $(xy)(yx) = 1$, we see that yx must be the inverse of xy, but since xy is an involution, $xy = yx$.

By standard results in group theory (the classification of finite abelian groups), any finite involution group G is isomorphic to $\mathbb{Z}_2^n$ for some $n \geq 0$ and thus is determined (up to isomorphism) by the number of elements it contains (its *order*), which must be a power of two.

Theorem 4.2 *The class of* (n, n) *classical channels contains a copy of the involution group of order* n *iff* $n = 2^k$ *for some integer* $k \geq 1$.

Proof. If $k = 1$, then the involution group is

$$G_1 = \left\{ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

This is the involution group of order 2^1. Now suppose we have the involution group G_k of order 2^k. Define G_{k+1} by specifying its element in block form as

$$G_{k+1} = \left\{ \begin{pmatrix} 0 & g \\ g & 0 \end{pmatrix} : g \in G_k \right\} \cup \left\{ \begin{pmatrix} g & 0 \\ 0 & g \end{pmatrix} : g \in G_k \right\}$$

where 0 refers to the $2^k \times 2^k$ zero matrix. Since G_k contains 2^k distinct elements, G_{k+1} contains $2^k + 2^k = 2^{k+1}$ distinct elements. Furthermore, G_{k+1} is closed under multiplication, which can be seen by considering the four possible forms of products that exist and recalling that G_k is closed under multiplication:

$$\begin{pmatrix} g & 0 \\ 0 & g \end{pmatrix} \cdot \begin{pmatrix} h & 0 \\ 0 & h \end{pmatrix} = \begin{pmatrix} gh & 0 \\ 0 & gh \end{pmatrix} \in G_{k+1}$$

$$\begin{pmatrix} g & 0 \\ 0 & g \end{pmatrix} \cdot \begin{pmatrix} 0 & h \\ h & 0 \end{pmatrix} = \begin{pmatrix} 0 & gh \\ gh & 0 \end{pmatrix} \in G_{k+1}$$

$$\begin{pmatrix} 0 & g \\ g & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & h \\ h & 0 \end{pmatrix} = \begin{pmatrix} gh & 0 \\ 0 & gh \end{pmatrix} \in G_{k+1}$$

$$\begin{pmatrix} 0 & g \\ g & 0 \end{pmatrix} \cdot \begin{pmatrix} h & 0 \\ 0 & h \end{pmatrix} = \begin{pmatrix} 0 & gh \\ gh & 0 \end{pmatrix} \in G_{k+1}$$

Finally, setting $h = g$ in the first and third expressions above, and recalling that each element of G_k is an involution, also shows that each element of G_{k+1} is an involution.

Conversely, suppose that (n, n) contains a copy of the involution group of order n. Then by standard results in group theory (the classification of finite abelian groups), n must be a power of two since it is the order of a finite involution group. □

The involution group G_n in $(2^n, 2^n)$ can also be derived by taking repeated tensor products of matrices. Specifically, setting $G_1 = \{\text{flip}, I\}$ and then using G_n to define $G_{n+1} = \{\text{flip} \otimes g : g \in G_n\} \cup \{I \otimes g : g \in G_n\}$.

Lemma 4.3 *Let $G_n = \{f_i : 1 \leq i \leq n\}$ be the involution group in $(2^n, 2^n)$. Then*

$$\sum_{f_i \in G_n} \frac{f_i}{2^n} = \bot \in (2^n, 2^n)$$

and thus the elements of G_n are independent: for any two probability distributions $x, y \in \Delta^{2^n}$,

$$\sum_{i=1}^{2^n} x_i \cdot f_i = \sum_{i=1}^{2^n} y_i \cdot f_i \implies x = y.$$

Proof. Proof of the first property is by induction. For G_1,

$$\frac{1}{2} \cdot \text{flip} + \frac{1}{2} \cdot I = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} = \bot \in (2, 2)$$

Assuming the property holds for G_n, we now prove it holds for G_{n+1}. We have

$$\begin{aligned}\sum_{f_i \in G_{n+1}} \frac{f_i}{2^{n+1}} &= \sum_{g_i \in G_n} \frac{1}{2^{n+1}} \begin{pmatrix} g_i & 0 \\ 0 & g_i \end{pmatrix} + \sum_{g_i \in G_n} \frac{1}{2^{n+1}} \begin{pmatrix} 0 & g_i \\ g_i & 0 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} \sum g_i/2^n & 0 \\ 0 & \sum g_i/2^n \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & \sum g_i/2^n \\ \sum g_i/2^n & 0 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} \bot & 0 \\ 0 & \bot \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & \bot \\ \bot & 0 \end{pmatrix} \\ &= \bot \in (2^{n+1}, 2^{n+1})\end{aligned}$$

This equation implies that the elements of G_n are independent. First, the only entries in $f_i \in G_n$ are either 0 or 1. Next, if some $f \in G_n$ has a 1 in position (i, j) then all other members of G_n have a zero at position (i, j) – otherwise, the equation would be violated. □

Definition 4.4 *Let M be a finite monoid of order n within a real algebra. Its* convex closure *is*

$$\langle M \rangle = \left\{ \sum_{i=1}^{n} x_i f_i : x \in \Delta^n \ \& \ f_i \in M \right\}$$

Because M has only n elements, any convex sum with more than n terms can always be rewritten as having exactly n terms by adding probabilities. Similarly, a convex sum of less than n elements can be extended to one having exactly n elements by adjoining zeroes. Thus, $\langle M \rangle$ contains all finite convex sums formed with elements of M and is itself a convex set.

Next, if we fix a particular enumeration of the elements of M, say $M = \{f_1, \ldots, f_n\}$, then any two elements $x, y \in \langle M \rangle$ can be written

$$x = \sum_{i=1}^{n} x_i f_i \quad \& \quad y = \sum_{i=1}^{n} y_i f_i$$

The reason for this is that we can rearrange terms because addition is commutative (we will be more precise about this shortly). This is a useful trick to remember. For instance, to prove that $\langle M \rangle$ is a monoid we simply calculate

$$xy = \left(\sum_{i=1}^{n} x_i f_i \right) \left(\sum_{i=1}^{n} y_i f_i \right) = \sum_{1 \le i,j \le n} x_i y_j f_i f_j = \sum_{k=1}^{n} z_k f_k \in \langle M \rangle$$

where $z_k = \sum_{f_i f_j = f_k} x_i x_j$. Notice that the f_k in the rightmost sum range over all of M because some $f_i = 1$. Thus, $\langle M \rangle$ is a *convex monoid*.

By Lemma 4.3, then, the elements of $\langle G_n \rangle$ are in one to one correspondence with Δ^n. This implies the existence of a 'universal property' possessed by classical channels:

Theorem 4.5 *Let M be a convex monoid that contains a copy of the involution group of order n and G denote the involution group in (n, n). Then there is a function from $\langle G\rangle$ to M such that for all $x, y \in \langle G\rangle$ and $p \in [0, 1]$ we have*

- $\varphi(xy) = \varphi(x)\varphi(y)$ *and*
- $\varphi(px + (1-p)y) = p\,\varphi(x) + (1-p)\varphi(y)$

That is, there is a convex linear homomorphism $\varphi : \langle G\rangle \to M$.

Proof. First, let $V \subseteq M$ denote the involution group of order n. Then because we know that G and V are both isomorphic to $\mathbb{Z}_2^k$ with $n = 2^k$, there is an isomorphism $\theta : G \to V$. Define $\varphi : \langle G\rangle \to M$ by

$$\varphi\left(\sum_{i=1}^{n} x_i f_i\right) = \sum_{i=1}^{n} x_i \cdot \theta(f_i)$$

The value assigned by φ belongs to M because M is a convex monoid. What must be proven though is that φ is a well-defined function. Suppose we have an element of $\langle G\rangle$ written in two different ways

$$\sum_{i=1}^{n} x_i f_i = \sum_{i=1}^{n} y_i g_i$$

For each f_i there is some $g_j = f_i$. This defines a permutation $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ such that $g_{\sigma(i)} = f_i$. Using the commutativity of addition,

$$\sum_{i=1}^{n} x_i f_i = \sum_{i=1}^{n} y_i g_i = \sum_{i=1}^{n} y_{\sigma(i)} g_{\sigma(i)} = \sum_{i=1}^{n} y_{\sigma(i)} f_i$$

and so by Lemma 4.3, $x_i = y_{\sigma(i)}$, for all i. Thus,

$$\varphi\left(\sum_{i=1}^{n} x_i f_i\right) = \sum_{i=1}^{n} x_i \theta(f_i) = \sum_{i=1}^{n} y_{\sigma(i)} \theta(g_{\sigma(i)}) = \sum_{i=1}^{n} y_i \theta(g_i) = \varphi\left(\sum_{i=1}^{n} y_i g_i\right)$$

where the third equality again uses commutativity of addition. This proves φ is well-defined.

Now let $x, y \in \langle G\rangle$ be two elements. As just seen, we can write

$$x = \sum_{i=1}^{n} x_i f_i \quad \& \quad y = \sum_{i=1}^{n} y_i f_i$$

To show that φ is convex linear, we calculate

$$\varphi(px + (1-p)y) = \varphi\left(\sum_{i=1}^{n} (px_i + (1-p)y_i) f_i\right) = p\,\varphi(x) + (1-p)\varphi(y)$$

for $p \in [0,1]$. To see that φ is a homomorphism,

$$\varphi(xy) = \sum_{i=1}^{n} z_k \theta(f_k) = \sum_{i=1}^{n} z_k \theta(f_i)\theta(f_j) = \varphi(x)\varphi(y)$$

where $xy = \sum_{k=1}^{n} z_k f_k$ and $z_k = \sum_{f_i f_j = f_k} x_i x_j$. □

The proof of the last result actually shows that the statement of the theorem remains true when G is *any* involution group that forms an independent set in $\langle G \rangle$. The consequences of this result for communication are quite profound: *any time* a collection of channels contains a copy of an involution group, we can study the convex closure of this involution group as though they were *classical channels* – even if the channels are in fact quantum. Let us now turn to the surprising usefulness of this idea.

5 How to factor zero in quantum information theory

The Pauli spin operators are $\{I, \sigma_x, \sigma_y, \sigma_z\}$, given by

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Each is self adjoint and unitary, so they are all *involutions*: $\sigma_i^2 = I$ for all $i \in \{x, y, z\}$. Each spin operator defines a unital quantum channel given by

$$\varepsilon_i(\rho) = \sigma_i \rho \sigma_i$$

Here are the Bloch representations of these channels:

Definition 5.1 *The involutions*

$$s_x := r_x(\pi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad s_y := r_y(\pi) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad s_z := r_z(\pi) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

are called the spin channels.

The reason that s_x, s_y and s_z are the Bloch representations of the unital channels corresponding to the spin operators $\{\sigma_x, \sigma_y, \sigma_z\}$ is that $e^{-i\pi\sigma_x/2} = -i\sigma_x$, $e^{-i\pi\sigma_y/2} = -i\sigma_y$, $e^{-i\pi\sigma_z/2} = -i\sigma_z\}$ and that Bloch representations do not depend on phases. In addition, $s_x s_y = s_z$, $s_y s_z = s_x$ and $s_x s_z = s_y$, so that $\{I, s_x, s_y, s_z\}$ is the involution group of order four, usually called the *Klein four group*. Because the set of unital channels $\mathcal{U}$ is a convex monoid that contains a copy of the Klein four

group, Theorem 4.5 gives a convex linear homomorphism

$$\varphi : \langle G \rangle \to \mathcal{U}$$

which sends $\varphi(I) = I$ with G denoting the involution group in $(4, 4)$. This homomorphism will now allow us to answer the question of how to randomly flip a quantum bit.

First, the channel $\perp \in (4, 4)$ is an algebraic zero in the monoid $\langle G \rangle$, so $e := \varphi(\perp)$ is an algebraic zero in the image of φ. Thus,

$$s_x \cdot e = e \;\&\; s_y \cdot e = e$$

If we call $v = e(r)$, this means that

$$(v_x, -v_y, -v_z) = (v_x, v_y, v_z) \quad \& \quad (-v_x, v_y, -v_z) = (v_x, v_y, v_z)$$

which means that e is the zero matrix in $\mathcal{U}$. Thus, $\varphi(\perp) = 0$.

Let us now define $x := \varphi^{-1}(s_x)$, $y := \varphi^{-1}(s_y)$, $z := \varphi^{-1}(s_z)$. By Prop. 4.3,

$$\perp = \frac{1}{4}\,(I + x + y + z)$$

so we can conclude

$$0 = \varphi(\perp) = \frac{1}{4}\,(\varphi(I) + \varphi(x) + \varphi(y) + \varphi(z)) = \frac{1}{4}(I + s_x + s_y + s_z).$$

That is, one way to flip a quantum bit is to randomly choose between the four types of unitary evolution represented by I, s_x, s_y, s_z. We can say more than this though.

Because $\{I, x, y, z\}$ is the involution group of order 4, we must have $z = xy$, so we can also write

$$\perp = \frac{1}{4}\,(I + x + y + z) = \left(\frac{I + x}{2}\right)\left(\frac{I + y}{2}\right)$$

and again using the fact that φ is a convex linear homomorphism, derive

$$0 = \left(\frac{I + s_x}{2}\right)\left(\frac{I + s_y}{2}\right)$$

That is, we can nontrivially factor zero into the product of two quantum channels. But what physical process do these two channels describe?

6 Experimental significance

In the last section, we obtained a nontrivial factorization of zero and did so without doing *any quantum mechanics*. In fact, all we had to do was work with $(4, 4)$ classical channels. On the surface, this seems very encouraging. However, this

approach leaves open the possibility that the solution we have obtained is of no use experimentally. Let us now show that this is emphatically *not* the case.

Lemma 6.1 *If x is an involution, then its average with the identity*

$$f := \frac{I + x}{2}$$

is an idempotent. That is, $f^2 = f$.

Proof. We have

$$\begin{aligned} f^2 &= \frac{I \cdot I}{4} + \frac{I \cdot x}{4} + \frac{x \cdot I}{4} + \frac{x \cdot x}{4} \\ &= \frac{I}{4} + \frac{x}{4} + \frac{x}{4} + \frac{I}{4} \\ &= \frac{I + x}{2} \\ &= f \end{aligned}$$

which finishes the proof. □

Thus, both of the channels

$$f := \frac{I + s_x}{2} \quad \& \quad g := \frac{I + s_y}{2}$$

are idempotents whose product is $f \cdot g = 0$. The canonical example of an idempotent quantum channel of course is a *projective measurement* since any complete set of projections $\{P_i\}$ defines an idempotent quantum channel

$$\varepsilon(\rho) = \sum_i P_i \rho P_i$$

which describes the manner in which the state of the system evolves as a result of the measurement $\{P_i\}$. This intuition leads us to wonder if the idempotents f and g can be understood as projective measurements.

Proposition 6.2 *The channel $f = (I + s_x)/2$ is the Bloch representation of a measurement in the $\{|+\rangle, |-\rangle\}$ basis, while $g = (I + s_y)/2$ is the Bloch representation of a measurement in the $\{|0\rangle, |1\rangle\}$ basis.*

Proof. Let us first recall that the density operator $\rho = |\psi\rangle\langle\psi|$ associated with any qubit $|\psi\rangle = a|0\rangle + b|1\rangle$ is

$$\rho = \begin{pmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 + r_z & r_x - ir_y \\ r_x + ir_y & 1 - r_z \end{pmatrix}$$

where $r = (r_x, r_y, r_z) \in \mathbb{B}^3$ is the Bloch vector associated to ρ. To prove the first statement, the quantum channel corresponding to a measurement in the $\{|+\rangle, |-\rangle\}$ basis is

$$\varepsilon(\rho) = |+\rangle\langle+|\rho|+\rangle\langle+| + |-\rangle\langle-|\rho|-\rangle\langle-|$$

Writing ρ in terms of its Bloch vector (r_x, r_y, r_z) and noting that

$$|+\rangle\langle+| = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \quad \& \quad |-\rangle\langle-| = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}$$

we calculate

$$\varepsilon(\rho) = \frac{1}{2}\begin{pmatrix} 1 & r_x \\ r_x & 1 \end{pmatrix}$$

which says that the Bloch representation of ε takes (r_x, r_y, r_z) to $(r_x, 0, 0)$ i.e. it is the idempotent f.

For the second case, the channel which describes a measurement in the $\{|0\rangle, |1\rangle\}$ basis is

$$\varepsilon(\rho) = |0\rangle\langle0|\rho|0\rangle\langle0| + |1\rangle\langle1|\rho|1\rangle\langle1|$$

Using

$$|0\rangle\langle0| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \& \quad |1\rangle\langle1| = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

we find that

$$\varepsilon(\rho) = \frac{1}{2}\begin{pmatrix} 1 + r_z & 0 \\ 0 & 1 - r_z \end{pmatrix}$$

where (r_x, r_y, r_z) is the Bloch vector associated to ρ. This means $g(r_x, r_y, r_z) = (0, 0, r_z)$ is the Bloch representation of a measurement in the $\{|0\rangle, |1\rangle\}$ basis. □

And so we have derived a *physical answer* to the question "how do we randomly flip a quantum bit?" The answer: first measure the system in the X basis and then measure it in the Z basis, or vice-versa. These are operations that are routinely used in laboratories today. For instance, realizations of QKD (BB84) require them. Notice that this answer from a security standpoint is very different from saying "randomly choose one of the spin operators" for the simple fact that we must then confront the issue of *how* to randomly choose between spin operators. The result above – that zero can be factored as a product of measurements – is quite different. The problem of randomness is entirely left to Nature. All we have to do is perform two successive measurements and Nature provides the randomness for free: our actions are entirely deterministic. This factorization is possible because the representations of the spin operators are closed under multiplication. Another take on the factorization of zero is that it amounts to bit flipping with probability 1/2 *followed by* phase flipping with probability 1/2 (or vice-versa).

7 A mathematical aside

The homomorphism $\varphi : \langle G \rangle \to \mathcal{U}$ is an *isomorphism* onto its image, the convex closure of the spin channels. The reason is that the spin operators also form an independent set within their convex closure. This implies an isomorphism between $\langle G \rangle$, $\mathrm{Im}(\varphi)$ and Δ^4. Because Δ^4 forms a 'domain' in the sense of [2], it is then possible to extend algebraic information theory to the important class of channels $\langle G \rangle$. Mysteriously, channels of exactly this type have also arisen in the analysis of spatial domain image steganography [3].

8 Summary

This paper makes contributions to three different areas:

- In quantum security, it gives us an *experimentally realizable* method for randomly flipping qubits, which can be used to interrupt communication schemes or to remove steganographic messages from quantum data,
- In information theory, it shows that some problems in quantum information can be solved by reasoning only about classical channels,
- In group theory, it shows that the involution group can be constructed using channels from Shannon's information theory, and establishes its significance in both classical and quantum communication

We find the interplay between these three areas fascinating.

References

[1] E.E. Majani and H. Rumsey, *Two results on binary-input discrete memoryless channels.* Proceedings of the IEEE International Symposium on Information Theory, p.104–104, 1991.

[2] K. Martin, I.S.Moskowitz and G. Allwein, *Algebraic information theory for binary channels.* Electronic Notes in Theoretical Computer Science, Vol. 158, MFPS 2006, Special Session on Security, p.289–306, 2006.

[3] I.S. Moskowitz, P. A. Lafferty and F. Ahmed, *On LSB spatial domain steganography and channel capacity.* Naval Research Laboratory Memorandum Report, 2008.

[4] M. Nielsen and I. Chuang, Quantum computation and quantum information. Cambridge University Press, 2000.

[5] C. E. Shannon. *A mathematical theory of communication.* Bell Systems Technical Journal 27, 379–423 and 623–656, 1948.