qitd212

Quantum Error Correction

Robert B. Griffiths

Version of 30 March 2010

References:

QCQI = *Quantum Computation and Quantum Information* by Nielsen and Chuang (Cambridge, 2000), Secs. 10.1, 10.2, 10.3

E. Knill, R. Laflamme, "Theory of quantum error-correcting codes," Phys. Rev. A 55 (1997) 900. quant-ph/9604034

Contents

1 Introduction

★ It seems very unlikely that quantum computation can be realized unless there is some means of correcting the errors which will inevitably arise when physical devices are constructed to carry out such a computation. The situation is far different from that in ordinary "classical" computers in which for most purposes the probabilities of errors are so small that they can be ignored.

• The absence of errors in ordinary computers is related to the fact that bits are embodied in devices which are thermodynamically irreversible: 0 and 1 correspond to local free energy minima in a thermodynamic sense. But thermodynamic irreversibility is a great enemy of quantum computing, since it tends to decohere qubits, thus introducing unwanted noise into the quantum computation.

• Effective techniques for quantum error correction were first developed in 1996 by Shor, and independently by Steane. Up till then many skeptical physicists regarded quantum computing as totally impractical. With the development of error correction techniques, "totally impractical" was replaced with "extremely difficult."

• Hopefully, there will be further improvements in error correction methods as various physical realizations of quantum computers are developed. As well as clever error correction methods, one should be on the lookout for quantum algorithms which are more error-tolerant than those known at present.

★ Quantum error correction was developed in analogy with classical error correcting codes, but in the quantum case one needs to add a number of clever tricks. Rather than introducing these in the abstract, it is helpful to explore some simple examples in which very limited types of errors are allowed, and one can get an appreciation for some of these clever tricks. Sections 5 and 6 contain a general theory of error correction, which will be much easier to understand after exploring the examples considered here.

★ Classical error correction is based on *redundancy*: making several copies of information in different signals or different physical objects, so that if one or a few of these are lost or corrupted, the original information can be recovered from the ones that remain. Quantum error correction is based on the same general principle, but simply copying the information in the classical sense will not work, in view of no-cloning arguments. Hence the need for tricks. Nonetheless, classical error correction provides a useful starting point.

• The simplest form of redundancy is simple duplication: make a copy. Then if you lose the original, you still have the copy. A *two bit* code accomplishes this: in place of 0 use 00, in place of 1 use 11. Here 11 may stand for two pieces of paper with "1" written on them, or two pieces of paper of the same color, or two signals of the appropriate sort sent in succession through a communication channel, etc.

★ Error correction is possible in the case of duplication provided it is clear which of the copies represents the original information. If one copy is lost, then it is the remaining copy which carries the information. However, one copy may be corrupted in such a way that it is not clear which of the two carries the original information. For instance, a signal going through a channel might change from 1 to 0, or vice versa.

• To get around the problem just mentioned, one can increase the redundancy; e.g., use a three-bit repetition code in which 000 represents 0 and 111 represents 1. Then if something happens to just one of the copies, e.g., the third bit is corrupted, it is possible to recover the original information using "majority rule": if the signals emerging from the channel are 0, 0, and 1, one assumes that the original codeword was 000.

2 Two Qubit Code

★ We begin the exploration of quantum error correction using a two qubit code in which the *logical* state $|0\rangle_L$ we wish to encode is represented by the $|00\rangle$ state of two *carrier* qubits or *carriers* (often referred to as *physical* qubits), and $|1\rangle_L$ by $|11\rangle$. Linear combinations of the type $\alpha|0\rangle_L + \beta|1\rangle_L$ are represented by $\alpha|00\rangle + \beta|11\rangle$.

• It is helpful to think of this code as produced by a *coding circuit* shown in Fig. 1. One can easily see that if the first qubit is in the state

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \tag{1}$$

and the second, or ancillary, qubit in the state $|0\rangle$ at the initial time t_0, then at time t_1 the combined state of the two qubits is

$$|\Psi_1\rangle = \alpha|00\rangle + \beta|11\rangle. \tag{2}$$

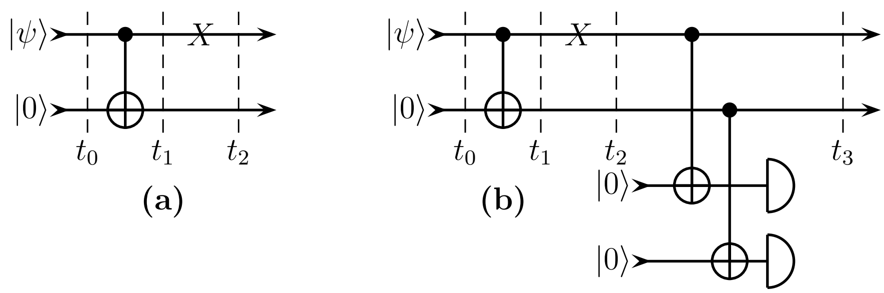

Figure 1: (a) Two-qubit coding circuit followed by a possible error X. (b) Nondestructive measurement scheme which will not recover input information.

★ Now consider a very simple sort of error. During the time interval between t_1 and t_2, the first qubit can either remain the same (no error) or be subjected to a unitary transformation X (σ_x) to produce a "bit flip error". On the figure this is indicated by an X placed over the line representing the qubit. (The same X inside a square box would indicate the corresponding 1-qubit gate as something happending every time the circuit is used.) Whether or not the error occurs could depend upon some interaction with the environment. Can we recover the original quantum state (1) when an error of this sort has occurred, or, to be more precise, when an error of this sort *might* have occurred?

• The "classical" solution would be to simply throw away the (possibly) corrupted first qubit and use the second. But this will not work in the quantum case, for if we ignore the first qubit the second qubit is described by a density operator

$$\rho = |\alpha|^2[0] + |\beta|^2[1]. \tag{3}$$

Only if $\alpha = 0$ or $\beta = 0$ is this a pure state, and in any case ρ contains no information about the relative phases of α and β.

• Measurements of the sort indicated in Fig. 1(b), where two ancillary qubits are used in order to allow nondestructive measurements of both code qubits in the standard basis, are not a good method for recovering from an error.

□ Exercise. Analyze Fig. 1(b) by working out the states of the two code qubits at t_3 conditional on the measurement outcomes, and show that one cannot, in general, recover the original $|\psi\rangle$.

★ There is, however, a solution to the problem based upon carrying out a measurement of the right sort. This is the first of the clever tricks associated with quantum error correction. To motivate it, note that the state $|\Psi_2\rangle$ at t_2 in Fig. 1 is the same as $|\Psi_1\rangle$ in (2) if no error occurs, whereas if a bit-flip error does occur, then it is

$$|\Psi_2'\rangle = \alpha|10\rangle + \beta|01\rangle. \tag{4}$$

A comparision of (4) with (2) shows that even though neither qubit has a definite value in either of these entangled states, they differ in that the labels are either identical in both kets making up the superposition, or they are opposite (1 vs. 0). This suggests carrying out a measurement of the property of "sameness" in order to determine whether an error has occurred.

∘ To be precise, "sameness" is a property associated with the Hermitian operator $Z_a Z_b$, where the subscripts refer to qubits a and b—we assume that a is above b in Fig. 2. Thus an eigenstate of $Z_a Z_b$ with the eigenvalue $+1$ has the property that the Z values are the same, and an eigenvalue -1 means the Z values are different. In spin-half terms, the values of S_z for the two particles are either the same, or they are opposite.

□ Exercise. Show that $|\Psi_1\rangle$ in (2) is an eigenstate of $Z_a Z_b$ with eigenvalue $+1$ whatever the values of α and β, so one can say that the Z values are the same, whereas $|\Psi_2'\rangle$ in (4) is an eigenstate with eigenvalue -1, again independent of α and β: the Z values are different.

• The measurement can be done using the arrangement shown in Fig. 2(a). The detector will register a 1 if an error has occurred, and a 0 if an error has not occurred. If an error has occurred, it can be corrected by applying an X gate to qubit 1.

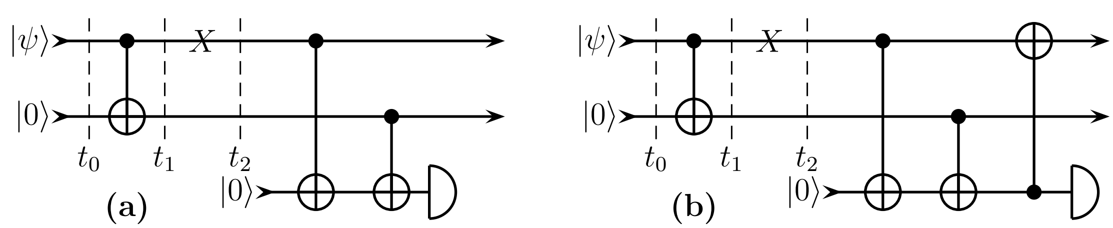

Figure 2: Quantum error correction. (a) Measurement outcome can be used to correct error. (b) Circuit automatically corrects error.

• The error correction can be implemented "automatically" using the quantum circuit in Fig. 2(b). In this case it is not necessary to carry out the measurement on the third qubit, which can be simply thrown away.

∘ Or the third qubit can be measured, in which case its value represents the "syndrome," and tells one whether or not the X error actually occurred between t_1 and t_2.

□ Exercise. Work out the unitary time transformation corresponding to Fig. 2(b), and verify that the initial $|\psi\rangle$ emerges in the first qubit after the final CNOT operation, whether or not the third qubit is measured. Show that if the third qubit is measured its value indicates whether or not the X error occurred.

• Does not a measurement always perturb a quantum system in an uncontrolled way? There is some justification behind this piece of folklore, but clear thinking requires greater precision. Figure 2 shows that it is sometimes possible to measure a *particular kind* of information about a system without producing an uncontrolled perturbation on some other type of information one is interested in.

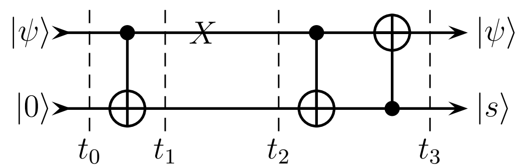

Figure 3: Quantum error correction. The last two CNOT gates constitute a unitary decoding mechanism that corrects the error.

★ The extra or *ancillary* third qubit in Fig. 2 is not really essential. The circuit in Fig. 3 will do just as well. The last two CNOT gates constitute a *decoding circuit* $\mathcal{D}$.

☐ Exercise. Check that the circuit in Fig. 3 will correct an X error between t_1 and t_2. How could one determine the syndrome?

★ That decoding is, indeed, possible using a unitary operator D can be seen by constructing a table of what one wants D to do, see Table 1.

Table 1: Method to obtain D

$$
\begin{array}{ccccccc}
t_0 & & t_1 & & t_2 & & t_3 \\
|00\rangle & \rightarrow & |00\rangle & \rightarrow & \left\{ \begin{array}{l} |00\rangle \\ |10\rangle \end{array} \right. & \begin{array}{c} \rightarrow \\ \rightarrow \end{array} & \begin{array}{l} |00\rangle \\ |01\rangle \end{array} \\
|10\rangle & \rightarrow & |11\rangle & \rightarrow & \left\{ \begin{array}{l} |11\rangle \\ |01\rangle \end{array} \right. & \begin{array}{c} \rightarrow \\ \rightarrow \end{array} & \begin{array}{l} |10\rangle \\ |11\rangle \end{array}
\end{array}
$$

• The kets at t_2 in Table 1 depend both on the input at t_0 and upon whether an X error has (lower) or has not (upper) occurred between t_1 and t_2. The kets at t_3 have been chosen so tha (i) the first or a qubit is the same as at t_0, i.e., the error has been corrected, and (ii) the second or b qubit is in state $|0\rangle$ if no error has occurred, and $|1\rangle$ if an error has occurred. One could equally well interchanged 0 and 1 for the second qubit. The fact that an orthonormal basis of 2 qubits in the t_2 column is mapped to an orthonormal basis in the t_3 column means the D operator is unitary, and a little guesswork yields the circuit in Fig. 3.

★ If in place of an X error on the first qubit in Fig. 3 there is a Z or "phase flip" error, this error cannot be corrected. In a Z or phase flip error one has $|0\rangle \rightarrow Z|0\rangle = |0\rangle$, $|1\rangle \rightarrow Z|1\rangle = -|1\rangle$.

• Such errors are not trivial. In quantum mechanics the *overall* phase of a ket of a quantum state has no physical significance, but *relative* phases inside a superposition are very important. Thus $\alpha|0\rangle - \beta|1\rangle$ does *not* represent the same thing as $\alpha|0\rangle + \beta|1\rangle$, except when $\alpha = 0$ or $\beta = 0$.

• An easy way to see that the phase flip error cannot be corrected is to note that if it occurs the state at t_2 will be

$$|\Psi_2''\rangle = \alpha|00\rangle - \beta|11\rangle. \tag{5}$$

But this is precisely the same as if in the initial input $|\psi\rangle$ β had had the opposite sign, and no error had occurred. That is to say, $|\Psi_2''\rangle$ carries no information indicating that an error has occurred, quite unlike $|\Psi_2'\rangle$ in (4). So error correction is impossible.

3 Three-Qubit Code

• See the description in QCQI Sec. 10.1.1. A single qubit is encoded using the circuit in Fig. 4(a) in three carrier qubits. As a result $|\Psi_1\rangle$ at t_1, compare (2), is

$$|\Psi_1\rangle = \alpha|000\rangle + \beta|111\rangle. \tag{6}$$

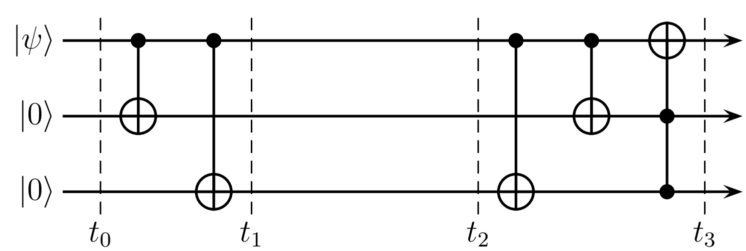

Figure 4: Three qubit encoding and decoding circuit corrects an X error on a single carrier if it occurs at a time between t_1 and t_2.

★ Let us now suppose that between t_1 and t_2 a bit flip error might occur on the first or second or third carrier. but not on more than one carrier. The result will be one of the four possibilities

$$\alpha|000\rangle + \beta|111\rangle, \quad \alpha|100\rangle + \beta|011\rangle, \quad \alpha|010\rangle + \beta|101\rangle, \quad \alpha|001\rangle + \beta|110\rangle \tag{7}$$

at t_2. If we know on which of the three carriers the bit flip occurred, we can correct it using an obvious extension of the method indicated in Sec. 2; see, in particular, Fig. 2(b). The situation where we don't know which carrier was affected, or whether an error actually occurred, is more complicated. Measuring the value of individual qubits obviously won't work. However, as in Sec. 2, measuring whether or not two qubits are the *same* or *different* in the standard basis provides a way of extracting information about where the error has occurred without disturbing the quantum information.

• Suppose the first two carriers are different in the sense that $Z_a Z_b = -1$. This means—take a look at (7)—that the error occurred either on carrier 1 or on carrier 2. We do not know which. However, if we determine "same" or "different" for *two different pairs* of carriers, this will tell us exactly where the error occurred, and having determined its location we can then correct it, by applying an X to the appropriate carrier.

□ Exercise. Design a circuit analogous to that in Fig. 2(b), but of course more complicated, which can be used with the help of ancillary bits (you can use three, but two suffice) to automatically correct a bit flip error on a *single* carrier. [Hint. The correction operations can be carried out fairly simply using Toffoli gates.]

★ Rather than using ancillary qubits, one can design a "compact" error correcting circuit by means of a suitable *decoding* operation shown in the circuit in Fig. 4 between t_2 and t_3. The corresponding unitary operator D acts in such a way that the desired information $|\psi\rangle$ emerges in the first qubit, while the ancillary qubits are left in a state that contains information about the syndrome—the nature of the error—but *no* information about $|\psi\rangle$ itself.

• Although we have the three qubit code in mind, it is helpful to think of Fig. 4 as representing in a schematic fashion a very general scheme of error correction, in which the number of ancillary qubits could be very large, and $|\psi\rangle$ might be a state on a Hilbert space of arbitrarily large dimension. The only thing special is that we assume that at the end the original information is perfectly restored: $|\psi\rangle$ out is the same as $|\psi\rangle$ in.

□ Exercise. Check that the decoding circuit in Fig. 4 does what it is supposed to do if there is an X error on one but not more than one of the three carriers. What happens if there is an X error on two carriers? A Z error on one carrier or two carriers? A Y (or XZ) error on one carrier?

□ Exercise. Instead of using Fig. 4 work out D yourself using a table similar to Table 1, but with 8 entries in the t_2 column corresponding to the different kets in (7). Make appropriate choices for entries in the t_3 column (there is more than one way to do this), and then check, using unitary time development for an initial $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, that your scheme actually works.

★ While the coding and decoding arrangement in Fig. 4 will correct an X error on any carrier, it will not correct a Z or phase flip error in which $|0\rangle \to |0\rangle$ and $|1\rangle \to -|1\rangle$, so that $\alpha|0\rangle + \beta|1\rangle$ is transformed to $\alpha|0\rangle - \beta|1\rangle$. The effect of a Z error on any one of three carriers in Fig. 4 during the time between t_1 and t_2 is to transform $|\Psi_1\rangle = \alpha|000\rangle + \beta|111\rangle$ into $|\Psi_2\rangle = \alpha|000\rangle - \beta|111\rangle$, which is just what $|\Psi_1\rangle$ would have been if the sign of β in the initial state $|\psi\rangle$ had been different. Obviously there is no way of correcting this kind of error, since there is no indication in the state $|\Psi_2\rangle$ itself that anything is wrong. Consequently, the 3 qubit code we are using is incapable of correcting Z errors.

• However, an alternative procedure can be used to correct any Z error on a single qubit: use the codewords $|+++\rangle$ and $|---\rangle$ to represent the logical states $|+\rangle_L$ and $|-\rangle_L$. Since $Z|+\rangle = |-\rangle$ and $Z|-\rangle = |+\rangle$, all we have done is to interchange the roles of X and Z.

• A simple way of constructing the coding and decoding circuit in this case is shown in Fig. 5, obtained by adding Hadamards at strategic points to the circuit in Fig. 4.

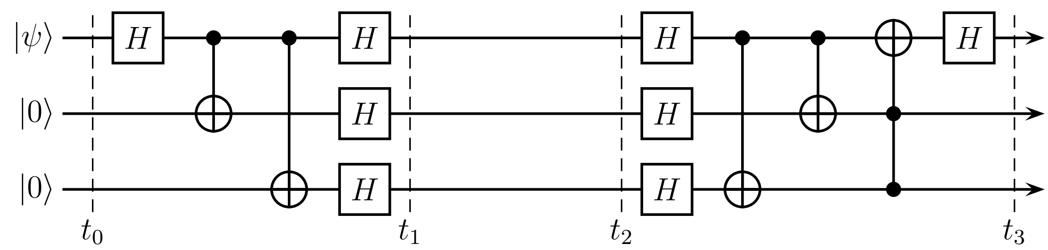

Figure 5: Three qubit encoding and decoding circuit corrects a Z error on a single carrier occurring during the interval $t_1 < t < t_2$.

☐ Exercise. Check that the encoding part of the circuit (up to t_1) in Fig. 5 does what it is supposed to, i.e., an initial $|+\rangle$ is encoded as $|+++\rangle$, and $|-\rangle$ as $|---\rangle$.

☐ Exercise. By working through the unitary transformations corresponding to the different gates, show that the circuit in Fig. 5 will correct a Z (phase flip) error occurring on one of the carriers between t_1 and t_2.

☐ Exercise. Show that the first and last H gates in Fig. 5 acting on the first qubit are not actually needed in terms of recovering from the effects of a Z error on a single qubit.

☐ Exercise. Suppose one of the carriers in Fig. 5 suffers an X error during $t_1 < t < t_2$. How does this affect what emerges as the first qubit at t_3?

4 Nine Qubit Code

★ We have seen in Sec. 3 how a three qubit code allows one to correct an X (bit flip) error on any carrier but not Z (phase flip) errors, while a different code on three qubits permits the correction of an X error on any carrier, but not Z errors. Neither code corrects both X and Z errors, and neither corrects Y errors, though of course we could design a different three qubit code that would correct Y, but not X or Z errors. A Y error is the same as an X error followed by a Z error or a Z error followed by an X error, since the difference in phase between Y, ZX, and XZ can for this purpose be ignored.

☐ Exercise. Construct the code that allows correction of a Y error if it occurs on only one carrier, and design the corresponding coding and decoding circuit. [Hint: One should replace H in Fig. 5 with something else. What should it be?]

• The shortest quantum code that will allow the correction of an X or Y or Z (or an arbitrary error, see Sec. 5) on any single carrier is a five qubit code, see QCQI Sec. 10.5.6.

★ Shor's nine qubit code was the first quantum code to be discovered that has the property that it allows recovery from an arbitary error on any one of the carriers. Though more efficient codes exist, QCQI Sec. 10.5.6, the nine qubit code is worth studying in that it allows one to "see" rather easily how the error recovery process works. It also illustrates the clever and important *concatenation* strategy for constructing error correcting codes.

• The code itself is easily written down

$$\begin{aligned} |0\rangle_L &= \Big(|000\rangle + |111\rangle\Big) \otimes \Big(|000\rangle + |111\rangle\Big) \otimes \Big(|000\rangle + |111\rangle\Big)/\sqrt{8} \\ |1\rangle_L &= \Big(|000\rangle - |111\rangle\Big) \otimes \Big(|000\rangle - |111\rangle\Big) \otimes \Big(|000\rangle - |111\rangle\Big)/\sqrt{8} \end{aligned} \tag{8}$$

Note how the nine qubits are divided into three blocks of three

★ To understand how the code works it is helpful to construct a circuit, the analog of Figs. 4 and 5, that does the coding and decoding, see Fig. 6.

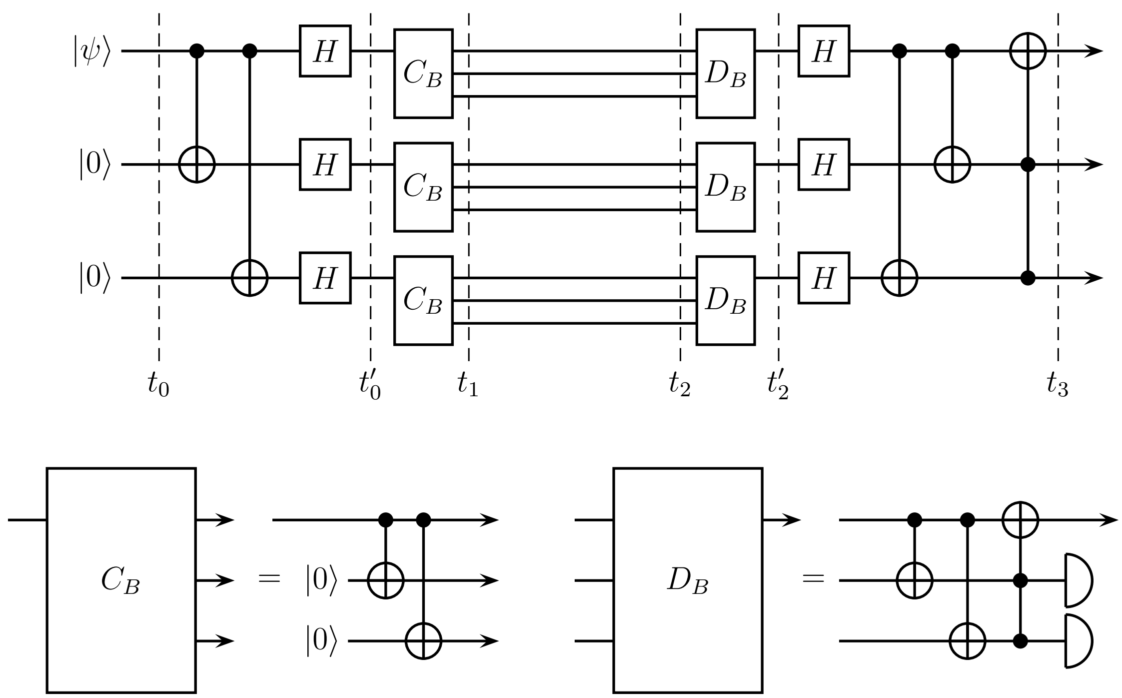

Figure 6: Nine qubit encoding and decoding circuit corrects any error on a single carrier, assuming it occurs during the interval $t_1 < t < t_2$.

• The 3 encoding boxes C_B include ancillary bits that are initially in the $|0\rangle$ state. Since these are fixed, one can regard C_B as an isometry ($C_B^\dagger C_B = I$) from the (variable) input qubit, Hilbert space dimension 2, entering the C_B box on the left, to the 3 qubits, Hilbert space of dimension 8, emerging on the right.

□ Exercise. Show that the encoding circuit in Fig. 6 produces the result in (8)

□ Exercise. Convince yourself that the decoding circuit in Fig. 6 works at least to the extent that if no errors occur between t_1 and t_2, an initial $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ in the first qubit at t_0 will emerge in the same state at t_3.

• The decoding D_B boxes contain measurements in the standard basis. These measurements are *not* needed for the decoding operation—one could simply throw the extra qubits away—but measuring them tells one somethng about the error syndrome.

★ To see how the nine qubit code works, suppose than an X error occurs on one of the nine carriers during the time interval between t_1 and t_2. Since this carrier lies between a C_B and a D_B, the error will be corrected (or eliminated) by the process described earlier in connection with the circuit in Fig. 4

◦ Indeed, one could tolerate up to three X errors provided they occur in different blocks. Thus even if X_1 and X_4 and X_9 occur simultaneously, they will all be corrected. But if X_1 and X_3 occur simultaneously the result will be an error that is not corrected by the circuit.

• Now suppose that a Z error occurs on one of the carriers, e.g., the first carrier in the first block. As noted in the discussion in Sec. 3, it will not be corrected by the encoding-decoding uperation represented by the first (uppermost) pair of C_B and D_B boxes in Fig. 6. Instead, it will be "passed along" and have exactly the same effect as if the top C_B and D_B boxes were missing and a Z error occurred on a single qubit carrier connecting the top two H gates. Thus if no errors occur in any of the other 8 carriers, a Z error on the first carrier has the same sort of effect as a Z error on the uppermost carrier in Fig. 5. But then the initial and final parts of the circuit in Fig. 6, those preceding t'_0 and following t'_2, will eliminate this error in the same way as the circuit in Fig. 5.

• What about a Y error on one of the 9 carriers? Assume this error occurs on the first carrier. So far as the inner part of the correction circuit in Fig. 6 is concerned, the part involving the top C_B and D_B boxes, the effect is the same as a Y error on the first carrier in Fig. 4. It is a straightforward exercise (did you do it already?) to show that that circuit will correct the X part but leave the Z part present. (There could also be an additional overall phase, but it does not matter.) But then the Z part will be eliminated, as already noted, by the outer part of the encoding/decoding circuit in Fig. 6.

★ In conclusion, we have shown that the nine qubit code when made part of the circuit in Fig. 6 will correct any of the 3 errors X or Y or Z, provided it occurs in a *single* carrier. Thus this code accomplishes in the quantum domain something very similar to the three bit repetition code of Sec. 1: the automatic correction of an error on any of the carriers.

• But why should errors be restricted to X or Y or Z? Cannot one imagine something that lies "in between," say some linear combination of X and Z? Yes one can, and errors need not even be represented by unitary operators. That is why we need to supplement the preceding examples with a general theory of (this kind of) error correction.

5 General Theory of Error Correction

★ The states $|0\rangle$ and $|1\rangle$ form an orthonormal basis of the Hilbert space $\mathcal{H}$ of a single qubit, but $\mathcal{H}$ itself consists of more than $|0\rangle$ and $|1\rangle$: it includes all linear combinations of these basis kets. In quantum mechanics it is the Hilbert space which is the "fundamental" mathematical structure, while there are many possible choices for bases, even orthonormal bases. The choice of basis is a matter of convenience.

• Similarly, a quantum code is best thought of not just as a collection of codewords, as in classical codes, but as a *subspace* $\mathcal{P}$ of the Hilbert space $\mathcal{H}_c$ of the code carriers, a subspace which is *spanned* by (made up of all linear combinations of) a collection of codewords $\{|c_j\rangle\}$, $1 \leq j \leq K$, which will hereafter be referred to as the *coding (sub)space* While it is customary and convenient to use a particular basis for this subspace, from the point of view of fundamental quantum mechanics, and of quantum error correction of the sort we are considering, the choice of basis is arbitrary; what counts is the subspace itself.

◦ In the examples in Secs. 2 to 4 we considered $K = 2$, thus two-dimensional subspaces. There are many interesting quantum codes with $K > 2$, and the general theory discussed here applies for arbitrary (finite) K.

★ We begin with the following model of encoding and errors. The quantum information of interest is a ket $|\psi\rangle$ in a Hilbert space $\mathcal{H}_a$ of dimension K. In addition there is an ancillary space $\mathcal{H}_b$ which is initially in a definite state $|b_0\rangle$. For example, $\mathcal{H}_a$ could be a single qubit, and $\mathcal{H}_b$ a set of ancillary qubits in a state $|b_0\rangle = |00\cdots 0\rangle$.

★ The information initially in $\mathcal{H}_a$ is *encoded* by a unitary transformation $\hat{C}$ which maps $\mathcal{H}_a \otimes \mathcal{H}_b$ to a space $\mathcal{H}_c$, the Hilbert space of the *code carriers* (or simply "carriers"). In particular, if $\{|a_p\rangle\}$ is an orthonormal basis of $\mathcal{H}_a$, the collection of kets

$$\hat{C}(|a_p\rangle \otimes |b_0\rangle = C|a_p\rangle = |c_p\rangle \tag{9}$$

span the coding subspace.

• There is no loss of generality, and formulas become a bit simpler, if one replaces $\hat{C}$ with the *isometry* C mapping $\mathcal{H}_a$ to $\mathcal{H}_c$, defined in (9) by its action on each of the basis states $|a_p\rangle$.

◦ The isometry $C : \mathcal{H}_a \to \mathcal{H}_c$ is like a unitary, except that it maps the smaller Hilbert space $\mathcal{H}_a$ onto the subspace $\mathcal{P}$ of $\mathcal{H}_c$, rather than onto all of $\mathcal{H}_c$. In particular it satisfies the conditions

$$C^\dagger C = I_a, \;\; CC^\dagger = P, \tag{10}$$

where P is the projector onto the subspace $\mathcal{P}$, thus in effect the identity operator on this subspace. In particular, C preserves inner products: $(C|a'\rangle)^\dagger C|a\rangle = \langle a'|a\rangle$, which justifies calling it an isometry (i.e., it preserves the metric).

◦ In what follows one could use the unitary $\hat{C}$ in place of the isometry C at the cost of carrying along the $|b_0\rangle$ from (9) in various formulas.

• Errors are introduced by interactions between the carriers, Hilbert space $\mathcal{H}_c$, and an environment. The effects of the environment on the quantum state of $\mathcal{H}_c$ can be represented by a collection $\mathcal{K}$ of Kraus operators $\{K_i\}$ satisfying the normalization condition $\sum_i K_i^\dagger K_i = I_c$, mapping $\mathcal{H}_c$ to itself, which we shall refer to informally as "errors." As a consequence the quantum state in the time interval of interest to us, from t_1 to t_2, is represented by the map

$$\rho_1 \to \rho_2 = \hat{\mathcal{K}}(\rho_1) := \sum_i K_i \rho_1 K_i^\dagger. \tag{11}$$

○ As usual, one can think of the transformation (11) as resulting from a unitary transformation mapping the Hilbert space $\mathcal{H}_c \otimes \mathcal{H}_e$ to itself, where $\mathcal{H}_e$ is the Hilbert space of the environment, assumed to be initially in some fixed pure state. After this the environment is ignored.

★ Next assume that there is a *decoding* operator D, an isometry mapping $\mathcal{H}_c$ to a space $\mathcal{H}_a \otimes \mathcal{H}_f$, such that for every $|\psi\rangle$ in $\mathcal{H}_a$ and every K_i in $\mathcal{K}$,

$$DK_iC|\psi\rangle = |\psi\rangle \otimes |s_i\rangle, \tag{12}$$

where $|s_i\rangle \in \mathcal{H}_f$ is a *syndrome*. Note that the syndromes are, in general, neither orthogonal nor normalized, and that $|s_i\rangle$ depends (or course) on K_i, but is independent of $|\psi\rangle$, i.e., (12) holds for any $|\psi\rangle$ in $\mathcal{H}_a$.

○ In the examples in Secs. 2, 3 and 4 D is a unitary operator, see Figs. 2, 3 and 4. However, in (12) we only require that it be an isometry. What this means is that the recovery operation may involve an additional ancillary system or ancilla prepared in a particular state, which is made to interact with the carriers through some unitary operation.

• Only for fairly special collections $\mathcal{K}$ of Kraus operators, i.e., special forms of interaction with the environment, will such a D exist. What one does in practice is to identify some subcollection of $\mathcal{K}$ as constituting the "important" or "most significant" errors, and then find a D which works for this subcollection.

○ There is a classical analog. Consider the three bit repetition code, 000 and 111, which is designed for correcting errors on a single bit, but cannot correct errors on two or more bits. This makes sense when the probabilities of errors on two or three bits are much smaller than the probability of an error on one bit alone. In this case the "important" errors are the one bit errors.

• The scheme in (12) is a quite general form of *perfect* error correction, in the sense that at the end the specified errors have no effect on the quantum information in $\mathcal{H}_a$. If an isometry of this form does not exist, then the information encoded in the carriers cannot be perfectly recovered. (Imperfect error recovery is outside the scope of these notes.)

★ We can turn (12) into a definition. Let the isometries C and D be given. Then a *correctable error* (relative to C and D) is any operator E on the Hilbert space $\mathcal{H}_c$ of carriers such that

$$DEC|\psi\rangle = |\psi\rangle \otimes |s(E)\rangle \tag{13}$$

for *every* $|\psi\rangle$ in $\mathcal{H}_a$, where the syndrome $|s(E)\rangle$ will in general depend upon the operator E, but *not* on $|\psi\rangle$.

• It is evident that if (13) holds for some E_1 with syndrome $|s_1\rangle$ and E_2 with syndrome $|s_2\rangle$, it will also hold for $\alpha E_1 + \beta E_2$, with a syndrome $|s\rangle = \alpha|s_1\rangle + \beta|s_2\rangle$. That is to say, the set of correctable errors forms a *linear vector space* $\mathcal{E}_c$ of operators on $\mathcal{H}_c$.

∘ Recall that for a Hilbert space $\mathcal{H}$ of dimension d, the collection of all operators is a linear vector space of dimension d^2, and it is itself a Hilbert space if one uses an operator inner product $\langle A, B\rangle = \mathrm{Tr}(A^\dagger B)$. The space $\mathcal{E}_c$ of correctable errors is a subspace of this space of operators for the Hilbert space $\mathcal{H}_c$.

∘ Note that $\mathcal{E}_c$ depends both on the encoding transformation C (which in turn depends on $|b_0\rangle$ and $\hat{C}$ in (9)) and on the decoding transformation D. For a given C there may be more than one decoding transformation D and thus more than one space of correctable errors. (See Sec. 6 below for a condition on $\mathcal{E}_c$ that guarantees the existence of a decoding operation D, and indicates how to construct it.)

∘ One usually assumes that the identity I is a member of $\mathcal{E}_c$, i.e., if there is no error, then D decodes things correctly. However, this is not essential.

★ As a particular (and very important) application, note that any operator on the space of one qubit can be written as a linear combination of the four operators

$$I, \quad X = \sigma_x, \quad Y = \sigma_y, \quad Z = \sigma_z, \tag{14}$$

which form a basis of the operator space. Thus if one can show that some error correction protocol, i.e., some D, corrects errors of the type X, Y, and Z on a particular qubit, and also gives the right answer if there is no error at all (the identity I), it will correct any and all errors on this qubit.

• Consider, in particular, Shor's 9-bit code with an appropriate D. One can show explicitly that it corrects errors X_j, Z_j and $X_j Z_j = -iY_j$ on the j'th qubit. Consequently it can correct any and all errors on the j'th qubit, denoted by a subscript, thus

$$X_1 = X \otimes I \otimes I \otimes \cdots, \quad Z_2 = I \otimes Z \otimes I \otimes \cdots, \tag{15}$$

and so forth.

★ However, the fact that $\mathcal{E}_c$ is a linear space of operators does *not* mean that *products* of operators in $\mathcal{E}_c$ are in $\mathcal{E}_c$. Thus it may well be the case that E and E' are members of $\mathcal{E}_c$, whereas EE' is not in $\mathcal{E}_c$.

∘ Again, Shor's 9-bit code provides an example. Both X_1 and X_2 are in D, so that a bit flip of qubit 1 is a correctable error, as is a bit flip of qubit 2. But the product $X_1 X_2 = X_1 \otimes X_2$, which means flipping both 1 and 2, is *not* in the space of correctable errors. On the other hand, $X_1 X_4$ *is* in $\mathcal{E}_c$, but this is something one has to work out in terms of the structure of the code; it does not follow from the fact that both X_1 and X_4 are in $\mathcal{E}_c$.

6 Subspace Condition

★ As noted in Sec. 5, code words of a quantum error correcting code span a linear subspace $\mathcal{P}$, the coding (sub)space of the Hilbert space $\mathcal{H}_c$ of the code carriers. One can use a condition due to Knill and Laflamme (1997) to help identify (operator) spaces $\mathcal{E}_c$ of correctable errors—there may be more than one interesting $\mathcal{E}_c$—using properties of $\mathcal{P}$ or, equivalently, the projector P onto $\mathcal{P}$, without first having to look for a decoding operation D.

◦ For the three qubit code of Sec. 3 with code words $|000\rangle$ an $|111\rangle$, $\mathcal{P}$ consists of all their linear combinations, and the projector is

$$P = |000\rangle\langle 000| + |111\rangle\langle 111|. \tag{16}$$

★ The Knill and Laflamme *projector condition* says that a linear space $\mathcal{E}_c$ of operators is correctable (i.e., there exists a decoding operation D in the sense of (13)) if and only if

$$PE^\dagger \bar{E} P = \alpha(E, \bar{E})P \tag{17}$$

whenever E and $\bar{E}$ are any two elements of $\mathcal{E}_c$, where $\alpha(E, \bar{E})$ is some complex number depending on the two operators.

• The projector P onto the coding space $\mathcal{P}$ does not depend upon the choice of a basis for $\mathcal{P}$, but it is often convenient to choose an orthonormal collection $\{|c_p\rangle\}$ of code words that span $\mathcal{P}$, and write

$$\langle c_p|E^\dagger \bar{E}|c_q\rangle = \alpha(E, \bar{E})\delta_{pq}. \tag{18}$$

□ Exercise. Prove the equivalence of (17) and (18). Hint: $P = \sum_p |c_p\rangle\langle c_p|$.

★ Since (17) or (18) holds for all operators in the linear space $\mathcal{E}_c$, they also holds when E and $\bar{E}$ are elements belonging to some basis $\{E_j\}$ of operators in $\mathcal{E}_c$. Conversely, it suffices to check (17) or (18) using operators belong to such a basis, i.e., it is enough to show that

$$PE_j^\dagger E_k P = \alpha_{jk} P \tag{19}$$

or

$$\langle c_p|E_j^\dagger E_k|c_q\rangle = \alpha_{jk}\delta_{pq}, \tag{20}$$

or where $\alpha_{jk} = \alpha(E_j, E_k)$ is a matrix of complex numbers.

□ Exercise. Show that (19) implies (17) or (18), i.e., it suffices to check the latter for operators belonging to the basis $\{E_j\}$.

□ Exercise. Show that α_{jk} is a positive matrix, i.e., a Hermitian matrix with nonnegative eigenvalues. [Hint. Use the adjoint of (19) to establish Hermiticity. Check positivity of eigenvalues by showing that for any collection of complex numbers $\{\beta_j\}$ it is the case that $\sum_{jk} \beta_j^* \alpha_{jk} \beta_k \geq 0$. Recall that for any operator B, $B^\dagger B$ is a positive operator.]

• Since α_{ij} is a Hermitian matrix it can be diagonalized using a unitary matrix, i.e.,

$$\alpha_{ij} = \sum_k u_{ik} d_k u^*_{jk}, \tag{21}$$

with eigenvalues $d_k \geq 0$. This means we can define a new set of operators

$$F_k = \sum_i u_{ik} E_i, \tag{22}$$

forming a basis of $\mathcal{E}_c$, and for which (19) takes the form

$$PF_k^\dagger F_l P = \delta_{kl} d_k P. \tag{23}$$

• In general some of the d_k will be zero, and we shall call these "null errors" (see below). For the cases with $d_k > 0$ define the *principal errors*

$$G_k = F_k/\sqrt{d_k} \tag{24}$$

which satisfy the simple relationship

$$PG_k^\dagger G_l P = \delta_{kl} P \tag{25}$$

or, equivalently, using the basis $\{|c_p\rangle\}$,

$$\langle c_p|G_k^\dagger G_l|c_q\rangle = \delta_{kl}\delta_{pq}. \tag{26}$$

◦ The "null errors" with $d_k = 0$ satisfy

$$\langle c_p|F_k^\dagger F_k|c_p\rangle = 0, \tag{27}$$

which implies that

$$F_k|c_p\rangle = 0. \tag{28}$$

While this does not mean that F_k is zero as an operator, it does mean that such "errors" occur with zero probability for any state in the code space $\mathcal{P}$. They are, nonetheless, a nuisance which needs to be taken account of when constructing proofs.

★ Let us explore the significance of (26) by defining

$$|c_p^k\rangle := G_k|c_p\rangle. \tag{29}$$

Then (26) becomes

$$\langle c_p^k|c_q^l\rangle = \delta_{kl}\delta_{pq}, \tag{30}$$

or, in other words, $\{|c_p^k\rangle\}$ is an orthonormal collection of vectors labeled by two sets of indices, p and k. In the case $k = l$, the significance of (30) is that G_k maps the code space $\mathcal{P}$ onto another subspace $\mathcal{P}^k$ of the Hilbert space, spanned by the $|c_p^k\rangle$ for $p = 1, 2, \ldots$, as an isometry (preserving inner products), in the same way as a unitary map. When $k \neq l$,

(30) tells us that the two subspaces $\mathcal{P}^k$ and $\mathcal{P}^l$ are mutually orthogonal. The general idea is illustrated schematically in the figure on p. 436 of QCQI.

• Using this picture we can think of an error correction process as follows. Let P^k be the projector onto the subspace $\mathcal{P}^k$, and construct a decomposition of the identity on $\mathcal{H}_c$ of the form

$$I = P^1 + P^2 + \cdots + (I - P^1 - P^2 - \cdots). \tag{31}$$

One can then imagine carrying out a measurement after one of the errors has occurred, to determine in which of these subspaces the system is to be found. If it is subspace $\mathcal{P}^k$, then map that subspace back to the original space $\mathcal{P}$ using an isometry that undoes the effects of G_k, and then decode by reversing the original encoding procedure.

★ These steps can be combined into a single unitary operation D constructed in the following way. Start with the orthonormal basis $\{|a_p\rangle\}$ of $\mathcal{H}_a$, and let the corresponding orthonormal basis $\{|c_p\rangle\}$ of the coding space $\mathcal{P}$, see (9). Now choose in $\mathcal{H}_f$ an orthonormal collection $\{|f^k\rangle\}$, one vector for each principal error. Then define D by requiring that

$$D|c_p^k\rangle = |a_p\rangle \otimes |f^k\rangle \tag{32}$$

for every k and p, and extending this to an isometry on mapping $\mathcal{H}_c$ to $\mathcal{H}_a \otimes \mathcal{H}_f$ — this is always possible, because the $\{|c_p^k\rangle\}$ form an orthonormal collection, as noted earlier, so (32) maps an orthonormal collection to another orthonormal collection. By combining (29) with (32) we obtain

$$DG_kC|\psi\rangle = |\psi\rangle \otimes |f^k\rangle, \tag{33}$$

since any $|\psi\rangle$ in $\mathcal{H}_a$ can be written as a linear combination of the basis elements $\{|a_p\rangle\}$. As this equation is of the form (13), it follows that the principal errors G_k belong to the space of correctable errors $\mathcal{E}_c(D)$. Now the original E_i satisfying (20) are linear combinations of the G_k along with the null errors, and since (33) is also satisfied when G_k is replaced by a null error—set $|f^k\rangle$ equal to the zero vector, and both sides are zero—it follows that all the E_i we started with belong to $\mathcal{E}_c(D)$.

★ We have shown that when the projector condition (17) is satisfied for every E and $\bar{E}$ in the space $\mathcal{E}_c$, then a decoding operation can be constructed. Now let us demonstrate the converse. Given a decoding isometry D, the linear space $\mathcal{E}_c(D)$ of operators E satisfying (13) written as

$$DEC|a_p\rangle = DE|c_p\rangle = |a_p\rangle \otimes |s(E)\rangle \tag{34}$$

has the property that for any E and $\bar{E}$ in $\mathcal{E}_c(D)$ it is the case that (18) holds. This follows by noting that the second equality in (34) implies that

$$\langle c_p|E^\dagger \bar{E}|c_q\rangle = \langle c_p|E^\dagger D^\dagger D\bar{E}|c_q\rangle = \langle s(E)|s(\bar{E})\rangle\delta_{pq}, \tag{35}$$

where we have used the fact that D is an isometry, so $D^\dagger D = I_c$. Thus in this case $\alpha(E,\bar{E}) = \langle s(E)|s(\bar{E})\rangle$ in (18) is the inner product of the syndromes.