

Chapter 5

Basic Quantum Information

One of the main applications of unitary quantum symmetry as discussed in this book is quantum information theory, especially in the context of topologically ordered phases of matter in theoretical condensed matter physics.

A primary goal of physics is to study the *state* of a physical system by measuring observable quantities. Depending on the physical context, the collection of states is called a *configuration space*, *phase space*, or in quantum mechanics, the *state space*. From approximately 1687 CE to 1900 CE, the primary framework by which to study physical systems was Newtonian mechanics which has a continuum phase space. Classically, we approximate our state by finitely many measurement outcomes, such as position and momentum, as a full description of the state of the system is unverifiable and theoretically intractable. Following Max Planck’s 1900 CE resolution of the black body radiation spectrum by introducing discrete energies, it was realized that the continuum phase space of classical mechanics was insufficient to describe these radiative phenomena.

In quantum mechanics, not only is the state space different than in classical mechanics, but the space of measurements is dramatically reduced. Given a generic quantum mechanical state, it is typically impossible to predict the outcome of an individual observation. The measurable quantities which are uniquely determined by states are instead given by expectation values of these observations. To further complicate matters, observations in quantum mechanics *change the state*, so classically independent measurements are now dependent on the order of observation.

In classical information theory, information is stored in two-state systems called *bits*. These bits may be in the “off/0” state or the “on/1” state. In quantum information theory, information is usually stored in *qubits* whose state space is a Hilbert space, and states are typically viewed as *superpositions* in the computational basis consisting of the “off” state $|0\rangle$ and the “on” state $|1\rangle$. Although experimental implementations of quantum information are beyond the scope of this book, to-date, there have been many exciting applications of unitary quantum symmetries to physical qubit systems.

5.1 Probability distributions and Shannon entropy

A *probability distribution* on a (finite) set of possible *event outcomes* Ω is a function $\mathbf{P} : \Omega \rightarrow [0, 1]$, such that $\sum_{x \in \Omega} \mathbf{P}_x = 1$. We say that outcome $x \in \Omega$ *occurs with probability* \mathbf{P}_x . A *random variable* with values in V is a function $X : \Omega \rightarrow V$, and we say that the probability that $X = v$ is given by

$$\mathbf{P}(X = v) = \sum \{\mathbf{P}_x \mid X(x) = v\}.$$

When our space of values $V = \mathbb{R}$, the *expected value* and *variance* of X are respectively given by

$$\mathbf{E}[X] := \sum_{x \in \Omega} \mathbf{P}_x \cdot X(x). \quad \text{and} \quad \Delta[X] := \mathbf{E}[X^2] - \mathbf{E}[X]^2.$$

The *standard deviation* is the square root of the variance: $\sigma[X] := \sqrt{\Delta[X]}$.

Here are some classic problems about computing probability distributions and expected values.

Exercise 5.1.1 (Monte Hall). Monte Hall hosts the game show *Let's Make a Deal*. In a specific game on this show, there are 3 doors, one of which hides a new car, and the other 2 hide goats. You select a door, after which Monte opens one of the other two doors, revealing a goat. You are then given the chance to change your selection to the other closed door. Should you?

Hint: What if there are 100 doors, one of which hides a car, and Monte opens up 98 of the doors you didn't pick revealing goats?

Exercise 5.1.2. A family has two children.

- (1) If you know one child is a boy, what is the probability that both children are boys?
- (2) If you know one child is a boy born in January, what is the probability that both children are boys?
- (3) If you know one child is a boy born on January 1, what is the probability that both children are boys?
- (4) If you know the *older* child is a boy, what is the probability that both children are boys?

Exercise 5.1.3. Alice is a mathematics PhD student and Bob is a physics PhD student, who are both taking Quantum Symmetries 101. They play the following game during the semester. Each brings either a red or blue stick of Hagoromo¹ chalk to class each day. If both sticks are red, Alice pays Bob \$1, and if both sticks are blue, Alice pays Bob \$3. If the sticks do not match, Bob pays Alice \$2. Would you prefer to be Alice or Bob? Why? Would your answer change if you only played this game only once?

¹Yes, it is important that the brand of chalk is Hagoromo. No, we did not receive any money for this ad placement.

Definition 5.1.4 — Let \mathbf{P} be a probability distribution on a finite set J . The *Shannon entropy* of this distribution is defined as

$$S(\mathbf{P}) := - \sum_{j \in J; \mathbf{P}_j \neq 0} \mathbf{P}_j \log(\mathbf{P}_j).$$

The next exercise motivates the above definition.

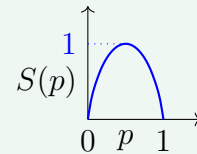
Exercise 5.1.5. A *Shannon information function* $\mathcal{I} : (0, 1] \rightarrow [0, \infty)$ is a function which measures the *information* or *surprise* $\mathcal{I}(p)$ of observing an event with probability $p \in (0, 1]$. We assume the function \mathcal{I} satisfies the following axioms:

$$p < q \implies \mathcal{I}(q) < \mathcal{I}(p), \quad \mathcal{I}(1) = 0, \quad \text{and} \quad \mathcal{I}(pq) = \mathcal{I}(p) + \mathcal{I}(q).$$

- (1) Prove that if \mathcal{I} is twice differentiable, then $\mathcal{I}(p) = -\log(p)$, up to a choice of base for the logarithm.
- (2) Deduce that the Shannon entropy of a distribution is the *expected/average information* one receives when making a measurement where outcomes are determined by that distribution.
- (3) Why might one assume the above axioms for an information function?
Hint: For the third axiom, consider performing two independent measurements.

Example 5.1.6 — A *Bernoulli trial* is a random variable with two possible outcomes, usually called *success* and *failure*, e.g., a coin flip where success is your chosen outcome. Such probability distributions correspond to a single value $p \in [0, 1]$ assigned to success. The entropy of this probability distribution is

$$S(p) = -p \log(p) - (1 - p) \log(1 - p)$$



One might interpret the graph of $S(p)$ as indicating that the most randomness occurs when $p = 1/2$, e.g., when flipping a fair coin. Indeed, when we have system with a high degree of randomness, we get a large amount of information from each individual measurement.

Warning 5.1.7 — In general, the variance of a random variable X is a measure of uncertainty in the outcome of X , whereas entropy is a measure of uncertainty of a probability distribution which is independent of any random variable. It may be the case that one probability distribution \mathbf{P} may have higher entropy than \mathbf{Q} on the same set of outcomes, but the variance of a random variable X might be higher for the opposite

distribution.

Exercise 5.1.8. Find two probability distributions \mathbf{P}, \mathbf{Q} and a random variable X on a three element set such that $S(\mathbf{P}) > S(\mathbf{Q})$, but $\Delta_{\mathbf{Q}}[X] > \Delta_{\mathbf{P}}[X]$. Can you find such an example for a two element set?

Facts 5.1.9. Here are some facts about probability distributions and their Shannon entropies.

- (S1) If \mathbf{P} and \mathbf{Q} are two probability distributions on the set $\{1, \dots, n\}$, the *convex combination* $t\mathbf{P} + (1-t)\mathbf{Q}$ defined by $t\mathbf{P}_i + (1-t)\mathbf{Q}_i$ is again a probability distribution.
- (S2) The multivariable function S on the space of probability distributions \mathbf{P} on $\{1, \dots, n\}$ is *strictly concave*, i.e.,

$$tS(\mathbf{P}) + (1-t)S(\mathbf{Q}) \leq S(t\mathbf{P} + (1-t)\mathbf{Q})$$

with equality if and only if $\mathbf{P} = \mathbf{Q}$.

Proof. On the interior of the space of probability distributions where $\mathbf{P}_i > 0$, the Hessian of S is negative definite, which is equivalent to strict concavity. \square

- (S3) When $|J| = N$, the uniform distribution $\mathbf{P}_j = 1/N$, $\forall j \in J$ is the unique distribution with maximal entropy $S(\mathbf{P}) = \log(N)$.

Proof. First, the uniform distribution $\mathbf{P}_j = 1/N$ for all j has entropy $S(\mathbf{P}) = \log(N)$.

Second, we show that the uniform distribution is the unique distribution with each $\mathbf{P}_j > 0$ which has extremal entropy. In order to use the method of Lagrange multipliers, we define

$$G(\mathbf{P}, \lambda) := - \sum_{j \in J} \mathbf{P}_j \log(\mathbf{P}_j) + \lambda \left(1 - \sum_{j \in J} \mathbf{P}_j \right). \quad (5.1.10)$$

Setting $\partial_j G = 0$, we see that critical points of G only occur when $\log(\mathbf{P}_j) = -1 - \lambda$, i.e., \mathbf{P}_j is *independent* of $j \in J$, and thus p is the uniform distribution.

Finally, if \mathbf{P} is a distribution with some $\mathbf{P}_j = 0$, then the problem reduces to finding an extremal entropy distribution on a proper subset with strictly positive probabilities. Since $\log(N) > \log(M)$ whenever $M < N$, we see that for any N , the extremal distribution where $\mathbf{P}_j = 1/N$ is, in fact, maximal. \square

5.2 State vectors and vector states: superposition and entanglement

We now introduce the notion of quantum state; we do so using the antinomy of *state vector* versus *vector state* to highlight and clarify the difference between a vector in a Hilbert space and the corresponding rank one projection.

Definition 5.2.1 — A *state vector* is a *unit* (length one) vector $|\psi\rangle \in H$. Its corresponding *vector state* is the rank one projection $|\psi\rangle\langle\psi| \in B(H)$. By the correspondence between projections and their images, a vector state also corresponds to the ray $\mathbb{C}|\xi\rangle \subset H$.

Observe that $\psi, \xi \in H$, $|\psi\rangle\langle\psi| = |\xi\rangle\langle\xi|$ if and only if $\xi = \lambda\psi$ for some $\lambda \in U(1) = \{z \in \mathbb{C} \mid |z| = 1\}$.

Example 5.2.2 (Qubits and Bloch Sphere) — A *qubit* is a state in \mathbb{C}^2 . There are several canonical state vectors that get special names in quantum information.

$$|0\rangle := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle := \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad |\pm\rangle := \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix}.$$

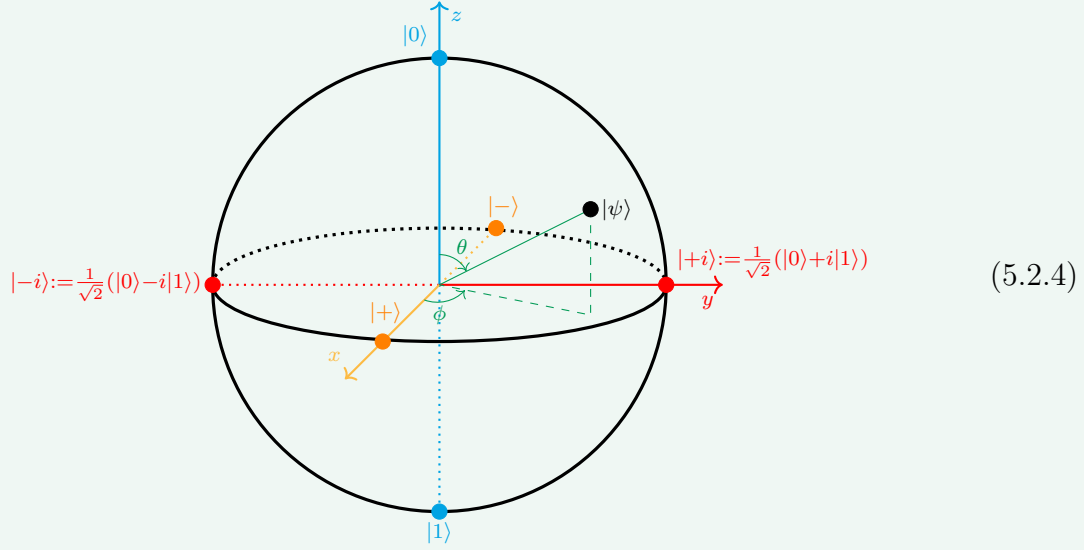
We call $\{|0\rangle, |1\rangle\}$ the *Z-computational basis* and $\{|+\rangle, |-\rangle\}$ the *X-computational basis*.

Given $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \in \mathbb{C}^2$, we may assume $\alpha \geq 0$ by multiplying by a phase in $U(1)$. We can therefore parametrize the space of vector states by

$$|\psi\rangle = \cos(\theta/2)|0\rangle + e^{i\phi}\sin(\theta/2)|1\rangle \quad \theta \in [0, \pi] \text{ and } \phi \in [0, 2\pi). \quad (5.2.3)$$

Thus the space of vector states in $M_2(\mathbb{C})$ is a 2-sphere, whose parametrization as above

is commonly referred to as the *Bloch sphere*.



Warning 5.2.5 — In (5.2.4) above, we have used the labeling convention from quantum information theory, where the quantum states are labeled by state vectors up to a phase, rather than vector states. This may lead to the confusion that $|0\rangle$ and $|1\rangle$ appear to be *colinear* in the Bloch sphere (5.2.4). However, the label $|0\rangle$ really means the rank one projection $|0\rangle\langle 0|$, and similarly $|1\rangle$ really means $|1\rangle\langle 1|$. These vector states are orthogonal in $M_2(\mathbb{C})$ under the GNS inner product from the trace, but the Bloch sphere lies entirely in the positive cone of $M_2(\mathbb{C})$.

Exercise 5.2.6. How would you parametrize vector states on the Bloch sphere in terms of the X -computational basis?

Remark 5.2.7. The identification of state vectors up to phase resulting in the Bloch sphere is known in algebraic topology as the *Hopf fibration*.

$$\begin{aligned}
 U(1) = \{z \in \mathbb{C} \mid |z| = 1\} &\longrightarrow S^3 = \{\text{unit vectors in } \mathbb{C}^2\} \\
 &\downarrow \\
 S^2 &= \text{Bloch sphere}
 \end{aligned}$$

Definition 5.2.8 (Superposition) — Suppose we have a state vector $|\psi\rangle \in H$ and a chosen ONB $|e_i\rangle$ for H . We may then uniquely express

$$|\psi\rangle = \sum_i \langle e_i | \psi \rangle \cdot |e_i\rangle$$

as a linear combination of the $|e_i\rangle$. This expression is often referred to as writing $|\psi\rangle$ as a *superposition* of the $|e_i\rangle$.

Example 5.2.9 — In Example 5.2.2, we discussed the Z and X computational bases. Observe that (5.2.4) writes our state vector as a superposition in the Z -computational basis.

We will reprise the notion of superposition further in the next section on quantum observables in Example 5.3.4. For the time being, we will warn the reader that it does not make sense to say a state vector is in a superposition without reference to a chosen ONB (more precisely, a quantum observable; see Warning 5.3.6).

We learned the following example from a talk of Greg Moore in July 2025 at CMSA.

Example 5.2.10 — Let us compute what a superposition of two vector states looks like as a rank one projector. If $|\psi_1\rangle$ and $|\psi_2\rangle$ are orthonormal, then for any state vector $|\psi\rangle = \alpha|\psi_1\rangle + \beta|\psi_2\rangle$, we have that the projector $|\psi\rangle\langle\psi|$ is of the form

$$|\psi\rangle\langle\psi| = tp_1 + wq + \bar{w}q^\dagger + (1-t)p_2$$

where $p_j = |\psi_j\rangle\langle\psi_j|$ for $j = 1, 2$, and $q = |\psi_1\rangle\langle\psi_2|$. Observe that q is completely characterized up to phase by $qq^\dagger = p_1$ and $q^\dagger q = p_2$. For what follows, it will be easier to use the following normalization:

$$p(z, r) := \frac{1}{1+r}(p_1 + zq + \bar{z}q^\dagger + rp_2).$$

Modeling p_1, p_2, q by

$$p_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad p_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad q = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

we can model

$$p(z, r) = \frac{1}{1+r} \begin{pmatrix} 1 & z \\ \bar{z} & r \end{pmatrix}$$

which is a projector exactly when $r = |z|^2$. As a function of $z \in \mathbb{C}$, the projectors which are a superposition of $|\psi_1\rangle\langle\psi_1|$ and $|\psi_2\rangle\langle\psi_2|$ are then given by the *Bott projector*

$$p(z) = \frac{1}{1+|z|^2} \begin{pmatrix} 1 & z \\ \bar{z} & |z|^2 \end{pmatrix}.$$

The Bott projector plays an important role in *complex K-theory*.

Entanglement refers to shared quantum information between two parts of a quantum system. When a quantum system is a composite of two distinct quantum systems A, B with

Hilbert spaces H_A, H_B , respectively, the total Hilbert space is given by $H_A \otimes H_B$. We will occasionally refer to this kind of composite system as a *bipartite* system.

Definition 5.2.11 — Suppose $|\psi\rangle \in H_A \otimes H_B$. We say that $|\psi\rangle$ is *separable* or a *product state* if $|\psi\rangle = |\eta\rangle \otimes |\xi\rangle$ for $|\eta\rangle \in H_A$ and $|\xi\rangle \in H_B$. If $|\psi\rangle$ is not separable, then $|\psi\rangle$ is called *entangled*.

Example 5.2.12 (Bell basis) — The following four states are called the *Bell basis* of $\mathbb{C}^2 \otimes \mathbb{C}^2$. Each of the Bell basis states is entangled.

$$\begin{aligned} |\Phi^+\rangle &:= |\beta_{00}\rangle := \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\ |\Phi^-\rangle &:= |\beta_{01}\rangle := \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \\ |\Psi^+\rangle &:= |\beta_{10}\rangle := \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \\ |\Psi^-\rangle &:= |\beta_{11}\rangle := \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \end{aligned}$$

The state $|\Psi^-\rangle$ is also called the *singlet state*.

Proposition 5.2.13 — The singlet state $|\Psi^-\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ is independent of the choice of ONB for \mathbb{C}^2 . That is, choosing another ONB $|e_1\rangle, |e_2\rangle$ for \mathbb{C}^2 , $|\Psi^-\rangle\langle\Psi^-| = |\psi\rangle\langle\psi|$ for

$$|\psi\rangle := \frac{1}{\sqrt{2}}(|e_1e_2\rangle - |e_2e_1\rangle).$$

Proof. Write $|e_1\rangle = \alpha|0\rangle + \beta|1\rangle$ as a superposition in the Z computational basis. Since we only care about $|\psi\rangle$ up to phase, we may take $|e_2\rangle = -\bar{\beta}|0\rangle + \bar{\alpha}|1\rangle$. We compute

$$|e_1e_2\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \otimes \begin{pmatrix} -\bar{\beta} \\ \bar{\alpha} \end{pmatrix} = \begin{pmatrix} -\alpha\bar{\beta} \\ |\alpha|^2 \\ -|\beta|^2 \\ \bar{\alpha}\beta \end{pmatrix} \quad \text{and} \quad |e_2e_1\rangle = \begin{pmatrix} -\alpha\bar{\beta} \\ -|\beta|^2 \\ |\alpha|^2 \\ \bar{\alpha}\beta \end{pmatrix}.$$

We then calculate that

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|e_1e_2\rangle - |e_2e_1\rangle) = \frac{1}{\sqrt{2}} \begin{pmatrix} -\alpha\bar{\beta} \\ |\alpha|^2 \\ -|\beta|^2 \\ \bar{\alpha}\beta \end{pmatrix} - \frac{1}{\sqrt{2}} \begin{pmatrix} -\alpha\bar{\beta} \\ -|\beta|^2 \\ |\alpha|^2 \\ \bar{\alpha}\beta \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} = |\Psi^-\rangle. \quad \square$$

Exercise 5.2.14. Show that the other Bell basis states in $\mathbb{C}^2 \otimes \mathbb{C}^2$ are not independent of the choice of ONB.

Remark 5.2.15. The fact that the singlet state is independent of the choice of basis in Proposition 5.2.13 above has an elegant explanation in terms of representation theory. Indeed, \mathbb{C}^2 is the standard representation of

$$SU(2) = \{x \in M_2(\mathbb{C})^\times \mid x \text{ is unitary and } \det(x) = 1\},$$

and $\mathbb{C}^2 \otimes \mathbb{C}^2$ is the direct sum of two unitary representations: the trivial representation and an irreducible 3-dimensional representation. This trivial representation is spanned by a singlet state vector, and the other Bell basis states span this 3-dimensional representation.

Warning 5.2.16 — The Bell basis is defined for $\mathbb{C}^2 \otimes \mathbb{C}^2$, not $\mathbb{C}^2 \otimes \overline{\mathbb{C}^2}$. In applications to quantum teleportation (see 5.9 below), people often identify $\mathbb{C}^2 \cong \overline{\mathbb{C}^2}$ by the *linear* map $|j\rangle \mapsto \langle j|$ for $j = 0, 1$, but this depends on the choice of basis.

Remark 5.2.17. Since $\mathbb{C}^2 \cong \overline{\mathbb{C}^2}$ as $SU(2)$ representations, there is still a unique 1-dimensional sub-representation in $\mathbb{C}^2 \otimes \overline{\mathbb{C}^2} \cong M_2(\mathbb{C})$ corresponding to the singlet state. Indeed, the action of $SU(2)$ is by conjugation on $M_2(\mathbb{C})$, and thus the invariant subspace is spanned by the identity. Thus the singlet state is now described by

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}} \left(|e_1\rangle \otimes \langle e_1| + |e_2\rangle \otimes \langle e_2| \right) \in \mathbb{C}^2 \otimes \overline{\mathbb{C}^2}.$$

5.3 Quantum observables and the Born rule

In quantum mechanics, an *observable* quantity (e.g., position, energy, frequency) corresponds to a self-adjoint operator, as we observe real values in measurements/experiments. However, the *outcome* of a measurement in a given state is probabilistic, and even worse (or better?), *no one knows why*.

Definition 5.3.1 (Born Rule) — Suppose $x \in B(H)$ is an observable/self-adjoint operator we wish to measure in the state $|\psi\rangle \in \mathcal{H}$. By the Spectral Theorem 1.7.9, we may write

$$x = \sum_{\lambda \in \text{spec}(x)} \lambda p_\lambda,$$

where $p_\lambda \in B(H)$ is the orthogonal projection onto the eigenspace

$$E_\lambda = \{\xi \in H \mid x\xi = \lambda\xi\}.$$

The probability of measuring the value λ for x in state $|\psi\rangle$ is given by

$$\mathbf{P}_{|\psi\rangle}(x = \lambda) := \langle \psi | p_\lambda | \psi \rangle = \text{Tr}(p_\lambda |\psi\rangle \langle \psi|).$$

After a measurement of $x = \lambda$ is observed (which requires $\mathbf{P}_{|\psi\rangle}(x = \lambda) \neq 0$), the quantum

state of the system is then given by the eigenstate of x closest to $|\psi\rangle$, namely

$$\frac{p_\lambda|\psi\rangle}{\|p_\lambda|\psi\rangle\|},$$

obtained by projecting $|\psi\rangle$ to E_λ and then normalizing.

Exercise 5.3.2. Show that $\mathbf{P}_{|\psi\rangle}(x = \lambda)$ is a probability distribution on $\text{spec}(x)$.

Exercise 5.3.3. Show that $\frac{p_\lambda|\psi\rangle}{\|p_\lambda|\psi\rangle\|}$ is the unique unit vector in E_λ closest to $|\psi\rangle$.

Example 5.3.4 (Superposition, reprise) — Suppose we have a state vector $|\psi\rangle \in H$ and we wish to measure a quantum observable x . Using the spectral decomposition of x as in the Born Rule 5.3.1, we can write $|\psi\rangle$ as a linear combination of eigenstates of x :

$$|\psi\rangle = 1 \cdot |\psi\rangle = \sum_{\lambda \in \text{spec}(x)} p_\lambda |\psi\rangle = \sum_{\lambda \in \text{spec}(x)} \|p_\lambda|\psi\rangle\| \cdot \frac{p_\lambda|\psi\rangle}{\|p_\lambda|\psi\rangle\|}.$$

This linear combination of eigenstates is called a *superposition*.

Definition 5.3.5 (Measurement bases) — Given a quantum observable $x \in B(H)$, a *measurement basis with respect to x* is an ONB of H consisting of eigenvectors of x . One can obtain a measurement basis by independently choosing an ONB for each of the subspaces $p_\lambda H$ for $\lambda \in \text{spec}(x)$. Given such a measurement basis $\{|e_i\rangle\}$ for x , the *amplitude* of $|e_i\rangle$ in $|\psi\rangle$ is the *Fourier coefficient* $\langle e_i|\psi\rangle$ of $|\psi\rangle$ with respect to $|e_i\rangle$. The probability of measuring $x = \lambda$ is then given by

$$\mathbf{P}_{|\psi\rangle}(x = \lambda) := \sum_{\substack{i: \\ x|e_i\rangle = \lambda|e_i\rangle}} |\langle e_i|\psi\rangle|^2.$$

Warning 5.3.6 — It does not type check to talk about a quantum state vector being in a *superposition* unless you have chosen a particular observable to measure. Our previous Definition 5.2.8 of superposition was motivated by ONBs, not by measuring quantum observables. While this is a mathematically convenient framework, it is not physically motivated. In particular, choosing to express a state vector in a superposition with respect to a chosen ONB corresponds to the superposition associated to any self-adjoint observable which is diagonal with distinct real eigenvalues in that ONB.

“Hot Take” 5.3.7 — The Born Rule is what we observe in measurements with quantum states, but it may be misleading to draw conclusions about the underlying reality of the universe. But just for fun, let us imagine an observer \mathcal{O} meets a qubit $|\psi\rangle$ walking down the street. The observer wants to measure $|\psi\rangle$ in the Z -computational basis and asks, “Are you $|0\rangle$ or $|1\rangle$?” At this point $|\psi\rangle$ looks at its position on the Bloch sphere (5.2.4). This position need not be measured as a linear combination of $|0\rangle$ and $|1\rangle$ as in (5.2.3), but by asking the state whether it is $|0\rangle$ or $|1\rangle$, the observer forces the state into a superposition of $|0\rangle$ and $|1\rangle$. If $\theta \notin \{0, \pi\}$, then $|\psi\rangle$ is neither $|0\rangle$ nor $|1\rangle$; such qubits are *inherently* non-binary. So $|\psi\rangle$ makes as unbiased a choice as possible, and answers $|0\rangle$ or $|1\rangle$ according to a probability density; $|\psi\rangle$ returns $|0\rangle$ with probability $\cos^2(\theta/2)$ and $|1\rangle$ with probability $\sin^2(\theta/2)$. In answering this question, the qubit *updates* itself to $|0\rangle$ or $|1\rangle$, as it must be *consistent* if it is asked if it is $|0\rangle$ or $|1\rangle$ again. However, if it now measured in another basis, e.g., the X -computational basis $\{|+\rangle, |-\rangle\}$, its reply will again become random, this time with a 50-50 chance of being $|+\rangle$ or $|-\rangle$.

The really mind-blowing part of the Born Rule is that we can actually *harness* this randomness for applications in quantum information and computation.

The final statement in the above hot take is a playful explanation of what one observes in repeated *Stern-Gerlach* experiments.

Example 5.3.8 (Stern-Gerlach) — The Stern-Gerlach experiment [GS22] was one of the first experiments which established that electrons have a quality best described as *spin* which is *quantized*. The experiment sent silver atoms, which have one unpaired $5s$ electron,^a through an inhomogeneous magnetic field. Instead of a single Gaussian distribution which one might expect from classical electrodynamics, the result showed two distinct peaks, corresponding to the two values *up* and *down* for the spin of the electron.

This experiment has also been carried out with polarized photons which also represent qubits $|\psi\rangle \in \mathbb{C}^2$, where Z -polarized light passes through an X -polarization filter. Here, X, Z refer to *Pauli operators*

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

which we will discuss in more detail in §5.4 below.^b For the time being, we will say being polarized in the Z -direction corresponds to being a $+1$ eigenstate $|0\rangle$ for Z , and the Z -polarization filter is the quantum observable corresponding to the spectral projection

$$p_{Z=1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Now rotating the filter by an angle $\theta \in (0, 2\pi)$ corresponds to conjugating our observable

by the rotation unitary

$$R_\theta := \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Thus $1 - p$ corresponds to rotation by $\pi/2$, and rotation by $\pi/4$ corresponds to the $+1$ spectral projection of the Pauli X operator:

$$R_\theta \cdot p_{Z=1} \cdot R_\theta^\dagger = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = p_{X=1}.$$

One can now apply repeated Stern-Gerlach experiments with subsequent filters rotated by some angle. We can calculate the intensity (photon throughput) of Z -polarized light after passing through an X -polarization filter as

$$\langle 0 | p_{X=1} | 0 \rangle = \begin{pmatrix} 1 & 0 \end{pmatrix} \cdot \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}.$$

That is, our intensity drops by half, as a Z -polarized has probability $1/2$ of passing through an X -polarization filter.

^aThe magnetic dipole moment from the nucleus is negligible compared to that of the unpaired electron.

^bWe warn the reader that Z and X do *not* correspond to the orientations of these filters in \mathbb{R}^3 .

Exercise 5.3.9. Find 2 light polarization filters and see what happens when you offset them by $\pi/2$. What do you think would happen if you insert a *third* polarization filter at angle $\pi/4$ to both of these? Surely adding more filters will block more light, right? Confirm or disprove your suspicion by computing the intensity of Z -polarized light after passing through:

- a $-Z$ -polarization filter, or
- an X -polarization filter and then a $-Z$ -polarization filter.

Definition 5.3.10 — For $x \in B(H)$ (not just self-adjoint observables!), we define the *expectation value* of x in state $|\psi\rangle$ by the vector state

$$\mathbf{E}_{|\psi\rangle}(x) := \langle \psi | x | \psi \rangle.$$

Facts 5.3.11.

(E1) The expectation value of a self-adjoint operator in the state vector $|\psi\rangle$ is equal to the weighted arithmetic mean of measurement outcomes:

$$\mathbf{E}_{|\psi\rangle}(x) = \langle \psi | x | \psi \rangle = \left\langle \psi \left| \sum_{\lambda \in \text{spec}(x)} \lambda p_\lambda \right| \psi \right\rangle = \sum_{\lambda \in \text{spec}(x)} \lambda \langle \psi | p_\lambda | \psi \rangle = \sum_{\lambda \in \text{spec}(x)} \lambda \mathbf{P}_{|\psi\rangle}(x = \lambda).$$

- (E2) As $\mathbf{E}_{|\psi\rangle} : B(H) \rightarrow \mathbb{C}$ is a state, it is completely positive.
- (E3) $\mathbf{E}_{|\psi\rangle}(x) = \text{Tr}(x \cdot |\psi\rangle\langle\psi|)$ for all state vectors $|\psi\rangle \in H$ and $x \in B(H)$.
- (E4) If $\mathbf{E}_{|\psi\rangle} = \mathbf{E}_{|\xi\rangle}$ for state vectors $|\psi\rangle, |\xi\rangle \in H$, then $|\psi\rangle\langle\psi| = |\xi\rangle\langle\xi|$. Physically, this means that vector states are uniquely defined by their expectation values.
- (E5) For $x, y \in B(H)$, $[x, y] = 0$ if and only if $\mathbf{E}_{|\psi\rangle}(xy) = \mathbf{E}_{|\psi\rangle}(yx)$ for all state vectors $|\psi\rangle \in H$. Although this statement is trivial mathematically, it has an important physical implication: if two observables do not commute, then there is some state vector such that the order in which the two observables are measured changes the measurement outcome.

Exercise 5.3.12. Let us revisit Alice and Bob in Quantum Symmetries 101 from Exercise 5.1.3. Suppose Alice and Bob can now each bring a superposition of their red and blue sticks of Hagoromo chalk to class, represented as $|\psi_A\rangle$ and $|\psi_B\rangle$ in $\mathbb{C}|R\rangle \oplus \mathbb{C}|B\rangle$. Each day, they measure the observable

$$\begin{pmatrix} 1 & & & \\ & -2 & & \\ & & -2 & \\ & & & 3 \end{pmatrix}$$

which is diagonal in the ONB $\{|RR\rangle, |RB\rangle, |BR\rangle, |BB\rangle\}$, in their product state $|\psi_A\rangle \otimes |\psi_B\rangle$. Alice pays Bob the outcome in \$ of the measurement, where a negative outcome means Bob pays Alice. Show that Alice has a winning strategy. Then reconsider your answers to Exercise 5.1.3 in the context of this winning strategy.

Exercise 5.3.13. Bob lost a lot over the course of the semester, so let us help Bob out a little. We now allow Bob to choose *any* unitary operator u to apply to the product state $|\psi_A\rangle \otimes |\psi_B\rangle$ before measuring, although Bob still does not know $|\psi_A\rangle$. How does this change the analysis in the previous exercise?

Remark 5.3.14. Historically, only self-adjoint operators $x = x^\dagger \in B(\mathcal{H})$ were considered observables, as they have real measurement outcomes. However, many authors use the term observable for *any* operator. Although not every operator has a measurement basis or a set of measurement outcomes, we are still able to define the expectation value of any operator as a generalization of the weighted arithmetic mean. This is partially due to the fact that every operator is a complex linear combination of self-adjoint operators.

Exercise 5.3.15. Let $x, y \in M_2(\mathbb{C})$ be self-adjoint, $[x, y] \neq 0$, and $x^2 = y^2 = 1$. Show that there are exactly two vector states $|\psi\rangle$ up to phase such that $\mathbf{E}_{|\psi\rangle}(x) = \mathbf{E}_{|\psi\rangle}(y) = 0$. *Hint: Without loss of generality, $x = Z$ and $y = u^* Z u$ for some unitary u . Set $|\psi\rangle = |0\rangle + \alpha|1\rangle$ for a phase $\alpha \in U(1)$ and directly compute $\langle\psi|y|\psi\rangle$.*

Definition 5.3.16 — The *variance* of a self-adjoint observable $x \in B(H)$ in a state $|\psi\rangle$ is

$$\Delta x := \mathbf{E}_{|\psi\rangle}(x^2) - \mathbf{E}_{|\psi\rangle}(x)^2.$$

The *standard deviation* of x is $\sigma_x := \sqrt{\Delta x}$. (It is common practice to omit the state dependence on the notation for the variance and the standard deviation.)

Exercise 5.3.17. Prove that $\Delta x \geq 0$ for a self-adjoint $x \in B(H)$ by proving $\Delta x = \langle \psi_x | \psi_x \rangle$ for $|\psi_x\rangle := (x - \mathbf{E}_{|\psi\rangle}(x))|\psi\rangle$.

In quantum physics, an ordered pair of (unbounded) self-adjoint operators (x, p) is called a *canonically conjugate* pair if $[x, p] = i1$. The traditional statement of the *Heisenberg Uncertainty Principle* (when \hbar is set to 1) is that if x, p are canonically conjugate, then

$$\sigma_x \sigma_p \geq \frac{1}{2}.$$

In quantum information, where our state space is finite dimensional, we lack canonically conjugate pairs, which leads to a generalized version of the Heisenberg Uncertainty Principle beyond canonically conjugate pairs of operators.

Exercise 5.3.18. Prove that there are no canonically conjugate pairs of operators in $M_n(\mathbb{C})$. *Hint:* What does $M_n(\mathbb{C})$ have that $B(\ell^2)$ does not?

Theorem 5.3.19 (Heisenberg Uncertainty Principle) — Let $x, y \in B(H)$ be self-adjoint operators. In any vector state,

$$\Delta x \cdot \Delta y \geq \frac{1}{4} |\mathbf{E}([x, y])|^2.$$

Proof. Let $|\psi\rangle \in H$ be a state vector, and let $|\psi_x\rangle, |\psi_y\rangle$ be as in Exercise 5.3.17. Observe that

$$\langle \psi_x | \psi_y \rangle = \langle \psi | (x - \mathbf{E}_{|\psi\rangle}(x))(y - \mathbf{E}_{|\psi\rangle}(y)) | \psi \rangle = \langle \psi | xy | \psi \rangle - \mathbf{E}_{|\psi\rangle}(x) \mathbf{E}_{|\psi\rangle}(y),$$

which implies that

$$\operatorname{Im}(\langle \psi_x | \psi_y \rangle) = \frac{1}{2i} (\langle \psi_x | \psi_y \rangle - \langle \psi_y | \psi_x \rangle) = \frac{1}{2i} \langle \psi | xy - yx | \psi \rangle = \frac{1}{2i} \mathbf{E}([x, y]).$$

For all $z \in \mathbb{C}$, $|\operatorname{Im}(z)| \leq |z|$, so setting $z = \langle \psi_x | \psi_y \rangle$, we have

$$\frac{1}{4} |\mathbf{E}([x, y])|^2 = |\operatorname{Im}(\langle \psi_x | \psi_y \rangle)|^2 \leq |\langle \psi_x | \psi_y \rangle|^2 \stackrel{\text{(C-S)}}{\leq} \langle \psi_x | \psi_x \rangle \langle \psi_y | \psi_y \rangle \stackrel{\text{(Exer. 5.3.17)}}{=} \Delta x \cdot \Delta y. \quad \square$$

Remark 5.3.20. In physical applications and experiments, measurable quantities have *units*, but operators in $B(H)$ are *unitless*. Quantum observables in experiments are elements of $B(H)$ multiplied by the appropriate units. As we focus on applications in quantum information, we drop all units, as our outcomes tend to be binary values. A physics oriented reader may wish to assume that we are using *natural units*, especially in the context of spin systems.

5.4 Mixed states and von Neumann entropy

In the last section, we saw probability distributions arise when measuring a quantum (self-adjoint) observable in a state vector. However, in order to execute such a measurement, we must have complete knowledge of our quantum state. In practice, we only have *partial* information of the quantum state, introducing another layer of *classical* probability on top of the inherent *quantum* probability of measurement. Before giving a formal definition of a *mixed state*, which should be viewed as a *classical ensemble* of vector states rather than a single completely known vector state, we discuss the following ‘paradox.’

Example 5.4.1 (Spooky action at a distance, [EPR35]) — Suppose Alice and Bob each control one qubit of the Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, and the two qubits are spatially separated. (How this state was prepared is not important.) The instant that Alice measures her qubit in the Z -computational basis, Bob’s state is completely determined; it appears that information has traveled faster than the speed of light. However, Bob has no knowledge of Alice’s observation, and thus his best guess as to his qubit’s value is probabilistic: a 50% chance of $|0\rangle$ and a 50% chance of $|1\rangle$. Note, however, that his qubit is either $|0\rangle$ or $|1\rangle$, and not a superposition. So no information has traveled faster than the speed of light, as Alice would need to relay the outcome for Bob to know his qubit’s value.

The above example illustrates that probabilities arise in two distinct ways in quantum information: (1) as measurement outcomes of a quantum state, and (2) the *classical mixture* of pure states which best describes our partial knowledge of the actual quantum state. The first is a truly quantum phenomenon, whereas the second uses classical probability theory to describe unknown, predetermined quantities.

The above example is part of the EPR paradox.

Example 5.4.2 (EPR paradox, [EPR35]) — Suppose Alice and Bob each control one qubit of an entangled singlet state $|\Psi^-\rangle$, and the two qubits are spatially separated. (How this state was prepared is not important.) Incorporating that the singlet state is independent of the choice of basis (see Proposition 5.2.13), Alice’s measurement in any computational basis immediately determines the state of Bob’s qubit in that basis. EPR produced a scheme which they thought would violate the uncertainty principle: Alice measures in the X -computational basis, which would determine Bob’s qubit in the X -computational basis, but before any information traveled from Alice to Bob, he measures in the Z -computational basis, which would allow him to know both the X and Z values of his qubit simultaneously, a contradiction to the Uncertainty Principle 5.3.19 as they do not commute.

As we have just explained why spooky action at a distance does not mean information has traveled faster than the speed of light, from a modern perspective, this phenomenon does not seem all that surprising. The point, however, is that the above scheme violates

local realism, the tenet that an isolated quantum system is completely described by a definite state vector in its Hilbert space. When the two qubits are separated, EPR viewed the two qubits as isolated quantum systems, but isolated systems need not be *independent*. This failure of local realism has been observed and even quantified since, especially in violations of Bell's inequalities (see §5.7 below).

We now introduce *mixed states* as a bookkeeping technique for when we only have partial information about our quantum state.

Definition 5.4.3 — A state $\varphi: B(H) \rightarrow \mathbb{C}$ is called *pure* if it is a vector state, i.e., its density $d = \frac{d\varphi}{d\text{Tr}}$ is a rank one projection. If a state is not pure, it is called a *mixed state*.

We often identify a state with its density matrix. In this sense, we can call a density matrix pure or mixed.

Example 5.4.4 — In the Spooky Action at a Distance Example 5.4.1, the mixed state which represents Bob's state after Alice's measurement is

$$d := \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}I.$$

Interestingly, this is also the mixed state which represents Bob's state in the EPR paradox Example 5.4.2, *independent* of the measurement Alice performs on the singlet state.

Exercise 5.4.5. Prove that the following are equivalent for a state $\varphi: B(H) \rightarrow \mathbb{C}$.

- (1) φ is pure.
- (2) The density d of φ is *extremal*, i.e., whenever we can write

$$d = td_1 + (1-t)d_0 \quad \text{for some} \quad t \in (0, 1)$$

for densities $d_0, d_1 \in B(H)$, then $d_0 = d_1 = d$.

- (3) The GNS representation $L^2(B(H), \varphi)$ is *irreducible*, i.e., $B(H)' = \mathbb{C}$.

Notation 5.4.6 — The *Pauli operators/matrices* in $M_2(\mathbb{C})$ are

$$X := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad Y := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad Z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Authors also use other notations, such as

$$\sigma^x = \sigma^1 := X \quad \sigma^y = \sigma^2 := Y \quad \sigma^z = \sigma^3 := Z.$$

It is also common to define $\sigma^0 := 1$ as the zeroth Pauli matrix. These four Pauli matrices

serve as an ONB for $M_2(\mathbb{C})$ with the trace inner product. They also serve as an ONB for the real Hilbert space of Hermitian matrices in $M_2(\mathbb{C})$.

Exercise 5.4.7. Prove that the commutation relations between the Pauli operators are given by

$$[\sigma^i, \sigma^j] = \sum_{k=1}^3 2\sqrt{-1} \cdot \epsilon^{ijk} \sigma^k$$

where ϵ^{ijk} the *Levi-Cevita symbol*, i.e., the anti-symmetric tensor determined by the rules

$$\epsilon^{123} = 1 \quad \text{and} \quad \epsilon^{ijk} = -\epsilon^{jik} = -\epsilon^{ikj} \quad \forall i, j, k.$$

(Here we use $\sqrt{-1}$ instead of i as i is the preferred index notation for the Levi-Cevita symbol.)

Remark 5.4.8. The unit sphere S^3 of state vectors in \mathbb{C}^2 is a torsor for the *unit quaternions*

$$\mathbb{H}_u := \{w1 + x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \mid w, x, y, z \in \mathbb{R} \text{ and } w^2 + x^2 + y^2 + z^2 = 1\}$$

where $\mathbf{i}, \mathbf{j}, \mathbf{k}$ multiply according to the Levi-Cevita symbol. One can represent the unit quaternions in $M_2(\mathbb{C})$ by

$$w1 + x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \mapsto \begin{pmatrix} w - iz & -y - ix \\ y - ix & w + iz \end{pmatrix}.$$

Thus the map $1 \mapsto 1$, $\mathbf{i} \mapsto -iX$, $\mathbf{j} \mapsto -iY$, $\mathbf{k} \mapsto -iZ$ determines a continuous (Lie) group isomorphism from the unit quaternions to $SU(2)$ which intertwines their actions on S^3 .

We saw in Example 5.2.2 that the Bloch sphere represents the pure states in \mathbb{C}^2 . In fact, the interior of the sphere corresponds to the mixed states in \mathbb{C}^2 .

Example 5.4.9 (Bloch vectors for mixed states in \mathbb{C}^2) — Given $\vec{n} = (x, y, z) \in \mathbb{R}^3$, we define

$$\vec{n} \cdot \vec{\sigma} := n_1\sigma^1 + n_2\sigma^2 + n_3\sigma^3 = xX + yY + zZ.$$

We can parametrize the density matrices in $M_2(\mathbb{C})$ as

$$d_{\vec{n}} := \frac{1}{2}(I + \vec{n} \cdot \vec{\sigma}) \quad \text{where } \|\vec{n}\| \leq 1.$$

Interpreting $\frac{1}{2}\vec{n}$ as a vector emanating from $\frac{1}{2}I$ viewed as the center of the Bloch sphere, we see that $d_{\vec{n}}$ is pure exactly when $\|\vec{n}\| = 1$ and mixed otherwise. (In this normalization, the Bloch sphere has radius $1/2$.)

Exercise 5.4.10. Let $r \in M_2(\mathbb{C})$ such that $r^2 = 1$, $r^\dagger = r$, and $r \neq \pm 1$. Show there exists a unique unit vector $\vec{n} \in \mathbb{R}^3$ such that $r = \vec{n} \cdot \vec{\sigma}$.

Exercise 5.4.11. Given $\vec{n} \in \mathbb{R}^3$ with $\|\vec{n}\| = 1$ and $\theta \in \mathbb{R}$, find $\alpha, \beta \in \mathbb{C}$ such that

$$e^{i\theta\vec{n}\cdot\vec{\sigma}} = \alpha 1 + \beta \vec{n} \cdot \vec{\sigma}.$$

Show that this operator is in $SU(2)$.

Exercise 5.4.12. Let $\vec{n}, \vec{m} \in \mathbb{R}^3$. Prove that

$$[\vec{n} \cdot \vec{\sigma}, \vec{m} \cdot \vec{\sigma}] = 2i(\vec{n} \times \vec{m}) \cdot \vec{\sigma}$$

where $\vec{n} \times \vec{m}$ is the usual cross product. Prove that if \vec{n}, \vec{m} are orthogonal, then

$$\{\vec{n} \cdot \vec{\sigma}, \vec{m} \cdot \vec{\sigma}\} = 0$$

where $\{A, B\} := AB + BA$.

Exercise 5.4.13. Let $\vec{n}, \vec{m} \in \mathbb{R}^3$ with $\|\vec{n}\| = 1$ and let $\theta \in \mathbb{R}$. Prove that

$$e^{-i\theta\vec{n}\cdot\vec{\sigma}} \frac{1}{2}(I + \vec{m} \cdot \vec{\sigma}) e^{i\theta\vec{n}\cdot\vec{\sigma}} = \frac{1}{2}(I + (R_{2\theta}\vec{m}) \cdot \vec{\sigma})$$

where $R_{2\theta} \in M_3(\mathbb{R})$ is the rotation matrix by an angle 2θ about \vec{n} .

Note: This exercise gives a double cover

$$SU(2) \longrightarrow SO(3) := \{w \in M_3(\mathbb{R}) \mid w^T = w^{-1} \text{ and } \det(w) = 1\}.$$

The notion of *von Neumann entropy* quantifies when a state is pure or mixed.

Definition 5.4.14 (von Neumann entropy) — Let $d \in B(H)$ be a density matrix. Its *von Neumann entropy* is

$$S(d) := -\text{Tr}(d \cdot \log(d)),$$

where $\log(d)$ is defined using the functional calculus, with the convention that $0 \cdot \log(0) = 0$. Using the Spectral Theorem 1.7.9, $S(d)$ equals the Shannon entropy of the probability distribution associated to the spectral decomposition of d

$$S(d) = - \sum_{\lambda \in \text{spec}(d)} \lambda \log(\lambda)$$

where the sum is taken with multiplicity.

Remark 5.4.15. The von Neumann entropy of a mixed state is the quantum mechanical analog of the classical Shannon entropy. Just as with Shannon entropy, von Neumann entropy is always non-negative.

Exercise 5.4.16. Prove that a density matrix $d \in B(H)$ is a rank one projection if and only if $S(d) = 0$.

Exercise 5.4.17. Let $d \in M_2(\mathbb{C})$ be density matrix. Compute $S(d)$ as a function of the distance between d and $\frac{1}{2}I$. (Recall that the radius of the Bloch sphere is $1/2$.) Deduce that if d_0, d_1 are two density matrices in $M_2(\mathbb{C})$, then

$$tS(d_1) + (1-t)S(d_0) \leq S(d_t) \quad \forall t \in [0, 1].$$

Definition 5.4.18 (Mixed state Born rule) — Let $d \in B(H)$ be a density matrix, and let $\varphi_d(\cdot) := \text{Tr}(d \cdot)$ be the corresponding state. Given an observable $x \in B(H)$ with spectral decomposition $x = \sum_{\lambda \in \text{spec}(x)} \lambda p_\lambda$, the probability of measuring $x = \lambda$ is given by

$$\mathbf{P}_d(x = \lambda) := \text{Tr}(dp_\lambda) = \varphi_d(p_\lambda).$$

After measuring $x = \lambda$, the density matrix of the system is then given by

$$\frac{1}{\text{Tr}(dp_\lambda)} p_\lambda dp_\lambda.$$

Exercise 5.4.19. Show that $\mathbf{P}_d(x = \lambda)$ is a probability distribution on $\text{spec}(x)$.

Exercise 5.4.20. Show that

$$\frac{1}{\text{Tr}(dp_\lambda)} p_\lambda dp_\lambda$$

is the unique density matrix d' with $d' \leq p_\lambda$ closest to d in the GNS norm with respect to Tr .

Remark 5.4.21. We may interpret the mixed state Born rule in terms of *Bayesian probability*, where $\mathbf{P}_d(x = \lambda)$ is a weighted average over the spectrum of d . Indeed, consider the spectral decomposition $d = \sum_{r \in \text{spec}(d)} r q_r$. Then

$$\mathbf{P}_d(x = \lambda) = \sum_{r \in \text{spec}(d)} \underbrace{\mathbf{P}_d(x = \lambda | d = r)}_{\text{Tr}(p_\lambda \frac{q_r}{\text{Tr}(q_r)})} \underbrace{\mathbf{P}_d(d = r)}_{r \text{Tr}(q_r)} = \sum_{r \in \text{spec}(d)} r \text{Tr}(p_\lambda q_r) = \text{Tr}(dp_\lambda).$$

Definition 5.4.22 — The *expectation value* of the observable x is the mixed state d is given by

$$\sum_{\lambda \in \text{spec}(x)} \lambda \cdot \mathbf{P}_d(x = \lambda) = \sum_{\lambda \in \text{spec}(x)} \lambda \text{Tr}(dp_\lambda) = \text{Tr}(dx) = \varphi_d(x).$$

Observe that this final quantity makes sense for *any* operator, not just self-adjoints.

Facts 5.4.23. We have the following facts about mixed states/density matrices and the expectation values they give for operators.

(ME1) Every density matrix is a convex combination of pure states. Writing such a d in this way is called a *pure state decomposition* of d .

(ME2) The space of density matrices in $B(H)$ is a convex set.

(ME3) For any pure state decomposition $d = \sum_i r_i |e_i\rangle\langle e_i|$ and any observable $x \in B(H)$,

$$\mathbf{P}_d(x = \lambda) = \sum_i r_i \mathbf{P}_{|e_i\rangle}(x = \lambda) \quad \text{and} \quad \mathbf{E}_d(x) = \sum_{\lambda \in \text{spec}(x)} \sum_i \lambda r_i \langle e_i | p_\lambda | e_i \rangle.$$

Hence $\mathbf{E}_d(p_\lambda) = \mathbf{P}_d(x = \lambda)$.

(ME4) Define $f(t) := -t \log(t)$ on $[0, 1]$. For a density matrix $d \in B(H)_+$ and a state vector $|\psi\rangle \in H$,

$$\mathbf{E}_{|\psi\rangle}(f(d)) \leq f(\mathbf{E}_{|\psi\rangle}(d))$$

with equality if and only if $|\psi\rangle$ is an eigenstate for d .

Proof. Write $|\psi\rangle = \sum_{\lambda \in \text{spec}(d)} \alpha_\lambda |e_\lambda\rangle$ as a superposition of eigenstates for d . Then

$$\mathbf{E}_{|\psi\rangle}(f(d)) = \langle \psi | f(d) | \psi \rangle = \sum |\alpha_\lambda|^2 f(\lambda) \leq f\left(\sum |\alpha_\lambda|^2 \lambda\right) = f(\langle \psi | d | \psi \rangle) = f(\mathbf{E}_{|\psi\rangle}(d))$$

by concavity of f (see Remark 5.4.24 below). In fact, f is *strictly concave* on $(0, 1)$, and thus equality holds if and only if only one $\alpha_\lambda \neq 0$. \square

(ME5) For a density $d \in B(H)_+$ and an ONB $\{|e_i\rangle\}$ of H ,

$$S(d) = \text{Tr}(f(d)) = \sum_i \langle e_i | f(d) | e_i \rangle = \sum_i \mathbf{E}_{|e_i\rangle}(f(d)) \stackrel{(\text{ME4})}{\leq} \sum_i f(\mathbf{E}_{|e_i\rangle}(d))$$

with equality if and only if each $|e_i\rangle$ is an eigenstate of d .

Proof. The only way the right hand side can sum to $S(d)$ is if we have equality of each of the terms $\mathbf{E}_{|e_i\rangle}(f(d)) = f(\mathbf{E}_{|e_i\rangle}(d))$. Now apply (ME4). \square

Remark 5.4.24. Recall that a function $f : [0, r] \rightarrow \mathbb{R}$ is called *concave* if

$$tf(\alpha) + (1-t)f(\beta) \leq f(t\alpha + (1-t)\beta) \quad \forall \alpha, \beta \in [0, r] \text{ and } 0 \leq t \leq 1,$$

and f is called *strictly concave* if equality holds above if and only if $\alpha = \beta$ or $t \in \{0, 1\}$. Fact (ME4) above can easily be modified to apply to a positive $x \in B(H_+)$ and a (strictly) concave function on $[0, r]$ for $r = \|x\|$.

Proposition 5.4.25 — Von Neumann entropy is strictly concave on the convex space of density matrices. That is, suppose $d_0, d_1 \in B(H)$ are density matrices. Then the density matrix $d_t := td_1 + (1-t)d_0$ for $t \in [0, 1]$ satisfies

$$tS(d_1) + (1-t)S(d_0) \leq S(d_t).$$

Moreover, S is constant along the line segment $t \mapsto d_t$ if and only if $d_0 = d_1$.

Proof. Consider the concave function $f(t) := -t \log(t)$ on $[0, 1]$. Fix $r \in (0, 1)$, and let $|e_i\rangle$ be an ONB of eigenvectors of d_r . By concavity of f ,

$$\begin{aligned}
rS(d_1) + (1-r)S(d_0) &\leq \sum_i r f(\mathbf{E}_{|e_i\rangle}(d_1)) + (1-r) f(\mathbf{E}_{|e_i\rangle}(d_0)) && \text{(ME5)} \\
&\leq \sum_i f(r\mathbf{E}_{|e_i\rangle}(d_1) + (1-r)\mathbf{E}_{|e_i\rangle}(d_0)) && \text{(Concavity)} \\
&= \sum_i f(\mathbf{E}_{|e_i\rangle}(\underbrace{rd_1 + (1-r)d_0}_{d_r})) && \text{(Linearity)} \\
&= \sum_i \mathbf{E}_{|e_i\rangle}(f(d_r)) && \text{(ONB of eigenstates)} \\
&= S(d_r).
\end{aligned}$$

If moreover $t \mapsto S(d_t)$ is constant along the line segment $[0, 1]$, then we must have $S(d_j) = \sum_i f(\mathbf{E}_{|e_i\rangle}(d_j))$ for $j = 0, 1$. Hence each $|e_i\rangle$ is an eigenvector for d_j for $j = 0, 1$ by (ME5). By (S2), we know that each d_t gives rise to the same probability distribution, i.e., $\text{spec}(d_t)$ is constant (with multiplicity) on $[0, 1]$. We conclude that $t \mapsto d_t$ is constant. \square

We record the following corollary for future use.

Corollary 5.4.26 — The entropy S is maximized at a unique point on every closed convex subset of density matrices. Moreover the minimum of S occurs at an *extreme point* in the sense of Exercise 5.4.5.

Proof. That the minimum occurs at an extreme point is immediate from Proposition 5.4.25. We focus on the unique maximum.

Existence: Observe that a closed convex set is compact and $S(d) = -\text{Tr}(d \log(d))$ is continuous by Proposition 1.7.15, so S attains its maximum by the Extreme Value Theorem.

Uniqueness: If d_0, d_1 both attain the maximum, then since $S(d_t) \geq tS(d_1) + (1-t)S(d_0)$, S is constant along the line segment $t \mapsto d_t = td_1 + (1-t)d_0$. We conclude that $d_0 = d_1$ by Proposition 5.4.25. \square

5.5 Separability and partial traces

Recall that a state vector $|\psi\rangle$ in a bipartite system $H_A \otimes H_B$ is *separable* if $|\psi\rangle = |\eta\rangle \otimes |\xi\rangle$ for some $|\eta\rangle \in H_A$ and $|\xi\rangle \in H_B$ and *entangled* otherwise. We now generalize these notions to mixed states.

Definition 5.5.1 (Separability for mixed states) — Consider a density matrix d for a bipartite system $H_A \otimes H_B$ and its corresponding state φ_d . We call d *simply separable* or a *product state* if $d = d_A \otimes d_B$ for some density matrices $d_A \in B(H_A)$ and $d_B \in B(H_B)$, and we call d *separable* if it is a convex combination of product states. If d is not separable, it is called *entangled*.

Proposition 5.5.2 — A pure state density matrix $d \in B(H_A \otimes H_B)$ is separable if and only if it is a product state.

Proof. It suffices to prove that if d is pure and separable, then d is simply separable. Since d is separable there are $|\eta_i\rangle \in H_A$ and $|\xi_i\rangle \in H_B$ for $i = 1, \dots, n$ and a probability distribution $\{\lambda_i\}$ such that

$$d = \sum \lambda_i |\eta_i\rangle\langle\eta_i| \otimes |\xi_i\rangle\langle\xi_i|.$$

Since d is pure, $S(d) = -\sum \lambda_i \log(\lambda_i) = 0$, which implies there is a unique i such that $\lambda_i = 1$ and all other λ_j are zero. We conclude d is a product state. \square

Example 5.5.3 — The mixed state $\frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|)$ is separable, but not simply separable.

Definition 5.5.4 (Partial trace and reduced density) — Consider a bipartite system with Hilbert space $H \otimes K$. Recall from Construction 3.7.10 that there is a unique trace-preserving conditional expectation $\mathbf{E}_{B(H)} : B(H \otimes K) \rightarrow B(H)$ satisfying

$$\mathrm{Tr}_{B(H)}(\mathbf{E}(x)y) = \mathrm{Tr}_{B(H \otimes K)}(xy) \quad \forall x \in B(H \otimes K), y \in B(H).$$

In quantum information theory, this conditional expectation is also called the *partial trace* and is often denoted (confusingly) by Tr_H . In this book, we will always use a bold font \mathbf{Tr}_H for the partial trace to help alleviate any confusion. Given a density matrix $d \in B(H \otimes K)$, observe that the *reduced density matrix* $d_H := \mathbf{Tr}_H(d) \in B(H)$ is again positive with trace 1, and is thus again a density. One may similarly define the partial trace \mathbf{Tr}_K and reduced density d_K .

When we use $H_A \otimes H_B$, we often write $\mathbf{Tr}_A, \mathbf{Tr}_B$ for $\mathbf{Tr}_{H_A}, \mathbf{Tr}_{H_B}$ and d_A, d_B for d_{H_A}, d_{H_B} respectively.

Exercise 5.5.5. Show that in the graphical calculus, \mathbf{Tr}_H is given by

$$\mathbf{Tr}_H(x) = \left(\begin{array}{c} H \\ | \\ \boxed{x} \\ | \\ H \end{array} \right) \overline{K} = \sum_i \left(\begin{array}{c} H \\ | \\ \boxed{x} \\ | \\ H \end{array} \right) \begin{array}{c} \boxed{\langle \xi_i |} \\ K \\ \boxed{|\xi_i \rangle} \\ K \end{array} \quad \forall \text{ ONB } \{|\xi_i\rangle\} \text{ of } K.$$

Remark 5.5.6. A quantum system H_S often interacts with its environment, usually represented by a large Hilbert space H_E . If this combined bipartite system is well-isolated, we describe the state of the entire system by some $|\psi\rangle \in H_S \otimes H_E$. Given that the environment usually has substantially many more degrees of freedom than our system of interest, and we have no way of observing all these degrees of freedom, it is generally not feasible to understand the dynamics of the total state $|\psi\rangle$. Thus the reduced density $d_S = \mathbf{Tr}_E(|\psi\rangle\langle\psi|)$ becomes the primary description of our quantum state rather than $|\psi\rangle$.

Definition 5.5.7 — Let $d \in B(H)$ be a density matrix. A *purification* of d is a pure density matrix (rank one projection) $p \in B(H \otimes L)$ where L is an ancillary Hilbert space such that $d = \mathbf{Tr}_L(p)$.

Construction 5.5.8 (Quantum state purification) — Every density matrix has a purification. Indeed, let $d \in B(H)$ be a density matrix and let $\{|e_i\rangle\}$ be an ONB of H consisting of eigenvectors for d with corresponding eigenvalues λ_i . Observe that the pure state

$$|\psi\rangle := \sum \sqrt{\lambda_i} |e_i\rangle \otimes |e_i\rangle \in H \otimes H$$

purifies d as

$$\begin{aligned} \mathbf{Tr}_{H_1}(|\psi\rangle\langle\psi|) &= \mathbf{Tr}_{H_1} \left(\sum \lambda_i |e_i\rangle\langle e_i| \otimes |e_i\rangle\langle e_i| \right) = \sum \lambda_i \mathbf{Tr}_{H_1} (|e_i\rangle\langle e_i| \otimes |e_i\rangle\langle e_i|) \\ &= \sum \lambda_i |e_i\rangle\langle e_i| = d. \end{aligned}$$

The next two constructions follow directly from the Spectral Theorem 1.7.9 and the polar decomposition; we have delayed them until now as this is the section in which these constructions will be used.

Construction 5.5.9 (Singular Value Decomposition) — Let $x \in M_{m \times n}(\mathbb{C})$. Setting $k = \min\{m, n\}$, There are unique $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ called the *singular values* of x and

orthonormal sets $\{|e_i\rangle\}_{i=1}^k \subset \mathbb{C}^n$ and $\{|f_i\rangle\}_{i=1}^k \subset \mathbb{C}^m$ (which are not unique) such that

$$x = \sum_{i=1}^k \lambda_i |f_i\rangle \langle e_i|.$$

This expression is called a *singular value decomposition* (SVD) of x . First, observe that the λ_i are necessarily the elements of $\text{spec}(|x|)$ counted with multiplicity, as

$$x^\dagger x = \sum_{i=1}^k \lambda_i^2 |e_i\rangle \langle e_i|.$$

This tells us how to reverse-engineer the SVD. Let $x = u|x|$ be the polar decomposition of x , and let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ be a monotone ordering, with multiplicity, of $\text{spec}(|x|)$. Choose an ONB $\{|e_i\rangle\}$ for \mathbb{C}^n diagonalizing $|x|$ so that $|x| \cdot |e_i\rangle = \lambda_i |e_i\rangle$ and setting $|f_i\rangle := u|e_i\rangle$ yields the result. Indeed, one observes that

$$\text{rank}(|x|) = \text{rank}(x^\dagger x) \leq k = \min\{m, n\},$$

so that $\lambda_i = 0$ whenever $i > k$.

The following closely related construction is used frequently in quantum information.

Construction 5.5.10 (Schmidt Decomposition) — For any state vector $|\psi\rangle \in \mathbb{C}^m \otimes \mathbb{C}^n$, setting $k := \min\{m, n\}$, there are non-negative numbers $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ and orthonormal sets $\{|\eta_i\rangle\}_{i=1}^k \subset \mathbb{C}^n$ and $\{|\xi_j\rangle\}_{j=1}^k \subset \mathbb{C}^m$ such that

$$|\psi\rangle = \sum_{i=1}^k \lambda_i |\eta_i\rangle \otimes |\xi_i\rangle$$

Indeed, choose an arbitrary unitary $w : \mathbb{C}^n \rightarrow \overline{\mathbb{C}^n}$, which gives an isomorphism

$$W : \mathbb{C}^m \otimes \mathbb{C}^n \xrightarrow{1 \otimes w} \mathbb{C}^m \otimes \overline{\mathbb{C}^n} \cong M_{m \times n}(\mathbb{C}).$$

Equipping $M_{m \times n}(\mathbb{C})$ with the inner product $\langle x|y\rangle := \text{Tr}_n(x^\dagger y)$ promotes W to a unitary. Use Construction 5.5.9 to find a SVD

$$W|\psi\rangle = \sum_{i=1}^k \lambda_i |f_i\rangle \langle e_i|, \quad \implies \quad |\psi\rangle = \sum_{i=1}^k \lambda_i W^\dagger |f_i\rangle \langle e_i|.$$

By construction, $W^\dagger |f_i\rangle \langle e_i| = |f_i\rangle \otimes |\xi_i\rangle$ for the orthonormal set $\{|\xi_i\rangle\}_{i=1}^k := \{w^\dagger \langle e_i|\}_{i=1}^k \subset$

\mathbb{C}^n , so

$$|\psi\rangle = \sum_{i=1}^k \lambda_i |f_i\rangle \otimes |\xi_i\rangle$$

as desired.

We close this section with the following useful application.

Definition 5.5.11 — Suppose $d \in B(H_A \otimes H_B)$ is a density matrix. The *bipartite entanglement entropy* of d with respect to A is $S_A(d) := S(d_A)$. Similarly, we define $S_B(d) := S(d_B)$. The *mutual quantum information* is

$$I_{A:B}(d) := S_A(d) + S_B(d) - S(d).$$

Exercise 5.5.12. Prove that the space of density matrices in $B(H_A \otimes H_B)$ with fixed reduced densities $d_A \in B(H_A)$ and $d_B \in B(H_B)$ is a closed convex set. Use Corollary 5.4.26 to deduce that S attains its maximum at a unique point in this closed convex set.

Exercise 5.5.13. Prove that for $d = d_A \otimes d_B$, $S(d) = S_A(d) + S_B(d)$.

Remark 5.5.14. It can be shown that for a density $d \in B(H_A) \otimes B(H_B)$,

$$|S_A(d) - S_B(d)| \leq S(d) \leq S_A(d) + S_B(d).$$

The second inequality is equivalent to the statement that $I_{A:B}(d) \geq 0$. Moreover, $I_{A:B}(d) = 0$ exactly when $d = d_A \otimes d_B$.

Proposition 5.5.15 — Given a density matrix $d \in B(H_A \otimes H_B)$, if d is pure, then $S_A(d) = S_B(d)$. In this case, d is (simply) separable if and only if $S_A(d) = S_B(d) = 0$.

Proof. Since d is pure, we may write $d = |\psi\rangle\langle\psi|$. We claim that $|\psi\rangle$ is common a purification of both d_A and d_B . Indeed, we may use the Schmidt Decomposition 5.5.10 to write

$$|\psi\rangle = \sum_{i=1}^k \lambda_i |e_i\rangle \otimes |f_i\rangle.$$

Observe that

$$d_A = \sum_{i=1}^k \lambda_i |e_i\rangle\langle e_i| \quad \text{and} \quad d_B = \sum_{i=1}^k \lambda_i |f_i\rangle\langle f_i|,$$

so the spectral decompositions of d_A and d_B yield identical probability distributions. It immediately follows that $S(d_A) = S(d_B)$. As $S(d_A) = S(d_B)$ is solely a function of the λ_i , the final statement is immediate. \square

The final statement in the above proposition holds more generally.

Proposition 5.5.16 — Suppose $d \in B(H_A \otimes H_B)$ is a density such that $d_B = \text{Tr}_A(d)$ is pure. Then $d = d_A \otimes d_B$ is simply separable.

Proof. Since d_B is pure, it is an orthogonal projection of rank 1. Writing $p = 1 - d_B$, we have

$$\text{Tr}_A((1 \otimes p)d(1 \otimes p)) = p\text{Tr}_A(d)p = pd_Bp = 0.$$

Since Tr_A is faithful, $d(1 \otimes p) = 0$ and $(1 \otimes p)d = 0$. Hence $d(1 \otimes d_B) = d = (1 \otimes d_B)d$. Expanding d as a sum of simple tensors in $B(H_A \otimes H_B)$, we see that d is of the form $d_A \otimes d_B$ as claimed. \square

5.6 Pure state error correction and stabilizer codes

The quantum information contained in a quantum state is subject to noise from the environment. We will explore this in terms of a bipartite system where one Hilbert space is our quantum system and one Hilbert space is the environment in §5.8 below. However, we begin this section with a more basic example.

Example 5.6.1 (Bit-flip) — Suppose we have a single noisy qubit state $|\psi\rangle \in \mathbb{C}^2$. We choose to measure in the Z computational basis, so we view $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ with $|\alpha|^2 + |\beta|^2 = 1$. Our noise takes the form of a random *bit-flip*, where $|0\rangle$ and $|1\rangle$ are flipped with probability $p \in (0, 1)$; observe this is achieved by applying the *error operator* X to $|\psi\rangle$. After this random error occurs, our qubit $|\psi'\rangle$ is only equal to $|\psi\rangle$ with probability $1 - p$. Even worse, there is no measurement or process we can perform to tell if an error occurred or to correct an error without destroying all the quantum information.

Now suppose we *encode* our one qubit in a 3-qubit system $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$ where we encode $|0\rangle$ as $|000\rangle$ and $|1\rangle$ as $|111\rangle$. We assume we are using the same hardware, so that each of these 3 qubits is still noisy with a probability p of a bit flip. Now we can correct for the optimistic scenario that at most one bit flip occurs. We make two quantum measurements, namely $Z_1Z_2 = Z \otimes Z \otimes 1$ and $Z_2Z_3 = 1 \otimes Z \otimes Z$ in our state $|\psi'\rangle$, which is manifestly an eigenvector for these operators assuming only bit-flip errors. Hence measuring Z_1Z_2 and Z_2Z_3 does not affect $|\psi'\rangle$. We claim that if at most one qubit has been flipped, we may recover our initial state $|\psi\rangle$ from these measurements.

Z_1Z_2	Z_2Z_3	Qubit flipped?	State $ \psi'\rangle$	Recovery operator
1	1	None	$\alpha 000\rangle + \beta 111\rangle$	1
1	-1	3rd qubit	$\alpha 001\rangle + \beta 110\rangle$	X_3
-1	1	1st qubit	$\alpha 100\rangle + \beta 011\rangle$	X_1
-1	-1	2nd qubit	$\alpha 010\rangle + \beta 101\rangle$	X_2

We may recover $|\psi\rangle$ from $|\psi'\rangle$ by applying the recovery operator indicated above.

The idea of a quantum error correction code is to use a large number of n qubits to encode our k qubits so that they are robust to errors. We write $H_n := \bigotimes_{k=1}^n \mathbb{C}^2$ for the n -qubit Hilbert space.

Definition 5.6.2 — A *quantum error correction code* (QECC) for a k -qubit system is an isometry $H_k \rightarrow H_n$ where typically $n \gg k$. More generally, a *code subspace* is a subspace $C \subset H$ of the Hilbert space H representing our quantum system.

A particularly well-behaved example of QECCs is the family of *stabilizer codes*, which includes the bit-flip Example 5.6.1. We write

$$\sigma_k^i := \underbrace{1 \otimes \cdots \otimes 1}_{k-1} \otimes \sigma^i \otimes \underbrace{1 \otimes \cdots \otimes 1}_{n-k}$$

for the i -th Pauli operator for $i = 1, 2, 3$ acting on site k ; we may similarly write X_k, Y_k, Z_k for these operators.

Definition 5.6.3 — The Pauli group P_N is the subgroup of the unitary group $U(N)$ generated by the operators X_k, Y_k, Z_k for $k = 1, \dots, N$. A *stabilizer group* is an abelian subgroup $S \subset P_N$ such that the *code subspace*

$$C_S := \{|\psi\rangle \in H_N \mid s|\psi\rangle = |\psi\rangle \text{ for all } s \in S\}$$

is non-zero. That is, there is some state $|\psi\rangle$ such that S *stabilizes* $|\psi\rangle$.

Warning 5.6.4 — Observe that $-1, \pm i \notin S$ for any stabilizer code.

Exercise 5.6.5. Compute how many elements P_N contains. Then show that all stabilizer groups are *elementary 2-groups*, i.e., $s^2 = 1$ for all $s \in S$. Deduce that $S \cong (\mathbb{Z}/2)^n$ for some $n \in \mathbb{N}$.

Exercise 5.6.6. Prove that the 3-qubit *GHSZ state*

$$|GHSZ\rangle := \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$$

spans the code subspace of the stabilizer group $S := \langle Z_1 Z_2, X_1 X_2 X_3, Z_2 Z_3 \rangle \subset P_3$.

Proposition 5.6.7 — Suppose $T \subset P_N$ is an abelian subgroup. Then there is a stabilizer group $S \subset P_N$ such that $\mathbb{C}[S] = \mathbb{C}[T]$.

Proof. Since $\mathbb{C}[T] \subset B(H_N)$ is a commutative unitary algebra, $\mathbb{C}[T] \cong \mathbb{C}^k$ for some k . There is an ONB of H_N in which all operators in $\mathbb{C}[T]$ are diagonal; indeed, we can find a single

$x \in \mathbb{C}[T] \cong \mathbb{C}^k$ which generates it as a unitary algebra and apply the Spectral Theorem 1.7.9. Pick $|\psi\rangle$ in this ONB, and consider

$$S := \{s \in T \mid s|\psi\rangle = |\psi\rangle\}.$$

Then clearly $S \subset P_N$ is a stabilizer group as $|\psi\rangle \in C_S \neq 0$. Moreover, $\mathbb{C}[S] = \mathbb{C}[T]$ as each $t \in T$ lies in S up to a scalar. \square

Remark 5.6.8. The physical content of Proposition 5.6.7 is that all information that can be obtained from measuring observables from T is already contained in measuring observables from S .

Stabilizer codes can easily correct for *Pauli error operators*, i.e., simple tensors of Pauli operators. Of course, they can correct for even more types of errors, but we focus on the Pauli error operators for simplicity.

Exercise 5.6.9. Prove that simple tensors of Pauli operators either commute or anticommute.

Definition 5.6.10 — Let $E \subset B(H_N)$ be a subset of Pauli error operators containing 1. For a stabilizer group $S \subset P_N$, since Pauli operators either commute or anticommute, for each $s \in S$ and $e \in E$, there is a sign $\sigma(s, e) \in \{\pm 1\}$ such that

$$es = \sigma(s, e)se.$$

The function $\sigma : S \times E \rightarrow \{\pm 1\}$ is called the *syndrome* of (S, E) . We call $e \in E$ *detectable* if $\sigma(s, e) = -1$ for some $s \in S$. We call two errors $e, f \in E$:

- *distinguishable* if there is an $s \in S$ such that $\sigma(s, e) \neq \sigma(s, f)$, and
- *equivalent* if every $|\psi\rangle \in C_S$ is an eigenvector for $e^\dagger f$.

A stabilizer group is called an *E-stabilizer code* if all indistinguishable errors $e, f \in E$ are equivalent.

Exercise 5.6.11. Show that for Example 5.6.1 of the bit-flip, the stabilizer group $S = \langle Z_1 Z_2, Z_2 Z_3 \rangle$ is an *E-stabilizer code* for $E = \{1, X_1, X_2, X_3\}$.

Exercise 5.6.12. Prove that the following are equivalent for Pauli error operators $e, f \in E$ with respect to a stabilizer group S .

- (1) $e, f \in E$ are indistinguishable, and
- (2) $e^\dagger f$ commutes with every $s \in S$.

Deduce that if e, f are indistinguishable, then

- $e^\dagger f$ commutes with p_{C_S} , and
- e, f are equivalent if and only if $p_{C_S} e^\dagger f p_{C_S} \in \mathbb{C} p_{C_S}$.

Exercise 5.6.13. Let $E \subset B(H_N)$ be a subset of Pauli error operators containing 1. Show that every $e \in E$ which is not multiplication by a constant phase on C_S is detectable.

We now give an error correction protocol for E -stabilizer codes.

Construction 5.6.14 (Error correction protocol for stabilizer codes) — Given a set E of Pauli operators containing 1 and an E -stabilizer code S , we can construct an explicit recovery operator based on the syndrome function $\sigma : S \times E \rightarrow \{\pm 1\}$ determined by

$$es = \sigma(s, e)se.$$

The above formula implies that for all $|\psi\rangle \in C_S$ and $e \in E$, $e|\psi\rangle$ is again an eigenvector for each s , and so we may measure s without altering $e|\psi\rangle$, which will always return the syndrome $\sigma(s, e)$ as the outcome. Thus even though we do not know which error $e \in E$ has been applied, we do know the function $\sigma(-, e) : S \rightarrow \{\pm 1\}$. Since our function $\sigma : S \times E \rightarrow \{\pm 1\}$ is completely known, we can now choose an arbitrary $f \in E$ such that $\sigma(-, e) = \sigma(-, f)$. Then e, f are indistinguishable, so e, f are equivalent, and thus $|\psi\rangle$ is an eigenvector for $f^\dagger e$, so

$$f^\dagger e|\psi\rangle\langle\psi|e^\dagger f = |\psi\rangle\langle\psi|.$$

Exercise 5.6.15 (Phase-flip). Find an E -stabilizer code for $E = \{Z_1, Z_2, Z_3\}$ on a 3-qubit quantum system.

Note: The name ‘phase-flip’ comes from the fact that $Z|j\rangle = (-1)^j|j\rangle$ in the Z -computational basis.

Example 5.6.16 (Shor’s code) — Suppose we have a noisy qubit which is now subject to bit-flips, phase-flips, and possibly both. Thus, our possible errors are 1, X , Z , XZ , which generate $M_2(\mathbb{C})$ as a vector space. (Note that $ZX = -XZ$, so ZX and XZ are clearly equivalent errors.) We now encode our logical qubit as a subspace in a 9-qubit system, and we assume we will have at most one noisy qubit in this system.

$$\bullet \quad \rightsquigarrow \quad \begin{array}{ccc} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{array}$$

We encode $\mathbb{C}^2 \rightarrow ((\mathbb{C}^2)^{\otimes 3})^{\otimes 3}$ by

$$\begin{aligned} |0\rangle &\mapsto \frac{1}{2\sqrt{2}}(|000\rangle + |111\rangle) \otimes (|000\rangle + |111\rangle) \otimes (|000\rangle + |111\rangle) \\ |1\rangle &\mapsto \frac{1}{2\sqrt{2}}(|000\rangle - |111\rangle) \otimes (|000\rangle - |111\rangle) \otimes (|000\rangle - |111\rangle). \end{aligned}$$

One checks that these two states span the codes subspace C_S for the stabilizer group generated by

$$\begin{aligned} &X_1X_2X_3X_4X_5X_6, \quad X_1X_2X_3X_7X_8X_9, \quad X_4X_5X_6X_7X_8X_9, \\ &Z_1Z_2, \quad Z_2Z_3, \quad Z_1Z_3, \quad Z_4Z_5, \quad Z_5Z_6, \quad Z_4Z_6, \quad Z_7Z_8, \quad Z_8Z_9, \quad Z_7Z_9 \end{aligned}$$

Notice this generating set can be reduced by omitting $X_4X_5X_6X_7X_8X_9$, Z_1Z_3 , Z_5Z_6 , and Z_8Z_9 .

In order to correct for one noisy qubit out of these 9, we do (at most) 4 separate sets of measurements. First, we measure the X -stabilizers.

$X_1X_2X_3X_4X_5X_6$	$X_1X_2X_3X_7X_8X_9$	Recovery Operator
1	1	1
1	-1	Z_7
-1	1	Z_4
-1	-1	Z_1

We then measure some of the Z -stabilizers

Z_1Z_2	Z_2Z_3	Recovery Operator
1	1	1
1	-1	X_3
-1	1	X_1
-1	-1	X_2

and similarly for the subsets Z_4Z_5, Z_5Z_6 and Z_7Z_8, Z_8Z_9 . Since we assumed only qubit was noisy, we should have only observed a -1 in two of these sets of stabilizer measurements. We leave the rest of the details to the reader.

The following theorem gives equivalent conditions for a code subspace C to correct an error set E . We defer a discussion of an error correction protocol for such (C, E) to Construction 5.6.22 below.

Theorem 5.6.17 (Knill-Laflamme [KL97]) — Suppose $E \subset B(H)$ is a subset of error operators. The following are equivalent for a code subspace $C \subset H$.

(KL1) For any state vectors $|\psi\rangle, |\phi\rangle \in C$,

$$\langle\psi|e^\dagger f|\phi\rangle = \lambda_{e,f} \cdot \langle\psi|\phi\rangle$$

where $\lambda_{e,f} \in \mathbb{C}$ is a constant independent of $|\psi\rangle, |\phi\rangle$.

(KL2) For all $e, f \in E$, $p_C e^\dagger f p_C \in \mathbb{C} p_C$.

(KL3) For all $e \in \text{span}(E)$ and state vectors $|\psi\rangle, |\phi\rangle \in C$,

$$\langle\psi|e^\dagger e|\psi\rangle = \langle\phi|e^\dagger e|\phi\rangle.$$

Proof.

(KL1) \Rightarrow (KL2): For an ONB $\{\xi_i\}$ of C and $e, f \in E$, we see that

$$p e^\dagger f p = \sum_{i,j} |\xi_i\rangle\langle\xi_i| e^\dagger f |\xi_j\rangle\langle\xi_j| = \lambda_{e,f} \sum_{i,j} |\xi_i\rangle\langle\xi_i| \xi_j \langle\xi_j| = \lambda_{e,f} \sum_i |\xi_i\rangle\langle\xi_i| = \lambda_{e,f} p.$$

(KL2) \Rightarrow (KL3): For $e, f \in E$, define $\lambda_{e,f} \in \mathbb{C}$ by $p e^\dagger f p = \lambda_{e,f} p$. Then observe that

$$\langle\psi|e^\dagger f|\psi\rangle = \langle\psi|p e^\dagger f p|\psi\rangle = \lambda_{e,f} \cdot \langle\psi|\psi\rangle = \lambda_{e,f} \cdot \langle\phi|\phi\rangle = \langle\phi|p e^\dagger f p|\phi\rangle = \langle\phi|e^\dagger f|\phi\rangle.$$

(KL3) \Rightarrow (KL1): By polarization for operators in Exercise 1.4.7, we see that for all state vectors $|\psi\rangle, |\phi\rangle$, $\langle\psi|e^\dagger f|\psi\rangle = \langle\phi|e^\dagger f|\phi\rangle$ for all $e, f \in E$; call this number $\lambda_{e,f}$. Then for all $|\eta\rangle \in C$,

$$\langle\eta|e^\dagger f|\eta\rangle = \|\eta\|^2 \cdot \left\langle \frac{\eta}{\|\eta\|} \left| e^\dagger f \right| \frac{\eta}{\|\eta\|} \right\rangle = \lambda_{e,f} \cdot \|\eta\|^2 \cdot \left\langle \frac{\eta}{\|\eta\|} \left| \frac{\eta}{\|\eta\|} \right\rangle = \lambda_{e,f} \cdot \langle\eta|\eta\rangle.$$

The result now follows by polarization for the sesquilinear form $(\eta, \xi) := \langle\eta|e^\dagger f|\xi\rangle$. \square

Exercise 5.6.18. Show that (KL3) implies that an error operator $e \in E$ is either zero or injective on C .

Remark 5.6.19. One strong type of error correcting code can correct for an error set E which is itself a unitary algebra. Having $1 \in E$ allows for the possibility that no error occurred at all, making the error correcting code more robust. Allowing for multiplication of operators in E allows for the simultaneous correction of several errors happening in succession. While there is perhaps not a physical reason to have E closed under adjoints, this is a common situation in practice, and it allows us to correct errors by using operators contained in E itself. We will construct such examples in Part [II] §[??] which are called *locally topologically ordered* spin systems.

Proposition 5.6.20 — Given a set E of Pauli error operators containing 1, a stabilizer code S is an E -stabilizer code if and only if the Knill-Laflamme conditions (KL1)-(KL3) hold for (E, C_S) .

Proof. Suppose S is an E -stabilizer code with syndrome σ , and consider $e, f \in E$. If e, f are distinguishable, pick $s \in S$ such that $\sigma(s, e) \neq \sigma(s, f)$, so $\sigma(s, e)\sigma(s, f) = -1$. Then

$$p_{C_S} e^\dagger f p_{C_S} = p_{C_S} e^\dagger f s p_{C_S} = \sigma(s, f) \cdot p_{C_S} e^\dagger s f p_{C_S} = \sigma(s, e)\sigma(s, f) \cdot p_{C_S} s e^\dagger e f p_{C_S} = -p_{C_S} e^\dagger f p_{C_S},$$

so $p_{C_S} e^\dagger f p_{C_S} = 0$. If e, f are indistinguishable, then by assumption e, f are equivalent, so $p_{C_S} e^\dagger f p_{C_S} \in \mathbb{C} p_{C_S}$. We conclude that (KL2) holds.

Conversely, if (E, C_S) satisfies (KL2), then $p_{C_S} e^\dagger f p_{C_S} \in \mathbb{C} p_{C_S}$ for all $e, f \in E$, in particular for indistinguishable e, f . The result now follows by Exercise 5.6.12. \square

In fact, we can widen our idea of error correction to processes which encode quantum information from system H to another system K .

Exercise 5.6.21. Show that the conditions in Theorem 5.6.17 are equivalent for a set of error operators $E \subset B(H, K)$ where K is another Hilbert space.

We now give an error correction protocol which corrects for error sets $E \subset B(H, K)$ satisfying the Knill-Laflamme conditions (KL1)-(KL3) for a code subspace C .

Construction 5.6.22 — Suppose (E, C) satisfies the Knill-Laflamme conditions (KL1)-(KL3). Starting in state $|\psi\rangle \in C$, suppose an error from E occurs; we do not know which one. We do, however, assume the error $e \in E$ acts injectively on C , and not as zero (see Exercise 5.6.18). We may also assume that E is a linear space, as the Knill-Laflamme conditions still hold replacing E with $\text{span}(E)$.

Consider the action map $a : E \otimes C \rightarrow K$ given by $e \otimes |\psi\rangle \mapsto e|\psi\rangle$. The image of this map is the subspace

$$\text{im}(a) = \text{span} \{e|\psi\rangle \mid |\psi\rangle \in C \text{ and } e \in E\} \subset K,$$

which is generated by the possible outcomes $e|\psi\rangle$ of an error operator $e \in E$ applied to a state $|\psi\rangle \in C$. The formula

$$\langle e|f\rangle := \lambda_{e,f}$$

where $\lambda_{e,f}$ is as in (KL1) defines a sesquilinear form on E ; indeed, this is the map given by

$$E \times E \mapsto E|\psi\rangle \times E|\psi\rangle \xrightarrow{\langle \cdot | \cdot \rangle} \mathbb{C}$$

for any state $|\psi\rangle \in C$. This sesquilinear form is not necessarily definite, but we can quotient out by the length zero vectors N to get a Hilbert space $\widehat{E} := E/N$ with inner product $\langle \widehat{e} | \widehat{f} \rangle := \lambda_{e,f}$. This space can be thought of as any of the subspaces $E|\psi\rangle \subset K$ for any state $|\psi\rangle \in C$ by (KL1).

Now observe that the map $a : E \otimes C \rightarrow K$ descends to a well-defined surjective linear map

$$\widehat{a} : \widehat{E} \otimes C \longrightarrow \text{im}(a) \quad \text{given by} \quad \widehat{e} \otimes |\psi\rangle \longmapsto e|\psi\rangle.$$

Moreover, since

$$\sum_{i,j} \langle \widehat{e}_i \otimes \psi_i | \widehat{e}_j \otimes \psi_j \rangle_{\widehat{E} \otimes C} := \sum_{i,j} \lambda_{e_i, e_j} \langle \psi_i | \psi_j \rangle_H \stackrel{\text{(KL1)}}{=} \sum_{i,j} \langle \psi_i | e_i^\dagger e_j | \psi_j \rangle_K,$$

this map is also isometric and thus unitary. Its adjoint is then given by $e|\psi\rangle \mapsto \widehat{e} \otimes |\psi\rangle$, which is automatically well-defined. We can thus recover $|\psi\rangle\langle\psi|$ up to the positive scalar $\mathbf{E}_{|\psi\rangle}(e^\dagger e) = \langle \psi | e^\dagger e | \psi \rangle_K = \langle \widehat{e} | \widehat{e} \rangle_{\widehat{E}}$ from

$$x = |\widehat{e}\rangle\langle\widehat{e}| \otimes |\psi\rangle\langle\psi|$$

by taking partial trace:

$$\mathbf{Tr}_{\widehat{E}}(|\widehat{e}\rangle\langle\widehat{e}| \otimes |\psi\rangle\langle\psi|) = \langle \widehat{e} | \widehat{e} \rangle_{\widehat{E}} \cdot |\psi\rangle\langle\psi| = \mathbf{E}_{|\psi\rangle}(e^\dagger e) \cdot |\psi\rangle\langle\psi|.$$

The Knill-Laflamme conditions exactly capture when an error correction code C can correct an error set E via an isometric recovery operator.

Definition 5.6.23 — Let $C \subset H$ be a subspace and $E \subset B(H, K)$ a subspace of error operators. Consider the action map $a : E \otimes C \rightarrow K$ given by $e \otimes |\psi\rangle \mapsto e|\psi\rangle$. A *recovery operator* is an isometry $v : K \rightarrow L \otimes H$ for some ancillary Hilbert space L such that

$$\mathbf{Tr}_A(v e |\psi\rangle\langle\psi| e^\dagger v^\dagger) = \mathbf{E}_{|\psi\rangle}(e^\dagger e) \cdot |\psi\rangle\langle\psi| \quad \forall |\psi\rangle \in C. \quad (5.6.24)$$

Remark 5.6.25. Observe that

$$d := \frac{1}{\mathbf{E}_{|\psi\rangle}(e^\dagger e)} v e |\psi\rangle\langle\psi| e^\dagger v^\dagger$$

is a pure density matrix whose reduced density d_C is again pure, i.e., $S_C(d) = 0$. By Propositions 5.5.2 and 5.5.15, it is necessarily the case that d is simply separable, i.e., there is a pure ancillary state $|\phi\rangle \in L$ such that

$$d = |\phi\rangle\langle\phi| \otimes |\psi\rangle\langle\psi|.$$

This happens if and only if

$$v e |\psi\rangle = \left(\gamma \sqrt{\mathbf{E}_{|\psi\rangle}(e^\dagger e)} \cdot |\phi\rangle \right) \otimes |\psi\rangle$$

where $\gamma \in U(1)$.

Lemma 5.6.26 — Suppose $x : C \rightarrow L \otimes C$ is injective and $x|\psi\rangle$ is of the form $|\phi_\psi\rangle \otimes |\psi\rangle$ for all $|\psi\rangle \in C$. Then $x = |\phi\rangle \otimes -$ for a universal $|\phi\rangle \in L$ independent of $|\psi\rangle$.

Proof. If $\dim(C) = 1$, the statement is obvious. Otherwise, choose two orthogonal states $|\psi_1\rangle, |\psi_2\rangle \in C$ and write $x|\psi_i\rangle = |\phi_i\rangle \otimes |\psi_i\rangle$ for $i = 1, 2$. For $|\psi\rangle := |\psi_1\rangle + |\psi_2\rangle$, there is a $|\phi\rangle \in LA$ such that

$$|\phi\rangle \otimes |\psi_1\rangle + |\phi\rangle \otimes |\psi_2\rangle = |\phi\rangle \otimes |\psi\rangle = x|\psi\rangle = x(\alpha|\psi_1\rangle + \beta|\psi_2\rangle) = |\phi_1\rangle \otimes |\psi_1\rangle + |\phi_2\rangle \otimes |\psi_2\rangle.$$

Applying the operators $1 \otimes \langle\psi_i|$ for $i = 1, 2$, we obtain $|\phi_1\rangle = |\phi\rangle = |\phi_2\rangle$. \square

Theorem 5.6.27 — There is a recovery operator v for (E, C) if and only if (E, C) satisfies the Knill-Laflamme conditions (KL1)-(KL3).

Proof. By Exercise 5.6.18 and Remark 5.6.25, whenever e is non-zero on C , the injective operator

$$x := ve : C \longrightarrow L \otimes C$$

satisfies the hypotheses of Lemma 5.6.26 and is therefore of the form $|\phi\rangle \otimes -$ for some fixed $|\phi\rangle \in L$. Thus for all states $|\psi_1\rangle, |\psi_2\rangle \in C$, we have

$$\langle\psi_1|e^\dagger e|\psi_1\rangle = \langle\psi_1|e^\dagger v^\dagger ve|\psi_1\rangle = \langle\phi|\phi\rangle = \langle\psi_2|e^\dagger v^\dagger ve|\psi_2\rangle = \langle\psi_2|e^\dagger e|\psi_2\rangle,$$

so (KL3) holds.

The other direction follows by Construction 5.6.22. Indeed, that construction gives a unitary

$$K \supset \text{im}(a) \xrightarrow{\hat{a}^\dagger} \hat{E} \otimes C \subset L \otimes H.$$

We may augment \hat{E} to an ancillary Hilbert space A and extend \hat{a}^\dagger to an isometry $v : K \rightarrow L \otimes H$ by arbitrarily defining v on $\text{im}(a)^\perp \subset K$ isometrically into $\hat{E}^\perp \otimes H$. Then since $ve|\psi\rangle \in \hat{E} \otimes C \subset L \otimes H$, we still have

$$\text{Tr}_L(ve|\psi\rangle\langle\psi|e^\dagger v^\dagger) = \mathbf{E}_{|\psi\rangle}(e^\dagger e) \cdot |\psi\rangle\langle\psi|$$

as desired. \square

5.7 Bell's inequalities and the GHSZ paradox

We now pause for an interlude to explore how quantum information both predicts and confirms the failure of *local realism*. The notion of *locality* means that an observable can only be affected by things nearby in spacetime. The notion of *realism* means that the universe consists of real objects which exist whether or not we observe them. The notion of *local realism* means that realism holds even at the local scale. Assuming local realism, we will derive the *CHSH inequality* [CHSH69], which is a Bell inequality [Bel64]. Quantum information theory predicts violations of the CHSH inequality, which have been experimentally observed many times [FC72]. We then discuss the *GHSZ paradox*, which again precludes a hidden variable theory, but this time without any inequalities [GHSZ90]. These results led to the Nobel Prize in 2022 of Aspect, Clauser,² and Zeilinger,³ showing local realism is simply false.

²the ‘C’ in ‘CHSH’

³the ‘Z’ in ‘GHSZ’

Example 5.7.1 (CHSH inequality) — Alice and Bob repeat the following experiment. They each control one half of a singlet state

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle).$$

Alice measure her qubit in one of two bases, corresponding to observables A_1, A_2 . Bob measures his qubit in one of two bases, corresponding to observables B_1, B_2 . The values of A_1, A_2 and B_1, B_2 will always be ± 1 .

We now impose *local realism*, which in this situation is the belief that each individual qubit has a *real hidden value* for either observable, a_1, a_2 and b_1, b_2 respectively, independent of which observables are measured. We consider the random variable

$$r := (a_1 + a_2)b_1 + (a_2 - a_1)b_2,$$

which always takes the value ± 2 , as only one of $a_1 + a_2$ and $a_1 - a_2$ is non-zero.

A single outcome in this experiment is not important, but rather the expected value over many repetitions. Since $r \in \{\pm 2\}$, we must have $\mathbf{E}[r] \in [-2, 2]$. As a_1, a_2 and b_1, b_2 are outcomes of observables A_1, A_2 and B_1, B_2 respectively, we should have the *CHSH inequality*:

$$-2 \leq \mathbf{E}[A_1 B_1] + \mathbf{E}[A_2 B_1] + \mathbf{E}[A_2 B_2] - \mathbf{E}[A_1 B_2] \leq 2. \quad (5.7.2)$$

However, setting $A_1 = X$, $A_2 = Z$, $B_1 = H := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, and $B_2 = ZHZ$ violates this inequality. Since

$$\begin{aligned} \langle \Psi^- | A_1 \otimes B_1 | \Psi^- \rangle &= -\frac{1}{\sqrt{2}} & \langle \Psi^- | A_2 \otimes B_1 | \Psi^- \rangle &= -\frac{1}{\sqrt{2}} \\ \langle \Psi^- | A_1 \otimes B_2 | \Psi^- \rangle &= \frac{1}{\sqrt{2}} & \langle \Psi^- | A_2 \otimes B_2 | \Psi^- \rangle &= -\frac{1}{\sqrt{2}}, \end{aligned}$$

setting $R = (A_1 + A_2)B_1 + (A_2 - A_1)B_2$, we have

$$\mathbf{E}_{|\Psi^-\rangle}[R] = 4 \cdot \frac{-1}{\sqrt{2}} = -2\sqrt{2} < -2,$$

contradicting (5.7.2).

The CHSH inequality above exploits *mutually unbiased bases*, which are pairs of ONBs $\{|e_i\rangle\}$ and $\{|f_j\rangle\}$ of \mathbb{C}^n satisfying

$$|\langle e_i | f_j \rangle|^2 = \frac{1}{n} \quad \forall i, j = 1, \dots, n.$$

Similarly, we say k ONBs of \mathbb{C}^n are mutually unbiased if they are pairwise mutually unbiased.

Exercise 5.7.3. How many mutually unbiased ONBs can you find in \mathbb{C}^2 ?⁴

The *GHSZ paradox* provides, perhaps, an even more unsettling failure of local realism. While the CHSH inequality shows the quantum world is *quantitatively* different than the classical world, the GHSZ inequality shows the quantum world is moreover *qualitatively* different.

Example 5.7.4 (GHSZ paradox [GHSZ90]) — Recall from Exercise 5.6.6 that the 3-qubit GHSZ state

$$|GHSZ\rangle := \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$$

spans the code subspace of the stabilizer code

$$S = \langle Z_1 Z_2, X_1 X_2 X_3, Z_2 Z_3 \rangle.$$

One way to see this is to first observe that

$$\{X_1^a Y_2^b Z_3^c |GHSZ\rangle \mid a, b, c \in \{0, 1\}\}$$

is an ONB which simultaneously diagonalizes $Z_1 Z_2$, $X_1 X_2 X_3$, and $Z_2 Z_3$, and $|GHSZ\rangle$ is the common $+1$ eigenspace for these three operators.

We now suppose that there are *local hidden variables* for X_i, Y_j, Z_k in $\{\pm 1\}$, which we label by x_i, y_j, z_k respectively for $i, j, k \in \{1, 2, 3\}$. Then when measuring a Pauli operator W in a state $|\psi\rangle$, the expectation $\langle\psi|W|\psi\rangle$ should equal the product of the hidden variables. In particular, in any state $|\psi\rangle$, we should have

$$\begin{aligned} \langle\psi|X_1 X_2 X_3|\psi\rangle &= x_1 x_2 x_3 & \langle\psi|Y_1 Y_2 X_3|\psi\rangle &= y_1 y_2 x_3 \\ \langle\psi|Y_1 X_2 Y_3|\psi\rangle &= y_1 x_2 y_3 & \langle\psi|X_1 Y_2 Y_3|\psi\rangle &= x_1 y_2 y_3 \end{aligned}$$

One now calculates the expectations of these operators in the GHSZ state:

$$\begin{aligned} \langle GHSZ|X_1 X_2 X_3|GHSZ\rangle &= 1 & \langle GHSZ|Y_1 Y_2 X_3|GHSZ\rangle &= -1 \\ \langle GHSZ|Y_1 X_2 Y_3|GHSZ\rangle &= -1 & \langle GHSZ|X_1 Y_2 Y_3|GHSZ\rangle &= -1. \end{aligned}$$

However, one quickly sees that the equations

$$x_1 x_2 x_3 = 1 \qquad y_1 y_2 x_3 = -1 \qquad y_1 x_2 y_3 = -1 \qquad x_1 y_2 y_3 = -1$$

⁴This is a famous open problem for \mathbb{C}^6 !

are inconsistent via a parity argument. Indeed, each of the variables $x_1, x_2, x_3, y_1, y_2, y_3$ appears twice on the left hand sides, so

$$-1 = 1 \cdot -1 \cdot -1 \cdot -1 = (x_1 x_2 x_3)(y_1 y_2 x_3)(y_1 x_2 y_3)(x_1 y_2 y_3) = x_1^2 x_2^2 x_3^2 y_1^2 y_2^2 y_3^2 = 1,$$

a contradiction.

Remark 5.7.5. The GHSZ state is more commonly called the *GHZ state* after the 4-qubit example from [GHZ89], but the above 3-qubit example has the additional author Shimony⁵ who should not be omitted.

5.8 Mixed state error correction and Kraus operators

In §5.6, we saw exactly when we can perform error correction for a pure state in our quantum system, and in doing so, we use an ancillary Hilbert space which is outside of the system. In general, our quantum system will be subject to errors from the environment, and so processes will appear to be noisy. Before we get further into it, we provide the following discussion.

“Hot Take” 5.8.1 — If we know all information about our quantum system, then our state is always pure and not mixed. Moreover, time evolution is always unitary. Thus if we consider the entire universe as a single quantum system, then we are in a single pure state, and we are ‘along for the ride,’ evolving unitarily in time. You were always going to read this paragraph; welcome to this moment.

So if time evolution is unitary why does the result of a quantum measurement appear to be probabilistic? We only ever have *partial information* of our quantum system, so we cannot possibly ‘know’ the entire pure state of the universe. Our best guess is the reduced density obtained by tracing out the environment which we cannot observe.

As an explicit example, suppose $|\psi\rangle = |\psi_0\rangle$ is the state of the universe at time $t = \text{now}$, H_A is the Hilbert space of the quantum information we have access to, and H_B is the rest of the Hilbert space we do not have access to, so that the Hilbert space of the universe is $H_A \otimes H_B$. As we evolve unitarily in time via a 1-parameter family of unitaries u_t , the state at time $\text{now} + t$ is given by

$$|\psi_t\rangle = u_t |\psi_0\rangle.$$

We can then trace out H_B by choosing an arbitrary ONB $\{|e_i\rangle\}$ for H_B , giving

$$d_A(t) = \text{Tr}_A(|\psi_t\rangle\langle\psi_t|) = \sum_i \underbrace{(1_A \otimes \langle e_i|) u_t}_{=: E_i(t)} |\psi_0\rangle\langle\psi_0| u_t (1_A \otimes |e_i\rangle) = \sum_i E_i(t) |\psi_0\rangle\langle\psi_0| E_i(t)^\dagger$$

We also observe that

$$\sum_i E_i(t)^\dagger E_i(t) = \sum_i u_t^\dagger (1_A \otimes |e_i\rangle\langle e_i|) u_t = u_t^\dagger (1_A \otimes 1_B) u_t = u_t^\dagger u_t = 1.$$

⁵the ‘S’ in both ‘CHSH’ and ‘GHSZ’

If we make the further (unreasonable) assumption that $|\psi_0\rangle = |\psi_0^A\rangle \otimes |\psi_0^B\rangle$ is a product state, then we may set $F_i(t) := E_i(t)(1_A \otimes |\psi_0^B\rangle)$ to see that

$$d_A(t) = \sum_i F_i(t) |\psi_0^A\rangle \langle \psi_0^A| F_i(t)^\dagger$$

is a function of $|\psi_0^A\rangle$ alone. Again, we calculate

$$\sum_i F_i(t)^\dagger F_i(t) = (1_A \otimes \langle \psi_0^B|) E_i(t)^\dagger E_i(t) (1_A \otimes |\psi_0^B\rangle) = 1_A \otimes \langle \psi_0^B| \psi_0^B \rangle = 1_A.$$

The above discussion motivates the definition of a *quantum channel* as a ‘noisy’ quantum operation. This ‘noise’ is *emergent* as our partial knowledge of our quantum system.

Definition 5.8.2 — A *quantum channel* is a completely positive trace-preserving (CPTP) map $\Phi : B(H) \rightarrow B(K)$.

Physical processes in finite quantum mechanical systems are described by quantum channels. In the context of quantum information, quantum channels usually represent intentional manipulations of quantum information as well as unintentional errors.

Example 5.8.3 (Kraus representation) — Suppose $\{E_1, \dots, E_n\} \subset B(H, K)$ such that $\sum E_i^\dagger E_i = 1_H$. Then the formula

$$\Phi(x) = \sum E_j x E_j^\dagger$$

defines a quantum channel. It is clearly completely positive as the sum of completely positive maps by Example 2.6.4, and we calculate

$$\text{Tr}(\Phi(x)) = \sum \text{Tr}_K(E_j x E_j^\dagger) = \sum \text{Tr}_H(E_j^\dagger E_j x) = \text{Tr}_H(x).$$

The set of $\{E_i\}$ is called a set of *Kraus operators* for Φ .

By the next result, all quantum channels admit Kraus representations. Such a representation is not unique, but we can interpolate isometrically between any two Kraus representations.

Lemma 5.8.4 — Every quantum channel $\Phi : B(H) \rightarrow B(K)$ admits a Kraus representation. Given any two sets of Kraus operators $\{E_1, \dots, E_m\}, \{F_1, \dots, F_n\} \subset B(H, K)$ for Φ , there is a partial isometry $V \in M_{m \times n}(B(K))$ such that $E_i = \sum_j V_{ij} F_j$ for all i and $F_j = \sum_i V_{ij}^\dagger E_i$ for all j .

Proof. Recall that Φ admits a Stinespring representation $\Phi(x) = v^\dagger \pi(x) v$ where $\pi : B(H) \rightarrow B(H \otimes \mathbb{C}^k)$ is the amplification $\pi(x) = x \otimes 1$ and $v : K \rightarrow H \otimes \mathbb{C}^k$. Since Φ is trace-preserving, for all $x \in B(H)$,

$$\text{Tr}_{B(H)}(x) = \text{Tr}_{B(K)}(\Phi(x)) = \text{Tr}_{B(K)}(v^\dagger \pi(x) v) = \text{Tr}_{B(H \otimes \mathbb{C}^k)}(v v^\dagger \pi(x)) = \text{Tr}_{B(H)}(\text{Tr}_{\mathbb{C}^k}(v v^\dagger) x),$$

which is the defining condition for the partial trace $\text{Tr}_{\mathbb{C}^k}(vv^\dagger) = 1$.

Now consider the standard ONB $\{|e_i\rangle\}$ of \mathbb{C}^k , and write $1_{\mathbb{C}^k} = \sum |e_i\rangle\langle e_i|$. Graphically, we can then write Φ as in Example 2.6.7 by

$$\begin{array}{c} \begin{array}{|c|} \hline K \\ \hline \end{array} \begin{array}{|c|} \hline \bar{K} \\ \hline \end{array} \\ \Phi \\ \begin{array}{|c|} \hline H \\ \hline \end{array} \begin{array}{|c|} \hline \bar{H} \\ \hline \end{array} \end{array} = \begin{array}{c} \begin{array}{|c|} \hline K \\ \hline \end{array} \begin{array}{|c|} \hline \bar{K} \\ \hline \end{array} \\ v^\dagger \quad \bar{v}^\dagger \\ \begin{array}{|c|} \hline \mathbb{C}^k \\ \hline \end{array} \begin{array}{|c|} \hline \bar{\mathbb{C}}^k \\ \hline \end{array} \\ \begin{array}{|c|} \hline H \\ \hline \end{array} \begin{array}{|c|} \hline \bar{H} \\ \hline \end{array} \end{array} = \sum_i \begin{array}{c} \begin{array}{|c|} \hline K \\ \hline \end{array} \begin{array}{|c|} \hline \bar{K} \\ \hline \end{array} \\ v^\dagger \quad \bar{v}^\dagger \\ \begin{array}{|c|} \hline \mathbb{C}^k \\ \hline \end{array} \begin{array}{|c|} \hline \bar{\mathbb{C}}^k \\ \hline \end{array} \\ \begin{array}{|c|} \hline H \\ \hline \end{array} \begin{array}{|c|} \hline \bar{H} \\ \hline \end{array} \end{array} =: \sum_i E_i \otimes \bar{E}_i$$

where i represents $|e_i\rangle : \mathbb{C} \rightarrow \mathbb{C}^k$ and \bar{i} represents $\langle e_i| : \mathbb{C} \rightarrow \bar{\mathbb{C}}^k$. One now verifies as in Example 2.6.7 that $\Phi(x) = \sum E_i x E_i^\dagger$ and

$$\sum E_i^\dagger E_i = \begin{array}{c} \begin{array}{|c|} \hline H \\ \hline \end{array} \begin{array}{|c|} \hline \mathbb{C}^k \\ \hline \end{array} \\ vv^\dagger \\ \begin{array}{|c|} \hline H \\ \hline \end{array} \begin{array}{|c|} \hline \mathbb{C}^k \\ \hline \end{array} \end{array} = \text{Tr}_{\mathbb{C}^k}(vv^\dagger) = 1.$$

For the final claim, the operator $V := (E_i F_j^\dagger) \in M_{m \times n}(B(K))$ is such a partial isometry. \square

Remark 5.8.5. Observe that if the Kraus operators $\{E_i\}_{i=1}^n$ and $\{F_i\}_{i=1}^n$ are obtained from choosing two distinct ONBs using the same Stinespring representation of Φ , then there is a unitary matrix $u \in M_n(\mathbb{C})$ such that $E_i = \sum_j u_{ij} F_j$ for all i .

Exercise 5.8.6. Show that a linear map $\Phi : B(H) \rightarrow B(K)$ is a quantum channel if and only if there is an ancillary Hilbert space L and an isometry $v : H \rightarrow K \otimes L$ such that $\Phi(x) = \text{Tr}_L(vxv^\dagger)$ for all $x \in B(H)$. Then show that if $w : H \rightarrow K \otimes L$ is another isometry such that $\Phi(x) = \text{Tr}_L(wxw^\dagger)$, then vw^\dagger commutes with $1 \otimes B(L)$ and thus lies in $B(K) \otimes 1$.

Recall that a code subspace is a subspace $C \subset H$. We say that a density matrix $d \in B(H)$ is *supported on* C if it is a convex combination of pure states $|\psi\rangle\langle\psi|$ where $|\psi\rangle \in C$; equivalently, in the spectral decomposition $d = \sum_{\lambda \in \text{spec}(d)} \lambda p_\lambda$, $p_\lambda H \subseteq C$ for all non-zero $\lambda \in \text{spec}(d)$.

Definition 5.8.7 — Let $\Phi : B(H) \rightarrow B(K)$ be a ‘noisy’ quantum channel, and let $C \subset H$ be a code subspace. An *error correction channel* for (Φ, C) is a quantum channel $\Psi : B(K) \rightarrow B(H)$ such that for all density matrices d supported on C , $\Psi(\Phi(d)) = d$, equivalently, $\Psi \circ \Phi = \text{id}$ when restricted to $p_C B(H) p_C \cong B(C)$. (Indeed, all operators in $B(C)$ are linear combinations of pure densities.)

Remark 5.8.8. In contrast to Definition 5.6.23 of a recovery operator, note that there is no mention of an ancillary Hilbert space in the definition of an error correction channel as the composite of a quantum channel with the partial trace Tr_A is again a quantum channel.

Example 5.8.9 — If (E, C) satisfy the Knill-Laflamme conditions (KL1)-(KL3), then we get an error correction channel by $\Psi(x) := \mathbf{Tr}_L(vxv^\dagger)$ where $v : K \rightarrow L \otimes H$ is a recovery operator as in Definition 5.6.23, which exists by Theorem 5.6.27. Indeed, given any choice of Kraus operators $E_1, \dots, E_n \in E$ defining a noisy quantum channel $\Phi(x) := \sum_i E_i x E_i^\dagger$, whenever $d = |\psi\rangle\langle\psi|$ for $|\psi\rangle \in C$,

$$\Psi(\Phi(|\psi\rangle\langle\psi|)) = \sum_i \mathbf{Tr}_L(v E_i |\psi\rangle\langle\psi| E_i^\dagger v^\dagger) \stackrel{(5.6.24)}{=} \sum_i \mathbf{E}_{|\psi\rangle}(E_i^\dagger E_i) |\psi\rangle\langle\psi| = |\psi\rangle\langle\psi|.$$

The result now follows by taking convex combinations of pure densities supported on C .

Lemma 5.8.10 — Suppose $\Theta : M_n(\mathbb{C}) \rightarrow M_n(\mathbb{C})$ is a linear map such that for every pure density $d \in M_n(\mathbb{C})$, $\Theta(d)$ is proportional to d . Then Θ is uniformly proportional to $\text{id}_{M_n(\mathbb{C})}$.

Proof. Since every operator in $M_n(\mathbb{C})$ is a linear combination of pure densities (use that every operator is a linear combination of 4 positive operators and the Spectral Theorem 1.7.9), it suffices to prove that whenever $1 = \sum p_i$ is a decomposition of 1 into minimal projections and $\lambda_i \geq 0$ such that $\Theta(p_i) = \lambda_i p_i$, then all the λ_i are equal, say to λ . Indeed, then

$$\Theta(1) = \sum \Theta(p_i) = n\lambda \sum p_i = n\lambda$$

implies that λ is independent of the choice of minimal projections p_i , and so $\Theta(p) = \lambda p$ for all minimal projections p .

Given our choice of $1 = \sum p_i$, we may extend (p_i) to a system of matrix units (e_{ij}) as in Exercise 1.4.16 by choosing an ONB $\{|e_i\rangle\}$ with $p_i = |e_i\rangle\langle e_i|$. Then for all $i \neq j$,

$$d_{ij} := \frac{1}{2}(e_{ii} + e_{ij} + e_{ji} + e_{jj}) \quad \text{and} \quad q_{ij} := \frac{1}{2}(e_{ii} - e_{ij} - e_{ji} + e_{jj})$$

are both pure densities (projections with rank 1). Let $\mu_{ij}, \nu_{ij} \geq 0$ such that $\Theta(d_{ij}) = \mu_{ij} d_{ij}$ and $\Theta(q_{ij}) = \nu_{ij} q_{ij}$. Then since $q_{ij} = -d_{ij} + p_i + p_j$, we have

$$-\nu_{ij} d_{ij} + \nu_{ij} p_i + \nu_{ij} p_j = \nu_{ij} q_{ij} = \Phi(q_{ij}) = \Phi(-d_{ij} + p_i + p_j) = -\mu_{ij} d_{ij} + \lambda_i p_i + \lambda_j p_j.$$

As d_{ij}, p_i, p_j are linearly independent, we must have $\lambda_i = \nu_{ij} = \lambda_j$ as desired. \square

The following exercise is the mixed state version of Lemma 5.6.26.

Exercise 5.8.11. Suppose $\Phi : B(C) \rightarrow B(L) \otimes B(C)$ is a completely positive map such that for every (pure) density $d \in B(C)$, $\mathbf{Tr}_L(\Phi(d))$ is proportional to d . Then $\Phi = x_L \otimes \text{id}$ for some fixed positive operator $x_L \in B(L)$. In particular, $\Phi \circ \Psi$ is proportional to $\text{id}_{B(C)}$.

Corollary 5.8.12 — Suppose (Ψ, C) is an error correction channel for $\Phi(x) = \sum_i E_i x E_i^\dagger$ where $E = \{E_1, \dots, E_n\} \subset B(H, K)$ is a set of Kraus operators. Then for each $i = 1, \dots, n$, the completely positive map $\Theta_i : B(C) \rightarrow B(H)$ given by $x \mapsto \Psi(E_i x E_i^\dagger)$ is proportional to $\text{id}_{B(C)}$.

Proof. Set $\lambda_i := \langle \psi | E_i^\dagger E_i | \psi \rangle$, and observe that $\sum_i \lambda_i = 1$. We calculate that whenever $\lambda_i \neq 0$,

$$d_i := \lambda_i^{-1} \Psi(E_i |\psi\rangle\langle\psi| E_i^\dagger) \in B(C)$$

is a density matrix such that

$$\sum \lambda_i d_i = \sum \Psi(E_i |\psi\rangle\langle\psi| E_i^\dagger) = \Psi(\Phi(|\psi\rangle\langle\psi|)) = |\psi\rangle\langle\psi|.$$

Since pure densities are extremal in the space of densities by Exercise 5.4.5, we conclude that $d_i = |\psi\rangle\langle\psi|$ whenever $\lambda_i \neq 0$. Hence for all pure densities $d \in B(C)$, we see that $\Theta_i(d)$ is proportional to d . Applying Lemma 5.8.10, we see that Θ_i is proportional to $\text{id}_{B(C)}$. \square

Theorem 5.8.13 — Given a ‘noisy’ quantum channel $\Phi(x) = \sum_i E_i x E_i^\dagger$ for a set of Kraus operators $E = \{E_1, \dots, E_n\} \in B(H, K)$, the subspace $C \subset H$ admits an error correction channel if and only if (E, C) satisfy the Knill-Laflamme conditions (KL1)-(KL3).

Proof. Suppose Ψ is an error correction channel for (Φ, C) . Then for all $i = 1, \dots, n$ and state vectors $|\psi\rangle \in C$,

$$\langle \psi | E_i^\dagger E_i | \psi \rangle = \text{Tr}_{B(C)}(\Psi(E_i |\psi\rangle\langle\psi| E_i^\dagger)) = \text{Tr}_{B(C)}(\Theta_i(|\psi\rangle\langle\psi|)) = \text{Tr}_{B(C)}(\lambda_i \cdot |\psi\rangle\langle\psi|) = \lambda_i$$

where λ_i is the scalar such that $\Theta_i = \lambda_i \cdot \text{id}_{B(C)}$ from Corollary 5.8.12. Clearly the above calculation is independent of $|\psi\rangle$, and thus (KL3) holds.

The converse direction is Example 5.8.9. \square

5.9 Quantum circuits and quantum teleportation

So far, we have tried to stay away basis-dependent proofs and explanations. However, the actual implementation of quantum protocols via quantum circuits is typically carried out in the Z -computational basis on an n -qubits system $H_n = \bigotimes^n \mathbb{C}^2$.

A *quantum circuit* is a concatenation of various unitary operators on H_n , along with measurements which turn qubits into classical bits in $\{0, 1\}$. Quantum circuits are usually represented in the graphical calculus reading *left-to-right* instead of bottom-to-top, similar to the conventions in computer science where methods are applied to objects in object oriented programming, or functions are applied to inputs in functional programming. So to think about a quantum circuit in the usual graphical calculus, just turn your head 90° to the right.

(Unfortunately, you will have to rotate the labels in the diagram as well.) Typically, each qubit is represented by a horizontal line, and the initial state of each qubit is typically an element of the Z -computational basis. More interesting states are then created via various state preparation protocols.

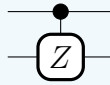
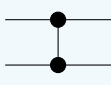
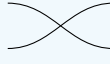
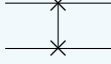
Warning 5.9.1 — Operator order in a quantum circuit is reversed with respect to the operator order of matrix multiplication. For example,

$$|0\rangle \text{---} \boxed{Z} \text{---} \boxed{X} \text{---} \quad \text{means} \quad XZ|0\rangle.$$

Unitary operators are called *quantum gates* or *unitary gates*, which can be applied to any number of qubits we like. Given a generating set of quantum gates, the *depth* of a circuit with respect to that generating set is the minimal number of simple tensors of generating gates needed to form the circuit.

Notation 5.9.2 — We provide one of the most common generating sets used for quantum circuits.

Name	Graphical notation	Operator
Identity 1	————	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
Pauli X	$\text{---} \boxed{X} \text{---}$ or $\text{---} \oplus \text{---}$	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$
Pauli Y	$\text{---} \boxed{Y} \text{---}$	$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$
Pauli Z	$\text{---} \boxed{Z} \text{---}$	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$
Hadamard H	$\text{---} \boxed{H} \text{---}$	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$
Phase S	$\text{---} \boxed{S} \text{---}$	$\begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$
Phase T	$\text{---} \boxed{T} \text{---}$	$\begin{pmatrix} 1 & 0 \\ 0 & \exp(i\pi/4) \end{pmatrix}$
CNOT or CX	$\begin{array}{c} \bullet \\ \\ \text{---} \boxed{X} \text{---} \\ \\ \oplus \end{array}$ or $\begin{array}{c} \bullet \\ \\ \oplus \\ \\ \text{---} \end{array}$	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$

CZ	 or 	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$
Swap	 or 	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

In the above table, CM stands for the *controlled M gate* for an operator $M \in M_2(\mathbb{C})$, which means that M is applied to the second qubit if and only if the first qubit is $|1\rangle$ in the Z -computational basis. That is, CM is a block diagonal matrix:

$$CM = \text{---} \begin{array}{c} \bullet \\ | \\ \boxed{M} \end{array} \text{---} = \left(\begin{array}{c|c} 1_2 & \\ \hline & M \end{array} \right) = p_{Z_1=1} \otimes 1 + p_{Z_1=-1} \otimes M.$$

In the Z -computational basis, this means

$$CM|ij\rangle = \begin{cases} |0j\rangle & \text{if } i = 0 \\ |1\rangle \otimes M|j\rangle & \text{if } i = 1. \end{cases}$$

Example 5.9.3 (Bell state preparation protocol) — We now give a state preparation protocol for the Bell basis states $|\beta_{ij}\rangle$ from Example 5.2.12. Starting with $|ij\rangle$, we get $|\beta_{ij}\rangle$ by the depth 2 quantum circuit

$$|\beta_{ij}\rangle = \begin{array}{c} |i\rangle \\ |j\rangle \end{array} \begin{array}{c} \boxed{H} \\ \oplus \end{array} \begin{array}{c} \bullet \\ | \\ \oplus \end{array}.$$

Observe that $|\beta_{ij}\rangle = (C_{ij} \otimes 1)|\beta_{00}\rangle$ where C_{ij} is the *correction factor*

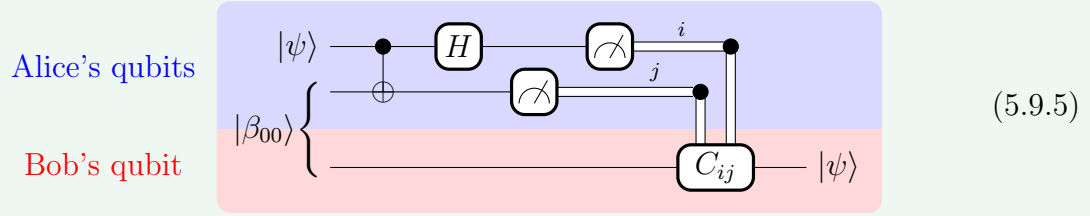
ij	C_{ij}
00	I_2
01	X
10	Z
11	ZX

Thus if we start only with the initial state $|00\rangle$, then we may prepare $|\beta_{ij}\rangle$ with the

depth ≤ 3 quantum circuit

$$|\beta_{ij}\rangle = \begin{array}{c} |0\rangle \\ |0\rangle \end{array} \begin{array}{c} \boxed{H} \\ \bullet \\ \oplus \end{array} \begin{array}{c} \boxed{C_{ij}} \\ \bullet \\ \oplus \end{array}.$$

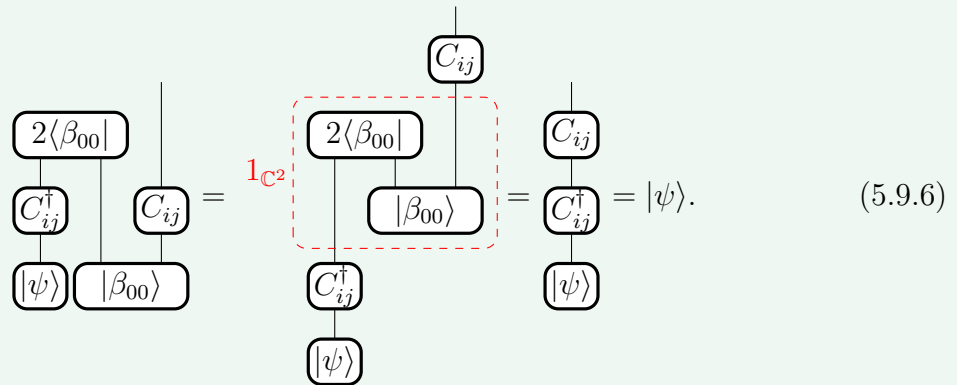
Example 5.9.4 (Quantum teleportation protocol [BBC⁺93]) — If Alice and Bob share an entangled 2-qubit state, then Alice may send a third qubit $|\psi\rangle$ directly to Bob only communicating 2 classical bits, which do not reveal the state $|\psi\rangle$.



In the protocol above, Alice and Bob share the Bell state $|\beta_{00}\rangle$, but this is not essential. Observe that Alice applies the Bell state preparation protocol in reverse, and then measures both qubits. When she measures the values ij for her qubits, Alice has effectively applied the bra/functional

$$2 \cdot \langle \beta_{ij} | = 2 \cdot \langle \beta_{00} | (C_{ij}^\dagger \otimes 1),$$

which includes the normalization factor to account for the fact that after the projective measurement, the 3-qubit state must have norm 1. (Exercise: work out this normalization factor!) This means that the above protocol can effectively be written in the graphical calculus as



One sees that dashed rectangle above is the operator $1_{\mathcal{C}^2}$ by expanding $1_{\mathcal{C}^2} = |0\rangle\langle 0| + |1\rangle\langle 1|$.

Exercise 5.9.7. Adapt the above protocol to the setting where Alice and Bob share one of the other Bell basis states.

Remark 5.9.8. Some researchers like to make a comparison between the quantum teleportation protocol and the zig-zag/snake axiom for ev and coev for \mathbb{C}^2 . This comparison necessitates identifying $\overline{\mathbb{C}}^2$ with \mathbb{C}^2 via a choice of ONB. Since quantum circuits favor working in the Z -computational basis, there is a preferred choice of basis, even if it is not canonical.

Exercise 5.9.9. Using the isomorphism $\overline{\mathbb{C}}^2 \cong \mathbb{C}^2$ via the *linear* map $\langle j| \mapsto |j\rangle$, quantify how the dashed rectangle in (5.9.6) relates to the zig-zag/snake axiom.

Exercise 5.9.10. Show that $(1_2 \otimes C_{ij})|\beta_{00}\rangle = (C_{ij}^\dagger \otimes 1_2)|\beta_{00}\rangle$. Use this fact to cancel the C_{ij} in (5.9.6) before using the zig-zag/snake axiom.

For more applications of the material in this chapter to quantum information theory, we suggest the reader look into the superdense coding protocol and quantum key distribution, especially the BB84 and E91 protocols.