MITOPENCOURSEWARE

Information and Entropy

by Professor Paul Penfield

Next, consider the application of this test to someone with an unknown family history, so that $p(A) = 0.9995$ and $p(B) = 0.0005$. Then, if the test is negative, the probability of that person carrying the defective gene $p(B \mid N)$ is

$$\frac{0.0005 \times 0.01}{0.0005 \times 0.01 + 0.9995 \times 0.98} = 0.000005105 \tag{8.9}$$

and the probability of that person carrying the normal gene $p(A \mid N)$ is

$$\frac{0.9995 \times 0.98)}{0.0005 \times 0.01 + 0.9995 \times 0.98} = 0.999994895 \tag{8.10}$$

On the other hand, if the test is positive, the probability of that person carrying the defective gene $p(B \mid P)$ is

$$\frac{0.0005 \times 0.99}{0.0005 \times 0.99 + 0.9995 \times 0.02} = 0.02416 \tag{8.11}$$

and the probability of not having the defect $p(A \mid P)$ is

$$\frac{0.9995 \times 0.02}{0.0005 \times 0.99 + 0.9995 \times 0.02} = 0.97584 \tag{8.12}$$

The test does not seem to distinguish the two possible inputs, since the overwhelming probability is that the person has a normal gene, regardless of the test results. In other words, if you get a positive test result, it is more likely to have been caused by a testing error than a defective gene. There seems to be no useful purpose served by testing people without a family history. (Of course repeated tests could be done to reduce the false positive rate.)

An information analysis shows clearly the difference between these two cases. First, recall that probabilities are subjective, or observer-dependent. The lab technician performing the test presumably does not know whether there is a family history, and so would not be able to infer anything from the results. Only someone who knows the family history could make a useful inference. Second, it is instructive to calculate the information flow in the two cases. Recall that all five information measures (I, L, M, N, and J) depend on the input probabilities. A straightforward calculation of the two cases leads to the information quantities (in bits) in Table 8.1 (note how much larger N is than M if there is no known family history).

	$p(A)$	$p(B)$	I	L	M	N	J
Family history	0.5	0.5	1.00000	0.11119	0.88881	0.11112	0.99993
Unknown Family history	0.9995	0.0005	0.00620	0.00346	0.00274	0.14141	0.14416

Table 8.1: Huntington's Disease Test Process Model Characteristics

Clearly the test conveys information about the patients' status by reducing the uncertainty in the case where there is a family history of the disease. On the other hand, without a family history there is very little information that could possibly be conveyed because there is so little initial uncertainty.

8.1.3 Berger's Burgers

A former 6.050J/2.110J student opened a fast-food restaurant, and named it in honor of the very excellent Undergraduate Assistant of the course. At Berger's Burgers, meals are prepared with state-of-the-art high-tech equipment using reversible computation for control. To reduce the creation of entropy there are no warming tables, but instead the entropy created by discarding information is used to keep the food warm. Because the rate at which information is discarded in a computation is unpredictable, the food does not always stay warm. There is a certain probability, different for the different menu items, of a meal being "COD" (cold on delivery).

The three original menu items are Value Meals 1, 2, and 3. Value Meal 1 (burger) costs $1, contains 1000 Calories, and has a probability 0.5 of arriving cold. Value Meal 2 (chicken) costs $2, has 600 Calories, and a probability 0.2 of arriving cold. Value Meal 3 (fish) costs $3, has 400 Calories, and has a 0.1 probability of being cold.

Item	Entree	Cost	Calories	Probability of arriving hot	Probability of arriving cold
Value Meal 1	Burger	$1.00	1000	0.5	0.5
Value Meal 2	Chicken	$2.00	600	0.8	0.2
Value Meal 3	Fish	$3.00	400	0.9	0.1

Table 8.2: Berger's Burgers

There are several inference questions that can be asked about Berger's Burgers. All require an initial assumption about the buying habits of the public, i.e., about the probability of each of the three meals being ordered $p(B)$, $p(C)$, and $p(F)$. Then, upon learning another fact, such as a particular customer's meal arriving cold, these probabilities can be refined to lead to a better estimate of the meal that was ordered.

Suppose you arrive at Berger's Burgers with your friends and place your orders. Assume that money is in plentiful supply so you and your friends are equally likely to order any of the three meals. Also assume that you do not happen to hear what your friends order or see how much they pay. Also assume that you do not know your friends' taste preferences and that the meals come in identical packages so you cannot tell what anyone else received by looking.

Before the meals are delivered, you have no knowledge of what your friends ordered and might assume equal probability of 1/3 for $p(B)$, $p(C)$, and $p(F)$. You can estimate the average amount paid per meal ($2.00), the average Calorie count (667 Calories), and the probability that any given order would be COD (0.267).

Now suppose your friend Alice remarks that her meal is cold. Knowing this, what is the probability she ordered a burger? (0.625) Chicken? (0.25) Fish? (0.125). And what is the expected value of the amount she paid for her meal? ($1.50) And what is her expected Calorie count? (825 Calories)

Next suppose your friend Bob says he feels sorry for her and offers her some of his meal, which is hot. Straightforward application of the formulas above can determine the refined probabilities of what he ordered, along with the expected calorie count and cost.

8.1.4 Inference Strategy

Often, it is not sufficient to calculate the probabilities of the various possible input events. The correct operation of a system may require that a definite choice be made of exactly one input event. For processes without loss, this can be done accurately. However, for processes with loss, some strategy must be used to convert probabilities to a single choice.

One simple strategy, "maximum likelihood," is to decide on whichever input event has the highest probability after the output event is known. For many applications, particularly communication with small error, this is a good strategy. It works for the symmetric binary channel when the two input probabilities are equal. However, sometimes it does not work at all. For example, if used for the Huntington's Disease test on people without a family history, this strategy would never say that the person has a defective gene, regardless of the test results.

Inference is important in many fields of interest, such as machine learning, natural language processing and other areas of artificial intelligence. An open question, of current research interest, is which inference strategies are best suited for particular purposes.

8.2 Principle of Maximum Entropy: Simple Form

In the last section, we discussed one technique of estimating the input probabilities of a process given that the output event is known. This technique, which relies on the use of Bayes' Theorem, only works if the process is lossless (in which case the input can be identified with certainty) or an initial input probability distribution is assumed (in which case it is refined to take account of the known output).

The Principle of Maximum Entropy is a technique that can be used to estimate input probabilities more generally. The result is a probability distribution that is consistent with known constraints expressed in terms of averages, or expected values, of one or more quantities, but is otherwise as unbiased as possible (the word "bias" is used here not in the technical sense of statistics, but the everyday sense of a preference that inhibits impartial judgment). This principle is described first for the simple case of one constraint and three input events, in which case the technique can be carried out analytically. Then it is described more generally in Chapter 9.

This principle has applications in many domains, but was originally motivated by statistical physics, which attempts to relate macroscopic, measurable properties of physical systems to a description at the atomic or molecular level. It can be used to approach physical systems from the point of view of information theory, because the probability distributions can be derived by avoiding the assumption that the observer has more information than is actually available. Information theory, particularly the definition of information in terms of probability distributions, provides a quantitative measure of ignorance (or uncertainty, or entropy) that can be maximized mathematically to find the probability distribution that best avoids unnecessary assumptions.

This approach to statistical physics was pioneered by Edwin T. Jaynes (1922–1998), a professor at Washington University in St. Louis, and previously Stanford University. The seminal publication was

- E. T. Jaynes, "Information Theory and Statistical Mechanics," Physical Review, vol. 106, no. 4, pp. 620-630; May 15, 1957.
  (http://bayes.wustl.edu/etj/articles/theory.1.pdf)

Other references of interest by Jaynes include:

- a continuation of this paper, E. T. Jaynes, "Information Theory and Statistical Mechanics. II," Physical Review, vol. 108, no. 2, pp. 171-190; October 15, 1957.
  (http://bayes.wustl.edu/etj/articles/theory.1.pdf)
- a review paper, including an example of estimating probabilities of an unfair die, E. T. Jaynes, "Information Theory and Statistical Mechanics," pp. 181-218 in "Statistical Physics," Brandeis Summer Institute 1962, W. A. Benjamin, Inc., New York, NY; 1963.
  (http://bayes.wustl.edu/etj/articles/brandeis.pdf)
- personal history of the approach, Edwin T. Jaynes, "Where Do We Stand on Maximum Entropy?," pp. 15-118, in "The Maximum Entropy Formalism," Raphael D. Levine and Myron Tribus, editors, The MIT Press, Cambridge, MA; 1979.
  (http://bayes.wustl.edu/etj/articles/stand.on.entropy.pdf)

The philosophy of assuming maximum uncertainty as an approach to thermodynamics is discussed in

- Chapter 3 of M. Tribus, "Thermostatics and Thermodynamics," D. Van Nostrand Co, Inc., Princeton, NJ; 1961.

Before the Principle of Maximum Entropy can be used the problem domain needs to be set up. In cases involving physical systems, this means that the various states in which the system can exist need to be identified, and all the parameters involved in the constraints known. For example, the energy, electric charge, and other quantities associated with each of the states is assumed known. Often quantum mechanics is needed for this task. It is not assumed in this step which particular state the system is in (or, as often

expressed, which state is actually "occupied"); indeed it is assumed that we do not know and cannot know this with certainty, and so we deal instead with the probability of each of the states being occupied. Thus we use probability as a means of coping with our lack of complete knowledge. Naturally we want to avoid inadvertently assuming more knowledge than we actually have, and the Principle of Maximum Entropy is the technique for doing this. In the application to nonphysical systems, the various events (possible outcomes) have to be identified along with various numerical properties associated with each of the events. In these notes we will derive a simple form of the Principle of Maximum Entropy and apply it to the restaurant example set up in Section 8.1.3.

8.2.1 Berger's Burgers

The Principle of Maximum Entropy will be introduced by means of an example. This example was described in Section 8.1.3. A fast-food restaurant, Berger's Burgers, offers three meals: burger, chicken, and fish. The price, Calorie count, and probability of each meal being delivered cold are as as listed in Table 8.2.

8.2.2 Probabilities

This example has been defined so that the choice of one of the three meals constitutes an outcome. If we do not know this outcome we may still have some knowledge, and we use probabilities to express this knowledge. The question is how to assign probabilities that are consistent with whatever information we may have.

In the case of Berger's Burgers, there are three probabilities which for simplicity we denote $p(B)$, $p(C)$, and $p(F)$ for the three meals. A probability distribution $p(A_i)$ has the property that each of the probabilities is between or equal to 0 and 1, and, since the input events are mutually exclusive and exhaustive, the sum of all the probabilities is 1:

$$\begin{aligned} 1 &= \sum_i p(A_i) \\ &= p(B) + p(C) + p(F) \end{aligned} \tag{8.13}$$

If any of the probabilities is equal to 1 then all the other probabilities are 0, and we know exactly which state the system is in; in other words, we have no uncertainty and there is no need to resort to probabilities.

8.2.3 Entropy

More generally our uncertainty is expressed quantitatively by the information which we do not have about the meal chosen, or the state occupied. This is

$$\begin{aligned} S &= \sum_i p(A_i) \log_2 \left(\frac{1}{p(A_i)} \right) \\ &= p(B) \log_2 \left(\frac{1}{p(B)} \right) + p(C) \log_2 \left(\frac{1}{p(C)} \right) + p(F) \log_2 \left(\frac{1}{p(F)} \right) \end{aligned} \tag{8.14}$$

Here, information is measured in bits because we are using logarithms to base 2.

In the context of physical systems this uncertainty is known as the entropy. In communication systems the uncertainty regarding which actual message is to be transmitted is also known as the entropy of the source. Note that in general the entropy, because it is expressed in terms of probabilities, depends on the observer. One person may have different knowledge of the system from another, and therefore would calculate a different numerical value for entropy. The Principle of Maximum Entropy is used to discover the probability distribution which leads to the highest value for this uncertainty, thereby assuring that no information is inadvertently assumed. The resulting probability distribution is not observer-dependent.

8.2.4 Constraints

It is a property of the entropy formula above that it has its maximum value when all probabilities are equal (we assume the number of possible states is finite). This property is easily proved using the Gibbs inequality. If we have no additional information about the system, then such a result seems reasonable. However, if we have additional information then we ought to be able to find a probability distribution that is better in the sense that it has less uncertainty.

For simplicity we consider only one such constraint, namely that we know the expected value of some quantity (the Principle of Maximum Entropy can handle multiple constraints but the mathematical procedures and formulas become more complicated). The quantity in question is one for which each of the states of the system has its own amount, and the expected value is found by averaging the values corresponding to each of the states, taking into account the probabilities of those states. Thus if there is an attribute for which each of the states has a value $g(A_i)$ and for which we know the actual value G, then we should consider only those probability distributions for which the expected value is equal to G

$$G = \sum_i p(A_i)g(A_i) \tag{8.15}$$

Note that this constraint cannot be achieved if G is less than the smallest $g(A_i)$ or larger than the largest $g(A_i)$.

For our Berger's Burgers example, suppose we are told that the average price of a meal is \$1.75, and we want to estimate the separate probabilities of the various meals without making any other assumptions. Then our constraint would be

$$\$1.75 = \$1.00p(B) + \$2.00p(C) + \$3.00p(F) \tag{8.16}$$

Note that the probabilities are dimensionless and so both the expected value of the constraint and the individual values must be expressed in the same units, in this case dollars.

8.2.5 Maximum Entropy, Analytic Form

Here we demonstrate the Principle of Maximum Entropy for the simple case in which there is one constraint and three variables. It will be possible to go through all the steps analytically.

Suppose you have been hired by Carnivore Corporation, the parent company of Berger's Burgers, to analyze their worldwide sales. You visit Berger's Burgers restaurants all over the world, and determine that, on average, people are paying \$1.75 for their meals. (As part of Carnivore's commitment to global homogeneity, the price of each meal is exactly the same in every restaurant, after local currencies are converted to U.S. dollars.)

After you return, your supervisors ask about the probabilities of a customer ordering each of the three value meals. In other words, they want to know $p(B)$, $p(C)$, and $p(F)$. You are horrified to realize that you did not keep the original data, and there is no time to repeat your trip. You have to make the best estimate of the probabilities $p(B)$, $p(C)$, and $p(F)$ consistent with the two things you do know:

$$1 = p(B) + p(C) + p(F) \tag{8.17}$$

$$\$1.75 = \$1.00p(B) + \$2.00p(C) + \$3.00p(F) \tag{8.18}$$

Since you have three unknowns and only two equations, there is not enough information to solve for the unknowns.

What should you do? There are a range of values of the probabilities that are consistent with what you know. However, these leave you with different amounts of uncertainty S

$$S = p(B)\log_2\left(\frac{1}{p(B)}\right) + p(C)\log_2\left(\frac{1}{p(C)}\right) + p(F)\log_2\left(\frac{1}{p(F)}\right) \tag{8.19}$$

If you choose one for which S is small, you are assuming something you do not know. For example, if your average had been \$2.00 rather than \$1.75, you could have met both of your constraints by assuming that everybody bought the chicken meal. Then your uncertainty would have been 0 bits. Or you could have assumed that half the orders were for burgers and half for fish, and the uncertainty would have been 1 bit. Neither of these assumptions seems particularly appropriate, because each goes beyond what you know. How can you find that probability distribution that uses no further assumptions beyond what you already know?

The **Principle of Maximum Entropy** is based on the reasonable assumption that you should select that probability distribution which leaves you the largest remaining uncertainty (i.e., the maximum entropy) consistent with your constraints. That way you have not introduced any additional assumptions into your calculations.

For the simple case of three probabilities and two constraints, this is easy to do analytically. Working with the two constraints, two of the unknown probabilities can be expressed in terms of the third. For our case we can multiply Equation 8.17 above by \$1.00 and subtract it from Equation 8.18, to eliminate $p(B)$. Then we can multiply the first by \$2.00 and subtract it from the second, thereby eliminating $p(C)$:

$$p(C) = 0.75 - 2p(F) \tag{8.20}$$
$$p(B) = 0.25 + p(F) \tag{8.21}$$

Next, the possible range of values of the probabilities can be determined. Since each of the three lies between 0 and 1, it is easy to conclude from these results that

$$0 \leq p(F) \leq 0.375 \tag{8.22}$$
$$0 \leq p(C) \leq 0.75 \tag{8.23}$$
$$0.25 \leq p(B) \leq 0.625 \tag{8.24}$$

Next, these expressions can be substituted into the formula for entropy so that it is expressed in terms of a single probability. Thus

$$S = (0.25 + p(F)) \log_2 \left(\frac{1}{0.25 + p(F)} \right) + (0.75 - 2p(F)) \log_2 \left(\frac{1}{0.75 - 2p(F)} \right) + p(F) \log_2 \left(\frac{1}{p(F)} \right) \tag{8.25}$$

Any of several techniques can now be used to find the value of $p(F)$ for which S is the largest. In this case the maximum occurs for $p(F) = 0.216$ and hence $p(B) = 0.466$, $p(C) = 0.318$, and $S = 1.517$ bits.

After estimating the input probability distribution, any averages over that distribution can be estimated. For example, in this case the average Calorie count can be calculated (it is 743.2 Calories), or the probability of a meal being served cold (31.8%).

8.2.6 Summary

Let's remind ourselves what we have done. We have expressed our constraints in terms of the unknown probability distributions. One of these constraints is that the sum of the probabilities is 1. The other involves the average value of some quantity, in this case cost. We used these constraints to eliminate two of the variables. We then expressed the entropy in terms of the remaining variable. Finally, we found the value of the remaining variable for which the entropy is the largest. The result is a probability distribution that is consistent with the constraints but which has the largest possible uncertainty. Thus we have not inadvertently introduced any unwanted assumptions into the probability estimation.

This technique requires that the model for the system be known at the outset; the only thing not known is the probability distribution. As carried out in this section, with a small number of unknowns and one more unknown than constraint, the derivation can be done analytically. For more complex situations a more general approach is necessary. That is the topic of Chapter 9.

Chapter 9

Principle of Maximum Entropy

Section 8.2 presented the technique of estimating input probabilities of a process that are unbiased but consistent with known constraints expressed in terms of averages, or expected values, of one or more quantities. This technique, the Principle of Maximum Entropy, was developed there for the simple case of one constraint and three input events, in which case the technique can be carried out analytically. It is described here for the more general case.

9.1 Problem Setup

Before the Principle of Maximum Entropy can be used the problem domain needs to be set up. In cases involving physical systems, this means that the various states in which the system can exist need to be identified, and all the parameters involved in the constraints known. For example, the energy, electric charge, and other quantities associated with each of the quantum states is assumed known. It is not assumed in this step which particular state the system is actually in (which state is "occupied"). Indeed it is assumed that we cannot ever know this with certainty, and so we deal instead with the probability of each of the states being occupied. In applications to nonphysical systems, the various possible events have to be enumerated and the properties of each determined, particularly the values associated with each of the constraints. In this Chapter we will apply the general mathematical derivation to two examples, one a business model, and the other a model of a physical system (both very simple and crude).

9.1.1 Berger's Burgers

This example was used in Chapter 8 to deal with inference and the analytic form of the Principle of Maximum Entropy. A fast-food restaurant offers three meals: burger, chicken, and fish. Now we suppose that the menu has been extended to include a gourmet low-fat tofu meal. The price, Calorie count, and probability of each meal being delivered cold are listed in Table 9.1.

9.1.2 Magnetic Dipole Model

An array of magnetic dipoles (think of them as tiny magnets) are subjected to an externally applied magnetic field H and therefore the energy of the system depends on their orientations and on the applied field. For simplicity our system contains only one such dipole, which from time to time is able to interchange information and energy with either of two environments, which are much larger collections of dipoles. Each

Item	Entree	Cost	Calories	Probability of arriving hot	Probability of arriving cold
Meal 1	Burger	\$1.00	1000	0.5	0.5
Meal 2	Chicken	\$2.00	600	0.8	0.2
Meal 3	Fish	\$3.00	400	0.9	0.1
Meal 4	Tofu	\$8.00	200	0.6	0.4

Table 9.1: Berger's Burgers

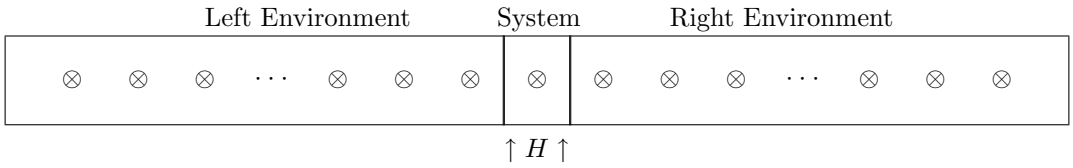

Figure 9.1: Dipole moment example.
(Each dipole can be either up or down.)

dipole, both in the system and in its two environments, can be either "up" or "down." The system has one dipole so it only has two states, corresponding to the two states for that dipole, "up" and "down" (if the system had n dipoles it would have 2^n states). The energy of each dipole is proportional to the applied field and depends on its orientation; the energy of the system is the sum of the energies of all the dipoles in the system, in our case only one such.

State	Alignment	Energy
U	up	$-m_d H$
D	down	$m_d H$

Table 9.2: Magnetic Dipole Moments

The constant m_d is expressed in Joules per Tesla, and its value depends on the physics of the particular dipole. For example, the dipoles might be electron spins, in which case $m_d = 2\mu_B\mu_0$ where $\mu_0 = 4\pi \times 10^{-7}$ henries per meter (in rationalized MKS units) is the permeability of free space, $\mu_B = \hbar e/2m_e = 9.272 \times 10^{-24}$ Joules per Tesla is the Bohr magneton, and where $\hbar = h/2\pi$, $h = 6.626 \times 10^{-34}$ Joule-seconds is Plank's constant, $e = 1.602 \times 10^{-19}$ coulombs is the magnitude of the charge of an electron, and $m_e = 9.109 \times 10^{-31}$ kilograms is the rest mass of an electron.

In Figure 9.1, the system is shown between two environments, and there are barriers between the environments and the system (represented by vertical lines) which prevent interaction (later we will remove the barriers to permit interaction). The dipoles, in both the system and the environments, are represented by the symbol $\otimes$ and may be either spin-up or spin-down. The magnetic field shown is applied to the system only, not to the environments.

The virtue of a model with only one dipole is that it is simple enough that the calculations can be carried out easily. Such a model is, of course, hopelessly simplistic and cannot be expected to lead to numerically accurate results. A more realistic model would require so many dipoles and so many states that practical computations on the collection could never be done. For example, a mole of a chemical element is a small amount by everyday standards, but it contains Avogadro's number $N_A = 6.02252 \times 10^{23}$ of atoms, and a correspondingly large number of electron spins; the number of possible states would be 2 raised to that power. Just how large this number is can be appreciated by noting that the earth contains no more than 2^{170} atoms, and the visible universe has about 2^{265} atoms; both of these numbers are way less than the number of states in that model. Even if we are less ambitious and want to compute with a much smaller sample, say

200 spins, and want to represent in our computer the probability of each state (using only 8 bits per state), we would still need more bytes of memory than there are atoms in the earth. Clearly it is impossible to compute with so many states, so the techniques described in these notes cannot be carried through in detail. Nevertheless there are certain conclusions and general relationships we will be able to establish.

9.2 Probabilities

Although the problem has been set up, we do not know which actual state the system is in. To express what we do know despite this ignorance, or uncertainty, we assume that each of the possible states A_i has some probability of occupancy $p(A_i)$ where i is an index running over the possible states. A probability distribution $p(A_i)$ has the property that each of the probabilities is between 0 and 1 (possibly being equal to either 0 or 1), and (since the input events are mutually exclusive and exhaustive) the sum of all the probabilities is 1:

$$1 = \sum_i p(A_i) \tag{9.1}$$

As has been mentioned before, two observers may, because of their different knowledge, use different probability distributions. In other words, probability, and all quantities that are based on probabilities, are subjective, or observer-dependent. The derivations below can be carried out for any observer.

9.3 Entropy

Our uncertainty is expressed quantitatively by the information which we do not have about the state occupied. This information is

$$S = \sum_i p(A_i) \log_2 \left(\frac{1}{p(A_i)} \right) \tag{9.2}$$

Information is measured in bits, as a consequence of the use of logarithms to base 2 in the Equation 9.2.

In dealing with real physical systems, with a huge number of states and therefore an entropy that is a very large number of bits, it is convenient to multiply the summation above by Boltzmann's constant $k_B = 1.381 \times 10^{-23}$ Joules per Kelvin, and also use natural logarithms rather than logarithms to base 2. Then S would be expressed in Joules per Kelvin:

$$S = k_B \sum_i p(A_i) \ln \left(\frac{1}{p(A_i)} \right) \tag{9.3}$$

In the context of both physical systems and communication systems the uncertainty is known as the entropy. Note that because the entropy is expressed in terms of probabilities, it also depends on the observer, so two people with different knowledge of the system would calculate a different numerical value for entropy.

9.4 Constraints

The entropy has its maximum value when all probabilities are equal (we assume the number of possible states is finite), and the resulting value for entropy is the logarithm of the number of states, with a possible scale factor like k_B. If we have no additional information about the system, then such a result seems reasonable. However, if we have additional information in the form of constraints then the assumption of equal probabilities would probably not be consistent with those constraints. Our objective is to find the probability distribution that has the greatest uncertainty, and hence is as unbiased as possible.

For simplicity we consider only one such constraint here. We assume that we know the expected value of some quantity (the Principle of Maximum Entropy can handle multiple constraints but the mathematical

procedures and formulas are more complicated). The quantity in question is one for which each of the states of the system has its own amount, and the expected value is found by averaging the values corresponding to each of the states, taking into account the probabilities of those states. Thus if there is a quantity G for which each of the states has a value $g(A_i)$ then we want to consider only those probability distributions for which the expected value is a known value $\widetilde{G}$

$$\widetilde{G} = \sum_i p(A_i)g(A_i) \tag{9.4}$$

Of course this constraint cannot be achieved if $\widetilde{G}$ is less than the smallest $g(A_i)$ or greater than the largest $g(A_i)$.

9.4.1 Examples

For our Berger's Burgers example, suppose we are told that the average price of a meal is \$2.50, and we want to estimate the separate probabilities of the various meals without making any other assumptions. Then our constraint would be

$$\$2.50 = \$1.00p(B) + \$2.00p(C) + \$3.00p(F) + \$8.00p(T) \tag{9.5}$$

For our magnetic-dipole example, assume the energies for states U and D are denoted $e(i)$ where i is either U or D, and assume the expected value of the energy is known to be some value $\widetilde{E}$. All these energies are expressed in Joules. Then

$$\widetilde{E} = e(U)p(U) + e(D)p(D) \tag{9.6}$$

The energies $e(U)$ and $e(D)$ depend on the externally applied magnetic field H. This parameter, which will be carried through the derivation, will end up playing an important role. If the formulas for the $e(i)$ from Table 9.2 are used here,

$$\widetilde{E} = m_d H[p(D) - p(U)] \tag{9.7}$$

9.5 Maximum Entropy, Analytic Form

The **Principle of Maximum Entropy** is based on the premise that when estimating the probability distribution, you should select that distribution which leaves you the largest remaining uncertainty (i.e., the maximum entropy) consistent with your constraints. That way you have not introduced any additional assumptions or biases into your calculations.

This principle was used in Chapter 8 for the simple case of three probabilities and one constraint. The entropy could be maximized analytically. Using the constraint and the fact that the probabilities add up to 1, we expressed two of the unknown probabilities in terms of the third.

Next, the possible range of values of the probabilities was determined using the fact that each of the three lies between 0 and 1. Then, these expressions were substituted into the formula for entropy S so that it was expressed in terms of a single probability. Then any of several techniques could be used to find the value of that probability for which S is the largest.

This analytical technique does not extend to cases with more than three possible states and only one constraint. It is only practical because the constraint can be used to express the entropy in terms of a single variable. If there are, say, four unknowns and two equations, the entropy would be left as a function of two variables, rather than one. It would be necessary to search for its maximum in a plane. Perhaps this seems feasible, but what if there were five unknowns? (Or ten?) Searching in a space of three (or eight) dimensions would be necessary, and this is much more difficult.

A different approach is developed in the next section, one well suited for a single constraint and many probabilities.

9.6 Maximum Entropy, Single Constraint

Let us assume the average value of some quantity with values $g(A_i)$ associated with the various events A_i is known; call it $\widetilde{G}$ (this is the constraint). Thus there are two equations, one of which comes from the constraint and the other from the fact that the probabilities add up to 1:

$$1 = \sum_i p(A_i) \tag{9.8}$$

$$\widetilde{G} = \sum_i p(A_i) g(A_i) \tag{9.9}$$

where $\widetilde{G}$ cannot be smaller than the smallest $g(A_i)$ or larger than the largest $g(A_i)$.

The entropy associated with this probability distribution is

$$S = \sum_i p(A_i) \log_2 \left(\frac{1}{p(A_i)} \right) \tag{9.10}$$

when expressed in bits. In the derivation below this formula for entropy will be used. It works well for examples with a small number of states. In later chapters of these notes we will start using the more common expression for entropy in physical systems, expressed in Joules per Kelvin,

$$S = k_B \sum_i p(A_i) \ln \left(\frac{1}{p(A_i)} \right) \tag{9.11}$$

9.6.1 Dual Variable

Sometimes a problem is clarified by looking at a more general problem of which the original is a special case. In this case, rather than focusing on a specific value of G, let's look at all possible values of G, which means the range between the smallest and largest values of $g(A_i)$. Thus G becomes a variable rather than a known value (the known value will continue to be denoted $\widetilde{G}$ here). Then rather than express things in terms of G as an independent variable, we will introduce a new **dual variable**, which we will call β, and express all the quantities of interest, including G, in terms of it. Then the original problem reduces to finding the value of β which corresponds to the known, desired value $\widetilde{G}$, i.e., the value of β for which $G(\beta) = \widetilde{G}$.

The new variable β is known as a **Lagrange Multiplier**, named after the French mathematician Joseph-Louis Lagrange (1736–1813)[1]. Lagrange developed a general technique, using such variables, to perform constrained maximization, of which our current problem is a very simple case. We will not use the mathematical technique of Lagrange Multipliers—it is more powerful and more complicated than we need.

Here is what we will do instead. We will start with the answer, which others have derived using Lagrange Multipliers, and prove that it is correct. That is, we will give a formula for the probability distribution $p(A_i)$ in terms of the β and the $g(A_i)$ parameters, and then prove that the entropy calculated from this distribution, $S(\beta)$ is at least as large as the entropy of any probability distribution that has the same expected value for G, namely $G(\beta)$. Therefore the use of β automatically maximizes the entropy. Then we will show how to find the value of β, and therefore indirectly all the quantities of interest, for the particular value $\widetilde{G}$ of interest (this will be possible because $G(\beta)$ is a monotonic function of β so calculating its inverse can be done with zero-finding techniques).

9.6.2 Probability Formula

The probability distribution $p(A_i)$ we want has been derived by others. It is a function of the dual variable β:

$$p(A_i) = 2^{-\alpha} 2^{-\beta g(A_i)} \tag{9.12}$$

[1] See a biography of Lagrange at http://www-groups.dcs.st-andrews.ac.uk/~history/Biographies/Lagrange.html

which implies

$$\log_2\left(\frac{1}{p(A_i)}\right) = \alpha + \beta g(A_i) \tag{9.13}$$

where α is a convenient abbreviation[2] for this function of β:

$$\alpha = \log_2\left(\sum_i 2^{-\beta g(A_i)}\right) \tag{9.14}$$

Note that this formula for α guarantees that the $p(A_i)$ from Equation 9.12 add up to 1 as required by Equation 9.8.

If β is known, the function α and the probabilities $p(A_i)$ can be found and, if desired, the entropy S and the constraint variable G. In fact, if S is needed, it can be calculated directly, without evaluating the $p(A_i)$—this is helpful if there are dozens or more probabilities to deal with. This short-cut is found by multiplying Equation 9.13 by $p(A_i)$, and summing over i. The left-hand side is S and the right-hand side simplifies because α and β are independent of i. The result is

$$S = \alpha + \beta G \tag{9.15}$$

where S, α, and G are all functions of β.

9.6.3 The Maximum Entropy

It is easy to show that the entropy calculated from this probability distribution is at least as large as that for any probability distribution which leads to the same expected value of G.

Recall the Gibbs inequality, Equation 6.4, which will be rewritten here with $p(A_i)$ and $p'(A_i)$ interchanged (it is valid either way):

$$\sum_i p'(A_i)\log_2\left(\frac{1}{p'(A_i)}\right) \le \sum_i p'(A_i)\log_2\left(\frac{1}{p(A_i)}\right) \tag{9.16}$$

where $p'(A_i)$ is any probability distribution and $p(A_i)$ is any other probability distribution. The inequality is an equality if and only if the two probability distributions are the same.

The Gibbs inequality can be used to prove that the probability distribution of Equation 9.12 has the maximum entropy. Suppose there is another probability distribution $p'(A_i)$ that leads to an expected value G' and an entropy S', i.e.,

$$1 = \sum_i p'(A_i) \tag{9.17}$$

$$G' = \sum_i p'(A_i) g(A_i) \tag{9.18}$$

$$S' = \sum_i p'(A_i)\log_2\left(\frac{1}{p'(A_i)}\right) \tag{9.19}$$

Then it is easy to show that, for any value of β, if $G' = G(\beta)$ then $S' \le S(\beta)$:

[2]The function $\alpha(\beta)$ is related to the **partition function** $Z(\beta)$ of statistical physics: $Z = 2^\alpha$ or $\alpha = \log_2 Z$.

$$\begin{aligned} S' &= \sum_i p'(A_i) \log_2 \left(\frac{1}{p'(A_i)} \right) \\ &\leq \sum_i p'(A_i) \log_2 \left(\frac{1}{p(A_i)} \right) \\ &= \sum_i p'(A_i)[\alpha + \beta g(A_i)] \\ &= \alpha + \beta G' \\ &= S(\beta) + \beta[G' - G(\beta)] \end{aligned} \tag{9.20}$$

where Equations 9.16, 9.13, 9.17, 9.18, and 9.15 were used. Thus the entropy associated with any alternative proposed probability distribution that leads to the same value for the constraint variable cannot exceed the entropy for the distribution that uses β.

9.6.4 Evaluating the Dual Variable

So far we are considering the dual variable β to be an independent variable. If we start with a known value $\widetilde{G}$, we want to use G as an independent variable and calculate β in terms of it. In other words, we need to invert the function $G(\beta)$, or find β such that Equation 9.9 is satisfied.

This task is not trivial; in fact most of the computational difficulty associated with the Principle of Maximum Entropy lies in this step. If there are a modest number of states and only one constraint in addition to the equation involving the sum of the probabilities, this step is not hard, as we will see. If there are more constraints this step becomes increasingly complicated, and if there are a large number of states the calculations cannot be done. In the case of more realistic models for physical systems, this summation is impossible to calculate, although the general relations among the quantities other than $p(A_i)$ remain valid.

To find β, start with Equation 9.12 for $p(A_i)$, multiply it by $g(A_i)$ and by 2^{α}, and sum over the probabilities. The left hand side becomes $G(\beta)2^{\alpha}$, because neither α nor $G(\beta)$ depend on i. We already have an expression for α in terms of β (Equation 9.14), so the left hand side becomes $\sum_i G(\beta)2^{-\beta g(A_i)}$. The right hand side becomes $\sum_i g(A_i)2^{-\beta g(A_i)}$. Thus,

$$0 = \sum_i [g(A_i) - G(\beta)]2^{-\beta g(A_i)} \tag{9.21}$$

If this equation is multiplied by $2^{\beta G(\beta)}$, the result is

$$0 = f(\beta) \tag{9.22}$$

where the function $f(\beta)$ is

$$f(\beta) = \sum_i [g(A_i) - G(\beta)]2^{-\beta[g(A_i) - G(\beta)]} \tag{9.23}$$

Equation 9.22 is the fundamental equation that is to be solved for particular values of $G(\beta)$, for example $\widetilde{G}$. The function $f(\beta)$ depends on the model of the problem (i.e., the various $g(A_i)$), and on $\widetilde{G}$, and that is all. It does not depend explicitly on α or the probabilities $p(A_i)$.

How do we know that there is any value of β for which $f(\beta) = 0$? First, notice that since $\widetilde{G}$ lies between the smallest and the largest $g(A_i)$, there is at least one i for which $(g(A_i) - \widetilde{G})$ is positive and at least one for which it is negative. It is not difficult to show that $f(\beta)$ is a monotonic function of β, in the sense that if $\beta_2 > \beta_1$ then $f(\beta_2) < f(\beta_1)$. For large positive values of β, the dominant term in the sum is the one that has the smallest value of $g(A_i)$, and hence f is negative. Similarly, for large negative values of β, f is positive. It must therefore be zero for one and only one value of β (this reasoning relies on the fact that $f(\beta)$ is a continuous function.)

9.6.5 Examples

For the Berger's Burgers example, suppose that you are told the average meal price is \$2.50, and you want to estimate the probabilities $p(B)$, $p(C)$, $p(F)$, and $p(T)$. Here is what you know:

$$1 = p(B) + p(C) + p(F) + p(T) \tag{9.24}$$

$$0 = \$1.00p(B) + \$2.00p(C) + \$3.00p(F) + \$8.00p(T) - \$2.50 \tag{9.25}$$

$$S = p(B)\log_2\left(\frac{1}{p(B)}\right) + p(C)\log_2\left(\frac{1}{p(C)}\right) + p(F)\log_2\left(\frac{1}{p(F)}\right) + p(T)\log_2\left(\frac{1}{p(T)}\right) \tag{9.26}$$

The entropy is the largest, subject to the constraints, if

$$p(B) = 2^{-\alpha}2^{-\beta\$1.00} \tag{9.27}$$
$$p(C) = 2^{-\alpha}2^{-\beta\$2.00} \tag{9.28}$$
$$p(F) = 2^{-\alpha}2^{-\beta\$3.00} \tag{9.29}$$
$$p(T) = 2^{-\alpha}2^{-\beta\$8.00} \tag{9.30}$$

where

$$\alpha = \log_2(2^{-\beta\$1.00} + 2^{-\beta\$2.00} + 2^{-\beta\$3.00} + 2^{-\beta\$8.00}) \tag{9.31}$$

and β is the value for which $f(\beta) = 0$ where

$$f(\beta) = \$0.50 \times 2^{-\$0.50\beta} + \$5.50 \times 2^{-\$5.50\beta} - \$1.50 \times 2^{\$1.50\beta} - \$0.50 \times 2^{\$0.50\beta} \tag{9.32}$$

A little trial and error (or use of a zero-finding program) gives $\beta = 0.2586$ bits/dollar, $\alpha = 1.2371$ bits, $p(B) = 0.3546$, $p(C) = 0.2964$, $p(F) = 0.2478$, $p(T) = 0.1011$, and $S = 1.8835$ bits. The entropy is smaller than the 2 bits which would be required to encode a single order of one of the four possible meals using a fixed-length code. This is because knowledge of the average price reduces our uncertainty somewhat. If more information is known about the orders then a probability distribution that incorporates that information would have even lower entropy.

For the **magnetic dipole example**, we carry the derivation out with the magnetic field H set at some unspecified value. The results all depend on H as well as E.

$$1 = p(U) + p(D) \tag{9.33}$$

$$\begin{aligned}\widetilde{E} &= e(U)p(U) + e(D)p(D) \\ &= m_d H[p(U) - p(D)]\end{aligned} \tag{9.34}$$

$$S = p(U)\log_2\left(\frac{1}{p(A)}\right) + p(D)\log_2\left(\frac{1}{p(D)}\right) \tag{9.35}$$

The entropy is the largest, for the energy $\widetilde{E}$ and magnetic field H, if

$$p(U) = 2^{-\alpha}2^{-\beta m_d H} \tag{9.36}$$
$$p(D) = 2^{-\alpha}2^{\beta m_d H} \tag{9.37}$$

where

$$\alpha = \log_2(2^{-\beta m_d H} + 2^{\beta m_d H}) \tag{9.38}$$

and β is the value for which $f(\beta) = 0$ where

$$f(\beta) = (m_d H - \widetilde{E})2^{-\beta(m_d H - \widetilde{E})} - (m_d H + \widetilde{E})2^{\beta(m_d H + \widetilde{E})} \tag{9.39}$$

Note that this example with only one dipole, and therefore only two states, does not actually require the Principle of Maximum Entropy because there are two equations in two unknowns, $p(U)$ and $p(D)$ (you can solve Equation 9.39 for β using algebra). If there were two dipoles, there would be four states and algebra would not have been sufficient. If there were many more than four possible states, this procedure to calculate β would have been impractical or at least very difficult. We therefore ask, in Chapter 11 of these notes, what we can tell about the various quantities even if we cannot actually calculate numerical values for them using the summation over states.

Chapter 10

Physical Systems

Until now we have ignored most aspects of physical systems by dealing only with abstract ideas such as information. Although we assumed that each bit stored or transmitted was manifested in some physical object, we focused on the abstract bits, and ignored any limitations caused by the laws of physics. This is the fundamental mantra of the information age.

It has not always been that way, and it will not be that way in the future. In past centuries, the physical manifestation of information was of great importance because of its great cost. To preserve or communicate information, books had to be written or even words cut into stone. For example, think of the process of creating a medieval manuscript during the middle ages. Pages were laboriously copied and illustrated. The results may be viewed today with great admiration for their artistry and cultural importance, in part because they were so expensive to create—society could only afford to deal with what it considered the most important information, and the cost of superb art work was not high compared with the other costs of production.

Advances over the years have improved the efficiency of information storage and transmission—think of the printing press, telegraph, telephone, radio, television, digital signal processing, semiconductors, and fiber optics. These technologies have led to sophisticated systems such as computers and data networks, and have shaped the methods used for the creation and distribution of information-intensive products by companies such as those in the entertainment business. As the cost of processing and distributing data drops, it is relevant to consider the case where that cost is small compared to the cost of creating, maintaining, and using the information. It is in this domain that the abstract ideas of information theory, bits, coding, and indeed all of computer science are dominant. All sectors of modern society are coping with the increasing amount of information that is available. Fundamental ideas of intellectual property, copyrights, patents, and trade secrets are being rethought in light of the changing economics of information processing. Welcome to the information age.

The model of information separated from its physical embodiment is, of course, an approximation of reality. Eventually, as we make microelectronic systems more and more complicated, using smaller and smaller components, we will need to face the fundamental limits imposed not by our ability to fabricate small structures, but by the laws of physics. All physical systems are governed by quantum mechanics.

Quantum mechanics is often believed to be of importance only for small structures, say the size of an atom. Although it is unavoidable at that length scale, it also governs everyday objects. When dealing with information processing in physical systems, it is pertinent to consider both very small systems with a small number of bits of information, and large systems with large amounts of information.

The key ideas we have used thus far that need to be re-interpreted in the regime where quantum mechanics is important include

- The **digital abstraction** made practical by devices that can restore data with small perturbations
- Use of **probability** to express our knowledge in the face of uncertainty
- The **Principle of Maximum Entropy** as a technique to estimate probabilities without bias

10.1 Nature of Quantum Mechanics

Quantum mechanics is weird. There seems to be no way to make it appear otherwise. Many of its predictions are not what one would expect from everyday experience with objects of the size normally encountered.

Quantum mechanics is mysterious, even to very good physicists. The underlying philosophy and interpretation of its equations and techniques are controversial.

Quantum mechanics is difficult to use. Relatively advanced mathematical skills are needed. The basic equation, although linear, is a partial differential equation that cannot be solved analytically except in a very few simple situations. Usually numerical solutions are necessary.

Quantum mechanics, like other physical theories, requires skill and judgement both in modelling and in mathematics. It is not generally taught in any depth before the graduate or advanced undergraduate level.

Quantum mechanics comes in different forms. It has many alternate formulations. Generally these are equivalent in the sense that they predict the same results of experiments, but are not equally easy to learn or to use for particular purposes.

In light of these disadvantages, why is quantum mechanics important? Because it works. It is the ONLY fundamental physical theory that works over such a wide range of situations. Its predictions have been verified experimentally time after time. It applies to everyday size objects, and to astronomical objects (although it is usually not necessary for them). It applies to atomic-size objects, to electromagnetic waves, and to sub-atomic objects. There is a version that is compatible with the theory of special relativity. About the only physical phenomenon not handled well at this time is gravity; quantum mechanics has not yet been extended to be compatible with the theory of general relativity.

In these notes we cannot cover quantum mechanics in much depth. For the purpose of examining information processing in physical systems, we only need to understand a few of the general features such systems must have. In particular, we need a model of physical systems in which there are many possible states, each with its own probability of being the one the system is actually in (i.e., the state "occupied"). These states all have physical properties associated with them, and energy is one of these. Quantum mechanics justifies this model.

We will use this model in two situations. The first (below) is one with many states, where the objective is to understand how the information associated with the occupancy of these states affects the flow of energy. The second (in a later chapter of these notes) is one with a very small number of states, where information is represented using the occupancy of these states, and the objective is to understand both the limits and opportunities in information processing afforded by quantum mechanics.

The next two sections, Section 10.2 "Introduction to Quantum Mechanics" and Section 10.3 "Stationary States," may be skipped by readers who are prepared to accept the state model without justification. They may proceed directly to Section 10.4 "Multi-State Model." Other readers may glean from the next two sections some indication of how quantum considerations lead to that model, and in the process may find some aspects of quantum mechanics less mysterious.

10.2 Introduction to Quantum Mechanics

Perhaps the first question to ask about a physical object is, "where is it?" In everyday experience, it is possible to answer that question with great precision, limited only by the quality of measurement apparatus. In the realm of very small objects, however, there are some fundamental limitations and quantum mechanics must be used to address that question.

At its heart, quantum mechanics deals with energy. Because of the equivalence of mass and energy (remember Einstein's famous formula $E = mc^2$ where c is the speed of light, 2.998×10^8 meters per second) quantum mechanics also deals with particles with mass. And because of the relationship between energy of a photon and its frequency ($E = hf$ where h is the Planck constant, 6.626×10^{-34} Joule-seconds) quantum mechanics deals with photons.

According to quantum mechanics, the question "where is it" cannot be answered with certainty. How do we deal with uncertainty? By assigning probabilities. It is a little more complicated because of the continuous nature of space, and because space is considered to be infinite in extent (at least if general relativity is ignored), but the idea is the same as for probabilities of a finite set of events. The probability density is nonnegative, and integrates over all space to 1 (this is like the sum of the probabilities of all events that are mutually exclusive and exhaustive adding up to 1).

Thus in quantum mechanics, an object is represented as a "probability blob" which evolves over time. How does it evolve? The underlying equation is not written in terms of the probability density, but rather in terms of another function of space and time from which the probability density can be found.

Consider the square root of the probability density, as a function of space and time. Then, for added generality, let the square root be either positive or negative—when you square it to get the probability density, either one will do. Next, for even more generality, allow this square root to have an arbitrary phase in the complex plane, so that it has both a real and an imaginary part. We will no longer call this the square root, but instead the "wave function" $\psi(r,t)$ which is a complex-valued function of space r and time t. The probability density is then the magnitude of the wave function squared

$$|\psi(r,t)|^2 = \psi(r,t)\psi^*(r,t) \tag{10.1}$$

where the asterisk * denotes the complex conjugate.

In dealing with probabilities earlier, we never expressed them in terms of anything more primitive. Why do we need to now? Because the fundamental equation of quantum mechanics deals with $\psi(r,t)$ rather than the probability density. Why is this? Don't ask. It's just one of many bizarre features of quantum mechanics.

The fundamental equation of quantum mechanics is the Schrödinger equation, published in 1926 by the Austrian physicist Erwin Schrödinger (1887–1961).[1]

$$i\hbar\frac{\partial\psi(r,t)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi(r,t) + V(r)\psi(r,t) \tag{10.2}$$

where i is the (imaginary) square root of -1, m is the mass of this object, $V(r)$ is the potential energy function, and $\hbar = h/2\pi = 1.054 \times 10^{-34}$ Joule-seconds. Note that this equation contains partial derivatives in both space and time. The derivative with respect to time is first order, and the spatial derivatives are second order. The Laplacian ∇^2 is defined as

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \tag{10.3}$$

where x, y, and z are the three spacial dimensions.

This equation 10.2 is frequently interpreted by multiplying both sides of it by $\psi^*(r,t)$ and integrating over space. Then the left-hand side is identified as the total energy, and the right-hand side as the sum of the kinetic and potential energies (assuming the wave function is normalized so that the space integral of $|\psi(r,t)|^2$ is 1, a property required for the interpretation in terms of a probability density). Along with this interpretation it is convenient to call $i\hbar\partial/\partial t$ the **energy operator**. It is an operator in the mathematical sense (something that operates on a function and produces as a result a function) and it has the correct dimensional units to be energy. Quantum mechanics is often formulated in terms of similar operators.

The Schrödinger equation is deceptively simple. It is a linear equation in $\psi(r,t)$ in the sense that if both ψ_1 and ψ_2 are solutions then so is any linear combination of them

$$\psi_{\text{total}} = \alpha_1\psi_1 + \alpha_2\psi_2 \tag{10.4}$$

[1] See a biography of Schrödinger at http://www-groups.dcs.st-andrews.ac.uk/~history/Biographies/Schrodinger.html

where α_1 and α_2 are complex constants (if the linear combination is to lead to a valid probability distribution then the values of α_1 and α_2 must be such that the integral over all space of $|\psi_{\text{total}}|^2$ is 1). However, except for the simplest cases of $V(r)$ the equation has not been solved in closed form.

Strictly speaking, the Schrödinger equation is really only correct if the object being described is the entire universe and $V(r) = 0$, in which case the equation is useless because it is so complicated. However, it is often used as an approximation in the case where the universe is considered in two pieces—a small one (the object) whose wave function is being calculated, and the rest of the universe (the "environment") whose influence on the object is assumed to be represented by the $V(r)$ term. Note that the object may be a single photon, a single electron, or two or more particles, i.e., it need not correspond to the everyday concept of a single particle.

An object might interact with its environment. Naturally, if the object changes its environment (as would happen if a measurement were made of some property of the object) then the environment in turn would change the object. Thus after a measurement, an object will generally have a different wave function, and some information about the object may no longer be accessible. It is a feature of quantum mechanics that this new wave function is consistent with the changes in the environment; whether this feature is a consequence of the Schrödinger equation or is a separate aspect of quantum mechanics is unclear.

10.3 Stationary States

Even though, for a given $V(r)$ term, the Schrödinger equation may be impossible to solve in closed form, much can be said about the nature of its solutions without knowing them in detail. This is done by expressing $\psi(r,t)$ as a sum of functions known as stationary states.

Stationary states are by definition solutions of the Schrödinger equation that are of a particular form, namely the product of a function of space times another function of time. It can be easily shown from the Schrödinger equation that the most general form stationary states can have is

$$\psi(r,t) = \phi(r)e^{-iEt/\hbar} \tag{10.5}$$

for some real constant E (real because otherwise $\psi(r,t)$ would grow without bound for very large or very small time), where $\phi(r)$ obeys the equation (not involving time)

$$E\phi(r) = -\frac{\hbar^2}{2m}\nabla^2\phi(r) + V(r)\phi(r) \tag{10.6}$$

and where the integral over all space of $|\phi(r)|^2$ is 1. This technique of separating the dependence of $\psi(r,t)$ on its two variables r and t is sometimes called "separation of variables."

Nonzero solutions for $\phi(r)$ cannot be obtained for all values of E. There may be some ranges in which any value of E is OK and other ranges in which only specific discrete values of E lead to nonzero wave functions. Generally speaking, solutions corresponding to discrete values of E become small far away (i.e., they "vanish at infinity") and are therefore localized in space, although their "probability blobs" may have large values at several places and might therefore be thought of as representing two or more particles.

These solutions are called "stationary states" because the magnitude of the wave function (and therefore the probability density as well) does not change in time; it is only a function of space.

For these stationary states, E has an interesting interpretation. If we multiply each side of Equation 10.6 by $\phi^*(r)$ and integrate over space, we see that (just as in the previous section) E is the sum of two terms from the right-hand side, interpreted as the kinetic and potential energies of the object. Thus E is the total energy associated with that solution.

Of course in general solutions to the Schrödinger equation with this potential $V(r)$ are not stationary states, i.e., do not have the special form of Equation 10.5. But remember that any linear combination of solutions to the Schrödinger equation is also a solution. We can use these stationary states as building blocks to generate more general solutions.

We are most interested in stationary states that are localized in space, so that the allowed values of E are discrete, although there could be many of them (perhaps even a countable infinite number). If we let j

be an index over the stationary states, then it is possible to define the resulting wave functions $\psi_j(r,t)$ so that they are both "normalized" in the sense that the space integral of the magnitude of each squared is 1 and "orthogonal" in the sense that the product of any one with the complex conjugate of another is zero when integrated over all space. We will then denote the values of E, which we have interpreted as the energy associated with that state, by e_j.

Then the general solutions to the Schrödinger equation are written as a linear combination of stationary states

$$\psi(r,t) = \sum_j a_j \phi_j(r) e^{-ie_j t/\hbar} \tag{10.7}$$

where a_j are known as expansion coefficients, and may be complex. If the wave function $\psi(r,t)$ is normalized then it is easily shown that

$$1 = \sum_j |a_j|^2 \tag{10.8}$$

and that the energy associated with the function can be written in terms of the e_j as

$$\sum_j e_j |a_j|^2 \tag{10.9}$$

From these relationships we observe that $|a_j|^2$ behaves like a probability distribution over the events consisting of the various states being occupied, and that this distribution can be used to calculate the average energy associated with the object.

The conclusion of our brief excursion into quantum mechanics is to justify the multi-state model given in the next section. Those readers who were willing to accept this model without any explanation have skipped over the past two sections and are now rejoining us.

10.4 Multi-State Model

Our model of a physical object, justified by the brief discussion of quantum mechanics in the previous two sections, is as follows. The object has a wave function ψ which, in principle, characterizes its behavior over time. This wave function may be difficult or impossible to calculate, and it can change, perhaps in unpredictable ways, when the object interacts with its environment.

The object has a finite (or perhaps countable infinite) number of "stationary states" that are easier to calculate (although for complicated objects finding them may still be impossible). Each of the stationary states has its own wave function ψ_j where j is an index over the stationary states. If the actual wave function is one of these stationary states (i.e., if this state is "occupied") then the object stays in that state indefinitely (or until it interacts with its environment). Each stationary state has its own energy e_j and possibly its own values of other physical quantities of interest.

The object's wave function can be expressed as a linear combination of the stationary states, in the form

$$\psi = \sum_j a_j \psi_j \tag{10.10}$$

where the a_j are complex numbers called expansion coefficients. If the object occupies one of the stationary states then all a_j are 0 except one of them. Without loss of generality the expansion coefficients can be defined so that the sum of their magnitudes squared is one:

$$1 = \sum_j |a_j|^2 \tag{10.11}$$

Measurement of an object's property, such as its energy, involves an interaction with the object's environment, and a change in the environment (if for no other reason than to record the answer). It is a

consequence of quantum mechanics that if the object is in one of its stationary states and its energy is measured, the result of the measurement is simply the energy of that state, and the state does not change (i.e., the expansion coefficients, all of which are 0 except one, are not changed by the measurement). On the other hand, if the object is not in one of the stationary states, then the result of the measurement is the energy of one of the stationary states, and the object immediately assumes that stationary state. Thus after each measurement the object ends up in a stationary state. Which state? The probability that state j is the one selected is $|a_j|^2$. Thus the expected value of the energy measured by an experiment is

$$\sum_j e_j |a_j|^2 \tag{10.12}$$

where e_j is the energy associated with the stationary state j. Measurement in quantum mechanics is thus not like measurement of everyday objects, where it is assumed that the energy or other physical properties can be measured with arbitrary accuracy, and that such measurements need not perturb the object. The nature of quantum measurement is one more of those aspects of quantum mechanics that must be accepted even though it may not conform to intuition developed in everyday life.

10.4.1 Energy Systems

An object that stores, transmits, or converts energy must have possible states. Such an object typically might consist of a large number (say Avogadro's number $N_A = 6.02 \times 10^{23}$) of similar or identical particles and therefore a huge number of stationary states. The Schrödinger equation cannot be solved in such circumstances. Interactions with the environment would occur often in order to transfer energy to and from the environment. It is impossible to know whether the system is in a stationary state, and even if it is known, unpredictable interactions with the environment make such knowledge irrelevant rapidly.

The most that can be done with such systems is to deal with the probabilities p_j of occupancy of the various stationary states

$$p_j = |a_j|^2 \tag{10.13}$$

The expected value of the energy E would then be

$$E = \sum_j e_j p_j \tag{10.14}$$

This model is set up in a way that is perfectly suited for the use of the Principle of Maximum Entropy to estimate the occupation probability distribution p_j. This topic will be pursued in Chapter 11 of these notes.

10.4.2 Information Systems

An object intended to perform information storage, transmission, or processing should avoid the errors that are inherent in unpredictable interactions with the environment. It would seem that the simplest such object that could process information would need two states. One bit of information could be associated with the knowledge of which state is occupied. More complex objects, with more than two states, could represent more than one bit of information.

Quantum information systems, including computers and communication systems, will be the topic of Chapter 13 of these notes.

Chapter 11

Energy

In Chapter 9 of these notes we introduced the Principle of Maximum Entropy as a technique for estimating probability distributions consistent with constraints.

A simple case that can be done analytically is that in which there are three probabilities, one constraint in the form of an average value, and the fact that the probabilities add up to one. There are, then, two equations in three unknowns, and it is straightforward to express the entropy in terms of one of the unknowns, eliminate the others, and find the maximum. This approach also works if there are four probabilities and two average-value constraints, in which case there is again one fewer equation than unknown.

Another special case is one in which there are many probabilities but only one average constraint. Although the entropy cannot be expressed in terms of a single probability, the solution in Chapter 9 is practical if the summations can be calculated.

In the application of the Principle of Maximum Entropy to physical systems, the number of possible states is usually very large, so that neither analytic nor numerical solutions are practical. Even in this case, however, the Principle of Maximum Entropy is useful because it leads to relationships among different quantities. In this chapter we look at general features of such systems.

Because we are now interested in physical systems, we will express entropy in Joules per Kelvin rather than in bits, and use the natural logarithm rather than the logarithm to the base 2.

11.1 Magnetic Dipole Model

Most of the results below apply to the general multi-state model of a physical system implied by quantum mechanics; see Chapter 10. However, an important aspect is the dependence of energy on external parameters. For example, for the magnetic dipole, the external parameter is the magnetic field H. Here is a brief review of the magnetic dipole so it can be used as an example below.

This model was introduced in section 9.1.2. Figure 11.1 shows a system with two dipoles and two environments for the system to interact with. (Of course any practical system will have many more than two dipoles, but the important ideas can be illustrated with only two.) The dipoles are subjected to an externally applied magnetic field H, and therefore the energy of the system depends on the orientations of the dipoles and on the applied field. Each dipole, both in the system and in its two environments, can be either "up" or "down," so there are four possible states for the system, "up-up," "up-down," "down-up," and "down-down." The energy of a dipole is $m_d H$ if down and $-m_d H$ if up, and the energy of each of the four states is the sum of the energies of the two dipoles.

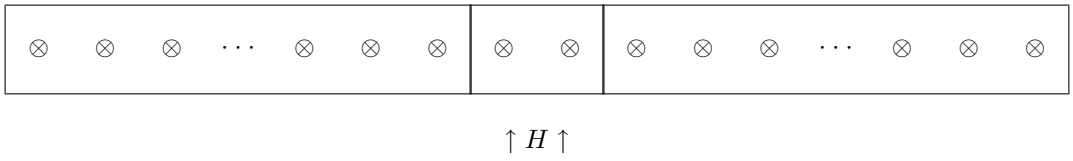

Figure 11.1: Dipole moment example. Each dipole can be either up or down

11.2 Principle of Maximum Entropy for Physical Systems

According to the multi-state model motivated by quantum mechanics (see Chapter 10 of these notes) there are a finite (or countable infinite) number of quantum states of the system. We will use i as an index over these states. The states have energy E_i, and might have other physical attributes as well. After these states are enumerated and described, the Principle of Maximum Entropy can be used, as a separate step, to estimate how likely each state is to be occupied.

We denote the occupancy of state i by the event A_i. The state i has probability $p(A_i)$ of being occupied. For simplicity we will write this probability $p(A_i)$ as p_i. We use the Principle of Maximum Entropy to estimate the probability distribution p_i consistent with the average energy E being a known (for example, measured) quantity $\widetilde{E}$. Thus

$$\widetilde{E} = \sum_i p_i E_i \tag{11.1}$$

$$1 = \sum_i p_i \tag{11.2}$$

The entropy is

$$S = k_B \sum_i p_i \ln\left(\frac{1}{p_i}\right) \tag{11.3}$$

where $k_B = 1.38 \times 10^{-23}$ Joules per Kelvin and is known as Boltzmann's constant.

The probability distribution that maximizes S subject to a constraint like Equation 11.2 was presented in Chapter 9, Equation 9.12. That formula was for the case where entropy was expressed in bits; the corresponding formula for physical systems, with entropy expressed in Joules per Kelvin, is the same except for the use of e rather than 2:

$$p_i = e^{-\alpha} e^{-\beta E_i} \tag{11.4}$$

so that

$$\ln\left(\frac{1}{p_i}\right) = \alpha + \beta E_i \tag{11.5}$$

The sum of the probabilities must be 1 and therefore

$$\alpha = \ln\left(\sum_i e^{-\beta E_i}\right) \tag{11.6}$$

As expressed in terms of the Principle of Maximum Entropy, the objective is to find the various quantities given the expected energy E. However, except in the simplest circumstances it is usually easier to do calculations the other way around. That is, it is easier to use β as an independent variable, calculate α in terms of it, and then find the p_i and then the entropy S and energy E.

11.2.1 General Properties

Because β plays a central role, it is helpful to understand intuitively how different values it may assume affect things.

First, if $\beta = 0$, all probabilities are equal. This can only happen if the number of states is finite.

Second, if $\beta > 0$, then states with lower energy have a higher probability of being occupied. Similarly, if $\beta < 0$, then states with higher energy have a higher probability of being occupied. Because of the exponential dependence on energy, unless $\mid \beta \mid$ is small, the only states with much probability of being occupied are those with energy close to the minimum possible (β positive) or maximum possible (β negative).

Third, we can multiply the equation above for $ln(1/p_i)$ by p_i and sum over i to obtain

$$S = k_B(\alpha + \beta E) \tag{11.7}$$

This equation is valid and useful even if it is not possible to find β in terms of E or to compute the many values of p_i.

Fourth, in Section 11.2.2 we will look at a small change dE in E and inquire how the other variables change. Such first-order relationships, or "differential forms," provide intuition which helps when the formulas are interpreted.

Fifth, in Section 11.2.3 we will consider the dependence of energy on an external parameter, using the magnetic dipole system with its external parameter H as an example.

The critical equations above are listed here for convenience

$$1 = \sum_i p_i \tag{11.8}$$

$$E = \sum_i p_i E_i \tag{11.9}$$

$$S = k_B \sum_i p_i \ln\left(\frac{1}{p_i}\right) \tag{11.10}$$

$$p_i = e^{-\alpha} e^{-\beta E_i} \tag{11.11}$$

$$\begin{aligned} \alpha &= \ln\left(\sum_i e^{-\beta E_i}\right) \\ &= \frac{S}{k_B} - \beta E \end{aligned} \tag{11.12}$$

11.2.2 Differential Forms

Now suppose E_i does not depend on an external parameter, and E changes by a small amount dE. We will calculate from the equations above the changes in the other quantities, keeping only first-order variations (i.e., neglecting terms like $(dE)^2$ which, for small enough dE, are insignificantly small)

$$0 = \sum_i dp_i \tag{11.13}$$

$$dE = \sum_i E_i \, dp_i \tag{11.14}$$

$$\begin{aligned} dS &= k_B \sum_i \ln\left(\frac{1}{p_i}\right) dp_i + k_B \sum_i p_i\, d\left[ln\left(\frac{1}{p_i}\right)\right] \\ &= k_B \sum_i \ln\left(\frac{1}{p_i}\right) dp_i - k_B \sum_i \left(\frac{p_i}{p_i}\right) dp_i \\ &= k_B \sum_i (\alpha + \beta E_i)\, dp_i \\ &= k_B \beta\, dE \end{aligned} \tag{11.15}$$

$$\begin{aligned} d\alpha &= \left(\frac{1}{k_B}\right) dS - \beta\, dE - E\, d\beta \\ &= -E\, d\beta \end{aligned} \tag{11.16}$$

$$\begin{aligned} dp_i &= p_i(-d\alpha - E_i\, d\beta) \\ &= -p_i(E_i - E)\, d\beta \end{aligned} \tag{11.17}$$

from which it is not difficult to show

$$dE = -\left(\sum_i p_i (E_i - E)^2\right) d\beta \tag{11.18}$$

$$dS = -k_B \beta \left(\sum_i p_i (E_i - E)^2\right) d\beta \tag{11.19}$$

These equations may be used in several ways. Note that all first-order variations are expressed as a function of $d\beta$ so it is natural to think of β as the independent variable. But this is not necessary; these equations remain valid no matter which change causes the other changes.

As an example of the insight gained from these equations, note that the formula relating dE and $d\beta$, Equation 11.18, implies that if E goes up then β goes down, and vice versa.

11.2.3 Differential Forms with External Parameters

Now we want to extend these differential forms to the case where the constraint quantities depend on external parameters. In our magnetic-dipole example, the energy of each state depends on the externally applied magnetic field H. Each E_i could be written in the form $E_i(H)$ to emphasize this dependence. Thus the constraint could be written to show this dependence explicitly:

$$E = \sum_i p_i E_i(H) \tag{11.20}$$

Then all the quantities (p_i, α, β, and S) can be thought of as depending on both E and H. In the case of our magnetic-dipole model, the energy $E_i(H)$ happens to be proportional to H with a constant of proportionality that depends on i but not on H. In other models, for other physical systems, E might depend on H or other parameters in different ways.

Consider what happens if both E and H vary slightly, by amounts dE and dH, from the values used to calculate p_i, α, β, and S. There will be small changes dp_i, $d\alpha$, $d\beta$, and dS in those quantities which can be expressed in terms of the small changes dE and dH. The changes due to dE have been calculated above. The changes due to dH enter through the change in the energies associated with each state, $dE_i(H)$ (formulas like the next few could be derived for changes caused by any external parameter, not just the magnetic field).

$$0 = \sum_i dp_i \tag{11.21}$$

$$dE = \sum_i E_i(H)\, dp_i + \sum_i p_i\, dE_i(H) \tag{11.22}$$

$$dS = k_B\beta\, dE - k_B\beta \sum_i p_i\, dE_i(H) \tag{11.23}$$

$$d\alpha = -E\, d\beta - \beta \sum_i p_i\, dE_i(H) \tag{11.24}$$

$$dp_i = -p_i(E_i(H) - E)\, d\beta - p_i\beta\, dE_i(H) + p_i\beta \sum_j p_j\, dE_j(H) \tag{11.25}$$

$$dE = -\left[\sum_i p_i(E_i(H) - E)^2\right] d\beta + \sum_i p_i(1 - \beta(E_i(H) - E))\, dE_i(H) \tag{11.26}$$

$$dS = -k_B\beta \left[\sum_i p_i(E_i(H) - E)^2\right] d\beta - k_B\beta^2 \sum_i p_i(E_i(H) - E)\, dE_i(H) \tag{11.27}$$

For the particular magnetic dipole model considered here, the terms involving $dE_i(H)$ can be simplified by noting that each state's energy $E_i(H)$ is proportional to the parameter H and therefore

$$dE_i(H) = \left(\frac{E_i(H)}{H}\right) dH \tag{11.28}$$

$$\sum_i p_i dE_i(H) = \left(\frac{E}{H}\right) dH \tag{11.29}$$

so these formulas simplify to

$$0 = \sum_i dp_i \tag{11.30}$$

$$dE = \sum_i E_i(H)\, dp_i + \left(\frac{E}{H}\right) dH \tag{11.31}$$

$$dS = k_B\beta\, dE - \left(\frac{k_B\beta E}{H}\right) dH \tag{11.32}$$

$$d\alpha = -E\, d\beta - \left(\frac{\beta E}{H}\right) dH \tag{11.33}$$

$$dp_i = -p_i(E_i(H) - E)(d\beta + \left(\frac{\beta}{H}\right) dH) \tag{11.34}$$

$$dE = -\left[\sum_i p_i(E_i(H) - E)^2\right] (d\beta + \left(\frac{\beta}{H}\right) dH) + \left(\frac{E}{H}\right) dH \tag{11.35}$$

$$dS = -k_B\beta \left[\sum_i p_i(E_i(H) - E)^2\right] (d\beta + \left(\frac{\beta}{H}\right) dH) \tag{11.36}$$

These formulas can be used to relate the trends in the variables. For example, the last formula shows that a one percent change in β produces the same change in entropy as a one percent change in H.

11.3 System and Environment

The formulas to this point apply to the system if the summations are over the states of the system, and they also apply to the system and its environment if the summations are over the larger number of states of the system and environment together. (The magnetic-dipole model of Figure 11.1 even shows a system capable of interacting with either of two environments, a feature that is needed if the system is used for energy conversion.) Next are some results for a system and its environment interacting.

11.3.1 Partition Model

Let us model the system and its environment (for the moment consider only one such environment) as parts of the universe that each have their own set of possible states, and which can be isolated from each other or can be in contact. That is, the system, considered apart from its environment, has states which, at least in principle, can be described. Each has an energy associated with it, and perhaps other physical properties as well. This description is separate from the determination of which state is actually occupied—that determination is made using the Principle of Maximum Entropy.

We also assume that the environment has its own set of states, each with its own energy and possibly other physical properties. Again this description of the states is independent of which states are actually occupied.

Our model for the interaction between these two (or what is equivalent, our model for the way the total combination is partitioned into the system and the environment) is that the combination has states each of which consists of one state from the environment and one from the system. Thus, for example, if the system has four states (as our simple two-dipole model does) and the environment has 1000 states, then the combination would have 4000 states. Each state of the combination corresponds to exactly one state of the system and exactly one state of the environment.

We need a notation to keep things straight. We will use the index i for the system and the index j for the environment. Then we can denote the states of the total combination using both i and j, in the form i,j just like the notation for joint probability (which is exactly what it is). A sum over the states of the total combination is then a sum over both i and j.

We will assume that it is possible for the system and the environment to be isolated from one another (the dipole drawing shows a vertical bar which is supposed to represent a barrier to interaction) and then, at other times, for the two to be interacting. Whether they are isolated or interacting does not affect the states or the physical properties associated with the states, although it may affect the probability of occupancy of the states.

11.3.2 Interaction Model

The reason for our partition model is that we want to control interaction between the system and its environment. Different physical systems would have different modes of interaction, and different mechanisms for isolating different parts. Here is described a simple model for interaction of magnetic dipoles that are aligned in a row. It is offered as an example.

Suppose that the apparatus that holds the magnetic dipoles allows adjacent dipoles to influence each other. This influence might be to cause one dipole to change from up to down or vice versa. Naturally, if one dipole influences its neighbor, then its neighbor at the same time influences it. It is reasonable to suppose that if one dipole changes its status from, say, up to down, then the neighbor that is interacting with it should change its status in the opposite direction. The effect is that the two dipoles exchange their orientations. The total number of dipoles oriented in each direction stays fixed.

Consider two adjacent dipoles that exchange their orientations—the one on the left ends up with the orientation that the one on the right started with, and vice versa. There are only a few different cases.

First, if the two dipoles started with the same orientation, nothing would change. On the other hand, if the two dipoles started with different orientations, the effect would be that the pattern of orientations has changed–the upward orientation has moved to the left or the right. This has happened even though the

dipoles themselves have not moved. Since the energy associated with the two possible alignments is different, there has been a small change in the location of the energy, even though the total energy is unchanged.

Second, if both dipoles are in the system, or both are in the environment, then energy may have shifted position within the system or the environment, but has not moved between them.

Third, if the two dipoles started with different alignment, and they are located one on each side of the boundary between the system and the environment, then energy has flowed from the system to the environment or vice versa. This has happened not because the dipoles have moved, but because the orientations have moved.

Energy that is transferred to or from the system as a result of interactions of this sort is referred to as **heat**. A formula for heat in terms of changes of probability distribution is given below.

Sometimes this kind of a process is referred to as "mixing" because the effect is similar to that of different kinds of particles being mixed together. However, in this analogy the dipoles do not move; it is their pattern of orientations or their microscopic energies that have moved and mixed.

Let us assume that we can, by placing or removing appropriate barriers, either inhibit or permit this process. For example, the process might be inhibited by simply moving the system away from its environment physically. Energy conversion devices generally use sequences where mixing is encouraged or discouraged at different times.

11.3.3 Extensive and Intensive Quantities

This partition model leads to an important property that physical quantities can have. Some physical quantities will be called "extensive" and others "intensive."

Whether the system is isolated from the environment or is interacting with it, and whatever the probability distributions $p_{s,i}$ of the system, $p_{e,j}$ of the environment, and $p_{t,i,j}$ of the combination, the energies of the system state and of the environment state add up to form the energy of the corresponding total state (subscripts s, e, and t mean system, environment, and total):

$$E_{t,i,j} = E_{s,i} + E_{e,j} \tag{11.37}$$

The probability of occupancy of total state k is the product of the two probabilities of the two associated states i and j:

$$p_{t,i,j} = p_{s,i} p_{e,j} \tag{11.38}$$

With this background it is easy to show that the expected value of the total energy is the sum of the expected values of the system and environment energies:

$$\begin{aligned}
E_t &= \sum_{i,j} E_{t,i,j} p_{t,i,j} \\
&= \sum_{i,j} \left[E_{s,i} + E_{e,j} \right] p_{s,i} p_{e,j} \\
&= \sum_{i} \sum_{j} \left[E_{s,i} + E_{e,j} \right] p_{s,i} p_{e,j} \\
&= \sum_{i} p_{s,i} \sum_{j} E_{e,j} p_{e,j} + \sum_{j} p_{e,j} \sum_{i} E_{s,i} p_{s,i} \\
&= \sum_{j} E_{e,j} p_{e,j} + \sum_{i} E_{s,i} p_{s,i} \\
&= E_e + E_s
\end{aligned} \tag{11.39}$$

This result holds whether the system and environment are isolated or interacting. It states that the energy of the system and the energy of the environment add up to make the total energy. It is a consequence

of the fact that the energy associated with each total state is the sum of the energies associated with the corresponding system and environment states.

A quantity with the property that its total value is the sum of the values for the two (or more) parts is known as an **extensive** quantity. Energy has that property, as was just demonstrated. Entropy is also extensive. That is,

$$\begin{aligned} S_t &= \sum_{i,j} p_{t,i,j} \ln\left(\frac{1}{p_{t,i,j}}\right) \\ &= \sum_{i,j} p_{s,i}p_{e,j}\left[\ln\left(\frac{1}{p_{s,i}}\right) + \ln\left(\frac{1}{p_{e,j}}\right)\right] \\ &= \sum_{i}\sum_{j} p_{s,i}p_{e,j}\left[\ln\left(\frac{1}{p_{s,i}}\right) + \ln\left(\frac{1}{p_{e,j}}\right)\right] \\ &= \sum_{i} p_{s,i}\sum_{j} p_{e,j}\ln\left(\frac{1}{p_{e,j}}\right) + \sum_{j} p_{e,j}\sum_{i} p_{s,i}\ln\left(\frac{1}{p_{s,i}}\right) \\ &= \sum_{j} p_{e,j}\ln\left(\frac{1}{p_{e,j}}\right) + \sum_{i} p_{s,i}\ln\left(\frac{1}{p_{s,i}}\right) \\ &= S_e + S_s \end{aligned} \tag{11.40}$$

Again this result holds whether or not the system and environment are isolated or interacting.

Not all quantities of interest are extensive. In particular, α and β are not. Consider β. This is an example of a quantity for which the values associated with the system, the environment, and the total configuration may or may not be related. If the system and environment are isolated, so that a separate application of the Principle of Maximum Entropy is made to each, then there is no reason why β_s and β_e would be related. On the other hand, if the system and environment are interacting so that they are exchanging energy, the distribution of energy between the system and the environment may not be known and therefore the Principle of Maximum Entropy can be applied only to the combination, not to the system and environment separately. Then, the same value of β would apply throughout.

Quantities like β that are the same throughout a system when analyzed as a whole are called **intensive**.

11.3.4 Equilibrium

The partition model leads to interesting results when the system and its environment are allowed to come into contact after having been isolated. In thermodynamics this process is known as the total configuration coming into equilibrium.

Let us suppose that the system and the environment have been in isolation and therefore are characterized by different, unrelated, values of energy, entropy, and other quantities. Then suppose they are allowed to interact, using a model of interaction in which the total energy is unchanged. Energy may flow from the system to the environment or vice versa because of mixing, and this flow of energy is called heat. As a result, the probabilities of occupancy will change, although the descriptions of the states and their properties, including their energies, do not change.

We have developed general formulas that relate small changes in probabilities, and in E, S, α, and β which can be used now. If the energy of the system is assumed to change somewhat (because of mixing), that fact could be incorporated into a new application of the Principle of Maximum Entropy to the system that would result in modified probabilities, E, S, α, and β. In particular, we saw earlier that the signs of dE and $d\beta$ are opposite, so that if E goes up, β then goes down, and vice versa.

Soon, the transfer of energy between the system and the environment may result in our not knowing the energy of each separately, but only the total energy (which does not change as a result of the mixing). In that case, it would be appropriate to use the Principle of Maximum Entropy on the total combination of

system and environment considered together. When that is done, there will be a new single value of β and a new total entropy. What can be said about these values?

First, the new entropy is the sum of the new entropy of the system and the new entropy of the environment, because entropy is an extensive quantity. Furthermore, the old total entropy (at the time the interaction started) is the sum of the old system entropy and the old environment entropy, for the same reason. However, what is interesting is the new total entropy compared with the old total entropy.

The new entropy, because it is evaluated with the probability distribution that comes from the Principle of Maximum Entropy, is the largest value consistent with the total energy. Any other probability distribution consistent with the same total energy would lead to a smaller (or possibly equal) entropy. One such probability distribution is the distribution prior to the mixing, the one that led to the old entropy value. Therefore the total entropy has increased (or at best stayed constant) as a result of the interaction between the system and the environment. It may be that the entropy of the system alone has gone down, but if so then the entropy of the environment must have gone up at least as much.

The energies of the system and the environment have changed, and as a result the values of β_s and β_e have changed, in opposite directions. Their new values are the same (each is equal to β_t), and therefore this new value lies between the two old values.

11.3.5 Energy Flow, Work and Heat

Let us return to the magnetic dipole model as shown in Figure 11.1.

In this section we will consider interactions with only one of the two environments. In Chapter 12 we will consider use of both environments, which will allow the machine to be used as a heat engine or refrigerator.

Consider first the case that the system is isolated from its environment, as shown in Figure 11.1 (the vertical bars represent barriers to interaction). The system is in some state, and we do not necessarily know which one, although the probability distribution p_i can be obtained from the Principle of Maximum Entropy. A change in state generally requires a nonzero amount of energy, because the different states have different energies. We can always imagine a small enough change dH in H so that the magnetic field cannot supply or absorb the necessary energy to change state. Then we can imagine a succession of such changes in H, none of which can change state, but when taken together constitute a large enough change in H to be noticeable. We conclude that changing H for an isolated system does not by itself change the state. Thus the probability distribution p_i is unchanged. Of course changing H by an amount dH does change the energy through the resulting change in $E_i(H)$:

$$dE = \sum_i p_i \, dE_i(H) \tag{11.41}$$

This change is reversible: if the field is changed back, the energy could be recovered in electrical or magnetic or mechanical form (there is nowhere else for it to go in this model). Energy flow of this sort, that can be recovered in electrical, magnetic, or mechanical form (or some other forms) is referred to as **work**. If $dE > 0$ then we say that work is positive, in that it was done by the external source on the system; if $dE < 0$ then we say that work is negative, in that it was done on the external source by the system. Naturally, in energy-conversion devices it is important to know whether the work is positive or negative. In many cases simply running the machine backwards changes the sign of the work; this is not always true of the other form of energy transfer, discussed below.

Changes to a system caused by a change in one or more of its parameters, when it cannot interact with its environment, are known as **adiabatic** changes. Since the probability distribution is not changed by them, they produce no change in entropy of the system. This is a general principle: adiabatic changes do not change the probability distribution and therefore conserve entropy.

First-order changes to the quantities of interest were given above in the general case where E and the various E_i are changed. If the change is adiabatic, then dE is caused only by the changes dE_i and the general equations simplify to

$$dp_i = 0 \tag{11.42}$$
$$dE = \sum_i p_i \, dE_i(H) \tag{11.43}$$
$$dS = 0 \tag{11.44}$$
$$d\alpha = -E \, d\beta - \beta \sum_i p_i \, dE_i(H) \tag{11.45}$$
$$0 = \left[\sum_i p_i (E_i(H) - E)^2 \right] d\beta + \beta \sum_i p_i (E_i(H) - E) \, dE_i(H) \tag{11.46}$$

If, as in our magnetic-dipole model, the energies of the states are proportional to H then these adiabatic formulas simplify further to

$$dp_i = 0 \tag{11.47}$$
$$dE = \left(\frac{E}{H} \right) dH \tag{11.48}$$
$$dS = 0 \tag{11.49}$$
$$d\alpha = 0 \tag{11.50}$$
$$d\beta = -\left(\frac{\beta}{H} \right) dH \tag{11.51}$$

Next, consider the system no longer isolated, but instead interacting with its environment. The interaction model permits heat to flow between the system and the environment, and by convention we will say the heat is positive if energy flows into the system from the environment, and negative if the energy flows the other way. Energy can be transferred by heat and work at the same time. Work is represented by changes in the energy of the individual states dE_i, and heat by changes in the probabilities p_i. Thus the formula for dE above becomes

$$dE = \sum_i E_i(H) \, dp_i + \sum_i p_i \, dE_i(H) \tag{11.52}$$

where the first term is heat and the second term is work.

11.3.6 Reversible Energy Flow

We saw in section 11.3.4 that when a system is allowed to interact with its environment, total entropy generally increases. In this case it is not possible to restore the system and the environment to their prior states by further mixing, because such a restoration would require a lower total entropy. Thus mixing in general is irreversible.

The limiting case where the total entropy stays constant is one where, if the system has changed, it can be restored to its prior state. It is easy to derive the conditions under which such changes are, in this sense, reversible.

From the formulas given earlier, specifically Equation 11.23, the change in system entropy is proportional to the part of the change in energy due to heat. Thus

$$dS_s = k_B\beta_s\, dE_s - k_B\beta_s \sum_i p_{s,i}\, dE_{s,i}(H) \tag{11.53}$$

$$= k_B\beta_s \left[dE_s - \sum_i p_{s,i}\, dE_{s,i}(H) \right] \tag{11.54}$$

$$= k_B\beta_s\, dq_s \tag{11.55}$$

where dq_s stands for the heat that comes into the system due to the interaction mechanism.

This formula applies to the system and a similar formula applies to the environment:

$$dS_e = k_B\beta_e\, dq_e \tag{11.56}$$

The two heats are the same except for sign

$$dq_s = -dq_e \tag{11.57}$$

and it therefore follows that the total entropy $S_s + S_e$ is unchanged (i.e., $dS_s = -dS_e$) if and only if the two values of β for the system and environment are the same:

$$\beta_s = \beta_e \tag{11.58}$$

Reversible changes (with no change in total entropy) can involve work and heat and therefore changes in energy and entropy for the system, but the system and the environment must have the same value of β. Otherwise, the changes are irreversible. Also, we noted in Section 11.3.4 that interactions between the system and the environment result in a new value of β intermediate between the two starting values of β_s and β_e, so reversible changes result in no change to β.

The first-order change formulas given earlier can be written to account for reversible interactions with the environment by simply setting $d\beta = 0$

$$0 = \sum_i dp_i \tag{11.59}$$

$$dE = \sum_i E_i(H)\, dp_i + \sum_i p_i\, dE_i(H) \tag{11.60}$$

$$dS = k_B\beta\, dE - k_B\beta \sum_i p_i\, dE_i(H) \tag{11.61}$$

$$d\alpha = -\beta \sum_i p_i\, dE_i(H) \tag{11.62}$$

$$dp_i = -p_i\beta\, dE_i(H) + p_i\beta \sum_j p_j\, dE_j(H) \tag{11.63}$$

$$dE = \sum_i p_i(1 - \beta(E_i(H) - E))\, dE_i(H) \tag{11.64}$$

$$dS = -k_B\beta^2 \sum_i p_i(E_i(H) - E)\, dE_i(H) \tag{11.65}$$

As before, these formulas can be further simplified in the case where the energies of the individual states is proportional to H

$$0 = \sum_i dp_i \tag{11.66}$$

$$dE = \sum_i E_i(H)\, dp_i + \left(\frac{E}{H}\right) dH \tag{11.67}$$

$$dS = k_B\beta\, dE - k_B\beta \left(\frac{E}{H}\right) dH \tag{11.68}$$

$$d\alpha = -\left(\frac{\beta E}{H}\right) dH \tag{11.69}$$

$$dp_i = -\left(\frac{p_i\beta}{H}\right) (E_i(H) - E)\, dH \tag{11.70}$$

$$dE = \left(\frac{E}{H}\right) dH - \left(\frac{\beta}{H}\right) \left[\sum_i p_i (E_i(H) - E)^2\right] dH \tag{11.71}$$

$$dS = -\left(\frac{k_B\beta^2}{H}\right) \left[\sum_i p_i (E_i(H) - E)^2\right] dH \tag{11.72}$$

These formulas will be used in the next chapter of these notes to derive constraints on the efficiency of energy conversion machines that involve heat.

Chapter 12

Temperature

In previous chapters of these notes we introduced the Principle of Maximum Entropy as a technique for estimating probability distributions consistent with constraints.

In Chapter 8 we discussed the simple case that can be done analytically, in which there are three probabilities, one constraint in the form of an average value, and the fact that the probabilities add up to one. There are, then, two equations and three unknowns, and it is straightforward to express the entropy in terms of one of the unknowns, eliminating the others, and find the maximum. This approach also works if there are four probabilities and two average-value constraints, in which case there is again one fewer equation than unknown.

In Chapter 9 we discussed a general case in which there are many probabilities but only one average constraint, so that the entropy cannot be expressed in terms of a single probability. The result previously derived using the method of Lagrange multipliers was given.

In Chapter 11 we looked at the implications of the Principle of Maximum Entropy for physical systems that adhere to the multi-state model motivated by quantum mechanics, as outlined in Chapter 10.

We found that the dual variable β plays a central role. Its value indicates whether states with high or low energy are occupied (or have a higher probability of being occupied). From it all the other quantities, including the expected value of energy and the entropy, can be calculated.

In this chapter, we will interpret β further, and will define its reciprocal as (to within a scale factor) the temperature of the material. Then we will see that there are constraints on the efficiency of energy conversion that can be expressed naturally in terms of temperature.

12.1 Temperature Scales

A heat engine is a machine that extracts heat from the environment and produces work, typically in mechanical or electrical form. As we will see, for a heat engine to function there need to be two different environments available. The formulas below place restrictions on the efficiency of energy conversion, in terms of the different values of β of the two environments. We will derive these restrictions.

First, however, it is useful to start to deal with the reciprocal of β rather than β itself. Recall that β is an intensive property: if two systems with different values of β are brought into contact, they will end up with a common value of β, somewhere between the original two values, and the overall entropy will rise. The same is true of $1/\beta$, and indeed of any constant times $1/\beta$. (Actually this statement is not true if one of the two values of β is positive and the other is negative; in this case the resulting value of β is intermediate but

the resulting value of $1/\beta$ is not.) Note that $1/\beta$ can, by using the formulas in Chapter 11, be interpreted as a small change in energy divided by the change in entropy that causes it, to within the scale factor k_B.

Let us define the "absolute temperature" as

$$T = \frac{1}{k_B \beta} \tag{12.1}$$

where $k_B = 1.381 \times 10^{-23}$ Joules per Kelvin is Boltzmann's constant. The probability distribution that comes from the use of the Principle of Maximum Entropy is, when written in terms of T,

$$p_i = e^{-\alpha} e^{-\beta E_i} \tag{12.2}$$

$$= e^{-\alpha} e^{-E_i/k_B T} \tag{12.3}$$

The interpretation of β in terms of temperature is consistent with the everyday properties of temperature, namely that two bodies at the same temperature do not exchange heat, and if two bodies at different temperatures come into contact one heats up and the other cools down so that their temperatures approach each other. In ordinary experience absolute temperature is positive, and the corresponding value of β is also. Because temperature is a more familiar concept than dual variables or Lagrange multipliers, from now on we will express our results in terms of temperature.

Absolute temperature T is measured in Kelvins (sometimes incorrectly called degrees Kelvin), in honor of William Thomson (1824–1907), who proposed an absolute temperature scale in 1848.[1] The Celsius scale, which is commonly used by the general public in most countries of the world, differs from the Kelvin scale by an additive constant, and the Fahrenheit scale, which is in common use in the United States, differs by both an additive constant and a multiplicative factor. Finally, to complete the roster of scales, William Rankine (1820–1872) proposed a scale which had 0 the same as the Kelvin scale, but the size of the degrees was the same as in the Fahrenheit scale.

More than one temperature scale is needed because temperature is used for both scientific purposes (for which the Kelvin scale is well suited) and everyday experience. Naturally, the early scales were designed for use by the general public. Gabriel Fahrenheit (1686–1736) wanted a scale where the hottest and coldest weather in Europe would lie between 0 and 100. He realized that most people can deal most easily with numbers in that range. In 1742 Anders Celsius (1701–1744) decided that temperatures between 0 and 100 should cover the range where water is a liquid. In his initial Centigrade Scale, he represented the boiling point of water as 0 degrees and the freezing point as 100 degrees. Two years later it was suggested that these points be reversed.[2] The result, named after Celsius in 1948, is now used throughout the world.

For general interest, Table 12.1 shows a few temperatures of interest on the four scales, along with β.

12.2 Heat Engine

The magnetic-dipole system we are considering is shown in Figure 12.1, where there are **two environments** at different temperatures, and the interaction of each with the system can be controlled by having the barriers either present or not (shown in the Figure as present). Although Figure 12.1 shows two dipoles in the system, the analysis here works with only one dipole, or with more than two, so long as there are many fewer dipoles in the system than in either environment.

Now let us rewrite the formulas from Chapter 11 with the use of β replaced by temperature. Thus Equations 11.8 to 11.12 become

[1] Thomson was a prolific scientist/engineer at Glasgow University in Scotland, with major contributions to electromagnetism, thermodynamics, and their industrial applications. He invented the name "Maxwell's Demon." In 1892 he was created Baron Kelvin of Largs for his work on the transatlantic cable. Kelvin is the name of the river that flows through the University.

[2] According to some accounts the suggestion was made by Carolus Linnaeus (1707–1778), a colleague on the faculty of Uppsala University and a protege of Celsius' uncle. Linnaeus is best known as the inventor of the scientific notation for plants and animals that is used to this day by botanists and zoologists.

	K	°C	°F	°R	$k_BT = \frac{1}{\beta}$ (J)	β (J^{-1})
Absolute Zero	0	-273.15	-459.67	0	0	∞
Outer Space (approx)	2.7	-270	-455	4.9	3.73×10^{-23}	2.68×10^{22}
Liquid Helium **bp**	4.22	-268.93	-452.07	7.6	5.83×10^{-23}	1.72×10^{22}
Liquid Nitrogen **bp**	77.34	-195.81	-320.46	139.2	1.07×10^{-21}	9.36×10^{20}
Water **mp**	273.15	0.00	32.00	491.67	3.73×10^{-21}	2.65×10^{20}
Room Temperature (approx)	290	17	62	520	4.00×10^{-21}	2.50×10^{20}
Water **bp**	373.15	100.00	212.00	671.67	5.15×10^{-21}	1.94×10^{20}

Table 12.1: Various temperatures of interest
(**bp** = boiling point, **mp** = melting point)

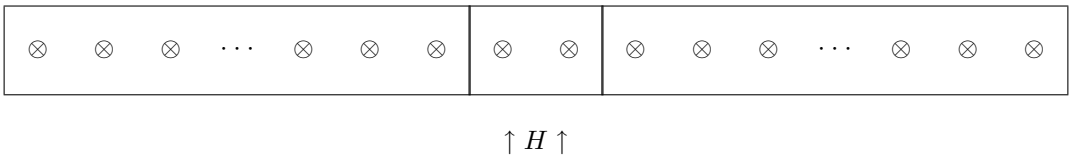

Figure 12.1: Dipole moment example.
(Each dipole can be either up or down.)

$$1 = \sum_i p_i \tag{12.4}$$

$$E = \sum_i p_i E_i \tag{12.5}$$

$$S = k_B \sum_i p_i \ln\left(\frac{1}{p_i}\right) \tag{12.6}$$

$$p_i = e^{-\alpha} e^{-E_i/k_BT} \tag{12.7}$$

$$\begin{aligned} \alpha &= \ln\left(\sum_i e^{-E_i/k_BT}\right) \\ &= \frac{S}{k_B} - \frac{E}{k_BT} \end{aligned} \tag{12.8}$$

The differential formulas from Chapter 11 for the case of the dipole model where each state has an energy proportional to H, Equations 11.30 to 11.36 become

$$0 = \sum_i dp_i \tag{12.9}$$

$$dE = \sum_i E_i(H)\, dp_i + \left(\frac{E}{H}\right) dH \tag{12.10}$$

$$TdS = dE - \left(\frac{E}{H}\right) dH \tag{12.11}$$

$$d\alpha = \left(\frac{E}{k_B T}\right)\left[\left(\frac{1}{T}\right) dT - \left(\frac{1}{H}\right) dH\right] \tag{12.12}$$

$$dp_i = p_i \left[\frac{E_i(H) - E}{k_B T}\right]\left[\left(\frac{1}{T}\right) dT - \left(\frac{1}{H}\right) dH\right] \tag{12.13}$$

$$dE = \left[\sum_i p_i (E_i(H) - E)^2\right]\left(\frac{1}{k_B T}\right)\left[\left(\frac{1}{T}\right) dT - \left(\frac{1}{H}\right) dH\right] + \left(\frac{E}{H}\right) dH \tag{12.14}$$

$$TdS = \left[\sum_i p_i (E_i(H) - E)^2\right]\left(\frac{1}{k_B T}\right)\left[\left(\frac{1}{T}\right) dT - \left(\frac{1}{H}\right) dH\right] \tag{12.15}$$

and the change in energy can be attributed to the effects of work dw and heat dq

$$dw = \left(\frac{E}{H}\right) dH \tag{12.16}$$

$$\begin{aligned} dq &= \sum_i E_i(H)\, dp_i \\ &= TdS \end{aligned} \tag{12.17}$$

12.3 Energy-Conversion Cycle

This system can act as a heat engine if the interaction of the system with its environments, and the externally applied magnetic field, are both controlled appropriately. The idea is to make the system change in a way to be described, so that is goes through a succession of states and returns to the starting state. This represents one cycle, which can then be repeated many times. During one cycle heat is exchanged with the two environments, and work is exchanged between the system and the agent controlling the magnetic field. If the system, over a single cycle, gets more energy in the form of heat from the environments than it gives back to them, then energy must have been delivered to the agent controlling the magnetic field in the form of work.

The cycle of the heat engine is shown below in Figure 12.2. Without loss of generality we can treat the case where H is positive. Assume that the left environment has a temperature T_1 which is positive but less (i.e., a higher value of β) than the temperature T_2 for the right environment (the two temperatures must be different for the device to work). This cycle is shown on the plane formed by axes corresponding to S and T of the system, and forms a rectangle, with corners marked a, b, c, and d, and sides corresponding to the values S_1, S_2, T_1, and T_2.

Since the temperatures are assumed to be positive, the lower energy levels have a higher probability of being occupied. Therefore, the way we have defined the energies here, the energy E is negative. Thus as the field gets stronger, the energy gets more negative, which means that energy actually gets delivered from the system to the magnetic apparatus. Think of the magnetic field as increasing because a large permanent magnet is physically moved toward the system. The magnetic dipoles in the system exert a force of attraction on that magnet so as to draw it toward the system, and this force on the magnet as it is moved could be

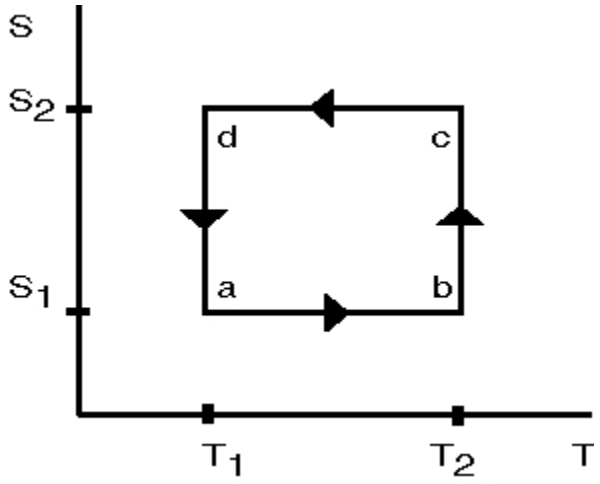

Figure 12.2: Temperature Cycle

used to stretch a spring or raise a weight against gravity, thereby storing this energy. Energy that moves into the system (or out of the system) of a form like this, that can come from (or be added to) an external source of energy is work (or negative work).

First consider the bottom leg of this cycle, during which the temperature of the system is increased from T_1 to T_2 without change in entropy. An operation without change in entropy is called **adiabatic**. By Equation 12.15 above, increasing T is accomplished by increasing H, while not permitting the system to interact with either of its two environments. (In other words, the barriers preventing the dipoles in the system from interacting with those in either of the two environments are in place.) The energy of the system goes down (to a more negative value) during this leg, so energy is being given to the external apparatus that produces the magnetic field, and the work done on the system is negative.

Next, consider the right-hand leg of this cycle, during which the entropy is increased from S_1 to S_2 at constant temperature T_2. This step, at constant temperature, is called **isothermal**. According to Equation 12.15, this is accomplished by decreasing H, while the system is in contact with the right environment, which is assumed to be at temperature T_2. (In other words, the barrier on the left in Figure 12.1 is left in place but that on the right is withdrawn.) During this leg the change in energy E arises from heat, flowing in from the high-temperature environment, and work from the external magnetic apparatus. The heat is $T_2(S_2 - S_1)$ and the work is positive since the decreasing H during this leg drives the energy toward 0.

The next two legs are similar to the first two except the work and heat are opposite in direction, i.e., the heat is negative because energy flows from the system to the low-temperature environment. During the top leg the system is isolated from both environments, so the action is adiabatic. During the left-hand isothermal leg the system interacts with the low-temperature environment.

After going around this cycle, the system is back where it started in terms of its energy, magnetic field, and entropy. The two environments are slightly changed but we assume that they are each so much larger than the system in terms of the number of dipoles present that they have not changed much. The net change is a slight loss of entropy for the high-temperature environment and a gain of an equal amount of entropy for the low-temperature environment. Because these are at different temperatures, the energy that is transferred when the heat flow happens is different—it is proportional to the temperature and therefore more energy leaves the high-temperature environment than goes into the low-temperature environment. The difference is a net negative work which shows up as energy at the magnetic apparatus. Thus heat from two environments is converted to work. The amount converted is nonzero only if the two environments are at different temperatures.

Table 12.2 summarizes the heat engine cycle.

Leg	Start	End	Type	dS	dT	H	E	Heat in	Work in
bottom	a	b	adiabatic	0	positive	increases	decreases	0	negative
right	b	c	isothermal	positive	0	decreases	increases	positive	positive
top	c	d	adiabatic	0	negative	decreases	increases	0	positive
left	d	a	isothermal	negative	0	increases	decreases	negative	negative
Total	a	a	complete cycle	0	0	no change	no change	positive	negative

Table 12.2: Energy cycle

For each cycle the energy lost by the high-temperature environment is $T_2(S_2 - S_1)$ and the energy gained by the low-temperature environment is $T_1(S_2 - S_1)$ and so the net energy converted is the difference $(T_2 - T_1)(S_2 - S_1)$. It would be desirable for a heat engine to convert as much of the heat lost by the high-temperature environment as possible to work. The machine here has efficiency

$$\left(\frac{\text{work out}}{\text{high-temperature heat in}}\right) = \frac{T_2 - T_1}{T_2} \tag{12.18}$$

This ratio is known as the Carnot efficiency, named after the French physicist Sadi Nicolas Léonard Carnot (1796 - 1832).[3] He was the first to recognize that heat engines could not have perfect efficiency, and that the efficiency limit (which was subsequently named after him) applies to all types of reversible heat engines.

The operations described above are reversible, i.e., the entire cycle can be run backwards, with the result that heat is pumped from the low-temperature environment to the one at high temperature. This action does not occur naturally, and indeed a similar analysis shows that work must be delivered by the magnetic apparatus to the magnetic dipoles for this to happen, so that more heat gets put into the high-temperature environment than is lost by the low-temperature environment. Heat engines run in this reverse fashion act as refrigerators or heat pumps.

[3]For a biography check out http://www-groups.dcs.st-andrews.ac.uk/~history/Mathematicians/Carnot_Sadi.html

Chapter 13

Quantum Information

In Chapter 10 of these notes the multi-state model for quantum systems was presented. This model was then applied to systems intended for energy conversion in Chapter 11 and Chapter 12. Now it is applied to systems intended for information processing.

The science and technology of quantum information is relatively new. The concept of the quantum bit (named the qubit) was first presented, in the form needed here, in 1995. There are still many unanswered questions about quantum information (for example the quantum version of the channel capacity theorem is not known precisely). As a result, the field is in a state of flux. There are gaps in our knowledge.

13.1 Quantum Information Storage

We have used the bit as the mathematical model of the simplest classical system that can store information. Similarly, we need a model for the simplest quantum system that can store information. It is called the "qubit." At its simplest, a qubit can be thought of as a small physical object with two states, which can be placed in one of those states and which can subsequently be accessed by a measurement instrument that will reveal that state. However, quantum mechanics both restricts the types of interactions that can be used to move information to or from the system, and permits additional modes of information storage and processing that have no classical counterparts.

An example of a qubit is the magnetic dipole which was used in Chapters 9, 11, and 12 of these notes. Other examples of potential technological importance are quantum dots (three-dimensional wells for trapping electrons) and photons (particles of light with various polarizations).

Qubits are difficult to deal with physically. That's why quantum computers are not yet available. While it may not be hard to create qubits, it is often hard to measure them, and usually very hard to keep them from interacting with the rest of the universe and thereby changing their state unpredictably.

Suppose our system is a single magnetic dipole. The dipole can be either "up" or "down," and these states have different energies. The fact that the system consists of only a single dipole makes the system fragile.

The reason that classical bits are not as fragile is that they use more physical material. For example, a semiconductor memory may represent a bit by the presence or absence of a thousand electrons. If one is missing, the rest are still present and a measurement can still work. In other words, there is massive redundancy in the mechanism that stores the data. Redundancy is effective in correcting errors. For a similar reason, it is possible to read a classical bit without changing its state, and it is possible for one bit to control the input of two or more gates (in other words, the bit can be copied).

However, there are at least three reasons why we may want to store bits without such massive redundancy. First, it would be more efficient. More bits could be stored or processed in a structure of the same size or cost. The semiconductor industry is making rapid progress in this direction, and before 2015 it should be possible to make memory cells and gates that use so few atoms that statistical fluctuations in the number of data-storing particles will be a problem. Second, sensitive information stored without redundancy could not be copied without altering it, so it would be possible to protect the information securely, or at least know if its security had been compromised. And third, the properties of quantum mechanics could permit modes of computing and communications that cannot be done classically.

A model for reading and writing the quantum bit is needed. Our model for writing (sometimes called "preparing" the bit) is that a "probe" with known state (either "up" or "down") is brought into contact with the single dipole of the system. The system and the probe then exchange their states. The system ends up with the probe's previous value, and the probe ends up with the system's previous value. If the previous system state was known, then the state of the probe after writing is known and the probe can be used again. If not, then the probe cannot be reused because of uncertainty about its state. Thus writing to a system that has unknown data increases the uncertainty about the environment. The general principle here is that discarding unknown data increases entropy.

The model for reading the quantum bit is not as simple. We assume that the measuring instrument interacts with the bit in some way to determine its state. This interaction forces the system into one of its stationary states, and the state of the instrument changes in a way determined by which state the system ends up in. If the system was already in one of the stationary states, then that one is the one selected. If, more generally, the system wave function is a linear combination of stationary states, then one of those states is selected, with probability given by the square of the magnitude of the expansion coefficient.

We now present three models of quantum bits, with increasingly complicated behavior.

13.2 Model 1: Tiny Classical Bits

The simplest model of a quantum bit is one which we will consider only briefly. It is not general enough to accommodate the most interesting properties of quantum information.

This model is like the magnetic dipole model, where only two states (up and down) are possible. Every measurement restores the system to one of its two values, so small errors do not accumulate. Since measurements can be made without changing the system, it is possible to copy a bit. This model of the quantum bit behaves essentially like a classical bit except that its size may be very small and it may be able to be measured rapidly.

This model has proven useful for energy conversion systems. It was used in Chapter 12 of these notes.

13.3 Model 2: Superposition of States (the Qubit)

The second model makes use of the fact that the states in quantum mechanics can be expressed in terms of wave functions which obey the Schrödinger equation. Since the Schrödinger equation is linear, any linear combination of wave functions that obey it also obeys it. Thus, if we associate the logical value 0 with the wave function ψ_0 and the logical value 1 with the wave function ψ_1 then any linear combination of the form

$$\psi = \alpha\psi_0 + \beta\psi_1 \tag{13.1}$$

where α and β are complex constants with $|\alpha|^2 + |\beta|^2 = 1$, is a valid wave function for the system. Then the probability that a measurement returns the value 0 is $|\alpha|^2$ and the probability that a measurement returns the value 1 is $|\beta|^2$. When a measurement is made, the values of α and β change so that one of them is 1 and the other is 0, consistent with what the measurement returns.

It might seem that a qubit defined in this way could carry a lot of information because both α and β can take on many possible values. However, the fact that a measurement will return only 0 or 1 along with the

fact that these coefficients are destroyed by a measurement, means that only one bit of information can be read from a single qubit, no matter how much care was exerted in originally specifying α and β precisely.

13.4 Model 3: Multiple Qubits with Entanglement

Consider a quantum mechanical system with four states, rather than two. Let us suppose that it is possible to make two different measurements on the system, each of which returns either 0 or 1. It is natural to denote the stationary states with two subscripts, one corresponding to the first measurement and the other to the second. Thus the general wave function is of the form

$$\psi = \alpha_{00}\psi_{00} + \alpha_{01}\psi_{01} + \alpha_{10}\psi_{10} + \alpha_{11}\psi_{11} \tag{13.2}$$

where the complex coefficients obey the normalization condition

$$1 = \mid \alpha_{00} \mid^2 + \mid \alpha_{01} \mid^2 + \mid \alpha_{10} \mid^2 + \mid \alpha_{11} \mid^2 \tag{13.3}$$

You may think of this model as two qubits, one corresponding to each of the two measurements. These qubits are not independent, but rather are **entangled** in some way. Then it is natural to ask what happens if one of them is measured. A measurement of, for example, the first qubit will return 0 with probability $\mid \alpha_{00} \mid^2 + \mid \alpha_{01} \mid^2$ and if it does the wave function collapses to only those stationary states that are consistent with this measured value,

$$\psi = \frac{\alpha_{00}\psi_{00} + \alpha_{01}\psi_{01}}{\sqrt{|\alpha_{00}|^2 + |\alpha_{01}|^2}} \tag{13.4}$$

(note that the resulting wave function was "re-normalized" by dividing by $\sqrt{|\alpha_{00}|^2 + |\alpha_{01}|^2}$).

There is no need for this system to be physically located in one place. In fact, one of the most interesting examples involves two qubits which are entangled in this way but where the first measurement is done in one location and the second in another. A simple case is one in which there are only two of the four possible stationary states initially, so $\alpha_{01} = 0$ and $\alpha_{10} = 0$. This system has the remarkable property that as a result of one measurement the wave function is collapsed to one of the two possible stationary states and the result of this collapse can be detected by the other measurement, possibly at a remote location.

It is possible to define several interesting logic gates which act on multiple qubits. These have the property that they are reversible; this is a general property of quantum-mechanical systems.

Among the interesting applications of multiple qubits are

- Computing some algorithms (including factoring integers) faster than classical computers
- Teleportation (of the information needed to reconstruct a quantum state)
- Cryptographic systems
- Backwards information transfer (not possible classically)
- Superdense coding (two classical bits in one qubit if another qubit was sent earlier)

These applications are described in several books and papers, including these three:

- T. P. Spiller, "Quantum Information Processing: Cryptography, Computation, and Teleportation," Proc. IEEE, vol. 84, no. 12, pp. 1719–1746; December, 1996. Although this article is now several years old, it is still an excellent introduction.
- Michael A. Nielsen and Isaac L. Chuang, "Quantum Computation and Quantum Information," Cambridge University Press, Cambridge, UK; 2000

- Hoi-Kwong Lo, Sandu Popescu, and Tim Spiller, "Introduction to Quantum Computation and Information," World Scientific, Singapore; 1998. The book is based on a lecture series held at Hewlett-Packard Laboratories, Bristol, UK, November 1996–April, 1997

13.5 Detail: Qubit and Applications

Sections 13.5 to 13.11 are based on notes written by Luis Pérez-Breva May 4, 2005.

The previous sections have introduced the basic features of quantum information in terms of the wave function. We first introduced the wave function in the context of physical systems in Chapter 10. The wave function is a controversial object, that has given rise to different schools of thought about the physical interpretation of quantum mechanics. Nevertheless, it is extremely useful to derive probabilities for the location of a particle at a given time, and it allowed us to introduce the multi-state model in Chapter 10 as a direct consequence of the linearity of Schrödinger equation.

In the first chapters of these notes, we introduced the bit as a binary (two-state) quantity to study classical information science. In quantum information science we are also interested in two-state systems. However, unlike the classical bit, the quantum mechanical bit may also be in a state of superposition. For example we could be interested in superpositions of the first two energy levels of the infinite potential well. The type of mathematics required for addressing the dynamics of two-state systems, possibly including superpositions, is linear algebra (that we reviewed in Chapter 2 in the context of the discrete cosine transformation.)

To emphasize that we are interested in the dynamics of two-state systems, and not in the dynamics of each state, it is best to abstract the wave function and introduce a new notation, the bracket notation. In the following sections, as we introduce the bracket notation, we appreciate the first important differences with the classical domain: the no-cloning theorem and entanglement. Then we give an overview of the applications of quantum mechanics to communication (teleportation and cryptography), algorithms (Grover fast search, Deutsch-Josza) and information science (error correcting codes).

13.6 Bracket Notation for Qubits

The bracket notation was introduced by P. Dirac for quantum mechanics. In the context of these notes, bracket notation will give us a new way to represent old friends like column and row vectors, dot products, matrices, and linear transformations. Nevertheless, bracket notation is more general than that; it can be used to fully describe wave functions with continuous variables such as position or momentum.[1]

13.6.1 Kets, Bras, Brackets, and Operators

Kets, bras, brackets and operators are the building bricks of bracket notation, which is the most commonly used notation for quantum mechanical systems. They can be thought of as column vectors, row vectors, dot products and matrices respectively.

$|\, Ket\rangle$

A ket is just a column vector composed of complex numbers. It is represented as:

$$|\, k\rangle = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} = \vec{k}\,. \tag{13.5}$$

The symbol k inside the ket $|\, k\rangle$ is the label by which we identify this vector. The two kets $|\, 0\rangle$ and $|\, 1\rangle$ are used to represent the two logical states of qubits, and have a standard vector representation

$$|\, 0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad |\, 1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{13.6}$$

[1] Readers interested in an extremely detailed (and advanced) exposition of bracket notation may find the first chapters of the following book useful: *"Quantum Mechanics volume I" by Cohen-Tannoudji, Bernard Diu and Frank Laloe, Wiley-Interscience (1996).*

Recall from Equation 13.1 that the superposition of two quantum states ψ_0 andψ_1 is

$$\psi = \alpha\psi_0 + \beta\psi_1 \tag{13.7}$$

where α and β are complex numbers. In bracket notation, this superposition of $\mid 0\rangle$ and $\mid 1\rangle$ can be written as

$$\begin{aligned} \mid \psi\rangle &= \alpha \mid 0\rangle + \beta \mid 1\rangle && (13.8)\\ &= \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \end{pmatrix} && (13.9)\\ &= \begin{pmatrix} \alpha \\ \beta \end{pmatrix} && (13.10) \end{aligned}$$

$\langle Bra \mid$

Bras are the Hermitian conjugates of kets. That is, for a given ket, the corresponding bra is a row vector (the transpose of a ket), where the elements have been complex conjugated. For example, the qubit from 13.10 has a corresponding bra $\langle\psi \mid$ that results from taking the Hermitian conjugate of equation 13.10

$$\begin{aligned} (\mid \psi\rangle)^\dagger &= (\alpha \mid 0\rangle + \beta \mid 1\rangle)^\dagger && (13.11)\\ &= \alpha^* (\mid 0\rangle)^\dagger + \beta^* (\mid 1\rangle)^\dagger && (13.12)\\ &= \alpha^* \begin{pmatrix} 1 \\ 0 \end{pmatrix}^\dagger + \beta^* \begin{pmatrix} 0 \\ 1 \end{pmatrix}^\dagger && (13.13)\\ &= \alpha^* \begin{pmatrix} 1 & 0 \end{pmatrix} + \beta^* \begin{pmatrix} 0 & 1 \end{pmatrix} && (13.14)\\ &= \begin{pmatrix} \alpha^* & \beta^* \end{pmatrix} && (13.15) \end{aligned}$$

The symbol † is used to represent the operation of hermitian conjugation of a vector or a matrix.[2] The star (*) is the conventional notation for the conjugate of a complex number: $(a + i\,b)^* = a - i\,b$ if a and b are real numbers.

$\langle Bra \mid Ket\rangle$

The dot product is the product of a bra (row vector) $\langle q \mid$, by a ket (column vector) $\mid k\rangle$, it is called bracket and denoted $\langle q \mid k\rangle$, and is just what you would expect from linear algebra

$$\langle q \mid k\rangle = \begin{pmatrix} q_1^* & q_2^* \end{pmatrix} \times \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} = \sum_j q_j^* k_j. \tag{13.16}$$

Note that the result of $\langle q \mid k\rangle$ is a complex number.

Brackets allow us to introduce a very important property of kets. Kets are always assumed to be normalized, which means that the dot product of a ket by itself is equal to 1. This implies that at least one of the elements in the column vector of the ket must be nonzero. For example, the dot product of an arbitrary qubit ($\mid \psi\rangle$) by itself, $\langle\psi \mid \psi\rangle = 1$, so

$$\begin{aligned} \langle\psi \mid \psi\rangle &= (\alpha^*\langle 0 \mid + \beta^*\langle 1 \mid) \cdot (\alpha \mid 0\rangle + \beta \mid 1\rangle)\\ &= \alpha^*\alpha\langle 0 \mid 0\rangle + \beta^*\alpha\langle 1 \mid 0\rangle\alpha^*\beta\langle 0 \mid 1\rangle + \beta^*\beta\langle 1 \mid 1\rangle\\ &= \alpha^*\alpha + \beta^*\beta\\ &= |\alpha|^2 + |\beta|^2 = 1 \end{aligned} \tag{13.17}$$

[2] This operation is known by several different names, including "complex transpose" and "adjoint."

This is precisely the result we postulated when we introduced the qubit as a superposition of wave functions. In Chapter 10, we saw that the product of a wavefunction by its complex conjugate is a probability distribution and must integrate to one. This requirement is completely analogous to requiring that a ket be normalized.[3]

The dot product can be used to compute the probability of a qubit of being in either one of the possible states $\mid 0\rangle$ and $\mid 1\rangle$. For example, if we wish to compute the probability that the outcome of a measurement on the qubit $\mid \psi\rangle$ is state $\mid 0\rangle$, we just take the dot product of $\mid 0\rangle$ and $\mid \psi\rangle$ and square the result

$$\begin{aligned}\Pr\left(\mid 0\rangle\right) &= \left|\langle 0 \mid \psi\rangle\right|^2 \\ &= \left|\alpha\langle 0 \mid 0\rangle + \beta\langle 0 \mid 1\rangle\right|^2 \\ &= \left|\alpha_1 + \beta_0\right|^2 \\ &= \left|\alpha\right|^2 \end{aligned} \tag{13.18}$$

Operators

Operators are objects that transform one ket $\mid k\rangle$ into another ket $\mid q\rangle$. Operators are represented with hats: $\widehat{O}$. It follows from our definition of ket that an operator is just a matrix,

$$\begin{aligned}\widehat{O} \mid k\rangle &= \begin{pmatrix} o_{11} & o_{12} \\ o_{21} & o_{22} \end{pmatrix} \times \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} \\ &= \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} \\ &= \mid q\rangle \end{aligned} \tag{13.19}$$

Operators act on bras in a similar manner

$$\langle k \mid \widehat{O}^\dagger = \langle q \mid \tag{13.20}$$

Equations 13.19 and 13.20 and the requirement that kets be normalized allow us to derive an important property of operators. Multiplying both equations we obtain

$$\langle k \mid \widehat{O}^\dagger \widehat{O} \mid k\rangle = \langle q \mid q\rangle \tag{13.21}$$

$$\langle k \mid \widehat{O}^\dagger \widehat{O} \mid k\rangle = 1, \tag{13.22}$$

the second line follows from assuming that $\widehat{O}$ preserves the normalization of the ket, and since $\langle k \mid k\rangle = 1$, it implies that $\widehat{O}^\dagger \widehat{O} = \mathbb{I}$. Operators that have this property are said to be unitary, and their inverse is equal to their adjoint. All quantum mechanical operators must be unitary, or else, the normalization of the probability distribution would not be preserved by the transformations of the ket. Note that this is the exact same reasoning we employed to require that time evolution be unitary back in Chapter 10. From the physical standpoint unitarity means that doing and then undoing the operation defined by $\widehat{O}$ should leave us with the same we had in origin (Note the similarity with the definition of reversibility).

There is an easy way to construct an operator if we know the input and output kets. We can use the exterior product, that is, the product of a column vector by a row vector (the dot product is often also called inner or interior product, hence the name of exterior product). We can construct the operator $\widehat{O}$ using the exterior product of a ket by a bra

$$\widehat{O} \mid k\rangle = (\mid q\rangle\langle k \mid) \mid k\rangle = \mid q\rangle\langle k \mid k\rangle = \mid q\rangle \tag{13.23}$$

[3] It may not be immediately obvious that $\int \Psi^* \Psi dx$ is a dot product. To see that it is, discretize the integral $\int \Psi^* \Psi dx \rightarrow \sum_i \Psi_i^* \Psi_i$ and compare to the definition of dot product. You may argue that in doing so we have transformed a function Ψ into a vector with elements Ψ_i; but we defined a ket as a vector to relate it to linear algebra. If the ket were to represent a function, then the appropriate definition of the dot product would be $\langle \Phi \mid \Psi\rangle = \int \Phi^* \Psi dx$.

note that this would not be possible if the kets were not normalized to 1; another way to put it is that the normalization of the ket inforces the fact that operators built in this way are unitary.

For example, to transform a qubit in the state $\mid 0\rangle$ into the qubit in the state $\mid \psi\rangle$ defined above, we construct the operator

$$\widehat{O}_2 = \alpha \mid 0\rangle\langle 0 \mid + \beta \mid 1\rangle\langle 0 \mid \tag{13.24}$$

we can verify that this operator produces the expected result

$$\begin{aligned} \widehat{O}_2 \mid 0\rangle &= \alpha \mid 0\rangle\langle 0 \mid 0\rangle + \beta \mid 1\rangle\langle 0 \mid 0\rangle \\ &= \alpha \mid 0\rangle 1 + \beta \mid 1\rangle 1 \\ &= \alpha \mid 0\rangle + \beta \mid 1\rangle \\ &= \mid \psi\rangle \end{aligned} \tag{13.25}$$

we have just performed our first quantum computation!

In quantum mechanics books it is custommary to drop the hat from the operators ($\widehat{O} \to O$) to "simplify notation." Often at an introductory level (and an advanced level as well), this simplification causes confusion between operators and scalars; in these notes we will try to avoid doing so.

13.6.2 Tensor Product—Composite Systems

The notation we have introduced so far deals with single qubit systems. It is nonetheless desirable to have a notation that allows us to describe composite systems, that is systems of multiple qubits. A similar situation arises in set theory when we have two sets A and B and we want to consider them as a whole. In set theory to form the ensemble set we use the cartesian product $A \times B$ to represent the ensemble of the two sets. The analogue in linear algebra is called tensor product and is represented by the symbol $\otimes$. It applies equally to vectors and matrices (i.e. kets and operators).

From a practical standpoint, the tensor product concatenates physical systems. For example, a two particle system would be represented as $\mid \textit{particle } 1\rangle \otimes \mid \textit{particle } 2\rangle$, and the charge and the spin of a particle would be represented also by the tensor product $\mid \textit{charge}\rangle \otimes \mid \textit{spin}\rangle$. If we have two qubits $\mid \psi\rangle$ and $\mid \phi\rangle$, the system composed by these two qubits is represented by $\mid \psi\rangle \otimes \mid \phi\rangle$.

Athough they share a similar goal, cartesian and tensor product differ in the way elements of the ensemble are built out of the parts. Cartesian product produces tuples. So if A and B are two sets of numbers, an element of their cartesian product $A \times B$ is a pair of numbers (a, b) such that a belongs to A and b belongs to B. It is a simple concatenation.

The elements of a tensor product are obtained from the constituent parts in a slightly different way. For example, consider two *2×2 matrices*

$$\mathbb{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \mathbb{B} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \tag{13.26}$$

the tensor product yields a *4×4 matrix*

$$\mathbb{A} \otimes \mathbb{B} = \begin{pmatrix} a\alpha & a\beta & b\alpha & b\beta \\ a\gamma & a\delta & b\gamma & b\delta \\ c\alpha & c\beta & d\alpha & d\beta \\ c\gamma & c\delta & d\gamma & d\delta \end{pmatrix} \tag{13.27}$$

Although it may not be obvious at first sight, this way of constructing the tensor product is consistent with the way we do matrix multiplication. As an operation, it has a very interesting feature, it outputs 4×4 matrices out of 2×2 matrices, but not all 4×4 matrices can be generated in this way (for the mathematically inclined reader, the tensor product operation is not surjective, or "onto"), it is this feature of the tensor

product that will motivate the discussion about entanglement, probably the most peculiar feature of quantum mechanics.

You should take some time to get used to the tensor product, and ensure you do not get confused with all the different products we have introduced in in the last two sections.

1. the dot product ($\langle k \mid q \rangle$) yields a complex number;
2. the exterior product ($\mid k\rangle\langle q \mid$) yields a square matrix of the same dimension that the Ket;
3. The tensor product ($\mid k\rangle\otimes \mid q\rangle$) is used to examine composite systems. It yields a vector (or matrix) of a dimension equal to the sum of the dimensions of the two kets (or matrices) in the product.

Tensor Product in bracket notation

As we mentioned earlier, the tensor product of two qubits $\mid q_1\rangle$ and $\mid q_2\rangle$ is represented as $\mid q_1\rangle\otimes \mid q_1\rangle$. Sometimes notation is abridged and the following four representations of the tensor product are made equivalent

$$\mid q_1\rangle\otimes \mid q_2\rangle \equiv\mid q_1\rangle \mid q_2\rangle \equiv\mid q_1, q_2\rangle \equiv\mid q_1 q_2\rangle \tag{13.28}$$

For n qubits, it is frequent to abbreviate notation giving to each qubit $\mid q\rangle$ an index:

$$\mid q_1\rangle\otimes \mid q_2\rangle \otimes \ldots \otimes \mid q_n\rangle = \bigotimes_{j=1}^{n} \mid q_j\rangle \tag{13.29}$$

The dual of a tensor product of kets is the tensor product of the corresponding bras. This implies that in the abridged notations, the complex conjugate operation turns kets into bras, *but the labels retain their order*

$$\begin{aligned}(\mid q_1 q_2\rangle)^\dagger &= (\mid q_1\rangle\otimes \mid q_2\rangle)^\dagger \\ &= \langle q_1 \mid \otimes\langle q_2 \mid) \\ &= \langle q_1 q_2 \mid).\end{aligned} \tag{13.30}$$

As a consequence, the result of the dot product of two composite systems is the multiplication of the individual dot products taken in order

$$\begin{aligned}\langle q_1 q_2 \mid w_1 w_2\rangle &= (\langle q_1 \mid \otimes\langle q_2 \mid)(\mid w_1\rangle\otimes \mid w_2\rangle) \\ &= \langle q_1 \mid w_1\rangle \otimes \langle q_2 \mid w_1\rangle\end{aligned} \tag{13.31}$$

confusion often arises in the second term, where in the absence of the parenthesis it is easy to get confused by $\langle q_2 \mid\mid w_1\rangle$ and interpret it as a $\langle\mid\rangle$ (note the two vertical separators in the correct form), and then try to to take the dot products inside out, instead of taking them in parallel as it should be done.

13.6.3 Entangled qubits

We have previously introduced the notion of entanglement in terms of wave functions of a system that allows two measurements to be made. Here we see that it follows from the propeerties of the tensor product as a means to concatenate systems. Consider two qubits $\mid \psi\rangle = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$ and $\mid \varphi\rangle = \begin{pmatrix}a\\ b\end{pmatrix}$, according to the definition of the tensor product,

$$\mid \psi\rangle\otimes \mid \varphi\rangle = \begin{pmatrix}\alpha a\\ \alpha b\\ \beta a\\ \beta b\end{pmatrix} \tag{13.32}$$

If we operate in the ensemble system (i.e. ignoring that it is composed by two subsystems), it is not unthinkable to reach a state described by the following ket

$$| \psi_{12}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \tag{13.33}$$

It turns out that the composite system $| \psi_{12}\rangle$ **cannot be expressed as a tensor product of two independent qubits**. That is, operating directly on the ensemble, it is possible to reach states that cannot be described by two isolated systems.

To see why it is so, try to equal equations 13.32 and 13.33: the first element of the ket requires $\alpha a = 0$; this implies that either α or a must be zero. However if $\alpha = 0$ the second element cannot be 1, and similarly, if $a = 0$ the third element would have to be zero instead of one. So there is no combination of α, β, a, and b that allows us to write the system described in equation 13.33 as a tensor product like the one described in equation 13.32. We conclude that

$$| \psi_{12}\rangle \neq | \psi\rangle \otimes | \varphi\rangle. \tag{13.34}$$

We have already encountered similar situations in the context of "mixing" in chapter 11. There, we noted that intensive variables could no longer be well defined in terms of the subsystems, furtermore, if the process of mixing two subsystems was not reversible, the entropy in the final system was bigger than the sum of the entropies, and it no longer made sense to try to express the composite system in terms of its original constituents. The two situations are different but there is certainly room for analogy.

Whenever this situation arises at the quantum level, we say that the two qubits (or any two systems) are entangled. This word is a translation from the German word "Verschränkung", that is often also translated as "interleaved", Schrödinger coined this word to describe the situation in which:

"Maximal knowledge of a total system does not necessarily include total knowledge of all of its parts, not even when these are fully separated from each other and at the moment are not influencing each other at all.".[4]

The most salient difference with what we saw in the mixing process arises from the fact that, as Schrödinger points out in the paragraph above, this "interleaving" of the parts remains even after separating them, and the measurement of one of the parts will condition the result on the other. This is what Einstein called "spooky action at a distance".

We can illustrate the effect of entanglement on measurements rewriting equation 13.33 by means of a superposition of two tensor products

$$| \psi_{12}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \tag{13.35}$$

$$= \frac{1}{\sqrt{2}} \left[\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right] \tag{13.36}$$

$$= \frac{1}{\sqrt{2}} [| 0\rangle_1 | 1\rangle_2 + | 1\rangle_1 | 0\rangle_2]$$

$$= \frac{1}{\sqrt{2}} [| 0_1 1_2\rangle + | 1_1 0_2\rangle] , \tag{13.37}$$

[4] Extract from "Die gegenwartige Situation in der Quantenmechanik," Erwin Schrödinger, Naturwissenschaftern. 23 : pp. 807-812; 823-823, 844-849. (1935). English translation: John D. Trimmer, Proceedings of the American Philosophical Society, 124, 323-38 (1980).

where we have added subindices to distinguish between the two qubits. This is also a good example to start appreciating the power of bracket notation to simplfy expressions.

If we measure the first qubit, we will obtain either $\mid 0\rangle_1$ or $\mid 1\rangle_1$. To compute the probability that the outcome is $\mid 0\rangle_1$, we take the dot product of the entangled state with $\mid 0_1, ?_2\rangle = \mid 0\rangle_1$

$$\langle 0_1, ?_2 \mid \psi_{12}\rangle = \langle 0_1, ?_2 \mid \cdot \frac{1}{\sqrt{2}} [\mid 0\rangle_1 \mid 1\rangle_2 + \mid 1\rangle_1 \mid 0\rangle_2] \tag{13.38}$$

$$= \frac{1}{\sqrt{2}} (\langle 0 \mid 0\rangle_1 \langle ? \mid 1\rangle_2 + \langle 0 \mid 1\rangle_1 \langle ? \mid 0\rangle_2) \tag{13.39}$$

$$= \frac{1}{\sqrt{2}} \langle ? \mid 1\rangle_2 \tag{13.40}$$

This results says that the outcome will be $\mid 0\rangle_1$ with a probability of 1/2, if and only if the second system collapses at the same time to the state $\mid 1\rangle_2$ (note that otherwise if the question mark represented a 0, the probability would be equal to zero). So the measurement on the first system conditions the value on the second system even if the systems are far apart.

To appreciate the spookyness of entanglement, it may be worth thinking of it in a more mundane setting. Imagine you have such a great connection with a colleague that every time he yawns you systematically yawn as well. Your common friends will pay no attention to it as it is a normal thing to happen, we know that when somebody yawns people in the surroundings tend to yawn. You would certainly scare them though if your colleague went to Europe and you remained in the US, and every once in a while you were driven to have a yawn, precisely when your friend had one. In fact, to be scared your friends would need the input from a referee at each side of the ocean to record the events and match time tags. The question would arise as to whether you and your colleague can use your the yawn-connection for the purpose of inmediate communication, since there appears to be a need for a referee. This cartoon example is certainly not quantum mechanical, however it illustrates what is it about entanglement that has fascinated and scared at the same time some of the greatest minds of our time. And at the same time, it introduces one caveat of quantum communication: the need for a classical exchange to verify that communication has existed (the referees). You will have a chance to appreciate this in a real quantum mechanical setting when we discuss teleportation.

13.7 No Cloning Theorem

One of the most natural operations in classical information is to copy bits, it happens all the time in our computers. Quantum logic diverges from classical logic already at this level. *Qubits cannot be copied*, or as it is usually stated: qubits cannot be cloned.

There are several intuitive arguments that can help us understand why it is so. Remember that in Chapter 10 we emphasized that the act of measuring changes the system being measured; if the system is in a superposition, the result of the measurement will be one of the states of the superposition. And the superposition is destroyed. Intuitively, if measuring is at all required to do the cloning, then it will be impossible to have two clones, since we cannot learn anything about the initial superposition. Furthermore, the superposition itself is destroyed by the act of measuring. This implies that a viable cloning device cannot use measurement.

Assume we have such a device and it operates without requiring measurement. One of the foundations of quantum mechanics is the uncertainty principle introduced by Heisenberg. The principle says that certain physical variables cannot be measured at the same time to an arbitrary precision. The example is position and momentum; if the position of a particle is measured with a given precision Δx, the precision with which its momentum is measured is limited: $\Delta p > \hbar/2\Delta x$. With the presumed cloning machine at hand it should be possible to clone the particle and measure momentum to arbitrary precision in one clone and position to arbitrary precision in the other, possibly violating Heisenberg's principle.

These arguments by themselves do not prove the impossibility of cloning, but suggest that the matter is by no means trivial.

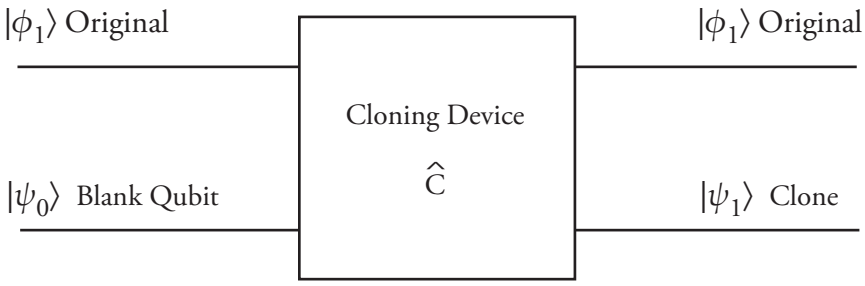

Figure 13.1: Suggested cloning device

To show that cloning is not possible, let us assume that it were possible to clone, and that we could set up a "machine" like the one in Figure 13.1. The cloning device takes the information of one qubit $|\ \phi_1\rangle$ and copies it into another "blank" qubit, the result is a qubit $|\ \psi_1\rangle$ identical to $|\ \phi_1\rangle$, and the original $|\ \phi_1\rangle$ is unmodified. According to what we saw in our overview of bracket notation, such a machine is an operator (we will call it $\widehat{C}$) because it transforms two qubits into two other qubits; and as an operator, $\widehat{C}$ must be unitary. Thus we define $\widehat{C}$

$$|\ Original\rangle \otimes |\ Blank\rangle \quad \xrightarrow{\widehat{C}} \quad |\ Original\rangle \otimes |\ clone\rangle \tag{13.41}$$

We are now ready to clone two arbitrary qubits $|\ \phi_1\rangle$ and $|\ \phi_2\rangle$ separately.

$$\widehat{C}\ |\ \phi_1\rangle\ |\ blank\rangle = |\ \phi_1\rangle\ |\ \psi_1\rangle \tag{13.42}$$

$$\widehat{C}\ |\ \phi_2\rangle\ |\ blank\rangle = |\ \phi_2\rangle\ |\ \psi_2\rangle \tag{13.43}$$

$$\tag{13.44}$$

where it is understood that $|\ \phi_1\rangle = |\ \psi_1\rangle$ and $|\ \phi_2\rangle = |\ \psi_2\rangle$, and we have given them different names to distinguish original from copy.

Since the cloning machine is unitary, it preserves the dot products, so we can compare the dot product before and after cloning

$$\langle \phi_2 |\ \langle blank\ ||\ \phi_1\rangle\ |\ blank\rangle = \langle \phi_2 |\ \langle \psi_2\ ||\ \phi_1\rangle\ |\ \psi_1\rangle \tag{13.45}$$

Recall the rules for taking the dot product of tensor products, each element in the tensor product of kets is multiplied by the bra in the same position in the tensor product of bras, therefore

$$\langle \phi_2 |\ \phi_1\rangle \langle blank |\ blank\rangle = \langle \phi_2 |\ \phi_1\rangle \langle \psi_2 |\ \psi_1\rangle \tag{13.46}$$

The requirements that kets be normalized imposes that $\langle blank |\ blank\rangle = 1$. The above equation can only be true in two cases:

- $\langle \phi_2 |\ \phi_1\rangle = 0$, which means that $|\ \phi_1\rangle$ and $|\ \phi_2\rangle$ are orthogonal. This means that we can clone states chosen at random from a set of orthogonal states. And is equivalent to say that we can clone $|\ 0\rangle$ and $|\ 1\rangle$, which we already knew since we do that classically all the time.
- $\langle \psi_2 |\ \psi_1\rangle = 1$, which means that $\psi_2 = \psi_1$, that is, that clones obtained in each operation are identical. If the two originals were different, as we had assumed, what this result says is that the clone is independent from the original, which is quite a bizarre property for a clone!.

This proof shows that perfect cloning of qubits cannot be achieved. We can certainly store the result of a measurement (this is another way of phrasing the first case), but we cannot clone the superpositions.

13.8 Representation of Qubits

The no-cloning theorem prepares us to expect changes in quantum logic with respect to its classical analog. To fully appreciate the differences, we have to work out a representation of the qubit that unlike the classical bit, allows us to picture superpositions. We will introduce two such representations, a pictorial one that represents the qubit bit as a point in the surface of a sphere, and an operational one that presents the qubit as a line in a circuit much like the logic circuits we explored in Chapter 1.

13.8.1 Qubits in the Bloch sphere

Consider a qubit in an arbitrary superposition

$$| \psi\rangle = \alpha \mid 0\rangle + \beta \mid 1\rangle. \tag{13.47}$$

If α and β were real, equation 13.17 would define a circle of radius one, and we would picture the qubit as a point in the boundary of the circle. However, α and β are complex numbers, so we will have to work a little bit harder to derive a similar intuition.

Every complex number can be represented by a phase and a magnitude, so we can rewrite α and β as:

$$\alpha = Ae^{ia} \qquad \beta = Be^{ib} \tag{13.48}$$

from the normalization of the kets (Equation 13.17), we can derive that

$$\begin{aligned} 1 &= |\alpha|^2 + |\beta|^2 \\ &= A^2 + B^2, \end{aligned} \tag{13.49}$$

and this now is the equation of a circle centered at the origin, so both A and B can be rewritten in terms of an angle[5] $\theta/2$.

$$A = \cos\frac{\theta}{2} \qquad B = \sin\frac{\theta}{2}. \tag{13.50}$$

let us introduce this result in the equation 13.47 of the original superposition:

$$| \psi\rangle = \cos\frac{\theta}{2}e^{ia} \mid 0\rangle + sin\frac{\theta}{2}e^{ib} \mid 1\rangle. \tag{13.51}$$

we can still do one more thing, take e^{ia} out as a common factor

$$\begin{aligned} | \psi\rangle &= e^{ia}\left(\cos\frac{\theta}{2} \mid 0\rangle + sin\frac{\theta}{2}e^{i(b-a)} \mid 1\rangle\right) \\ &= e^{ia}\left(\cos\frac{\theta}{2} \mid 0\rangle + sin\frac{\theta}{2}e^{i\varphi} \mid 1\rangle\right) \end{aligned} \tag{13.52}$$

where we have renamed $\varphi = b - a$. If we ignore the global phase factor(e^{ia}), the two angles θ and φ define a point in a unit sphere. This sphere is called the Bloch Sphere, and is shown in Figure 13.2.

Each point in its surface represents one possible superposition of the states $\mid 0\rangle$ and $\mid 1\rangle$. For example, consider the qubit in the state $\mid \eta\rangle = \frac{1}{\sqrt{2}}(\mid 0\rangle + \mid 1\rangle)$, we can compare this state with equation 13.52, and conclude that then $\theta/2 = \pi/4$, $\varphi = 0$, so that the qubit $\mid \eta\rangle$ is represented by a vector parallel to the x-axis of the Bloch Sphere.

[5]The choice of the angle as $\theta/2$ instead of θ is a technical detail for us, it is adequate for using the spin of an electron as a qubit, if instead, we used the polarization of the photon, then the adequate choice would be θ. This is related to fermions and bosons of which you may have heard, but whose nature is irrelevant to us here.

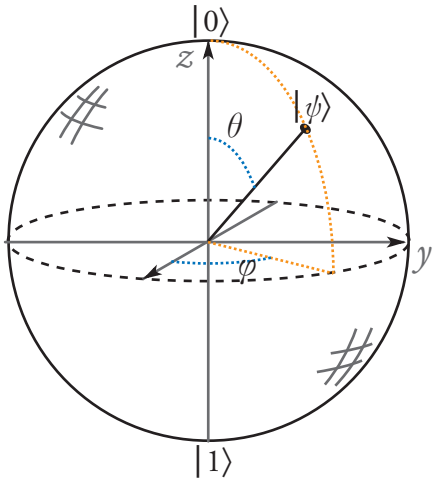

Figure 13.2: Geometrical representation of a Qubit: Bloch Sphere

When we introduced operators we said that they transformed the superposition state of qubits while preserving their normalization. In the context of the Bloch Sphere, this means that operators move dots around the unit sphere, i.e., define trajectories.

Back to equation 13.52, we still need to consider what happens to the global phase factor e^{ia} that we had ignored. This factor would seem to imply that the dot in the Bloch sphere can rotate over itself by an angle a. However, as we are ultimately interested in the probability of each state (because it is the state and not the superpositions what we measure), we should see what happens to this factor as we take the square of the dot products. For example let us examine the probability that measuring the qubit from equation 13.52 yields $\mid 1\rangle$ as an answer

$$\begin{aligned}
|\langle 1 \mid \psi\rangle|^2 &= \left|\langle 1 \mid \cdot e^{ia}\left(\cos\frac{\theta}{2} \mid 0\rangle + sin\frac{\theta}{2}e^{i\varphi} \mid 1\rangle\right)\right|^2 \\
&= \left|e^{ia}\right|^2 \times \left|\cos\frac{\theta}{2}\langle 1 \mid 0\rangle + sin\frac{\theta}{2}e^{i\varphi}\langle 1 \mid 1\rangle\right|^2 \\
&= 1 \times \left|0 + sin\frac{\theta}{2}e^{i\varphi} \times 1\right|^2 \\
&= \left|sin\frac{\theta}{2}e^{i\varphi}\right|^2 .
\end{aligned} \tag{13.53}$$

We see that the global phase factor squares to one, and so plays no role in the calculation of the probability. It is often argued that global phase factors disappear when computing probabilities, and so, are not measurable.

13.8.2 Qubits and symmetries

The Bloch sphere depicts each operation on a qubit as a trajectory on the sphere. However, any trajectory on the sphere can be represented by means of a sequence of rotations about the three axis. So, one way to address the definition of the operations on the qubit is to study rotations about the axis of the bloch sphere. This is intimately connected with the study of symmetries. Thomas Bohr was the first to suggest to use symmetries to interpret quantum mechanics.

We say that an object has a certain symmetry if after applying the corresponding symmetry operation (for example a rotation) the object appears to not have changed; then we say the object is invariant to that symmetry operation. In general, symmetry operations are rotations, reflections and inversions; and invariant

means that start and end position of the object are indistinguishable. For example, Figure 13.3 shows a

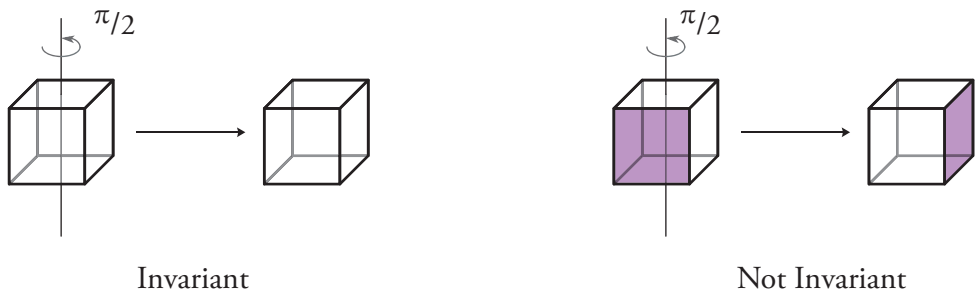

Figure 13.3: Concept of invariance. Symmetry operation: rotation of $\pi/2$ about the vertical axis.

cube, invariant to a rotation of $\pi/2$ about an axis in the center of any of its faces. To distinguish start and end position, you would have to paint one face of the cube, as in the picture to the right of Figure 13.3. Then the cube is no longer invariant to a rotation of $\pi/2$. We would then say that the group of symmetries of the cube to the left of Figure 13.3, contains, among others, the group of rotations of $\pi/2$ about the axis drawn in the figure.

The way physicists use symmetries is to characterize objects by studying the operations that best describe how to transform the object and what are its invariances. Then the representation of the qubit in the Bloch sphere is particularly useful, since it tells us to focus on the group of spatial rotations. In the remainder of this section we will reconcile both views, our perspective of operators as matrices, and the symmetries of the Bloch sphere.

We have already seen that operators on qubits are 2×2 unitary matrices, the additional technical requirement we have to impose to have the group of spatial rotations is that the determinant of the matrices is +1 (as opposed to -1). This group has a name, it is called SU(2)[6]. We can build all of the matrices of SU(2) combining the following four matrices:

$$\mathbb{I} \stackrel{def}{=} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \sigma_x \stackrel{def}{=} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y \stackrel{def}{=} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z \stackrel{def}{=} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{13.54}$$

these matrices are known as the Pauli matrices in honor of Wolfgang Pauli. Note that technically they do not belong to SU(2), to be mathematically rigorous we need to multiply each of them by i, the imaginary number (to verify that this is so, compute their determinant with and without multiplying by i).

Action of Pauli matrices on an arbitrary qubit

The best way to capture the intuition behind Pauli matrices, is to apply each of them to a qubit in an arbitrary superposition state

$$|\,\psi\rangle = \alpha\,|\,0\rangle + \beta\,|\,1\rangle = \cos\frac{\theta}{2}\,|\,0\rangle + \sin\frac{\theta}{2}e^{i\varphi}\,|\,1\rangle \tag{13.55}$$

[6]Groups are conventionally named with letters like O,U, SU, SO, etc. Each of this letters has a meaning. SU(2) stands for the special (S) group of unitary (U) matrices of dimension 2. Special means that the determinant of the matrices is +1, and unitary has here the same meaning it had in the discussion of the operators.

and interpret the result

$$\begin{aligned}\sigma_x \mid 0\rangle &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \\ &= \begin{pmatrix} \beta \\ \alpha \end{pmatrix} \quad \rightarrow \text{ Rotation of } \pi \text{ about x axis} \end{aligned} \tag{13.56}$$

$$\begin{aligned}\sigma_y \mid 0\rangle &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \\ &= i \begin{pmatrix} -\beta \\ \alpha \end{pmatrix} \quad \rightarrow \text{ Rotation of } \pi \text{ about y axis} \end{aligned} \tag{13.57}$$

$$\begin{aligned}\sigma_z \mid 0\rangle &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \\ &= \begin{pmatrix} \alpha \\ -\beta \end{pmatrix} \quad \rightarrow \text{ Rotation of } \pi \text{ about z axis} \end{aligned} \tag{13.58}$$

Figure 13.4 illustrates the operation of σ_y on a qubit in an arbitrary superposition on the Bloch sphere.

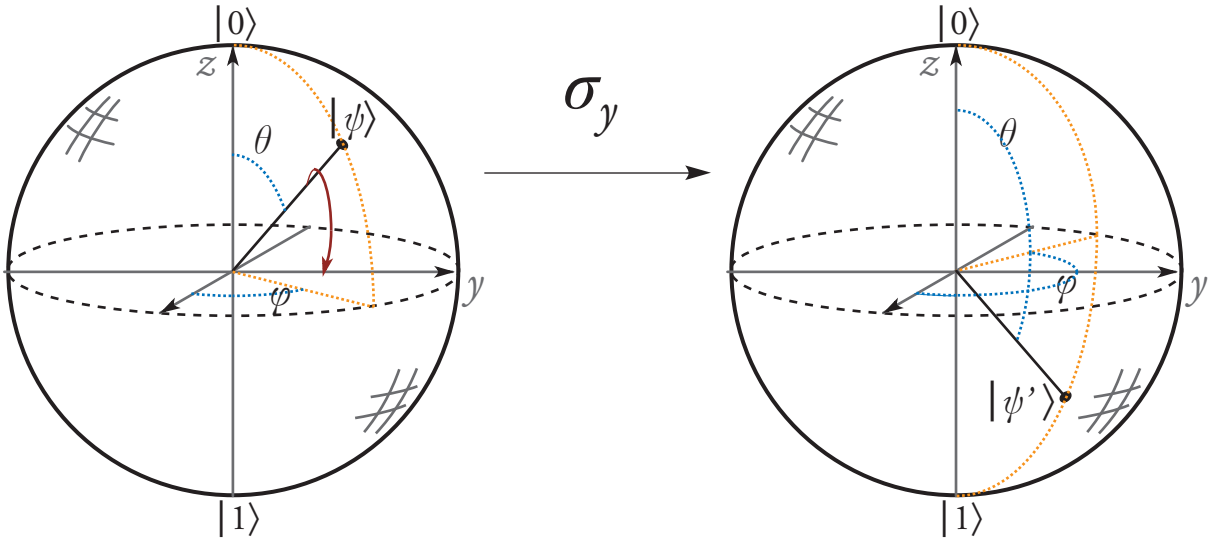

Figure 13.4: Operation of σ_y, on the Bloch sphere

Hence Pauli matrices are rotations of π about each of the axes of the bloch sphere (this motivates the names we gave them). However, to fully explore the surface of the Bloch sphere we need to be able to define arbitrary rotations (not just multiples of π). To do so we use the neat trick of exponentiating Pauli Matrices. Recall Euler's formula relating the exponential function to sine and cosine,

$$e^{ix} = \cos x + i \sin x. \tag{13.59}$$

Euler's formula applies when x is a real number. But we are interested in obtaining a similar result for Pauli matrices. We can prove the equivalent to Euler's formula for Pauli matrices by replacing x by $\frac{\theta}{2}\sigma_x$, and

expanding the exponential as a Taylor series (note that $\sigma_x\sigma_x = \mathbb{I}$)

$$e^{i\sigma_x\theta/2} = 1 + i\frac{\theta}{2}\sigma_x - \frac{1}{2}\left(\frac{\theta}{2}\right)^2\mathbb{I} - i\frac{1}{3}\left(\frac{\theta}{2}\right)^3\sigma_x + \frac{1}{4}\left(\frac{\theta}{2}\right)^4\mathbb{I} + \cdots \tag{13.60}$$

$$= \left(1 - \frac{1}{2}\left(\frac{\theta}{2}\right)^2 + \frac{1}{4}\left(\frac{\theta}{2}\right)^4 + \cdots\right)\mathbb{I} + i\left(0 + \left(\frac{\theta}{2}\right) - \frac{1}{3}\left(\frac{\theta}{2}\right)^3 + \cdots\right)\sigma_x \tag{13.61}$$

$$= \cos\frac{\theta}{2}\mathbb{I} + i\sin\frac{\theta}{2}\,\sigma_x. \tag{13.62}$$

This result shows us how to do arbitrary rotations of an angle θ about the x axis, the resulting operator is often called $R_x(\theta) = e^{i\sigma_x\theta/2}$. The cases of R_y and R_z are completely analogous.

Summing up, we have shown how to represent any qubit as a point in the Bloch sphere and we have learnt how to navigate the bloch sphere doing arbitrary rotations about any of the three axis. It follows that we have obtained an expression for the group of operations of symmetry that allow us to write the form of any operator acting on a single qubit.

13.8.3 Quantum Gates

In the first chapter of these notes, we explored all the possible functions of one and two input arguments and then singled out the most useful boolean functions *NOT*, *AND*, *NAND*, *NOR*, *OR*, *XOR*. Then, we associated it to a pictogram that we called gate, and reviewed the mechanisms to build logic circuits to do computations.

In the previous section we did the same thing for qubits, we characterized all the operators that may transform the value of a single qubit: we defined Pauli's matrices and explained how to do arbitrary rotations. By analogy with the classical case, Pauli matrices and arbitrary rotations are the gates of a quantum circuit. In more advanced treatises on quantum computing you would want to prove a variety of results about the quantum gates, such as the minimum set of gates necessary to do any quantum computation and various results on the generalization from 1 to n qubits. Here we will limit ourselves to summarizing the main details of the quantum algebra, its symbolic representation, and its properties.

Elementary Quantum Gates

The 5 elementary quantum gates are listed in Table 13.1. Their symbolic representation is much simpler

Pauli X	$X =$	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \equiv \sigma_x$	It is equivalent to doing a NOT or bit flip
Pauli Y	$Y =$	$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \equiv \sigma_y$	
Pauli Z	$Z =$	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \equiv \sigma_z$	Changes the internal phase
Hadamard	$H =$	$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$	
Phase	$S =$	$\begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$	

Table 13.1: Elementary quantum gates.

than that of their classical counterparts, it is shown in Figure 13.5.

Table 13.2 enumerates some of the properties of the elementary quantum gates from Table 13.1. These properties are the quantum counterpart to the properties of classical bit functions that we enumerated in

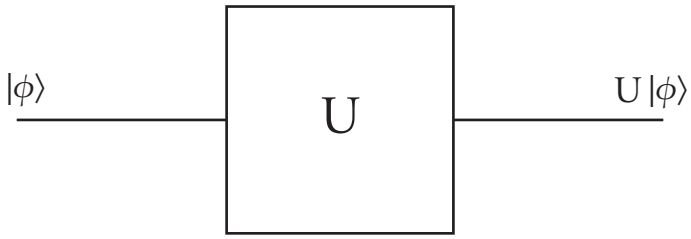

Figure 13.5: Generic quantum gate. Where U is the name given to the generic unitary matrix that this gate represents.

Chapter 1. These and other more advanced rules, help simplify quantum circuits much as deMorgan's law helps in simplifying classical circuits.

$H = \frac{1}{\sqrt{2}}(X + Z)$	$HXH = Z$
$XYX = -Y$	$HYH = -Y$
$XZX = -Z$	$HZH = X$
$XR_y(\theta)X = R_y(-\theta)$	$XR_z(\theta)X = R_y(-\theta)$

Table 13.2: Some of the main properties of single qubit gates.

Two-qubit gates. Controlled Gates

The first thing to note about multiple qubit gates is that the operators are unitary and square, and so unlike classical gates, quantum gates will always have the same number of inputs and outputs. Another way to say it is that all quantum gates are naturally reversible, as it should be expected from the fact that operators are unitary.

The most important two qubit gates are the controlled gates. In a controlled gate the first input qubit is a control qubit, with the same meaning than the classical control bit. If it is in the $|\ 1\rangle$ state, it will trigger the gate that acts on the second qubit, otherwise, it will not trigger it and the second qubit will remain unaltered. A generic example is shown in Figure 13.6, the gate in that example would be named C-U. There

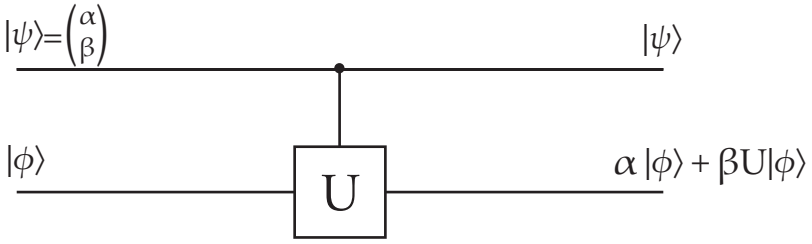

Figure 13.6: Generic quantum controlled gate (C-U). Where U is the name given to the generic unitary matrix that this gate represents.

are two controlled gates that are very relevant to the algorithms we will describe later on, the C-X also known as C-NOT and the C-Z also known as C-Phase. The popularity of the CNOT gate has awarded it a symbol of its own, shown in Figure 13.7.

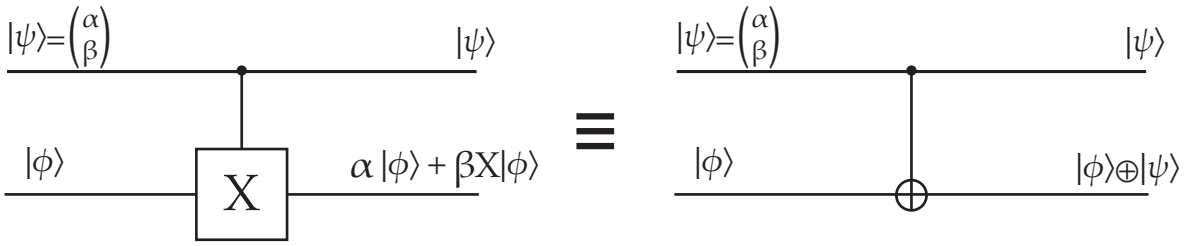

Figure 13.7: CNOT gate.

Finally, it is worth reviewing the matrix representation of the C-Z gate

$$C-Z=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&|1&0|\\0&0&|0&-1|\end{pmatrix} \tag{13.63}$$

where we have emphasized that the bottom right square is the Z matrix.

13.9 Quantum Communication

Two of the most tantalizing applications of quantum mechanics applied to information theory come from the field of communications. The first is the possibility that a quantum computer could break classical cryptographic codes in virtually no time. The second is the idea that quantum bits can be teleported.

In this section we will review the principles behind both teleportation and quantum cryptography (the solution to the problem of code breaking). As we are about to see, entanglement is key in both applications. The two most famous characters in communications: Alice and Bob, will play the main role in teleportation. For quantum cryptography, Eve will play a supporting role.

13.9.1 Teleportation - Alice and Bob's story

Alice and Bob, entangled a pair of qubits $\mid \phi_{AB}\rangle$ when they first met, as is the tradition in the 21st century, people no longer shake hands.

$$\mid \phi_{AB}\rangle = \frac{1}{\sqrt{2}}\left(\mid 0_A\rangle\otimes \mid 0_B\rangle + \mid 1_A\rangle\otimes \mid 1_B\rangle\right) \tag{13.64}$$

They shared a good time, but after school, life took each of them through separate paths. However, they each kept their piece of the entangled pair (it is unlawful not to do so in the 21st century). Now, the pair looks like

$$\mid \phi_{AB}\rangle = \frac{1}{\sqrt{2}}\left(\mid 0_A\rangle \overset{\text{far}}{\otimes} \mid 0_B\rangle + \mid 1_A\rangle \overset{\text{far}}{\otimes} \mid 1_B\rangle\right) \tag{13.65}$$

Alice has now decided to come forward and confess to Bob her love for him. However, she is afraid of rejection and she has heard that qubits can be "teleported instantly", so she decides that the best course of action is to send him a love letter in a qubit $\mid \psi_L\rangle = \alpha \mid 0_L\rangle + \beta \mid 1_L\rangle$ (it could not have been otherwise, love is in the ket).

To do so, Alice carefully puts the qubit of the pair she once entangled with Bob in a composite system with her love-ket-letter $\mid \psi_L\rangle$. The complete three-qubit system can be represented using tensor products

$$\mid \phi_A\psi_L\phi_B\rangle = \frac{1}{\sqrt{2}}\left(\mid 0_A\rangle \otimes (\alpha \mid 0_L\rangle + \beta \mid 1_L\rangle) \overset{\text{far}}{\otimes} \mid 0_B\rangle + \mid 1_A\rangle \otimes (\alpha \mid 0_L\rangle + \beta \mid 1_L\rangle) \overset{\text{far}}{\otimes} \mid 1_B\rangle\right) \tag{13.66}$$

Note that the order in the cross product is not relevant. It only matters when multiplying two composite systems, there we must ensure that each subsystem appears in the same position in each of the kets we multiply.

Alice has now a two-qubit system, and Bob, far away as he is, has one qubit which is entangled with one of the two qubits Alice has. Next, she takes the Bell-Analyzer-1001 (see Figure 13.8), a gift from Bob that she has cherished all this time, and uses it on her two qubits. The Bell analyzer does the following , starting

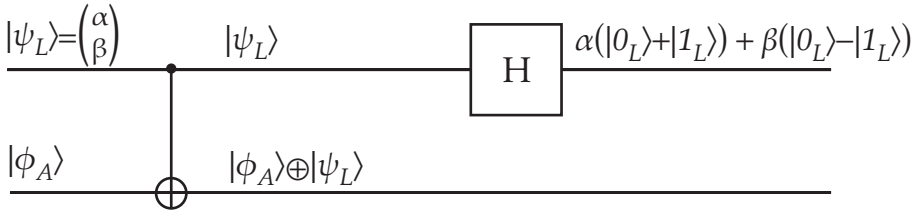

Figure 13.8: Bell analyzer.

from the initial state

$$|\,\phi_A\psi_L\phi_B\rangle = \frac{1}{\sqrt{2}}\left[|\,0_A\rangle\otimes(\alpha\,|\,0_L\rangle+\beta\,|\,1_L\rangle)\overset{\text{far}}{\otimes}|\,0_B\rangle+|\,1_A\rangle\otimes(\alpha\,|\,0_L\rangle+\beta\,|\,1_L\rangle)\overset{\text{far}}{\otimes}|\,1_B\rangle\right]$$

The CNOT gate has the effect of coupling the love-ket-letter with Alice's qubit, and indirectly, since Alice's qubit is entangled with Bob's, with Bob's as well. If ψ_L where equal to $|\,0\rangle$, Alice's other qubit would go unmodified (this is the first line in the equation below). If on the contrary it were $|\,1\rangle$, it would be added modulo 2 to Alice's otherqubit, or, in other words, Alice's other qubit would be flipped (this is what happens in the second line below). Since ψ_L is itself a superposition, what it ends up happening is that a fraction of the qubit remains unmodified with amplitude α and the other fraction is flipped, with amplitude β. So in practice what the CNOT does is transfer the superposition to Alice's Qubit

$$\begin{aligned}&=\frac{1}{\sqrt{2}}\alpha\left(|\,0_A\rangle\overset{\text{far}}{\otimes}|\,0_B\rangle+|\,1_A\rangle\overset{\text{far}}{\otimes}|\,1_B\rangle\right)\otimes|\,0_L\rangle\\ &+\frac{1}{\sqrt{2}}\beta\left(|\,1_A\rangle\overset{\text{far}}{\otimes}|\,0_B\rangle+|\,0_A\rangle\overset{\text{far}}{\otimes}|\,1_B\rangle\right)\otimes|\,1_L\rangle\end{aligned}$$

At this point Alice's and Bob's qubit have both the information of the superposition that was originally in the love-ket-letter. The Hadamard gate produces a new superposition out of the love-ket-letter as follows

$$\begin{aligned}&=\frac{1}{\sqrt{2}}\alpha\left(|\,0_A\rangle\overset{\text{far}}{\otimes}|\,0_B\rangle+|\,1_A\rangle\overset{\text{far}}{\otimes}|\,1_B\rangle\right)\otimes\frac{1}{\sqrt{2}}(|\,0_L\rangle+|\,1_L\rangle)\\ &+\frac{1}{\sqrt{2}}\beta\left(|\,1_A\rangle\overset{\text{far}}{\otimes}|\,0_B\rangle+|\,0_A\rangle\overset{\text{far}}{\otimes}|\,1_B\rangle\right)\otimes\frac{1}{\sqrt{2}}(|\,0_L\rangle-|\,1_L\rangle)\end{aligned}$$

At this point the information about the superposition in the original ket-love-letter is no longer in Alice's hands. However, to appreciate that it is so, we need to make some manipulations and reordering of the cross products. The Expression above can be separated in two sides, what Alice has and what Bob has, and then it breaks down into four terms that have a clearer interpretation

$$\begin{aligned}
&= \frac{1}{2} \mid 0_A\rangle \otimes \mid 0_L\rangle \overset{\text{far}}{\otimes} (\alpha \mid 0_B\rangle + \beta \mid 1_B\rangle) \\
&+ \frac{1}{2} \mid 0_A\rangle \otimes \mid 1_L\rangle \overset{\text{far}}{\otimes} (\alpha \mid 0_B\rangle - \beta \mid 1_B\rangle) \\
&+ \frac{1}{2} \mid 1_A\rangle \otimes \mid 0_L\rangle \overset{\text{far}}{\otimes} (\alpha \mid 1_B\rangle + \beta \mid 0_B\rangle) \\
&+ \frac{1}{2} \mid 1_A\rangle \otimes \mid 1_L\rangle \overset{\text{far}}{\otimes} (\alpha \mid 1_B\rangle - \beta \mid 0_B\rangle)
\end{aligned} \tag{13.67}$$

The manipulation we did leaves no doubt, all the information about the original superposition is in Bob's side. However the information reached bob as a superposition of the right love-ket-letter and all the possible errors produced by phase change (sign) and bit flips!.

Alice realizes now that there is no way for her to avoid talking with Bob about her love-ket-leter. Bob has the information, but to recover the exact superposition from the love-ket-letter he needs to know how to unscramble it, and only Alice can tell him. The next steps are the key to faultless teleportation

- Alice measures her two qubits, she will obtain either of $\mid 0_A 0_L\rangle, \mid 0_A 1_L\rangle, \mid 1_A 0_L\rangle$, or $\mid 1_A 1_L\rangle$ with equal probability.
- Upon Alice's measurement, Bob's qubit takes the value of one of the four possible superpositions. And so, the result of her measurement can help Bob unscramble his bit
- If she measured $\mid 0_A 0_L\rangle$, she will tell Bob not to do anything to his qubit. If she measured $\mid 0_A 1_L\rangle$, Bob will have to correct for the phase (that can be done with a Z gate). If she measured $\mid 1_A 0_L\rangle$, the states in the message have been flipped, and to unflip them Bob will have to use a bit-flip (a.k.a not, a.k.a X) gate. Finally if she measured $\mid 1_A 1_L\rangle$, Bob will have to correct both for the phase and the bit flip.

In total, Alice tells Bob to follow one of 4 possibilities, so she needs to communicate to him 2 classical bits, and then Bob will be able to read the love-ket-letter.

Now Bob better be ready to understand the love letter, because Alice no longer has it! Notice how quantum teleportation transfers the information about the superposition instantly, but Alice needs to measure her system and tell to Bob the result for him to unscramble his bit. In a sense, after teleportation, Bob is doing error correction based on the syndrome measured by Alice; the first bit of Alice informs about bit-flip errors and the second about phase errors. When we get to talk about quantum error correction, we will see that these are the two types of errors that appear in quantum communication.

It is important to appreciate that Alice can only communicate the syndrome to Bob classically, hence, instantaneous teleportation is impossible, Bob will not be able to read the love letter in less time than "far"/"speed of light". Remember the example of the yawn-connection?, there too observers had to exchange data to ensure that communication had occurred.

Quantum teleportation has served us as a verification of the no-cloning theorem, (Alice no longer has a copy of the love letter, much like it happens with "snail" mail), and has helped us introduce the notion of error and error correction. We have also seen that instantaneous communication is not possible, because classical communication is needed to correct for the errors.

13.9.2 Quantum Cryptography

This section is still under construction. Sorry.

13.10 Quantum Algorithms

This section is still under construction. Sorry.

13.10.1 Deutsch Josza

This section is still under construction. Sorry.

13.10.2 Grover

This section is still under construction. Sorry.

13.11 Quantum Information Science

This section is still under construction. Sorry.

Quantum Error Correcting Codes

This section is still under construction. Sorry.