

Information and Entropy in Physical Systems

Craig S. Lent

Abstract The Landauer Principle connects the information theoretic notion of entropy to the physics of statistical mechanics. When a physical system performs a logical operation that erases or loses information, without a copy being preserved, it must transfer a minimum amount of heat, $k_B T \log(2)$, to the environment. How can there be such a connection between the abstract idea of information and the concrete physical reality of heat? To address this question, we adopt the Jaynes approach of grounding statistical mechanics in the Shannon notion of entropy. Probability is a quantification of incomplete information. Entropy should not be conceived in terms of disorder, but rather as a measure on a probability distribution that characterizes the amount of missing information the distribution represents. The thermodynamic entropy is a special case of the Shannon entropy applied to a physical system in equilibrium with a heat bath so that its average energy is fixed. The thermal probability distribution is obtained by maximizing the Shannon entropy, subject to the physical constraints of the problem. It is then possible to naturally extend this description to include a physical memory device, which must be in a nonequilibrium long-lived metastable state. We can then explicitly demonstrate how the requirement for a fundamental minimum energy dissipation is tied to erasure of an unknown bit. Both classical and quantum cases are considered. We show that the classical thermodynamic entropy is in some situations best matched in quantum mechanics, not by the von Neumann entropy, but by a perhaps less familiar quantity—the quantum entropy of outcomes. The case of free expansion of an ideal quantum gas is examined in this context.

Craig S. Lent
Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA,
e-mail: lent@nd.edu

Contents

1	Introduction: What is information?	4
	1.1 Raw information	5
	1.2 Encoded information	7
	1.3 Present strategy	9
2	Probability	9
3	Information theory	12
	3.1 SMI: The Shannon measure of information	12
	3.2 SMI and the question game	13
	3.3 Information gain	17
	3.4 Shannon measure for continuous distributions	17
	3.5 Jaynes maximum entropy principle	18
	3.6 The microcanonical ensemble	19
	3.7 The canonical ensemble	20
4	Classical statistical mechanics	24
	4.1 Statistical mechanics of the canonical ensemble	25
	4.2 Statistical mechanics of the grand canonical ensemble	30
	4.3 Exploring the system microstates	32
5	The Landauer principle	34
	5.1 The many-to-one argument	34
	5.2 Argument from the second law of thermodynamics ..	37
	5.3 Direct calculation of erasure in minimal system	38
6	Quantum mechanics, entropy, and the nature of the	
	physical law	49
	6.1 Quantum formalism and probabilities	49
	6.2 Quantum mechanical SMI for an observable	51
	6.3 Open quantum systems and density operators	54
	6.4 Non-equilibrium quantum system: free expansion	
	of an ideal quantum gas	58
7	Discussion	63
	References	65

1 Introduction: What is information?

When we look out into the physical world, we do not see information. Rather, we see the physical world and the way things are. Or at least that is what we can hope to see, discern, reconstruct, or model, from evidence and experimentation. According to our best theories, what we find in the physical world are particles and fields, or perhaps more correctly: fields and the particles that are the quanta of those fields.

One use of the term “information” refers to this *raw information*—the state of the physical world or part of it. Physics is naturally particularly concerned with raw information. The second and actually more common use of the term “information” denotes *encoded information*.² This is information that supervenes on the raw information and can be expressed in a set of symbols, or most fundamentally, in bits. Encoded information is the domain of information theory. Connecting the two,

representing encoded information and information processing in physical systems, is often the work of applied physics and electrical engineering. The questions addressed here have principally to do with the nature of the application of information theory to physical systems and the consequences of the physical law for information processing procedures.

1.1 Raw information

The particles and fields of the physical world may exist in a number of states, permitted by the physical law, and the particular state of a specific physical system can be specified by a set of numbers. An electron can have spin up (+1) or spin down (-1) relative to a given magnetic field. Perhaps this particular rock was found on the lunar surface at this particular lunar latitude and longitude with this mass and composition. The x-component of electric field at a particular point in space has a specific value this time. The electron may be found in the left well or the right well, etc. The values which describe the state of a physical system are the *raw information* contained in the physical system itself. The raw information is often quite a lot of information. It might include, for example, the position, electron configuration, and nuclear state of each and every atom (or subatomic particle) in a piece of material.

Separability. In order to discuss the raw information present in a particular physical system, it is necessary to conceptually separate the system from the rest of the physical world. The simplest case of such a separation is one in which the system is in fact completely isolated, with no physical coupling or entanglement to any other system. More commonly, we rely on an approximate separation, wherein the interaction with the environment may be minimal or at least can be reasonably well characterized. In many cases, for example, the optical field couples the system to the environment. Some of the details of the raw information about a rock on the moon is flowing out into space as photons. The rock is also in thermal contact with the lunar surface, so the details of the motion of its individual atoms are being affected by thermal fluctuations from the underlying material.

When a system is not perfectly isolated, like the moon rock, raw information can flow to or from the environment. Some information is lost and other information is gained. Micro-bombardment has perhaps altered the moon rock so that some chemical information about its earlier constitution is no longer available. That information may have moved out into the environment carried by the raw information in photons and surface vibrations, for example. Moreover, information present in the rock now perhaps includes historical information about the environment recorded through the interactions with the environment over millennia. To the trained lunar geologist, the fine structure still visible in the sample might preserve a record of a previous cataclysm that occurred 4 billion years ago. So some raw information about earlier events in the solar system has been transferred to and stored in the rock. The in-

formation in the rock may exclude a vast number of historical scenarios. The mere existence of the moon rock means that many conceivable historical sequences, such as the sun exploding or a Mars-size planet colliding with the moon, did *not* occur.

Quantum mechanics makes separability of physical systems even more challenging—we can only point to some of the issues here. Even an isolated system will generally be in a quantum superposition state. Strictly speaking, such a system has no values of dynamical variables like position, momentum, spin, or energy, until one of these is measured. An isolated need not be in an energy eigenstate, for example—in which case it is inaccurate to say that it “has” a particular value of energy. Moreover a physical system can be quantum mechanically entangled with the some or many parts of the environment. No complete description of the quantum state of one part of an entangled system can be given. Some quantum information is shared with other subsystems and is not local to any.

Is information conserved? If the physical system is not completely isolated, then information is clearly not conserved in the system itself. As discussed above, information can flow into or out from the system and so we cannot from the current information reconstruct the past state of the system. It may be possible then for two different past system states to evolve into one present state, for example two different levels of excitation might relax to a single ground state. Many-to-one dynamics are possible because of the environment which can, in this case for example, absorb the excitation energy and with it information about the prior state. Of course it may be that *enough* raw information is retained that a partial reconstruction is possible. When we use a physical system as a memory device, it is a requirement that some important aspects of the past can be inferred from the current state, *e.g.*, what was the bit most recently recorded?

In classical physics, if we imagine a complete description of an entirely isolated system, or of the whole universe conceived as an isolated system, then raw information is indeed conserved by the physical law. The classical mechanical worldview of the world as reversible machinery in motion was famously expressed by Laplace describing an “intellect” subsequently known as *Laplace’s Demon*:

*We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.*¹

Because the microscopic laws of classical physics are reversible, we can solve the equations of motion forward or backward in time. In this sense, for an isolated

¹ Pierre Simon Laplace, *A Philosophical Essay on Probabilities*, 1814

system the raw information is conserved. No new raw information is generated internal to the system, and in virtue of being isolated, no raw information flows in or out. For example, an isolated container of classical gas molecules has a current state consisting of the positions and momenta of all the molecules. From this raw information about the present, the past positions and momenta can be inferred by solving the equations of motion backward in time.

Quantum mechanically for a fully isolated system, information is conserved by unitary evolution of the quantum state vector and this is time-reversible. One important caveat is that measurement of any quantity (which would presumably require interaction with another system) breaks the isolation and thus destroys the reversibility. Yet, measurements seem to happen all the time independent of humans, though we do not understand in detail what is required to produce a measurement event rather than just entanglement of the target system with the measurement system.² Measurement, which can be triggered by small environmental interactions, forces a quantum system to choose a new state—an eigenstate of the operator that corresponds to the measured quantity. New raw information, in the form of measurement outcomes, is created, and old quantum information is destroyed.

1.2 Encoded information

By the term “information” we most often mean what we refer to here as *encoded information*.” Consider a clay tablet on which someone has impressed arrow-shaped indentations in different orientations, or a row of capacitors each of which holds either zero charge or $+Q$, or a street sign on which has been painted the word “Stop” or “Slow.” The raw information consists of the precise shape of the tablet with its indentations, the presence or absence of electrons on each capacitor, the configuration of paint pigment on the sign. The encoded information is also present, but not as additional raw information. Encoded information supervenes on the physical, raw information, through another element—the *encoding scheme*.

An encoding scheme consists of a partition of the possible states of the physical system, the raw information, and an association between each element of the partition and abstract symbols. A particular arrangement of paint pigments is associated with the symbol “S”. The partition is broad enough to include variations in the precise shape of the pigment. The binary “1” might be associated with a certain amount of positive charge stored on the capacitor, give or take a margin of error. Some regions of the systems state space have no associated symbol—the pigment is in an indiscernible pattern, or the amount of charge is too low to be clearly significant. The usual encoding scheme partitions the space of possible raw information

² This is the famous Measurement Problem in quantum mechanics. The term is immediately misleading because prior to the measurement, a quantum system does not in general have an underlying value of the measured result.

states into areas representing symbols in a generalized alphabet, and a broad region representing *invalid*, meaning nothing is encoded there.

Encoded information is deliberate. Encoded information is therefore observer-dependent. For the information to be accessible requires access to both (a) the physical system containing the raw information, and (b) the encoding scheme to map the raw information onto a set of symbols. One or two lanterns are in the bell tower. The raw information includes their detailed construction and precise position, potentially down to the atomic level. The encoding scheme consists of the mapping “one lantern” → “The British are coming by land” and “two lanterns” → “The British are coming by sea.”

If one lacks knowledge of the encoding scheme, encoded information is at least unavailable information and some would argue it is not information at all, even though the raw information is present. Prior to discovering the Rosetta stone, Egyptian hieroglyphics were just raw information—patterns on walls.

A standard disclaimer: information theory is not about the *semantic content* of a string of symbols, it is only concerned with the “size” of the container. Consider the two sentences below.

That man wears red hats.

All men wear black hats.

The second sentence conveys much more information than the first, in the colloquial sense of meaningful and consequential knowledge. But because both sentences contain the same number of symbols, they have the same information theoretic size (at least prior to any further possible compression). Neither are we concerned with the truth or falsity of the information as it is connected with the way things actually are. The second sentence is certainly false. Information theory is not concerned with categories like false information or disinformation.

Reversibility of an operation on bits (a computation) is a mathematical feature of the operation. If one can correctly infer from the output of the operation what the input symbols were, then we say the process is logically reversible. The Landauer principle connects logical reversibility (input symbols can be inferred from output symbols) to physical reversibility (the physical process can be run backwards to reconstruct the input state). If raw information is transferred from the computational system to the large and complex environment, it cannot be reconstructed and so has been irreversibly lost.

Biological information. DNA encodes information for the synthesis of proteins with an encoding scheme involving “codons” composed of three-nucleotide sequences. It is now common to describe many processes in living systems as information systems—signaling, replication, sensing, transduction, etc. Information here is usually encoded in structure (as in DNA or RNA) or through the varying concentration of specific molecules. This is, of course, just a way of speaking at a higher level about raw information in chemical reactions. We normally understand this to

be information conceived by *analogy* to that deliberately encoded by humans, which is taken to be encoded information *sensu stricto*.

1.3 Present strategy

Our goal here is to connect what we know about the evolution of raw information, guided by the physical law, and the encoded information that supervenes on it. The particular focus here is on the Landauer Principle that connects a logical erasure of encoded information with the physical transfer of heat to the environment. Why should that be? Both the logical process and the physical process involve the concept of entropy. But entropy was defined as a physical thermodynamic and statistical mechanical quantity by Clausius, Boltzmann, Gibbs and von Neumann, and only later defined by Shannon as an information theoretic quantity. Some argue that information theoretic entropy has nothing to do with thermodynamic entropy so that the Landauer Principle makes a category error simply because the two words are identical. [1] Norton argues that mistaking unknown bits for a “bit gas” and thereby confusing the two entropy concepts is simply silly. [2]

To disentangle this requires several steps. The next section attempts to articulate carefully the concept of probability, which has both information theoretic and physical uses. Section 3 introduces the Shannon notion of entropy, here called the Shannon measure of information (SMI) as a measure on a probability distribution. The Jaynes principle of maximum entropy is then used for the information theoretic problem of constructing a probability distribution given limited knowledge about outcomes that can be expressed as mathematical constraints. The results are of immediately familiar to anyone acquainted with statistical mechanics. Section 4 follows Jaynes path in making the connection to physics. This can then be applied to the Landauer Principle as discussed in Section 5. A key result here is a concrete and specific calculation of entropy and heat flow in a minimal physical system. The quantum formulation in Section 6 requires extension of the basic formalism to open systems. The connection to Shannon entropy is made both through the usual von Neumann entropy and through the less-familiar “entropy of outcomes.” The quantum calculation of free expansion of a gas is revealing in this regard. By grounding statistical mechanics explicitly in the information theoretic notion of entropy, we can firmly establish the connections that make the Landauer Principle clear and compelling.

2 Probability

We first consider two classical systems.

System 1: A fair die. A fair six-sided die is randomly cast on a table top. The possible results are 1, 2, ..., 6, and the probability of obtaining each result are identical.

$$P_1 = P_2 = P_3 = P_4 = P_5 = P_6 = 1/6 \quad (1)$$

System 2: An ideal gas. We consider a monatomic gas with a very large number N of non-interacting atoms (e.g., argon) in a volume V with pressure P . Let us assume the system is in thermal contact with a heat bath with temperature T . A heat bath is a system with thermodynamically many degrees of freedom that has long-since stabilized all average measures. If the accessible microstates of the system are enumerated and have energies E_i , then the well-known Boltzmann result (to which we will return) is that the probability of finding the system in state j is

$$P_j = \frac{e^{-E_j/k_B T}}{\sum_j e^{-E_j/k_B T}}. \quad (2)$$

It is worth noting a few features of this basic description. We take it as understood in the classical case that at any particular time a specific system indeed *has* a specific state, and that the state which follows is determined by the previous state and the physical law. At the moment I toss the die into the air it has a certain position, velocity, and angular momentum about some axis. Knowing that, as well as perhaps the details of the air currents in the room and the landing surface properties, one could imagine calculating the precise trajectory including any bouncing on the table. The resulting motion, right through until the die settles on the surface, could in principle be calculated, and was certainly determined at the moment the die left my hand.

Similarly, for the ideal gas: the positions and momenta of all the N particles constitutes the actual, objective, state of the system at a particular time. There is a "fact of the matter" as to what the microstate of the gas (this liter of argon on my desk) is right now. It is a practical impossibility for us to measure all these dynamical quantities, particularly at the same instant, but they presumably exist.

We use the language and calculus of probabilities because we lack a complete knowledge of the state and its change over time. The probabilities are therefore "observer-relative." A robotic die tosser with fairly precise launching techniques might be able to predict, say, that the outcomes would more likely be a 4, 5 or 6. An observer who knew more microscopic information about the ideal gas could reasonably assign a different set of probabilities to the microstates of the system. Equation (2) represents the probabilities that *any* observer who knew only the macroscopic quantities T , N , and V , should assign to the microstates of the specific system in front of them. It is not in that sense "subjective." It does not depend on who the observer is or on their emotional or mental state.

Laplace's demon, who knows the position and momentum of each particle in the universe, has no need of probabilities. The physical law as understood in classical

mechanics enables the computation of all future states from a complete description of the present state. It is a practical impossibility to make such a calculation, given human limitations and also the limitations of the physical resources available in the universe for computation. But the point of principle is important. The classical universe is simply solving its equations of motion forward in time.

We are adopting a Bayes/Jaynes approach here that probabilities are to be understood as numerical expressions of partial knowledge, incomplete information, of a present state, or a future event. A probability $P = 1$ represents certain knowledge that the event *will* occur, a probability $P = 0$ means the event certainly *will not* occur, and a real value between 1 and 0 represents greater or lesser partial knowledge that the event will occur. Equations (1) and (2) specify the probability for a future measurements of the state of each system.

On the classical account, a measurement of the system (e.g. looking at the die) reveals an existing fact of the system's state that was true the instant before the measurement occurred. Therefore, we do not need to distinguish between a the probability of a measurement event having a certain outcome and the system having a certain state. We can equally well talk about the probability of the die being on the surface with a 5 showing and the die being seen to be a 5, or revealed to be a 5 when a shaker cup is lifted. The quantum account, discussed in Sec. 6, is different.

Confirmatory evidence that a probability distribution was correct would be the relative frequencies of many such measurements on many essentially identically prepared systems. For the die of (1), that would take many tosses of a fair die. For the thermodynamic case of (2), that means with the same macroscopic variables—an ensemble average in the limit of many trials.

Another feature of this probabilistic analysis is revealed in the phrase “accessible microstate.” There is always a background knowledge of the system which precedes the assignment of probabilities and limits the set of possibilities considered to what we will call the accessible region (AR). In the case of the die, for instance, we are assuming that the die will in fact land on the table and have a face up. We decide to ignore other possible sequences of events. Perhaps it falls off the table, bounces and lands tilted up against a table leg with a corner of the die facing upward. Perhaps a meteor impacts the table with catastrophic results before the die can land. For the gas, we assume of course that the container doesn't leak, that a passing ultra-high energy cosmic ray doesn't deposit extra energy in the gas, etc. A set of extremely low-probability possibilities are removed from consideration at the outset, normally without comment. The AR must be kept in mind because what constitutes “complete ignorance” about the outcome is, as in the case of the die above, uniform probability over the AR, not uniform probability over *every conceivable* outcome.³ We always begin some background knowledge.

³ We will not wade into the subtler issues involved, but refer the reader to Chapter 12 of Jaynes. [3] The quantum treatment in Sec. 6 actually makes the choice of basis explicit, and therefore clarifies the question: "Ignorance with respect to what?".

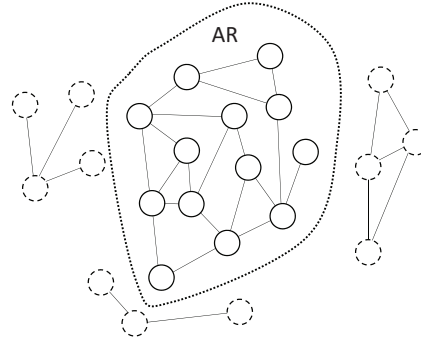


Fig. 1 Schematic of accessible region (AR) of state space. Circles represent possible system state and line represent possible transitions. Some state, those shown outside the dotted line, are reasonable to practically ignore because they are either too rare or difficult to access.

3 Information theory

3.1 SMI: *The Shannon measure of information*

Claude Shannon, considering the transmission of symbols in communication, introduced a measure on a probability distribution which he called the entropy. Using the term “entropy” was well-motivated, and was the course of action advised by von Neumann, but it has resulted in some confusion. We will adopt the strategy of Ben Naim and call this measure the Shannon Measure of Information (SMI). [4, 5]

The SMI characterizes a probability distribution $P = [P_1, P_2, \dots, P_k, \dots, P_N]$ by a real non-negative number, measured in bits, computed from the distribution.

$$\boxed{\text{SMI}[P] = - \sum_{k=1}^N P_k \log_2(P_k)} \quad (3)$$

The SMI is a measure of the the amount of information, in bits, that one is missing if all one knows about the current state or future outcomes is the probability distribution P . If one outcome, say event 2, is certain, then $P_2 = 1$ and $P_k = 0$ for all other k . In that case the SMI is 0; there is no missing information. If all the probabilities are equal, $P_k = 1/N$ for all k and the SMI is $\log_2(N)$ bits. If N is an even power of 2, this is clear: $N = 4$ corresponds to 2 bits missing; $N = 8$ corresponds to 3 missing bits, etc.

Figures 2a and 2b show graphically the cases of a probability distribution among 8 outcomes for the case when the outcome is certain to be event 2

$$P_{(2)} = [0, 1, 0, 0, 0, 0] \quad \text{SMI} = 0 \text{ bits}, \quad (4)$$

and when all outcomes are equally likely:

$$P_{\text{uniform}} = \left[\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8} \right] \quad \text{SMI} = 3 \text{ bits}. \quad (5)$$

Figure 2c shows the case when the probability distribution is

$$P = [0.15, 0.20, 0.30, 0.15, 0.05, 0.05, 0.05, 0.05] \quad \text{SMI} = 2.67 \text{ bits}. \quad (6)$$

The SMI is intermediate between the uniform $N = 4$, SMI= 2 and the uniform $N = 8$, SMI=3 cases. With this probability distribution we know *something* about which events are likely to occur. There is an 80% chance that the result will be events 1–4, for example. We have somewhat less *missing information* that if we only knew P_{uniform} . It is convenient that SMI could also stand for “Shannon Missing Information.”

3.2 SMI and the question game

To understand how the phrase “missing information” can have a precise meaning, it is helpful to consider a variation of the game of twenty questions and see how the SMI functions both to play the game and to make predictions. We consider the Question Game in which a player called the *chooser* selects one of set of N numbered items and the player called the *questioner* asks a series of yes/no questions with the object of deducing the index of the item chosen with the fewest number of questions.

Suppose, for example, $N = 8$ and the chooser picks an item at random, *i.e.* the probability for each choice is $1/8$ as in Fig. 2b. One strategy for the questioner is to ask “Is it 1?”, then “Is it 2?”, then “Is it 3?” and so on. On average the questioner would ask $N/2$ questions before learning the choice. This is, of course, a poor strategy.

The optimal strategy for a uniform probability distribution is the familiar binary search using repeated bipartitions. The questioner asks “Is the item in the set $\{1, 2, 3, 4\}$?”, and if the answer is yes, asks “Is it in $\{3, 4\}$?”, and if the answer is no, asks “Is it item 1?”, and thereby has determined the choice using only 3 questions. This will work every time. The SMI of the uniform probability distribution over 8 choices, $\log_2(8) = 3$ bits, is the number of yes/no questions one needs to ask to determine the choice using the optimal strategy. The amount of *missing information* was initially 3 bits. With the answer to each question, the questioner received an additional 1 bit of information, until finally there was no information missing and the identity of chosen item was certain to the questioner.

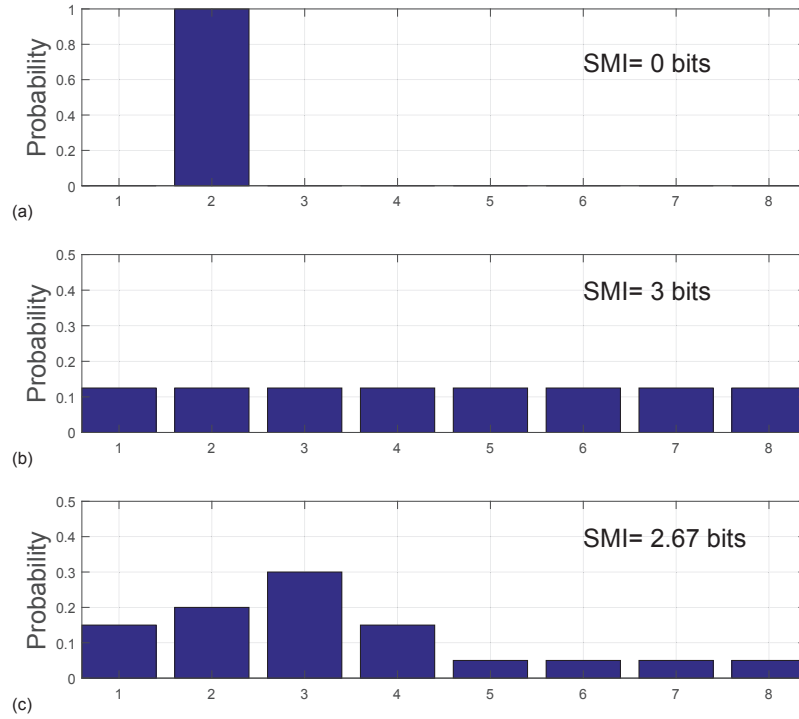


Fig. 2 Probability distributions and associated Shannon Measure of Information (SMI). (a) If exactly one state has unit probability, then there is certainty about which state the system will be found in and no information is missing; the SMI is 0. (b) The case of uniform probability for 8 possible states has an $\text{SMI} = \log_2(8) = 3$. (c) In general the probability distribution reflects some missing information, but less than complete ignorance. The amount of missing information is quantified by the SMI.

Suppose the selector was not selecting entirely at random, but was making the choice according to the probability distribution of Equation (6) shown in Fig. 2c. We

could imagine that the chooser is randomly drawing numbered balls from a large container. Each ball has a number $[1, \dots, N]$ on it but there are more 3's than 2's and so forth according to the ratios in (6). The questioner knows the probability distribution. The binary search as above is now not the optimal strategy. The set $\{1, 2, 3, 4\}$ has a total probability of 80%, so asking the first question as above seems like almost wasting a question—the answer is not providing as much information.

The optimal strategy is now as follows:

1. Let the set S be the set of possible items $\{1, 2, \dots, N\}$ and $P_k, k = 1, 2, \dots, N$ be the probabilities that each item is selected.
2. Consider all possible bipartitions of the set S into two non-empty sets, S_{left} and S_{right} .

For each bipartition $S \rightarrow \{S_{\text{left}}, S_{\text{right}}\}$:

- a. Sum the probabilities of the individual events in each set and renormalize to get two numbers: P_{left} and P_{right} that sum to 1.
 - b. Calculate the SMI of the probability distribution for the bipartition $P_{bp} = [P_{\text{left}}, P_{\text{right}}]$ using Equation (3).
3. Choose the bipartition with the largest SMI and ask the chooser the question: “Is the item in S_{left} ?”.
 4. If the answer is yes, replace S with S_{left} .
If the answer is no, replace S with S_{right} .
 5. Repeat from step (2) until there is only one item in the set S .

For our example using (6), an initial bipartition

$$S \rightarrow \{\{1, 2, 3, 5\}, \{4, 6, 7, 8\}\}$$

has an SMI of 0.88129 but

$$S \rightarrow \{\{3, 4, 5\}, \{1, 2, 6, 7, 8\}\}$$

has an SMI of 1.0, making the corresponding question a very productive question whose answer yields a full bit of information.

Now we imagine the chooser and the questioner playing the game many times with the same probability distribution P (we may suppose the numbered balls are replaced as they are drawn). Many times the optimal bipartition of the remaining set has an SMI of less than 1. Sometimes the questioner gets lucky and is able to deduce the chosen item in 2 questions, and sometimes it takes 3 or 4. Over very many games, what is the average number of questions required? One might have hoped the answer would be $\text{SMI}[P]$, but it is not quite that simple. In one simulation,

an average over 50,000 games yields an average number of questions $\langle N_q \rangle = 2.70$, whereas the SMI is 2.67. The constraint is actually

$$\text{SMI}[P] < \langle N_q \rangle < \text{SMI}[P] + 1 \quad (7)$$

which of course becomes relatively tight for large N .

We can interpret the series of yes/no answers in the game as 1's and 0's encoded in a string stored in a binary register. (We will assume that we have compressed the string using an optimal scheme—a Huffman code—so that common question sequences are encoded with fewer bits than rare sequences.) The average number of questions is bounded by $\text{SMI}[P] + 1$. Since the SMI need not be integer and we can only have an integer number of bit positions in the register, we need need to round up to the nearest integer. We conclude that the average size of a binary register necessary to specify a particular item choice, given probability distribution P , is

$$N_{(\text{binary register})} = \text{ceil}(\text{SMI}[P] + 1). \quad (8)$$

The SMI is a quantitative measure of missing information. It is helpful to keep in mind which party is missing the information. Who is missing the information? The chooser is holding the item in her hand. She is not missing any information, for her there is complete certainty about what the chosen item is and the SMI is 0. It is the questioner who is missing the information and relying on knowledge of the probability distribution to ask good questions. The SMI in this case is a measure of *questioner's* missing information about the item, when his information about which item is chosen is incomplete and characterized by probability distribution P .

Jaynes makes this point about Shannon's original problem of a sender transmitting a message, encoded in a set of symbols, through a communication channel to a receiver.⁴ The sender knows everything about the content of the message; there is no probability distribution involved and no missing information for him. Prior to getting the message, the receiver, by contrast, knows nothing about the content of message, perhaps not even the language that the message will be in. Normally the *designer* of the communication system does not know in advance what the specific messages will be, but suppose the designer does know something about the message, for example the language of the message and the probabilities of the occurrence of each letter in that language. He can then use the SMI of that probability distribution to create an efficient encoding scheme (i.e., a data compression algorithm). The SMI of the language characterizes the incomplete information of the designer of the information system. If the designer knows more about the messages, the SMI is less and he can make an even more efficient system. The sender has complete information, the receiver has no information (yet), and the designer has partial information characterized by a probability distribution and its associated SMI.

⁴ Jaynes [3], p. 634.

3.3 Information gain

For the Question Game above, the SMI characterized the initial missing information of the receiver when all he knew was the probability distribution P (Eq. (6), Fig. 2c). If the receiver had no information at all, the probability distribution he would have to use is just the uniform probability P_{uniform} of Equation (5). We can therefore ask how much information did he gain when he was given P . We define the information gained from the knowledge of the probability distribution P as the difference between the SMI (missing information) of P_{uniform} and the SMI of P .

$$I[P] \equiv \text{SMI}[P_{\text{uniform}}] - \text{SMI}[P] \quad (9)$$

$$I[P] = \log_2(N) + \sum_{k=1}^N P_k \log_2(P_k) \quad (10)$$

We are here using the uniform probability distribution over the AR to represent complete ignorance and asking how much information was gained by having been given the probability distribution P . In information theory, this quantity is known as the relative entropy of P with respect to $P_{(\text{uniform})}$, or the Kullback–Leibler divergence between the two. [6]

3.4 Shannon measure for continuous distributions

Can we define the SMI of a probability density $P(x)$ defined for a continuous outcome x ?

A natural approach is to consider the SMI of a finite discretization of $x \in [0, L]$ at each point $x_k = k\Delta x$. We note that the probability density $P(x)$ will now have the units inverse to the units of x . The probability of the event occurring in the small interval of width Δx around x_k is $P_k = P(x_k)\Delta x$. The SMI on this discrete set of $N = (L/\Delta x) + 1$ outcomes can be written

$$\text{SMI}[P^{\Delta x}(x)] = - \sum_k (P(x_k)\Delta x) \log_2[(P(x_k)\Delta x)]. \quad (11)$$

If we try to take the limit of this expression as $\Delta x \rightarrow 0$, however, this quantity diverges.

Shannon suggested, but did not derive, the following expression, usually called the *differential entropy* for a probability density $P(x)$.

$$S^{\text{diff}}[P(x)] = - \int P(x) \log_2[P(x)] dx \quad (12)$$

This normally converges, but turns out to have several problems if it is interpreted as a direct analogy to the discrete SMI: (a) the units of the expression are not correct, (b) it can be negative for some distributions, (c) it is not invariant under a change of variables, and (d) it does not smoothly match the discrete case in the usual Riemann limit of (11).

The fundamental problem with formulating the amount of missing information, an SMI, for a continuous distribution is simply that a particular value of x from a continuous range takes an infinite amount of information to specify. There are an infinite number of yes/no questions required to specify an arbitrary point on the real axis. So the answer to the question “*How much information is missing if all I know is the continuous probability distribution $P(x)$?*” turns out to be an infinite number of bits.

It can be argued that for any physical system, the precision of measurement (or quantum effects) limit the distinguishable values of the measurable x to a finite discrete set, so (11) is always the relevant measure.

We can, however, clearly establish the related measure $I[P(x)]$ for a continuous distribution by analogy with (9). We take the probability density reflecting complete ignorance to be the uniform distribution on the accessible regions $x = [0, L]$ to be $P_{\text{uniform}}(x) = 1/L$. We can then define the information gain of $P(x)$ for both the continuous probability density and its finite discretization.

$$I[P(x)] = S^{\text{diff}}[P_{\text{uniform}}(x)] - S^{\text{diff}}[P(x)] \quad (13)$$

$$I[P^{\Delta x}(x)] = \text{SMI}[P_{\text{uniform}}^{\Delta x}(x)] - \text{SMI}[P^{\Delta x}(x)] \quad (14)$$

Taking the difference removes the problems mentioned above and (14) is numerically equivalent to a trapezoidal integration of (13) with a discretization of Δx .

3.5 Jaynes maximum entropy principle

Probability is an expression of incomplete information. Given that we have *some* information, how should we construct a probability distribution that reflects that knowledge, but is otherwise unbiased? The best general procedure, known as Jaynes Maximum Entropy Principle (better would be: Maximum SMI Principle), is to choose the probabilities p_k to maximize the SMI of the distribution, subject to constraints that express what we do know.

3.6 The microcanonical ensemble

The simplest case is the one in which we know nothing but the rule for probabilities—that they must add up to 1.

Let us define the Shannon measure with the natural logarithm as the base, a simple matter of multiplying by $\log(2)$ (we take $\log(x)$ to denote $\log_e(x)$). The quantity of missing information represented by S_{MI} is then measured in *nats* rather than *bits*.

$$S_{MI} \equiv \log(2)\text{SMI} = -\sum_k p_k \log(p_k) \quad (15)$$

We want to write a probability density $P = \{p_k\}$ that maximizes $S_{MI}(P)$ subject only to the constraint:

$$\sum_k p_k = 1 \quad (16)$$

Using the method of Lagrange multipliers we construct the Lagrangian

$$\mathcal{L}(P, \lambda_0) = -\sum_k p_k \log(p_k) - (\lambda_0 - 1) \left(\sum_k p_k - 1 \right) \quad (17)$$

where $1 - \lambda_0$ is the Lagrange multiplier.⁵ We maximize \mathcal{L} by setting the partial derivatives with respect to each of the p_k to 0. The equation $\partial\mathcal{L}/\partial\lambda_0 = 0$ just recovers Equation (16).

$$\frac{\partial}{\partial p_k} \mathcal{L}(P, \lambda_0) = -\log(p_k) - 1 + 1 - \lambda_0 = 0 \quad (18)$$

The solution is then

$$p_k = e^{-\lambda_0} \quad (19)$$

which is true for all k , so each probability is the same and, using (16) again, we have

$$\sum_k p_k = \sum_k e^{-\lambda_0} = N e^{-\lambda_0} \quad (20)$$

$$e^{-\lambda_0} = \frac{1}{N} \quad (21)$$

$$\lambda_0 = \log(N) \quad (22)$$

$$p_k = \frac{1}{N}. \quad (23)$$

Thus our intuition that if we know nothing about the probability distribution we should make all probabilities equal is recovered from the maximum entropy princi-

⁵ The factor $(1 - \lambda_0)$ is used instead of λ_0 to simplify the form of the result.

ple. The value of S_{MI} at this maximum is:

$$S_{MI}^{max} = -\sum_k \frac{1}{N} \log\left(\frac{1}{N}\right) \quad (24)$$

$$S_{MI}^{max} = \log(N) \quad (25)$$

The reader will recognize in Equation (25) the famous Boltzmann expression for entropy ($S = k_b \log W$), without the Boltzmann constant. It also connects very simply to the case in Fig. 2b where for N results with equal probability we have an SMI = $\log_2(N) = 3$ bits, the size of the binary register needed to specify one outcome, and the average number of yes/no questions needed to determine one result.

In terms of choosing balls from an urn our picture is this. An urn contains a very large number of balls (many more than N), each of which is labeled with a number $k \in \{1, 2, \dots, N\}$. There are the same large number of balls with each index, so drawing a ball randomly from the urn picks one of possible results with equal probability.

3.7 The canonical ensemble

Here we consider that each ball in the urn has written on it both the index k and a value of another quantity we will call A . The values of A are denoted $[a_1, a_2, a_3, \dots, a_N]$. Every ball with a 1 on it has a_1 written on it, and so for each of the other indices. Suppose that we know the average value of the quantity A , denoted $\langle A \rangle$. By average here we mean simply an average of the values obtained from many repeated drawing of balls from the urn. What is the optimal (maximum SMI) probability distribution p_k that will yield the given average $\langle A \rangle$?

Following the maximization procedure, we maximize S_{MI} (15) subject to the two constraints:

$$1 = \sum_k p_k \quad (26)$$

$$\langle A \rangle = \sum_k p_k a_k. \quad (27)$$

We construct the Lagrangian, which now has Lagrange multipliers λ_0 and λ_1

$$\mathcal{L} = -\sum_k p_k \log(p_k) - (\lambda_0 - 1) \left(\sum_k p_k - 1 \right) - \lambda_1 \left(\sum_k p_k a_k - \langle A \rangle \right) \quad (28)$$

Maximizing \mathcal{L} with respect to each p_k , we obtain:

$$\frac{\partial \mathcal{L}}{\partial p_k} = -\log(p_k) - 1 + 1 - \lambda_0 - \lambda_1 a_k = 0 \quad (29)$$

with the result

$$p_k = e^{-\lambda_0} e^{-\lambda_1 a_k}. \quad (30)$$

We define

$$Z \equiv e^{\lambda_0} \quad (31)$$

or

$$\lambda_0 = \log(Z) \quad (32)$$

We call Z the partition function. The probabilities can therefore be written

$$p_k = \frac{e^{-\lambda_1 a_k}}{Z}. \quad (33)$$

From the constraint (26) we require

$$\sum_k p_k = 1 = \frac{1}{Z} \sum_k e^{-\lambda_1 a_k} \quad (34)$$

so

$$Z = \sum_k e^{-\lambda_1 a_k} \quad (35)$$

and

$$p_k = \frac{e^{-\lambda_1 a_k}}{\sum_k e^{-\lambda_1 a_k}} \quad (36)$$

which is the well-known Boltzmann distribution.

If we take the logarithm of Z we obtain a way to express the constraint that the average value of A is fixed.

$$\begin{aligned} \frac{\partial}{\partial \lambda_1} \log(Z) &= \left(\frac{1}{Z}\right) \frac{\partial}{\partial \lambda_1} \sum_k e^{-\lambda_1 a_k} \\ &= \left(\frac{1}{Z}\right) \sum_k (-a_k) e^{-\lambda_1 a_k} = -\langle A \rangle \end{aligned} \quad (37)$$

hence:

$$\langle A \rangle = -\frac{\partial}{\partial \lambda_1} \log(Z). \quad (38)$$

We define

$$F \equiv -\frac{1}{\lambda_1} \log(Z). \quad (39)$$

Now we can substitute the probability distribution (36) in the resulting probability distribution to evaluate the value S_{MI} at this maximum.

$$\begin{aligned}
S_{MI}^{\max} &= -\sum_k p_k \log(p_k) \\
&= -\frac{1}{Z} \sum_k e^{-\lambda_1 a_k} \log\left(\frac{e^{-\lambda_1 a_k}}{Z}\right) \\
&= -\frac{1}{Z} \sum_k e^{-\lambda_1 a_k} \left[\log\left(e^{-\lambda_1 a_k}\right) - \log(Z)\right] \\
&= -\frac{1}{Z} \sum_k e^{-\lambda_1 a_k} (-\lambda_1 a_k) + \underbrace{\frac{1}{Z} \sum_k e^{-\lambda_1 a_k}}_1 \log(Z) \\
&= \lambda_1 \frac{1}{Z} \sum_k e^{-\lambda_1 a_k} a_k + \log(Z)
\end{aligned}$$

$S_{MI}^{\max} = \log(Z) + \lambda_1 \langle A \rangle$ (40)

Notice the contrast between (40) and (25). For the case with no constraints, the maximum SMI was $\log(N)$. The term $\log(Z)$ in (40) is the sum over the N exponentials shown in (35), which rapidly decrease in magnitude. This term $\log(Z)$ has a correspondingly much smaller value than $\log(N)$; the missing information is much less. Equation (25) is recovered from (40) when $\lambda_1 = 0$ and the distribution is again uniform.

Writing Equation (40) in terms of F we obtain

$$F = \langle A \rangle - \frac{1}{\lambda_1} S_{MI}^{\max}.$$

$F = \langle A \rangle - \frac{1}{\lambda_1} S_{MI}^{\max}$ (41)

As an example, take the case of $N = 8$ and let the values of A be

$$a = [0.0, 1.0, 1.3, 2.1, 2.8, 3.4, 4.0, 6.0].$$

(42)

Suppose we know that $\langle A \rangle = 1.54$. The probability distribution p_k which maximizes S_{MI} is the exponential distribution (36) with $\lambda_1 = 0.4$. This is the probability which represents just the known facts: the probabilities must add to one and the average value of A is given. Figure 3 shows the probabilities p_k associated with each of the values of a_k .

There is a one-to-one relationship between λ_1 and $\langle A \rangle$ given by Equations (27) and (36) and shown in Fig. 4. As λ_1 becomes large, the probability accumulates in the lowest values of A . When $\lambda_1 = 0$ we recover the uniform distribution. A large negative λ_1 results in more probability at the higher values of A . If all that is known

is $\langle A \rangle$, one must just read off the associated λ_1 from Fig. 4 and assign the probability distribution appropriately using (36). Any other probability distribution would implicitly assume knowledge one does not actually have—it would put in an incorrect, if unintentional, bias. (Tribus⁶ calls the Lagrange multiplier λ_1 the “temper” of the distribution, an act of dramatic foreshadowing. [7])

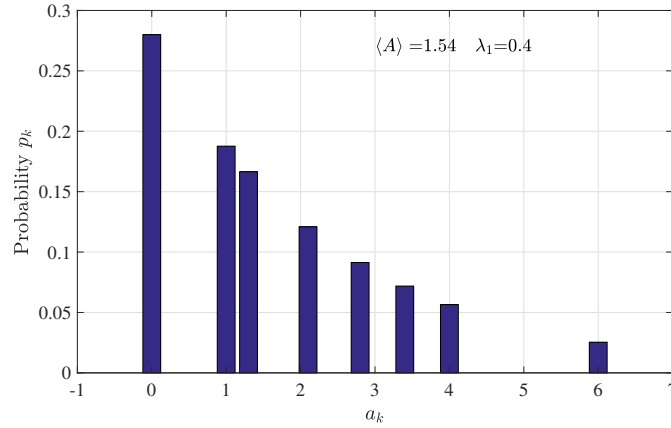


Fig. 3 The probability distribution p_k which maximizes the Shannon entropy. In this example we assume that each possible outcome of the quantity A is in the set $a = [0.0, 1.0, 1.3, 2.1, 2.8, 3.4, 4.0, 6.0]$. The distribution shown is the one that maximizes the Shannon information theoretic entropy S_{MI} , subject to the constraint that the average outcome $\langle A \rangle$ is known to be 1.54. The result is a Boltzmann distribution (36) with the Lagrange multiplier λ_1 (see (28)) equal to 0.4. This probability distribution uniquely captures only the known information. Any other distribution with the same $\langle A \rangle$ would implicitly, and incorrectly, represent more knowledge than simply knowing the average.

The procedure above can straightforwardly be extended to the case when each result is labeled with the values of additional quantities that characterize the outcome. If we have another quantity B with values b_k and a known average value $\langle B \rangle$, then we would obtain the corresponding exponential distribution with an additional parameter λ_2 .

$$p_k = \frac{e^{-(\lambda_1 a_k + \lambda_2 b_k)}}{\sum_k e^{-(\lambda_1 a_k + \lambda_2 b_k)}} \quad (43)$$

with

$$Z = \sum_k e^{-(\lambda_1 a_k + \lambda_2 b_k)} \quad (44)$$

and

⁶ Myron Tribus’s thermodynamics text for engineers was an early attempt to popularize Jaynes grounding of the field on Shannon information theory.

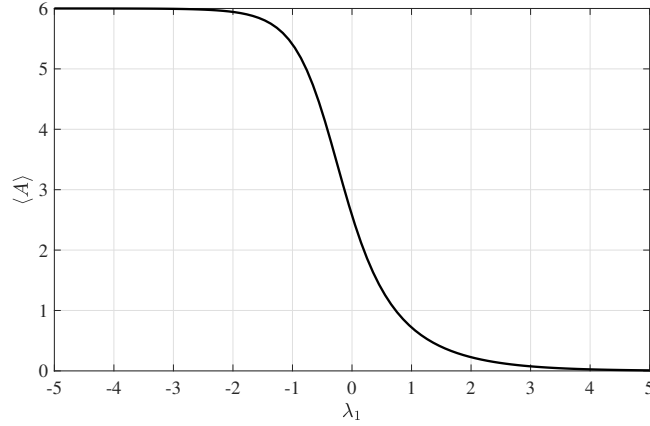


Fig. 4 For the example shown in Fig. 3, the expectation value (average) of the quantity A is shown as a function of the Lagrange multiplier λ_1 . Equations (27) and (36) fix the relationship between these quantities. When λ_1 is positive, the probability distribution is weighted toward small values of A . When λ_1 is negative, the probability distribution is weighted toward large values of A . When λ_1 is 0, the probability is uniform for all values of A .

$$\langle B \rangle = -\frac{\partial}{\partial \lambda_2} \log(Z). \quad (45)$$

$$S_{MI}^{max} = \lambda_1 \langle A \rangle + \lambda_2 \langle B \rangle + \log(Z) \quad (46)$$

The extension to any number of such quantities proceeds in the same way.

4 Classical statistical mechanics

We now turn to the application of the previous section to physical systems. The main results of the previous section are familiar mathematical forms from statistical mechanics. We consider now a physical system in equilibrium with a much larger system and apply the analysis to derive thermodynamic results. It will then be possible to see how to extend the analysis to metastable memory systems and dynamic systems far from equilibrium.

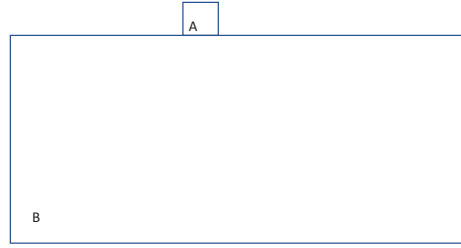


Fig. 5 A physical system A and very large system B in thermal contact so that energy can flow between them. The larger system B acts as a heat bath. In addition, Sect. 4.2 considers the case of diffusive contact, in which both energy and particles can be exchanged between the systems.

4.1 Statistical mechanics of the canonical ensemble

Equilibrium with a thermal bath. Consider a physical system A in thermal contact and equilibrium with a second physical system B as shown schematically in Fig. 5. By thermal contact we mean that the two systems can exchange energy with each other. By equilibrium we mean that whatever transients occurred when they were put in contact are now over and the expectation value of all physical quantities are now time-independent. That this happens is based on our empirical experience of the physical world. System B will represent a thermal bath with very many (*e.g.*, 10^{23}) degrees of freedom.

Suppose A can exist in N states with energies $[E_1, E_2, \dots, E_k, \dots, E_N]$ and similarly for the bath B. We allow that different states k and j may have the same energy. The energy E_k of each state plays the role of the label on each ball a_k in urn described in the previous section. The energy of neither A nor B is fixed because energy can fluctuate between them. Because of the fluctuations in energy, system A can be found in states with different energies at different times. The probabilities of finding system A in the k^{th} state with energy E_k are denoted $P^A = [p_1, p_2, \dots, p_k, \dots, p_N]$. We define the average energy $U_A = \langle E \rangle_A$ and $U_B = \langle E \rangle_B$ for each system.

The key assumption connecting information theory to physical systems is this: We assume the probability of finding the physical system in the state E_k is the same as the probability of randomly selecting from a set of E_k 's with a probability distribution which maximizes the SMI for each system, given the constraints. Here that the constraint is that the average energy is U .

The probabilities for each *physical* system are therefore given by the Boltzmann probability distribution, Equation (36), which we derived from applying Jaynes Principle, a purely information theoretic result.

$$p_{k_A}^{(A)} = \frac{e^{-\lambda_1^{(A)} E_k^{(A)}}}{\sum_{k_A} e^{-\lambda_1^{(A)} E_{k_A}^{(A)}}} \quad (47)$$

$$p_{k_B}^{(B)} = \frac{e^{-\lambda_1^{(B)} E_k^{(B)}}}{\sum_{k_B} e^{-\lambda_1^{(B)} E_{k_B}^{(B)}}} \quad (48)$$

Note that λ_1 has the units of inverse energy. From (27) we have

$$U_A = \sum_{k_A} p_{k_A} E_{k_A}^{(A)} \quad (49)$$

$$U_B = \sum_{k_B} p_{k_B} E_{k_B}^{(B)} \quad (50)$$

We now *define* the thermodynamic entropy $S(U)$ of each physical system, A or B, as $k_B \log(2)$ times the maximal SMI.

$$\boxed{S(U) \equiv k_B \log(2) \text{ SMI}^{\max} \equiv k_B S_{MI}^{\max}} \quad (51)$$

The entropy S is a thermodynamic quantity defined at equilibrium. The SMI, by contrast can be calculated for any probability distribution whatsoever. In words, (51) says:

The value of the thermodynamic entropy $S(U)$ is $k_B \log(2)$ times the amount of missing information in the probability distribution that maximizes the (information theoretic) SMI, given the constraint that the average energy is U .

The entropy is a so-called state function. It depends on U , the average energy but is not determined by the history of the system prior to coming to equilibrium (we can extend the dependence to other state variables like N and V).

The conversion factor between the SMI (in units of bits) and the entropy S (in units of energy/temperature) is $k_B \log(2)$. We can think of this as the entropy associated with 1 bit of missing information. The factor $\log(2)$ simply converts the base of the logarithms from the bit-oriented \log_2 to the more conventional natural logarithm. The Boltzmann factor k_B reflects the historical and convenient choice of a unit for temperature (which we will introduce below) in Kelvins rather than, say, in Joules.

If the system A and the bath B are not strongly coupled together we can assume that the entropy S (and SMI) for the composite A+B system is the sum of the entropy for each system.

$$S_{AB} = S_A + S_B \quad (52)$$

This would not be true, for example, if the state of B simply mirrored the state of A. This lack of correlation (in Shannon's terms, *mutual information*) is part of what we mean by being a thermal bath or reservoir—it has a vast number of degrees of freedom that are independent of the system degrees of freedom.

Conservation of energy in the composite A+B system gives us the constraint

$$U_{AB} = U_A + U_B. \quad (53)$$

Relying on identification of the thermodynamic entropy with the maximum value of SMI from Jaynes principle, we can apply Equation (40) to each system:

$$S_A = k_B \log(Z_A) + k_B \lambda_A U_A \quad (54)$$

$$S_B = k_B \log(Z_B) + k_B \lambda_B U_B \quad (55)$$

Consider now a small energy fluctuation ΔU that increases the average energy of A, and therefore must decrease the average energy of B by the same amount.

$$S_{AB} = S_A(U_A + \Delta U) + S_B(U_B - \Delta U). \quad (56)$$

We require that S_{AB} be maximal under this variation and so expand each term to first order.

$$S_{AB} = S_A(U_A) + \left(\frac{\partial S_A}{\partial U_A} \right) \Delta U + S_B(U_B) + \left(\frac{\partial S_B}{\partial U_B} \right) (-\Delta U) \quad (57)$$

Requiring that the first order change be zero then yields the stationary condition:

$$\left(\frac{\partial S_A}{\partial U_A} \right) = \left(\frac{\partial S_B}{\partial U_B} \right). \quad (58)$$

Using equations (54) and (55) to evaluate the partial derivatives, we find

$$\lambda_1^{(A)} = \lambda_1^{(B)}. \quad (59)$$

At this point we *define* the temperature to be inversely proportional the Lagrange multiplier λ_1 associated with the average energy constraint.

$$\boxed{\frac{1}{k_B T} \equiv \lambda_1} \quad (60)$$

So (59) gives us that in equilibrium between the two systems

$$T_A = T_B. \quad (61)$$

and, using (54) and (55) again, gives us

$$\frac{1}{T} = \left(\frac{\partial S}{\partial U} \right). \quad (62)$$

The Boltzmann distribution and the thermodynamic partition function for each system are then given by

$$p_k = \frac{1}{Z} e^{-E_k/k_B T} \quad (63)$$

where

$$Z = \sum_k e^{-E_k/k_B T}. \quad (64)$$

The average energy U is given by

$$U = \langle E \rangle = - \frac{\partial}{\partial \beta_1} \log(Z) \quad (65)$$

or

$$U = k_B T^2 \frac{\partial}{\partial T} \log(Z) \quad (66)$$

The information theoretic expression in (41) now becomes the definition of the Helmholtz free energy,

$$F \equiv -k_B T \log(Z) \quad (67)$$

and (40) becomes

$$F = U - TS. \quad (68)$$

If we consider differential changes at constant temperature (and volume) we have

$$dF = dU - T dS \quad (69)$$

which is a key thermodynamic identity. At equilibrium we have from (62),

$$dU = T dS \quad (70)$$

or,

$$dS = \frac{dU}{T}, \quad (71)$$

so at equilibrium $dF = 0$, that is, the free energy is at a minimum.

For a system in thermal equilibrium with a large heat bath the free energy is a minimum. A large mechanical system that can dissipate energy minimizes its total energy U , settling down to an energy minimum. The free energy $F = U - TS$ is

the corresponding quantity for a system in thermal equilibrium with a heat bath. It reflects the interplay between lowering the energy and thermal excitation.

We note that Equation (71) is the original Clausius definition of entropy. Thus, starting with the Shannon information theoretic definition of entropy (3) we have arrived at both the Boltzmann expression (25) and the thermodynamic expression of Clausius. The fact that the historical order was exactly opposite this logical order has been one source of confusion.

External Work. The energy of each state of the system may depend on externally imposed parameters like the volume V , applied magnetic field B , an applied electric field, etc. For example, for an ideal gas we take $E_k = E_k(V)$, then seek the average value of the differential energy shift with volume. The shift is fundamentally because the single-particle quantum energy levels move up in energy as the volume is decreased. Classically, it suffices for us to note that a piston must apply a force in the direction of its motion to squeeze a gas into a smaller volume, thus doing positive work on the gas. First, we take the derivative of $\log(Z)$ with respect to the volume.

$$\frac{\partial}{\partial V} \log(Z) = \frac{\frac{\partial}{\partial V} \sum_k e^{-E_k/k_B T}}{\sum_k e^{-E_k/k_B T}} = -\frac{1}{k_B T} \left\langle \frac{\partial E_k}{\partial V} \right\rangle \quad (72)$$

This average value of the energy shift, using the probabilities in (63) is then defined to be the pressure.

$$p \equiv \left\langle -\frac{\partial E_k}{\partial V} \right\rangle \quad (73)$$

The minus sign is because when the volume is decreased the energy increases as does the pressure. Using (72) we then have:

$$\boxed{p = -\left(\frac{\partial F}{\partial V}\right)_T} \quad (74)$$

Similarly, for the expectation value of the magnetization

$$M = -\left(\frac{\partial F}{\partial B}\right)_T \quad (75)$$

and so forth.

Applying an external force to the system mechanically, electrically, or magnetically, is another way to increase the average energy of the system. Equation (73) can be written in terms of this shift.

$$dU = -pdV \quad (76)$$

This kind of direct transfer of energy from outside the system to, or from, the system is called *work*. So there are two ways the average energy of the system can be changed: by heat transfer or by work. Heat is the transfer of a certain amount of energy Q accompanied by a change in entropy. To include both kinds of energy change, we need to modify (70) to:

$$dU = dW + dQ. \quad (77)$$

This is the first law of thermodynamics, the conservation of energy.

In the case of compressing the volume with a frictionless piston, for example, $dW = -pdV$. We will be concerned in section 5 with doing electrical work. If a voltage source transfers differential charge dq across a voltage difference V then, neglecting the resistance of conductors, it does work

$$dW = Vdq. \quad (78)$$

We will interpret the symbol V as voltage or volume by context.

4.2 Statistical mechanics of the grand canonical ensemble

We can extend the application of the information theoretic results of the previous section to the grand canonical ensemble by considering a system and bath both comprised of particles. Up to now we did not need that assumption so the results have even broader applicability. If in addition to energy, particles can flow between the system and the bath, we can label states of the system with both energy E_k and the number of particle N_k . The number of particles can fluctuate and in equilibrium we have a constant expectation value $\langle N \rangle$. This constraint gives us a distribution of the form in equation (44), where we now identify $\lambda_2 = -\mu/k_B T$, defining the chemical potential μ ,

$$p_k = \frac{e^{-(E_k - \mu N_k)/k_B T}}{\sum_k e^{-(E_k - \mu N_k)/k_B T}} = \frac{e^{-(E_k - \mu N_k)/k_B T}}{Z_G} \quad (79)$$

The constraint that the average particle number is $\langle N \rangle$ gives us, in analogy to (38):

$$\langle N \rangle = -\frac{\partial}{\partial \lambda_2} \log(Z_G) \quad (80)$$

The corresponding free energy expression is obtained by the same procedure that connected (40) and (41) to the free energy (67), now using (46) to yield

$$F = -k_B T \log(Z_G) \quad (81)$$

and

$$\boxed{F = U + \mu \langle N \rangle - TS.} \quad (82)$$

The chemical potential μ is the driver for particle exchange between systems in diffusive contact.

Non-interacting Fermions and Bosons. In the special case of a group of non-interacting particles in thermal and diffusive contact with a reservoir at temperature T , we can obtain the standard results for fermion and boson statistics using the probability distribution (79), derived from the information theoretic result (44), and the basic rules for state occupancy. For this case we consider a fixed set of single-particle energy levels e_i with the i^{th} level occupied by n_i particles. The total energy of a particular configuration of occupancies will be

$$E = \sum_i n_i e_i \quad N = \sum_i n_i \quad (83)$$

For fermions, we need only the fact that each level can have occupancy of either 0 or 1 but no greater. The partition function can be written in a factored form with each factor corresponding to the possible occupancies of each level.

$$Z_G = \left(1 + e^{-(e_1 - \mu)/k_B T}\right) \left(1 + e^{-(e_2 - \mu)/k_B T}\right) \left(1 + e^{-(e_3 - \mu)/k_B T}\right) \dots \quad (84)$$

$$= \prod_i \left(1 + e^{-(e_i - \mu)/k_B T}\right) \quad (85)$$

The average occupancy of the j^{th} level is then given by:

$$\langle n_j \rangle = \frac{\left(0 + 1e^{-(e_j - \mu)/k_B T}\right) \prod_{i \neq j} \left(1 + e^{-(e_i - \mu)/k_B T}\right)}{\prod_i \left(1 + e^{-(e_i - \mu)/k_B T}\right)} \quad (86)$$

$$= \frac{e^{-(e_j - \mu)/k_B T}}{1 + e^{-(e_j - \mu)/k_B T}} \quad (87)$$

$$\boxed{\langle n_j \rangle = \frac{1}{e^{(e_j - \mu)/k_B T} + 1}.} \quad (88)$$

Equation (88) is, of course, the famous Fermi-Dirac distribution function

The factorization of the partition function and subsequent cancellation is a general feature of composites of non-interacting systems. Each single-particle energy level acts like a separate system in thermal and diffusive equilibrium with the reservoir.

For bosons, each level can be occupied by any number of particles.

$$\langle n_j \rangle = \frac{e^{-(e_j - \mu)/k_B T} + 2e^{-2(e_j - \mu)/k_B T} + 3e^{-3(e_j - \mu)/k_B T} + \dots}{1 + e^{-(e_j - \mu)/k_B T} + e^{-2(e_j - \mu)/k_B T} + e^{-3(e_j - \mu)/k_B T} + \dots} \quad (89)$$

The denominator is a geometric series in $x = e^{-(e_j - \mu)/k_B T}$ yielding $1/(1 - x)$. In terms of the same x , the numerator is $S = x + 2x^2 + 3x^3 + \dots$. We note $S - xS = x + x^2 + x^3 + \dots$, which is the geometric series minus 1 or $S = x/(1 - x)^2$. We arrive at the Bose-Einstein distribution function:

$$\langle n_j \rangle = \frac{1}{e^{(e_j - \mu)/k_B T} - 1}. \quad (90)$$

4.3 Exploring the system microstates

Consider a monatomic classical ideal gas (no interactions between particles) with N particles of mass m in a volume V at temperature T in the dilute limit. The dilute limit is when the density of particles is low enough that the average occupancy of each energy level is much less than 1. In this case it makes no difference whether the particles are fermions or bosons. A noble gas is well approximated this way. The thermodynamic entropy S , and the corresponding SMI, are given by the Sakur-Tetrode equation:

$$SMI = \frac{1}{k_B \log(2)} S(N, V, T) = N \log_2 \left[\frac{V}{N} \left(\frac{mk_B T}{2\pi\hbar^2} \right)^{3/2} \right] + \frac{5}{2} N \quad (91)$$

This equation can be derived from the Jaynes Maximum Entropy principle [5] or from standard thermodynamics. Though describing a classical gas, the expression contains Planck's constant because it's necessary to enumerate the smallest volumes in phase space (limited by the uncertainty relationship) to give an absolute number.

For a liter of Argon gas at standard temperature and pressure, the SMI from (91) is about 10^{23} bits. Recall that this is the average length of the binary register necessary to specify all the accessible microstates of the gas.⁷ The number of possible microstates is therefore

$$N_{\text{microstates}} \approx 2^{SMI} = 2^{(10^{24})} \approx 10^{(10^{23})}. \quad (92)$$

Let us imagine this liter of argon moving from one accessible microstate to another according to its internal dynamics. We can imagine a binary register holding the (compressed) index of the current microstate of the gas which keeps clicking from

⁷ Note that we assume that state indices (a series of 1's and 0's specifying each particular state) are chosen in an optimal way, employing a so-called Huffman code, that uses fewer bits to specify more probable states and longer bit sequences for rarer states. The average register length is the average index length weighted by the state probabilities.

one number to the next as the gas particle move from state to state. How much time does it take to go from one state to another? Well, to change states requires the atomic nuclei to move positions. Suppose the shortest time to move nuclear positions establishes the “tic” of the clock at which point the microstate register changes to record the next microstate index. We are looking for an *upper bound* to the number of microstates explored in a given time, so we will take the shortest possible clock tic, the light travel time across a nucleus $T_{\text{tic}} \approx 10^{-25}$ seconds. The nuclei could hardly have changed positions faster than that. How many microstates could the liter of argon have explored? The time since the big bang is $T_{\text{universe}} \approx 10^{18}$ seconds. An upper bound on the number of microstates that the liter of argon could possibly have explored since the beginning of the universe is then

$$N_{\text{microstates explored}} \leq \frac{T_{\text{universe}}}{T_{\text{tic}}} = \frac{10^{18} \text{ s}}{10^{-25} \text{ s}} = 10^{43}. \quad (93)$$

Therefore, the *fraction* of the accessible microstates that could possibly have been explored since the universe began is

$$\frac{N_{\text{microstates explored}}}{N_{\text{microstates}}} = \frac{10^{43}}{10^{(10^{23})}} = 10^{(-10^{23}+43)} \approx 10^{-(10^{23})}. \quad (94)$$

This is an extremely small number. We conclude that any actual gas is visiting only an extraordinarily small fraction of its possible microstates. The vast majority of terms in the sums like the partition function represent states that *have never* been realized by the system and *will never* be realized by the system.

The connection established by Jaynes using Shannon entropy (SMI) put statistical mechanics on a much firmer footing for both classical and quantum cases. But it fair to wonder: why does this work so well in so many situations? We can view the system averages, like the average energy, as time averages. Or we can view these averages as ensemble averages over many systems with identical macroscopic properties but different microscopic configurations. In either case, the microscopic state of a particular system is the result of its particular *dynamics*, the equations of motion and its past history or initial state. The connection with information theory is fundamentally made by mapping one problem: (a) the physical process of dynamical motion, onto a second problem (b) the statistics of selecting a ball with an energy label on it from an urn with an optimal (maximum SMI) distribution of balls given the average energy.

The number of balls in the urn that corresponds to the possibilities for the physical system is staggeringly large. Whatever sample we use for the average is, in an important sense, a very small sample. We have just seen that the system state space is vastly larger than it could explore in any reasonable time.

The probability distribution is determined by the physical dynamics (classical or quantum) and yet those dynamics do not matter to statistical mechanics. Statistical mechanics is remarkably independent of mechanics. We know some dynamical model systems pull systems into attractors rather than distributing them uniformly

across state space. So the question becomes: what properties of real physical dynamical systems make them so amenable to very accurate description by selecting “typical” states from the incredibly vast state space available? This is an active research area. An example of sustained and sophisticated work on this issue in the quantum thermodynamics is the work of Gemmer, Michel, and Mahler. [8]

5 The Landauer principle

The Landauer Principle (LP) asserts that for a physical system representing an information state, loss of one bit of information necessarily entails dissipation to the environment of a minimum amount of heat equal to $k_B T \log(2)$. If information is not lost, there is no minimum amount of heat dissipation necessary.

Any logically irreversible operation, AND, OR, SUM, etc., involves a loss of information in the sense that inputs cannot be logically inferred from the outputs. The archetypal irreversible operation is erasure, so we will focus our attention on that.

For specific devices the heat dissipation may, of course, be much more than the fundamental minimum. Modern CMOS transistors operate with orders of magnitude more energy dissipated by each transition. If the device and associated architecture is designed optimally (adiabatic logic) it may be possible to lower the dissipation by switching its state more smoothly and slowly. The Landauer Principle places a fundamental lower limit on how much heat dissipation must occur, depending on the amount of information that is lost.

We discuss three arguments for the Landauer Principle. The first, the many-to-one argument is the one Landauer himself usually employed, though he most often simply asserted the principle as self-evident. The second grounds the argument on the Second Law of Thermodynamics. The third is a calculation on a minimal model system. This has the advantage of being a mathematical result with clear assumptions and clear conclusions. Of course it is susceptible to the objection that there might be another way of constructing a memory that would violate LP. Be that as it may, it is very clarifying to see LP simply emerge from a simple calculation. It also forces the issue of how to define the entropy of memory states that cannot be equilibrium states, using the same approach as in Sec. 4.

5.1 *The many-to-one argument*

This argument for LP is based on the time-reversal symmetry of the physical law at the microscale. We will consider a physical system that has three states (or re-

gions of state space) that encode a binary 1, binary 0, or a null state holding no information. Why not just use the two binary states? Choosing to define erasure as ‘set-to-zero’ results in a morass of ambiguity because we cannot physically distinguish the erasure process from the physical process of writing a 0. Consider the erasure of a single bit of information that is represented by the physical system. Let us assume that we do not know the value of the initial bit, 1 or 0, but need to create with *one protocol*, a series of externally controlled operations, that will result in the physical system being set to the null state.

This is shown schematically in Fig. 6. The proposed protocol would be such that given an initial physical configuration of the system that corresponds to *either* a 0 or a 1 bit, the protocol would be such that the physical system would evolve in time under the relevant physical law to the state representing the null bit, as shown in the figure.

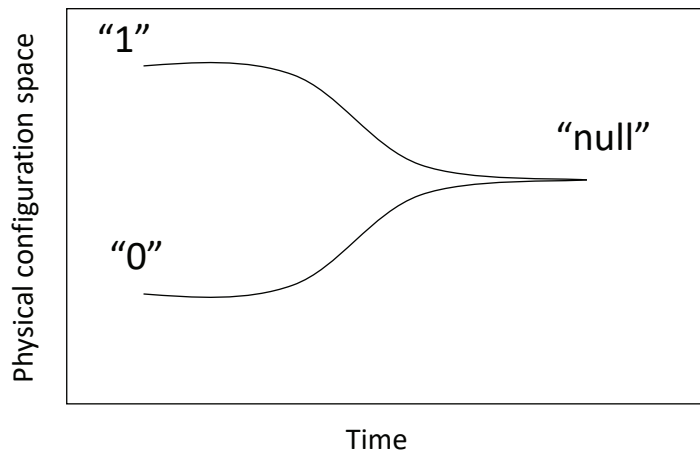


Fig. 6 The many-to-one quality of bit erasure to a null state. A general-purpose physical erasure procedure would have to be able to take either a 1 or a 0 state and move it into the null state. It would have to be the same procedure irrespective of the initial state and so work on an unknown bit. Reversing the temporal sequence of the procedure should be possible because the microscopic laws of physics are reversible and one-to-one. There cannot therefore be a unique backward evolution that would restore the state to its original 1 or 0.

But, the LP arguments objects, something like Fig. 6 cannot occur. If it did, then one could start at the null state, and run the protocol backwards. But which state would then result, 1 or 0? Running a movie of the whole physical process backward

should be a valid physical process.⁸ Fundamental physics takes one (classical or quantum) state forward in time to a unique future quantum state.

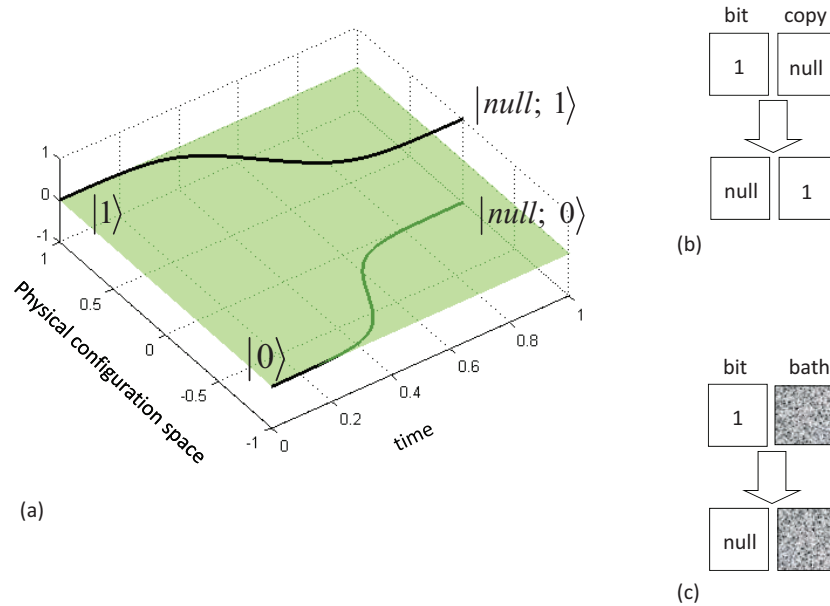


Fig. 7 The reality of an apparent many-to-one bit erasure process. (a) If a system *appears* to evolve as in Fig. 6 under the reversible laws of physics, there must actually be another physical system coupled to the bit-storing system. The physical state of the copy system is here represented on the z-axis of the graph. (b) The auxiliary bit could be simply another physical system that makes a copy of the original bit, for example a neighboring bit or one in a measurement apparatus. Having a copy allows there to be a different physical process to erase a 1 than to erase a 0, shown as separate curves in the figure, so the process is reversible. (c) Alternatively, the copy could be contained in the large system of a thermal bath in contact with the bit system. The copy in that case is encoded in the many degrees of freedom of the bath and is unrecoverable, leading to an irreversible erasure. This process transfers entropy and heat to the bath. After erasure, the increased energy of the bath means there are twice as many accessible bath states—those corresponding to the 1 bit having been erased, and those corresponding to the 0 bit having been erased.

Therefore, if it *looks* like the situation of Fig. 6 is occurring, careful examination will reveal that there is at least one other system involved that is not being properly accounted for. As shown in Fig. 7a, there must be at least one other degree of freedom in the process, shown in the figure as the z-coordinate that, is different between

⁸ The weak interaction responsible for the decay of the neutral B meson has been directly shown to violate time reversal symmetry. See J. P. Lees et al., “Observation of Time-Reversal Violation in the B^0 Meson System,” *Phys. Rev. Lett.* 109, 211801 (2012). We will restrict our considerations to systems not involving B or K mesons, or the weak interaction.

the two outcomes. The figure shows the z -coordinate of this auxiliary system initially at 0, and then moving to ± 1 depending on the initial state of the system. The auxiliary degree of freedom could be, for example, another system which is initially in the null state and then put by the protocol in the same 1 or 0 state as the primary system was initially. Of course, it could store the inverse state as well. In this case the information of the enlarged system, including the original system and the copy, has not lost the original information.

The auxiliary system could also be a very large complex system like a heat bath. In this that case the bath, taken as a whole, retains, in an entirely inaccessible way, a copy of the original bit. Again, the enlarged system+bath has not lost the information fundamentally, thus preserving the time-reversibility of the system. There are two different final states for the bath, one in which it has interacted with the system in a 1 state and another in which it has interacted with the system in the 0 state. The SMI of the bath has now increased because, knowing only the macrostate (P, V, T, \dots) of the bath, the information about the which bit was originally stored in the system is missing. If the bath is at thermal equilibrium at temperature T , the increase in thermodynamic entropy must be $\Delta S = k_B \log(2)$ with the corresponding heat transfer to the environment $\Delta Q = T \Delta S = k_B T \log(2)$.

If we deliberately make a copy of the bit, physics does not prevent us from exploiting the fact that we have a copy to create different erase-one and erase-zero protocols. We only are forced to pay the Landauer price if there is no copy, or if there is one and we just fail to make use of it.

5.2 Argument from the second law of thermodynamics

Consider again the small system A in thermal equilibrium with a very large system (a heat bath) B as shown in Fig. 5. We will assume that B has temperature T and that system B is in thermal equilibrium with A and so has the same temperature. We now suppose that there are a set of macroscopic controls that allow us to externally manipulate the state of B. In examples that we will flesh out below, these will be electrodes whose electrostatic potential we can change. For the often-considered example of a perfect single-particle gas, manipulation is typically by moving pistons and inserting and removing barriers. However it is accomplished, suppose the entropy S_A of A is *reduced* by the equivalent of one bit of information (see Equation(51)). This is what we mean by erasure.

$$SMI_A(\text{final}) - SMI_A(\text{initial}) = -1 \text{ bit} \quad (95)$$

$$S_A(\text{final}) - S_A(\text{initial}) = -k_B \log(2) = -\Delta S_{\text{bit}} \quad (96)$$

The heat bath is large so that this has no effect on its temperature. We will again assume that the entropy of the global system comprising A and B is the sum of the

entropies of the components. The internal motion of the heat bath B is not correlated to the motion of A .

$$S_{AB} = S_A + S_B \quad (97)$$

The manipulations of A we will assume do not perform a measurement of the microstate of B (or of A). We assume that the manipulation does not give us any new information about AB . Therefore, the change in entropy of the global system must be either zero or positive. The amount of information missing about the microstate of the composite system AB can only increase since we (or anyone or anything doing the manipulation) have not reduced the missing information about the microstate of AB . This is the heart of the Second Law of thermodynamics.

$$\Delta S_{AB} = \Delta S_A + \Delta S_B \geq 0 \quad (98)$$

Therefore:

$$\Delta S_B \geq -\Delta S_A \quad (99)$$

$$\Delta S_B \geq k_B \log(2) = \Delta S_{(\text{bit})} \quad (100)$$

After the manipulation is complete, the bath system B is in an equilibrium state with an increased entropy and an energy which is larger by at least the energy corresponding to one bit.

$$dU_B = T dS_B \geq k_B T \log(2). \quad (101)$$

This is, as we have stressed, a fundamental minimum, not a characteristic energy.

5.3 Direct calculation of erasure in minimal system

Representing *encoded information* with the *raw information* of a physical system involves both the dynamics of the physical system and the encoding scheme. The encoding scheme maps areas of the accessible region of the physical state space of the system to specific information states. Figure 8 represents a physical system with three states for a single particle which can be in one of three state (dots) indexed 1, 2, and 3. The figure illustrates (a) the particle on the right representing a “1”, (b) the particle on the left representing a “0”, and (c) the particle in the middle representing a “null” state containing no binary information.

The encoding scheme for the three logical states can be defined in terms of the probability of the occupancy of each state. The probability of the system being found in state 1, 2, or 3, we denote $[P_1, P_2, P_3]$, and the energy of each state is denoted $[E_1, E_2, E_3]$. We can choose a threshold value P_{th} and encode a binary 1 by a state with $P_3 > P_{th}$, a binary 0 with $P_1 > P_{th}$, and the null state with $P_2 > P_{th}$. If no state has probability above the threshold, the result is not yet a valid state. This in normal

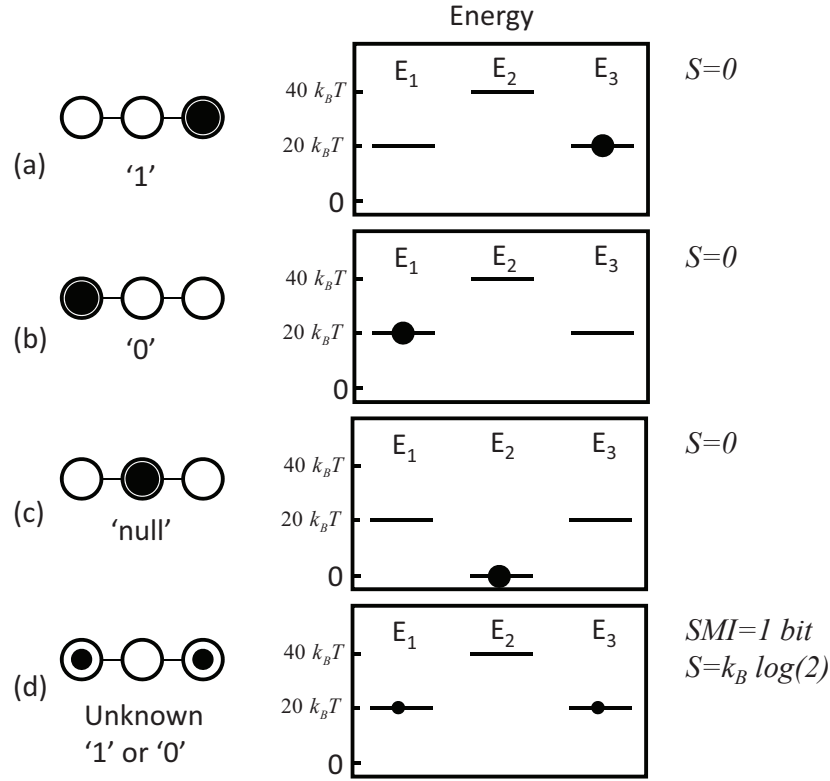


Fig. 8 Encoding information with three physical states. A particle can be in one of three dots encoding (a) a logical 1, (b) a logical 0, or (c) a logical null state. The energy landscape and probability density for each configuration are shown on the right. The solid circles represent the probability of finding the particle on a particular dot. Dot 2 acts as the barrier for holding the particle in the 1 or 0 state in (a) and (b). It acts as the low energy null state in (c). If the bit value is an unknown 1 or 0, as in (d), then the energy landscape is the same as in (a) or (b), but we must assign a probability distribution that is evenly divided between the 1 and 0 states.

in the switching regime. A robust system is designed so the logical state is valid and unambiguous when it is read or copied.

The dots can represent abstract states of the system or literal dots. In the quantum-dot cellular automata (QCA) scheme, they correspond to localized electron states on literal quantum dots. We will treat the system completely classically in this section, and will for convenience assume a single positive charge; the quantum treatment is taken up in Sec. 6. Actual QCA three-dot cells have been fabricated in metal-dot systems and synthesized in single-molecules. [9, 10] The threshold probability P_{th} in these systems is not set simply arbitrarily, but only needs to be large enough for

the next stage in logical operations to reset the bit strongly to a logical 1 or 0. Power gain from stage to stage means that P_{th} could be 0.8, for example, and still be strong enough to be effective as transmittable bit information.

We are interested in the process of bit erasure in this system when it is in thermal contact with a heat bath of temperature T . We can control the energy of each state with a set of control voltages capacitively coupled to each dot. The energetic landscape is shown in (a-c) of (8). We choose a $20k_B T$ energy separation between low, active (0 or 1) and high energy state. When state 2 is high, it acts as a barrier to hold the 1 or 0 bit. When state 2 is low it acts as a well to hold the particle in a neutral position. Thus the energy E_2 acts a clock which can latch a bit by being raised, or erase a bit by being lowered, returning it to the null state.

In the following, we will for convenience simply refer to the fully localized states: $P = [0, 0, 1]$ for the 1 state, $P = [1, 0, 0]$ for the 0 state, and $P = [0, 1, 0]$ for the null state. Examining Fig. 8, reveals an important point:

A physical memory containing information cannot be in a thermal equilibrium state.

The null state is an equilibrium state satisfying the Boltzmann distribution (63), but neither the 1 nor the 0 state shown in Fig. 8a,b can be. The reason is clear enough—to be a memory is to hold information stored at a previous time. The state of a physical memory *must* depend on the past and not just the current temperature and applied voltages, and whatever macro constraints are relevant. A thermal equilibrium state, by contrast, *cannot* depend on the past, but only on the present conditions.

A physical memory must be in a long-lived metastable state. When E_2 is high it must create a barrier that is sufficiently opaque to hold the particle in the 1 or 3 states for the relevant timescale—microseconds to years. Beyond that, the details of the physical dynamics that allow state transitions between states 1, 2, and 3, do not concern us. The 1 and 0 states of Fig. 8a,b are certainly low energy states—the problem is they preferentially occupy one active dot and not the other though it has the same energy. We will assume in our example that the $20k_B T$ barrier of E_2 for these states is indeed adequate to hold a 1 or 0 bit long enough to be a memory. If a higher barrier was needed, it could be created by raising the potential on dot 2 further.

Entropy for a representing a known or unknown bit. For the 1 or 0 states shown in Fig. 8a,b, we must take care to return to our derivation of the thermodynamic entropy S in Sec. 4.1 and now include the fact that we know the value of the bit. This should be treated as an additional constraint in the Jaynes maximum entropy principle as developed in Sec. 3.7.

Consider the case when we know the bit stored is a binary 0. We again require that the thermodynamic entropy be $k_B \log(2)$ times the SMI of that probability distribution which maximizes the Shannon entropy, subject to the given constraints. The constraints are now (a) the probabilities sum to 1, (b) the average energy $\langle E \rangle = U$,

and (c) $P_3 = 0$. To find the maximum entropy distribution, we construct the Lagrangian, which now has the additional Lagrange multiplier λ_3 .

$$\mathcal{L} = -\sum_k P_k \log(P_k) - (\lambda_0 - 1) \left(\sum_k P_k - 1 \right) - \lambda_1 \left(\sum_k P_k E_k - \langle E \rangle \right) - \lambda_3 P_3 \quad (102)$$

The Lagrange equation obtained from requiring the extremum with respect to λ_3 , $\partial \mathcal{L} / \partial \lambda_3 = 0$, yields simply $P_3 = 0$ with λ_3 arbitrary. We find the extremum of (102) with respect to all the other P_k 's as before. The derivations of all thermodynamic quantities derived in Sec. 4.1 go through as before, simply omitting occupancy of dot 3 as a possibility. The Boltzmann distribution then applies as before to all probabilities but P_3 . The thermodynamic entropy is then: $S = k_B \log(2) \text{SMI}([P_1, P_2])$. If the bit was a 1, the constraint would be $P_1 = 0$. In this straightforward way we can apply the definition of thermodynamic entropy to include a state storing a known bit, even though it represents a nonequilibrium metastable state—as it must.

By contrast, the state shown in Fig. 8d represents a reliably stored but *unknown* bit. Since we do not know the value we must assign probabilities $P = [0.5, 0, 0.5]$, the probabilities we would get for the equilibrium state with the barrier high (hence $P_2 = 0$). The associated $\text{SMI} = 1$ bit (one bit of missing information) and the thermodynamic entropy $S = k_B \log(2)$. In terms of thermodynamic quantities, it is indistinguishable from an equilibrium state. But because the barrier is sufficient to hold the unknown, but definite, bit for the relevant timescale, it should not be imagined to be switching back and forth. It is not a “bit gas”, as Norton has characterized it in [2], but is simply unknown.

The null state is an equilibrium state with entropy $S = k_B \log(2) \text{SMI}[P] = 0$.

Thermodynamic quantities during bit operations. We examine below three basic operations: writing a bit, erasing an unknown bit, and erasing a known bit. In each case we will manipulate the three dot energies in time, $E_1(t), E_2(t), E_3(t)$, according to a protocol designed to accomplish the task. Unless otherwise noted, at each point in time, we assume the system is in thermal equilibrium with a bath at temperature T . We assume that the variation in time is at gradual enough that the system is always in its equilibrium state, except as noted for a stored known bit. Therefore, temporal dynamics play no essential role here and we use arbitrary units spanning the event with $t = [0, 1]$. During the operation we calculate the following four thermodynamic quantities and plot them as functions of time:

1. The equilibrium probabilities $P = [P_1(t), P_2(t), P_3(t)]$ for finding the particle on each dot.

$$P_i(t) = \frac{e^{-E_i(t)/k_B T}}{\sum_k e^{-E_k(t)/k_B T}} \quad (103)$$

The corresponding expectation value of the charge on the dot is $q_i(t) = P_i(t)e$, where e is the elementary charge. As discussed above, for a state representing a known bit, we simply set P_1 or P_3 to zero.

2. The thermodynamic entropy is given by:

$$S(t) = -k_B \log(2) \sum_k P_k(t) \log(P_k). \quad (104)$$

3. The cumulative amount of heat transferred ΔQ_{bath} to the thermal bath from the beginning of the operation until the time t .

$$\Delta Q(t) = - \int_0^t dQ = - \int_0^t T dS = T(S(0) - S(t)) \quad (105)$$

The sign is chosen so that net heat flowing from the system to the bath is positive, and net heat flowing from the bath into the system is negative.

4. The work done on the system by the external control electrodes. The electrical potential of each dot i is $V_i(t) = E_i(t)/e$, and the differential work done by the external circuit is

$$dW_i(it) = -V_i(t)dq_i(t). \quad (106)$$

The minus sign is because when the external circuit raises the dot energy, it actually decreases the dot charge because the thermal occupancy of the dot is reduced via (103).

$$W(t) = \sum_i \int_0^t dW_i(t') = - \sum_i \int_0^t V_i(t') \frac{dq_i(t')}{dt'} dt' \quad (107)$$

This does not include the work done by the voltage sources on the gate electrodes that are capacitively coupled to the dots. That is most of the work that the external circuit does, pushing charge and off the gate capacitors. This motion is dissipationless if we neglect the residual resistance of conductors; gradual charging and discharging a capacitor can be done quasi-adiabatically. In any case, dissipation in the gating circuit is not what we are interested in here. Therefore, the only contributions to W are when charge flows on or off one of the dots in the system itself.

The bit operation is determined by the temperature T and the control protocol defined by $[E_1(t), E_2(t), E_3(t)]$. The calculations using (103)-(107) give us in $P(t), S(t), \Delta Q(t)$ and $W(t)$.

Writing a bit. Starting from the null state we write a 0 bit using the protocol shown in Fig. 9. The potential energy of dot 3 is raised first. Then the energy of dot 2, which holds the particle initially, is ramped up smoothly. As E_2 crosses E_1 , Fig. 9c, the particle transfers to dot 1. When E_2 reaches the high energy state, E_3 can be lowered and the particle is held in the state representing a bit value

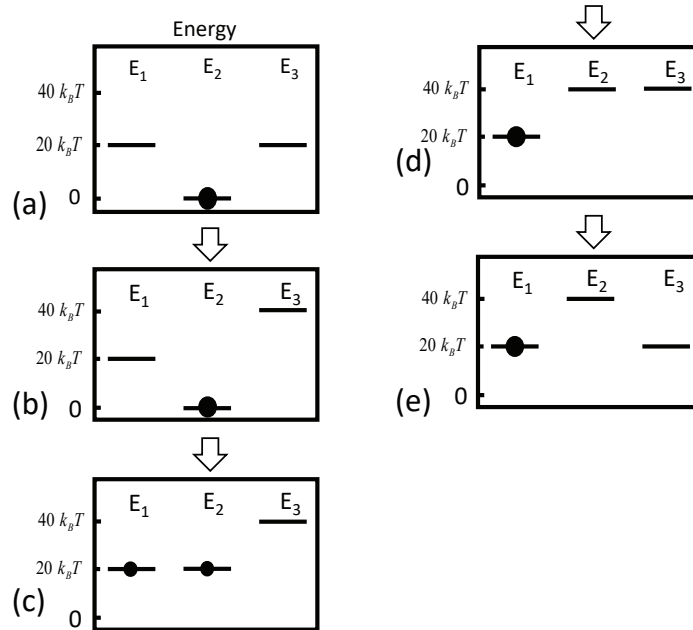


Fig. 9 Protocol for writing a 0 bit. (a) The initial state is null, with the particle on dot 2. (b) Energy E_3 is raised, biasing the system toward the 0 state. (c) The null state energy E_2 is smoothly ramped up, passing the energy E_1 . At this point the probability of occupying dots 1 or 2 are equal. (d) The energy E_2 is now high and the particle is localized in dot 1, representing a 0 bit. (e) E_3 is lowered again and the particle is held on dot 1 by the barrier. This is a memory storing a 0 bit in a long-lived metastable state. A memory-storing state cannot be an equilibrium state; it must depend on the past.

of 0, as shown in Fig. 9d,e. This is, as discussed above, a long-lived metastable state in which E_2 acts as a barrier trapping the particle in dot 1 forming a one-bit memory.

Figure 10a shows the energies $E_1(t)$, $E_2(t)$, and $E_3(t)$. The probabilities of occupancy of each state are shown in Fig. 10b. At each time $0 < t < 0.9$ the probabilities are thermal equilibrium values given by (103). For $0.9 < t < 1$, the system is in the nonequilibrium metastable state where $P_3 = 0$ by assumption. Figure 10c shows the thermodynamic entropy $S(t)$ in units of $k_B \log(2)$ (equivalent to the SMI). As the levels E_1 and E_2 cross, the entropy increases because there is less information about which dot the particle is on. At the crossing point the missing information is 1 bit. What we do not know is the details of the momentary thermal fluctuations which have put the system in state 1 or state 2. Figure 10d

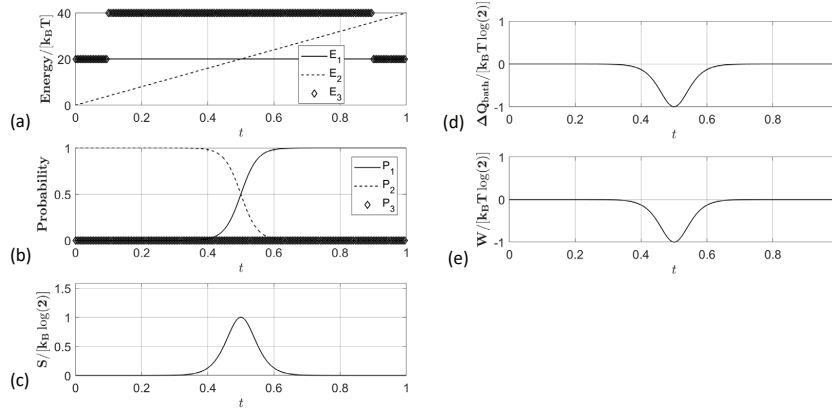


Fig. 10 Thermodynamic quantities during the process of writing a 0 bit using the protocol of Fig. 9. Time is in arbitrary units from 0 to 1. The system is in thermal equilibrium from $t = 0$ to $t = 0.9$, after which it is in a metastable memory state storing a 0 bit. (a) The dot energies E_1, E_2, E_3 are shown as functions of time. At $t = 0.1$, E_3 is raised to bias the system toward the 0 state. The energy of dot 2 is ramped up throughout the process, eventually forming a high barrier to hold the stored bit information. At $t = 0.9$ the bias is removed and the bit is firmly latched by $t = 1$. (b) The probabilities for dot occupancy P_1, P_2, P_3 are shown as functions of time. (c) The thermodynamic entropy of the system calculated from equation (104). The peak occurs as E_2 nears and then passes E_1 . Thermal excitations between dots 1 and 2 make the location of the particle 1 bit less certain. The peak corresponds to the moment shown in Fig. 9c when $E_2 = E_1$. (d) The net heat transferred to the bath ΔQ_{bath} up to time t , calculated from equation (105). As E_2 approaches E_1 , heat is drawn from the bath; as E_2 moves above E_1 , the heat energy is returned to the bath. (e) The net work done by the control circuit on the system W , calculated from equation (107). When the write process is complete, no net work has been done. This, of course, neglects any heat dissipated within the control circuit that changes the dot energies (e.g., due to nonzero resistance of conductors).

shows the heat transferred to the environments $\Delta Q(t)/(k_B T \log(2))$ calculated from (105). As the crossing point is approached the system takes energy from the environment to excite thermal occupancy in the higher energy dot (here dot 1). That energy (heat) is returned to the environment as E_2 continues to increase. Figure 10e shows the work done on the system from 0 to t by the external circuit $W(t)/(k_B T \log(2))$ calculated from (107). This is initially negative because energy is being drawn in from the thermal bath, but nets to zero as the energy is returned.

Erasing an unknown bit. Figure 11a shows the energy states for a stored bit. Suppose that this is an unknown bit—there is no other physical copy of the bit that we can use to bias the system into the state its already in. Therefore, all we

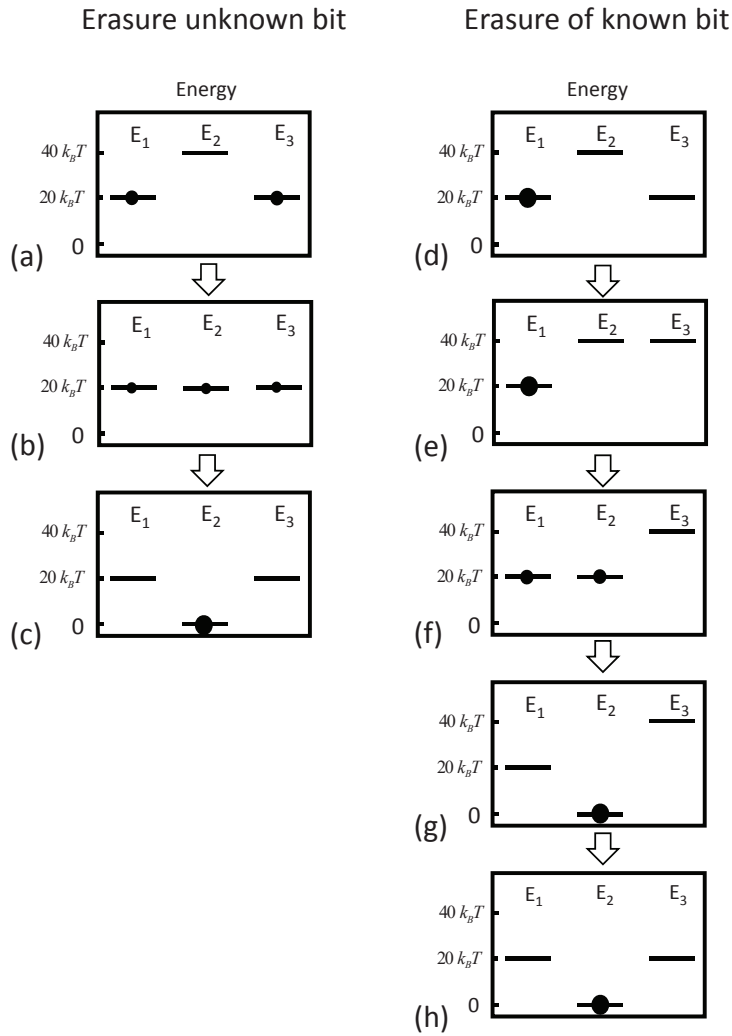


Fig. 11 Erasure protocols for known and unknown bits. Diagrams (a)-(c) represent the case of an unknown bit and diagrams (d)-(h) represent the case of a known bit. (a) If the bit value stored is unknown, the probability of dot occupancy of dots 1 and 3 are equal. Since we do not know in which direction to bias the system, we can only lower E_2 smoothly in time. (b) As E_2 is lowered it passes E_1 and E_3 . When they are degenerate the particle could be in any of the three dots. (c) Finally the middle state has captured the occupancy and the system is in the null state. (d,e) For the known bit, here shown to be 0, the system is biased into the state it is already in by raising E_3 . (f) As E_2 is lowered it passes E_1 . (g) The system is now in the low energy neutral state. (h) The bias can then be removed.

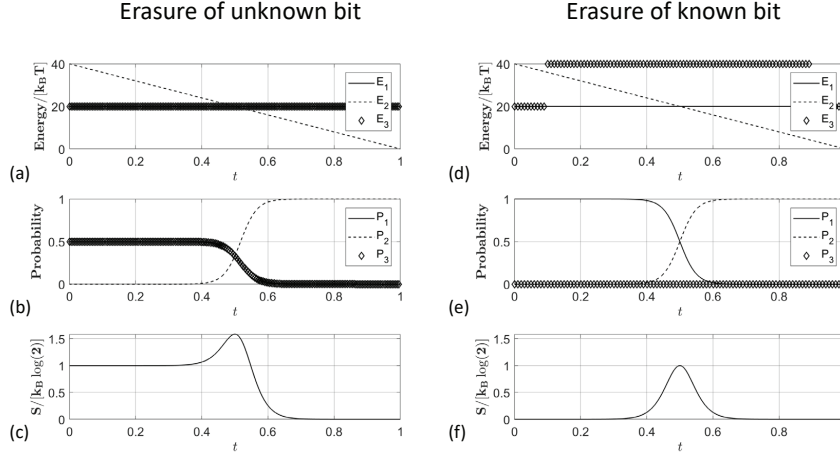


Fig. 12 Energy levels, probabilities, and entropy during bit erasure. Plots (a)-(c) are for the case of an unknown bit and (d)-(f) are for a known bit. (a) For an unknown bit, E_2 is simply lowered to cause the transition to the null state. (b) The probabilities for occupancy of dots 1, 2, and 3 are shown. Initially $P_1 = P_3$ because the bit value is unknown. (c) The entropy of the system is initially 1 bit. At the point where all three energy levels are degenerate $S/(k_B \log(2)) = \log_2(3) \approx 1.58$. The entropy then decreases as the system localizes completely on dot 2. The net decrease in entropy is 1 bit. (d) For a known bit, the system can be biased into the state it is in, by raising E_3 in this case, until the switching is complete. (e) The probabilities for occupancy of dots 1, 2, and 3 are shown for the case of a bit known to be initially 0. Initially only P_1 is nonzero because the system is known to be in the 0 state. (f) The entropy for the case of a known bit increases around the level crossing of E_1 and E_2 , but is initially and finally 0.

can do to erase the bit, i.e., set it to the null state, is to lower the barrier E_2 until it localizes the charge in the middle dot. We lower it smoothly, crossing the active-state energies, Fig. 11b, and arrive at the null state shown in Fig. 11c. Again, the system is assumed to be always in thermal equilibrium with the heat bath.

Figure 12a shows the dot energies as a function of time and Fig. 12b shows the probabilities. Initially both P_1 and P_3 are $1/2$ and finally $P_2 = 1$. The entropy shown in Fig. 12c is initially one bit, $k_B \log(2)$. The entropy increases as E_2 passes E_1 and E_3 . At the crossing point the SMI is equal to $\log_2(3)$ because each of the three states is equally probable. The entropy drops to zero as the particle becomes completely localized in dot 2. The erasure of an unknown bit involves a lowering of the system entropy by 1 bit. Fig. 13a shows the heat transfer to the environment calculated from (105) for this erasure process. It is initially zero and then becomes negative as E_2 is lowered and thermal excitation increases the probability of dot 2 being occupied. The heat transfer swings positive as E_2 drops

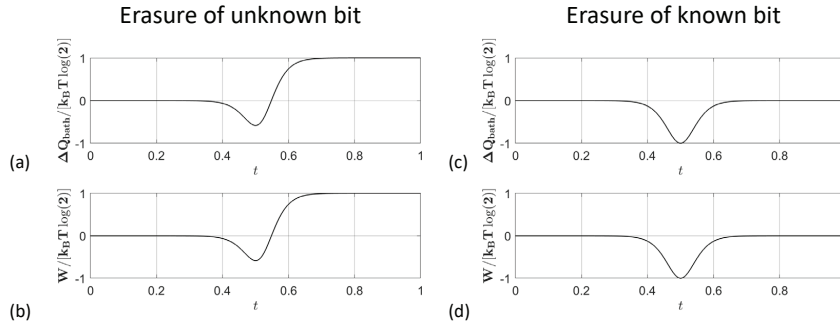


Fig. 13 Heat transfer and work during bit erasure. For the same two protocols of switching as in Figs. 11 and 12, two calculated thermodynamic quantities are shown. Plots (a) and (b) are for the case of erasure of an unknown bit, and plots (c) and (d) are for erasure of a known bit. (a) For erasure of an unknown bit, the plot shows the net heat transferred to the bath ΔQ_{bath} up to time t , calculated from equation (105). As E_2 is lowered, heat is drawn *from* the bath as thermal fluctuations excite the system from E_1 or E_3 to E_2 . As E_2 moves below the level of E_1 and E_2 , heat energy flows out *to* the bath as the system de-excites and occupies E_2 . The net heat transferred to the environment (i.e., dissipated) in this case is $k_B T \log(2)$. (b) The net work done by the control circuit on the system W , calculated from equation (107), for the case of unknown bit erasure. This is the source of the energy which ends up dissipated as heat to the thermal bath. (c) In the case of a known bit, switched according to Fig. 11(d)-(h), heat drawn in from the bath as E_2 approaches E_1 is returned to the bath as the system moves into the null state. The knowledge of the existing bit state affects the amount of heat dissipated to the bath precisely because it permits a different erasure protocol to be operated. (d) Similarly, the net work done by the control circuit on the system W , calculated from equation (107), is zero by end of the switching event.

below the crossing point. The final net value of the heat transferred to the environment is the Landauer Principle limit of $k_B T \log(2)$. Where did this energy come from? It came from the external circuit as shown in Fig. 13b, calculated from (107). The net work done by the circuit during the erasure process is precisely $k_B T \log(2)$.

Erasing a known bit. Now consider the situation of erasing a known bit. In this case we can execute a different set of operations shown in Fig. 11d-h. Figure 11d shows the dot energies when the system holds a known 0 bit (in a metastable rather than equilibrium state). Because it's known to be a 0, we can raise the energy of the other state by increasing the potential energy of dot 3 as shown in

Fig. 11d. Then the energy of dot 2 can be lowered as in Fig. 11f,g. Finally, the energy of dot 3 is lowered to restore the whole configuration to the null state.

The energy levels, probabilities, and entropy for this process are shown in Fig. 12d-f. During the initial time $0 < t < 0.1$, corresponding to the configuration of Fig. 12d, the probabilities are not the equilibrium values but rather correspond to the non-equilibrium state of the particle known to be in dot 1. Thereafter thermal equilibrium is assumed at all times. The difference between the initial probabilities in Figures 12b and 12e is precisely the difference between knowing the value of the bit stored and not knowing it. The entropy in known-bit erasure, Fig. 12f, rises as E_2 approaches E_1 , but then falls to zero again as the particle is confined on dot 2. The reversible heat transfer shown in Fig. 13a is similar; the system draws in heat from the bath as E_2 comes close enough for thermal excitations from dot 1 to dot 2. But as the system localizes by several $k_B T$, the heat is returned to the bath. The work done on the system, calculated from (107) and shown in Fig. 13a, nets to zero by the end of the switching. This is entirely consistent with the experimental result that erasing a known bit can dissipate orders of magnitude less than $k_B T$. [11]

The issue of a “known” versus “unknown” bit of course has nothing to do with consciousness. “Known” simply means there is another physical representation of the bit that can be used to bias the system toward the state its already in, as in Fig. 9e. This could be accomplished by a circuit, a computer, or a human brain. In QCA shift registers, the bias is accomplished by having the neighboring bit holding a copy of the target bit. [12] The neighbor is Coulombically coupled to the target bit so it provides an electrostatic bias. Needless to say, one could have a copy of the bit but *not* use it to employ the optimum erasure protocol. In that case one would dissipate at least $k_B T \log(2)$ amount of heat for each erasure.

Critics sometimes ask: How can “subjective” knowledge of the bit state change a physical quantity like heat dissipated? The answer is simply because knowing the existing state allows us to use a different protocol for the erasure, one which initially biases the system into the state it is already in. In the case illustrated in Figures 9–13, we initially raise E_3 to the high level to erase a known 0 bit, but would initially raise E_1 to a high level to erase a known 1 bit.

The free energy $F = U - TS$ is similarly higher for a known state, because S is lower, than for an unknown state with higher S . The Szilard engine is an example of exploiting knowledge of a system state (in that case a single molecule gas) to draw more energy out of the system. [13] A recent physical realization of this sort of “information engine” was reported in [14].⁹

Figure 13a is a clear demonstration of the Landauer Principle in the simplest single-particle system. It requires only equations (103)-(105) above, that is, the Boltzmann thermal occupancy of energy states, the careful definition of entropy

⁹ This sort of engine does not, needless to say, violate the second law of thermodynamics or create a perpetual motion machine of the second kind.

for memory states, and the calculation of heat flow. The dissipation of $k_B T \log(2)$ as heat to the bath is unavoidable.

The erasure of an unknown bit above is at each point in time an equilibrium process. It is not, however, time-reversible. Going backwards in the sequence shown in Fig. 11c,b,a would result in latching a *random* bit, determined by a momentary bath fluctuation, not the unknown but definite bit one started with. Imagine the original unknown bit was the result of a long calculation. It is unknown to us until we read it (and thereby make a copy), but its definite value has genuine information about the calculation's result. The time-reversed protocol would not restore the result, but replace it with a random bit. The probability distribution for an unknown bit holding the important result of a long calculation (e.g., target is friend or foe) is identical with the probability distribution of a meaningless random bit latched from a thermal fluctuation.¹⁰

By contrast the erasure protocol of Fig. 11d-h is time reversible. It is in fact just the time-reversed version of the writing protocol shown in Fig. 9. Both assume we have another copy of the bit to determine bias applied in the writing sequence (do we raise the potential on dot 1 or dot 3?) and the subsequent erasure.

The main result of this calculation is the difference between the heat dissipated to the environment for erasure of an unknown bit, shown in Fig. 13a, and that for erasure of a known bit, shown in Fig. 13c. The Landauer Principle result is clear—heat energy of $k_B T \log(2)$ is dissipated for erasure of an unknown bit, but there is no minimum dissipation required for a known bit. If the bit is known, of course, there is a copy of the bit somewhere and the information is not truly lost. If the copy is eventually erased, heat dissipation must occur.

6 Quantum mechanics, entropy, and the nature of the physical law

6.1 Quantum formalism and probabilities

In quantum mechanics a physical system is described by a state vector $|\psi\rangle$, a so-called Dirac “ket” in a Hilbert space.¹¹ The state vector is a *complete description* of the system. It contains all there is to know about the system—all the physical world itself “knows” about the system.

¹⁰ “Meaningless” is, of course, a question of context—maybe it was meant as a probe to measure bath fluctuations.

¹¹ A Hilbert space is a complex linear vector space with an inner product that produces a norm on the space. Using this norm, all Cauchy sequences of vectors in the space converge to a vector in the space.

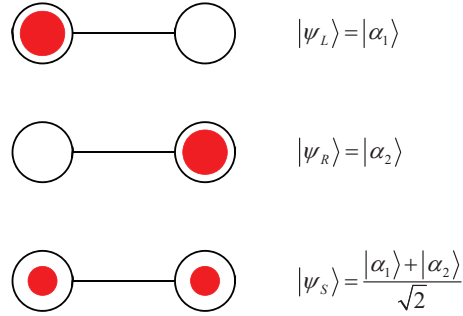


Fig. 14 Quantum state of a particle in two coupled quantum dots. The particle could be in a state that is fully localized on the left dot, $|\psi_L\rangle$, fully localized on the right dot, $|\psi_R\rangle$, or a symmetric superposition of the two $|\psi_S\rangle$. (Any normalized linear combination of $|\psi_L\rangle$ and $|\psi_R\rangle$ is possible.) If the particle is in the superposition state $|\psi_S\rangle$, the result of a measurement of position is not knowable in advance of the measurement. The indeterminacy is not a reflection of the experimenter's ignorance, but is a fundamental feature of the physical world.

The inner product between two state vectors is denoted $\langle\phi|\psi\rangle$ and gives the probability amplitude that a system in state $|\psi\rangle$ will be measured to be in state $|\phi\rangle$. The probability is the absolute square of the probability amplitude.

$$p_\phi = |\langle\phi|\psi\rangle|^2 \quad (108)$$

Equation (108) is the Born rule.

Dynamical observables, such as position x and momentum p , are represented in quantum mechanics by Hermitian operators that map one ket onto another ket. Measurements of an observable Q always and only yield eigenvalues q_k of the associated operator \hat{Q} .

$$\hat{Q}|\phi_{q_k}\rangle = q_k|\phi_{q_k}\rangle \quad (109)$$

The eigenvalue spectrum of an operator may be discrete or continuous. The eigenstates $|\phi_{q_k}\rangle$ are states which have a definite value of the property Q . The probability that a measurement of Q on the system in state $|\psi\rangle$ will yield a specific eigenvalue q_k is therefore:

$$p_{q_k} = |\langle\phi_{q_k}|\psi\rangle|^2 \quad (110)$$

This probability is different in kind from the probabilities we have dealt with heretofore. Here the probabilistic nature is not because we lack any information that we could otherwise have. The probabilities here reflect the fact that the physical law and the current state of the system taken together are fundamentally insufficient to determine the result of the measurement. This indeterminacy is a feature of the physical law revealed by quantum mechanics.

6.2 Quantum mechanical SMI for an observable

We can define the SMI (Shannon Measure of Information) of the probability distribution for measurements of eigenvalues q_k when the system is in the state ψ .

$$SMI_Q[\psi] \equiv - \sum_k p_k \log_2(p_k) \quad (111)$$

or

$$SMI_Q[\psi] = - \sum_k |\langle \phi_{q_k} | \psi \rangle|^2 \log_2 \left(|\langle \phi_{q_k} | \psi \rangle|^2 \right) \quad (112)$$

This represents the amount of missing information (in bits) about the outcome of measurements of Q . This information is missing from the quantum state $|\psi\rangle$ itself. It is not simply missing because of our incomplete knowledge. The physical world itself does not have this information present until measurement is actually made and one particular eigenvalue of \hat{Q} is obtained.

Consider the two-state system illustrated in Fig. 14. A particle can be in the dot on the left, a state represented by $|\psi_L\rangle$, or in the dot on the right, represented by $|\psi_R\rangle$. The symmetric superposition of the two is the state represented by the state

$$|\psi_S\rangle = (|\psi_L\rangle + |\psi_R\rangle)/\sqrt{2}. \quad (113)$$

This superposition state is a unique feature of quantum mechanics. A measurement of the position of the particle in state (113) will always find it in either the left dot or the right dot, each with probability 1/2. We can define an operator corresponding to the observable position of the particle.

$$\hat{X} \equiv |\psi_R\rangle \langle \psi_R| - |\psi_L\rangle \langle \psi_L| \quad (114)$$

The eigenvalues of \hat{X} are $+1$ and -1 corresponding to the particle on the right or on the left. The SMI for this operator corresponds to the amount of missing information about the position of the particle when its in each of these states.

$$\begin{aligned} SMI_X[\psi_L] &= 0 \\ SMI_X[\psi_R] &= 0 \\ SMI_X[\psi_S] &= 1 \text{ bit} \end{aligned} \quad (115)$$

In the first two cases, the position is definite and there is no information about it that is missing. For the symmetric superposition state, knowing the state tells us nothing about the position. There is 1 bit of position information missing, even though the quantum mechanical description of the state is complete.

Note the the SMI depends on the choice of observable as well as the state itself. For the same state $|\psi_S\rangle$, we could consider the parity operator:

$$\hat{\Pi} \equiv |\psi_L\rangle \langle \psi_R| + |\psi_R\rangle \langle \psi_L|. \quad (116)$$

The state $|\psi_S\rangle$ is an eigenstate of Π with eigenvalue 1—it has a definite value of parity. So there is no missing parity information and SMI_{Π} is 0. The SMI is basis dependent; it is not the eigenvalue of a Hermitian operator.

We return to considering the information the quantum state gives us about the particles position. It is helpful to ask the question in terms of the amount of information provided by the physical law. Let us suppose that, given an initial state of the system and a precise description of all the potentials, fields, and interactions that are subsequently applied to the particle, the physical law prescribes that at a particular time the system will be in the state $|\psi_L\rangle$. We may then ask: *How much information about the particle's position (by which we mean the results of a measurement of position) does the physical law yield?* The answer is 1 bit. For the state $|\psi_R\rangle$ the physical law also yields 1 bit of information, completely determining where the particle will be found (on the right). But if the physical law tells us that the particle is in the state $|\psi_S\rangle$, it gives us 0 bits of information about the position that will be measured. The information provided by the physical law (evaluating Eq. (9)) is zero. In general, if there are N eigenvalues of the operator \hat{Q} , then the physical law gives us a finite amount of information I about the outcome of a measurement of Q , where

$$I[\psi] = \log_2(N) - SMI_Q[\psi] \text{ bits.} \quad (117)$$

Continuous eigenvalues. Consider a wavefunction defined on a *continuous* range of positions. Let $\psi(x)$ be the probability amplitude for finding a particle at position $x \in [0, L]$. If the state for which the particle is exactly found at x is denoted $|x\rangle$, then $\psi(x) = \langle x | \psi \rangle$ and the probability density is $P(x) = |\psi(x)|^2$. Consider the wavefunction and probability distribution shown in Fig. 15 describing a particle in this interval. We can use the expression for information given by Equation (13) to calculate the amount of position information which we obtain from the wavefunction. In this case that is 1.32 bits. It is somewhat more constrained than if it was localized over half the distance, which would correspond to 1 bit of information gain. The wavefunction gives us *some* information about the position, but not complete information. If our detection apparatus gave us discretized information that the particle's position was in a particular bin of width Δx , we could use Equation (14) in a similar way. Note that we are implicitly assuming, as with any probability distribution, that there is an accessible region (AR) (see the discussion in Section 2) that we know the particle is in; here that is the interval $[0, L]$

Given the quantum state $\psi(x)$ we could as well ask about the results of measurements of the particle momentum p and use the same formalism to calculate the information about momentum we receive from knowing the wavefunction, I_p . This value is not the same as the information about position, and again depends on the range of momentum values considered to be the AR.

Time dependence. The operator \hat{H} is the Hamiltonian operator representing the total energy of the system. For an isolated system described by a time-independent

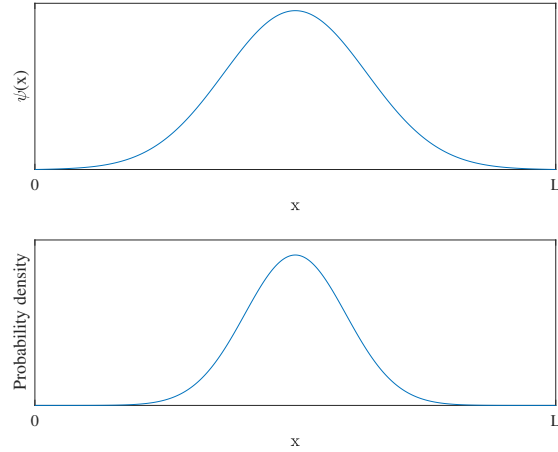


Fig. 15 Quantum wavefunction and probability density defined for a particle with a position on the continuous interval $x \in [0, L]$. The top figure shows a wavefunction $\psi(x)$ and the bottom figure the associated probability density. Given a probability density $P(x)$, one can ask how much information about the *position* of the particle does the wavefunction provide. In this case the result, from equation (13), is 1.32 bits. That is the Shannon measure of the information (SMI) gained by knowing the probability distribution compared to a uniform distribution over the accessible region, which is here the interval $[0, L]$.

Hamiltonian H , the time development of the state vector is determined by the Schrödinger equation. If the state of the system at $t = 0$ is given by $|\psi(0)\rangle$, then at any future time we can determine the state by solving the differential equation

$$\boxed{i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = \hat{H} |\psi(t)\rangle} \quad (118)$$

or by evaluating the integral form

$$|\psi(t)\rangle = e^{-i\frac{\hat{H}}{\hbar}t} |\psi(0)\rangle. \quad (119)$$

Although the time evolution of the quantum state is deterministic, the results of measurements of the state are not, and measurements happen all the time. We lack a good account of how to describe precisely what circumstances create a measurement—the “measurement problem.” But it is clear that measurements do occur, transcript of the history of the physical world is not determined by just the initial state and the physical law, but in addition by a vast number of events that could have gone one way or the other, with prescribed probabilities. One particular outcome actually occurred and was, so to speak, written into the transcript of history.

The physical law, in this case quantum mechanics, yields probabilities for possible outcomes of measurements and the Shannon measure gives us a concrete way of quantifying how much information about those outcomes the physical law provides. In the clockwork universe of Laplace's demon, the physical law provided certainty about outcomes, given a complete description of the physical state. It turns out that the physical law simply does not do that. The physical law constrains, but does not completely constrain, the outcome of measurements. This is a profound fact about the nature of the physical law. Nor is this just a feature of the present state of quantum theory. Recent Bell test experiments confirm to astonishing accuracy that this is a feature of the physical world quite independent of quantum mechanics. [15–18] Any future successor theory would have to contain this feature—the future of the physical world is not completely constrained by the physical law. It retains some “freedom.”

6.3 *Open quantum systems and density operators*

Pure isolated quantum states of a physical system are described by a state vector $|\psi\rangle$. Often we are dealing with a system A that is not isolated but interacting with a very large system B, which could be a thermal bath or simply the rest of the universe. We can then no longer describe the system with a state vector but must employ the formalism of the density matrix. The density matrix folds in two kinds of probability: that due to fundamental quantum indeterminacy and that due to our practical ignorance of the details of the state of a large system. It is helpful to derive it here so we can see exactly where this quantum probability and classical probability are brought together.

The starting point is writing the quantum state for the global system. We can write the state of the A system as a linear combination of basis states $|\alpha_i\rangle$. The basis states for the large system B are $|\beta_m\rangle$. The state describing the combined global system of A and B can then be written

$$|\psi\rangle = \sum_{i,m} C_{im} |\alpha_i; \beta_m\rangle \quad (120)$$

where

$$\langle \alpha_i; \beta_m | \alpha_j; \beta_n \rangle = \delta_{ij} \delta_{mn}. \quad (121)$$

The sum over m here is over a very large number of possible states for the bath (or universe). We define the global density operator

$$\hat{\rho} = |\psi\rangle\langle\psi| \quad (122)$$

$$= \sum_{i,j,m,n} \underbrace{C_{im}C_{jn}^*}_{\rho_{im;jn}} |\alpha_i; \beta_m\rangle\langle\alpha_j; \beta_n| \quad (123)$$

$$= \sum_{i,j,m,n} |\alpha_i; \beta_m\rangle \rho_{im;jn} \langle\alpha_j; \beta_n| \quad (124)$$

We now focus on the target system A. Any operator \hat{Q}^A which acts only on the A subsystem can be written

$$\hat{Q}^A = \sum_{i,j,m} |\alpha_i; \beta_m\rangle Q_{ij} \langle\alpha_j; \beta_m|. \quad (125)$$

The expectation value of Q we can write

$$\langle Q^A \rangle \equiv \langle \psi | Q^A | \psi \rangle = \sum_{\substack{i,j,k,\ell \\ m,n,p,q}} C_{kp}^* C_{\ell q} \langle \alpha_k; \beta_p | \alpha_i; \beta_m \rangle Q_{ij}^A \langle \alpha_j; \beta_m | \alpha_\ell; \beta_q \rangle. \quad (126)$$

We do the sums using (121) and exchange the symbols i and j to obtain

$$\begin{aligned} \langle Q^A \rangle &= \sum_{i,j,m} C_{im} C_{jm}^* Q_{ji}^A \\ &= \sum_{i,j} \underbrace{\left(\sum_m C_{im} C_{jm}^* \right)}_{\equiv \rho_{ij}^A} Q_{ji}^A \end{aligned} \quad (127)$$

$$= \sum_{i,j} \rho_{ij}^A Q_{ji}^A \quad (128)$$

Therefore we can write

$$\boxed{\langle Q^A \rangle = \text{Tr}(\hat{\rho}^A \hat{Q}^A)} \quad (129)$$

We have defined the reduced density operator ρ_{ij}^A for system A as the sum over the very large number of basis states for the environment. Our practical ignorance about the details of the large system B and its interaction with A are all hidden in this sum over the $C_{im}C_{jm}^*$ terms. Equation (129) defines the operator $\hat{\rho}^A$ as the operator on the A system which has matrix elements ρ_{ij}^A in the $|\alpha_i\rangle$ basis.

Note that comparing Equation (123) and (128) we have

$$\rho_{ij}^A = \sum_m \rho_{im;jm}, \quad (130)$$

which we write compactly as a partial trace of the global density matrix over the B degrees of freedom represented by the β_m states.

$$\hat{\rho}^A = \text{Tr}_B(\hat{\rho}). \quad (131)$$

Several properties of the density matrix can be stated briefly. It can be easily shown that the density matrix is Hermitian and has unit trace.

$$\hat{\rho}^\dagger = \hat{\rho} \quad (132)$$

$$\text{Tr}(\hat{\rho}) = 1 \quad (133)$$

If ρ describes a state of a system, the probability of measuring the system and finding it to be in the state ϕ is given by the expectation value of $|\phi\rangle\langle\phi|$, the projection operator for ϕ .

$$p_\phi = \text{Tr}(|\phi\rangle\langle\phi| \hat{\rho}) \quad (134)$$

The diagonal elements of the density matrix $\rho_{i,i}$ are the probabilities of the system being found in the basis state $|\alpha_i\rangle$.

Because $\hat{\rho}$ is Hermitian it can be diagonalized by a unitary transformation. Denote the eigenvalues of $\hat{\rho}$ as ρ_v and the eigenvectors $|v\rangle$. If the system is in a “pure state” that could be described by a single state vector, then all the eigenvalues will be zero except for one. If not, the state is described as “mixed.” In that case the eigenvalues ρ_v are the probability of the system being found in the state $|v\rangle$.

The von Neumann Entropy. Von Neumann defined the quantum entropy S_{vN} to be a measure of this “mixedness.”

$$\boxed{S_{vN}(\hat{\rho}) \equiv -\text{Tr}(\hat{\rho} \log(\hat{\rho}))} \quad (135)$$

The von Neumann entropy is equivalent to the SMI of the eigenvalues of $\hat{\rho}$ times $\log(2)$, the conversion factor between bases for the logarithm. Alternatively, we can express the von Neumann entropy in bits, in which case it is identical to the Shannon measure of the density matrix eigenvalues.

$$S_{vN}(\hat{\rho}) = -\log(2) \sum_v \rho_v \log_2(\rho_v) \quad (136)$$

$$= \log(2) \text{SMI}([\rho_1, \rho_2, \rho_3, \dots, \rho_v, \dots]) \quad (137)$$

$$S_{vN}^{(\text{bits})}(\hat{\rho}) \equiv S_{vN}(\hat{\rho}) / \log(2) \quad (138)$$

$$= \text{SMI}([\rho_1, \rho_2, \rho_3, \dots, \rho_v, \dots]) \quad (139)$$

We have a set of probabilities ρ_v and the Shannon measure tells us how much information is missing given this probability distribution. It is information about the extent to which the system could be found to be in various of the density matrix eigenstates. Fortunately, this measure is invariant under unitary transformations, so S_{vN} is the same regardless of which basis set we use.

Time development of the density matrix. We can consider the case of a system that is in a mixed state, presumably because of contact with another system in the past, but which is now isolated. From the Schrödinger equation for $|\Psi\rangle$ in (123),

one can derive the von Neumann equation for the time development of $\hat{\rho}$:

$$\boxed{i\hbar \frac{\partial \hat{\rho}}{\partial t} = [\hat{H}, \hat{\rho}].} \quad (140)$$

where the square brackets denote the anti-commutator. This holds of course for pure-state density matrices as well. Equation (140) is also known as the quantum Liouville equation. If the Hamiltonian is time-independent we can equivalently write the time development as a unitary transformation from the initial state directly:

$$\boxed{\hat{\rho}(t) = e^{-i\hat{H}t/\hbar} \hat{\rho}(0) e^{+i\hat{H}t/\hbar}} \quad (141)$$

The case where the system is in continuous contact with the bath is, as one might suppose, much more difficult because one has to adopt some approximate model of the behavior of the bath. The most straight-forward is the Lindblad formalism, but we will not explore that here.

We can say something qualitative about the evolution of the density matrix in contact with the larger system. The off-diagonal elements of the density matrix are called “coherences” because they express quantum mechanical coherence between the system A and the bath B. If the system and bath are initially in a direct product state, these coherences will quickly vanish (either exponentially or with a Gaussian shape) and the density matrix will become diagonal in the basis of energy eigenstates. The reason for this is that the system will become quantum mechanically entangled with the many degrees of freedom of the larger system and the quantum complex phases in the sum (127) will average to zero. This can be seen in moderately sized system where the global A+B system can be solved exactly. [19] If the mean interaction energy between the system and environment is E_{se} , then the coherences vanish and a time of the order of \hbar/E_{se} .

The density matrix is the best local description we can have of the state of the system. The reality is that the system has no state—only the combined system+bath really have a quantum state. There is a loss of information from the subsystem as it interacts and entangles with the larger system, reflected by the loss of the off-diagonal elements of the density matrix. The global state is pure and the global information is undiminished, but it cannot be found in any of its subsystems. This is a uniquely quantum mechanical feature.

Statistical mechanics for open quantum systems. Quantum statistical mechanics starts with the quite reasonable assumption that in thermal equilibrium all the coherences have vanished and the density matrix is diagonal in the energy basis with probabilities given by the Boltzmann distribution. [20]

$$\hat{\rho}_{eq} = \frac{e^{-\hat{H}/k_B T}}{Z} \quad \text{where} \quad Z = \text{Tr}(e^{-\hat{H}/k_B T}) \quad (142)$$

Therefore, somewhat anticlimactically, the quantum treatment of the write and erase bit operations adds little to the classical analysis in Sec. 5.3. The Hamiltonian in the basis of dot states $[|1\rangle, |2\rangle, |3\rangle]$ is given by

$$\hat{H}(t) = \sum_{k=1}^3 |k\rangle E_k(t) \langle k| - \gamma \sum_{k=1}^2 \left[|k\rangle \langle k+1| + |k+1\rangle \langle k| \right] \quad (143)$$

In order to hold the bit in the metastable state a memory requires, we want the tunneling energy γ to be small and the barrier height of E_2 in the high state to be large. If $E_1 = E_3$, we can define the barrier height as $E_b \equiv E_2 - E_1$. The effective barrier to tunneling is through this barrier is then [21]

$$\gamma_{eff} = \frac{\sqrt{E_b^2 + 8\gamma^2} - E_b}{4}. \quad (144)$$

The additional memory design requirement is then to make γ small and E_b large enough to suppress quantum tunneling for the required bit hold time. The characteristic tunneling time can be taken to be h/γ_{eff} .

With γ small there will be only slight anti-crossing of the otherwise degenerate energy levels as E_2 moves up and down to latch or erase a bit. The Hamiltonian eigenenergies will be very close to the on-site energies of each dot E_k . The energy scale of the switching is much larger than the tunneling energy by design. If the system is switched slowly enough to always keep it in the thermal ground state, which was our assumption, then the density matrix is always diagonal in the energy eigen-basis because the off-diagonal coherences are all 0. The thermodynamic entropy $S(t)$ is then identical to the von Neumann entropy $S_{vN}(t)$. As a result, each of the Figures (9) through (13) are essentially the same for both the classical and quantum cases. Where we could expect a difference would be if the switching speed were fast enough to drive the system out of equilibrium, or stress the ability of the system to tunnel to the equilibrium state. [22]

6.4 Non-equilibrium quantum system: free expansion of an ideal quantum gas.

Finally, we will look at a very-far from equilibrium situation to see the differing roles of the von Neumann entropy and the quantum entropy of outcomes. We first extend the concept of the entropy of outcomes (111) from pure states to mixed states described by density matrices. We will now call this quantity S_Q , remembering that it is not the thermodynamic equilibrium entropy S , a state function, but rather is a dynamical measure of missing information about a particular observable. We start with a Hermitian operator \hat{Q} representing an observable, its eigenvalues q_i ,

eigenstates $|q_i\rangle$, and the projection operator onto a particular eigenstate $|q_i\rangle\langle q_i|$. For a system described by the density matrix $\hat{\rho}$ we define:

$$p_{q_i} = \text{Tr}(\hat{\rho} |q_i\rangle\langle q_i|) \quad (145)$$

$$S_Q(\hat{\rho}) \equiv -\sum_{q_i} p_{q_i} \log_2(p_{q_i}) = \text{SMI}([p_{q_1}, p_{q_2}, p_{q_3}, \dots]). \quad (146)$$

For a dynamic system we will denote $S_Q(t) = S_Q(\hat{\rho}(t))$ and see to what degree this extends the equilibrium notion of the thermodynamic entropy.

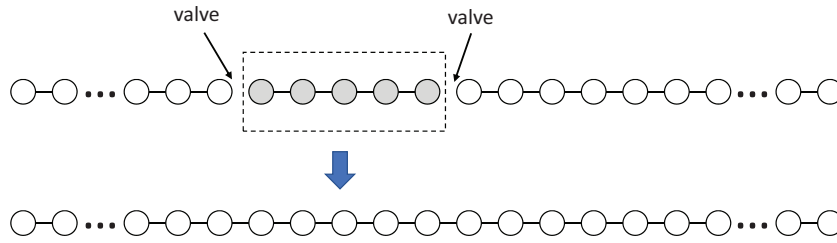


Fig. 16 Schematic view of free expansion of single-particle quantum gas on a 1D chain of quantum dots. The chain consists of a linear array of N quantum dots with near-neighbor Hamiltonian coupling energies γ . For $t < 0$ (top diagram) the particle is confined to a small initial segment of N_{init} dots in the chain. The coupling elements to the left and right of the segment are set to zero—the “valves” are closed. The initial state is in thermal equilibrium. At $t = 0$ the system is isolated from the bath and the values of coupling elements on the left and right of the segment are switched from 0 to γ (bottom diagram). The one-particle gas is therefore free to expand into the larger container. The unitary non-equilibrium evolution of the quantum density matrix is given by equation (141).

We consider a very simple system of a single-particle gas in one dimension in the low density limit. The one-particle gas models an ideal non-interacting gas in the low-density limit for which one need not take into account the Fermi-Dirac or Bose-Einstein character of the occupancy statistics. Fig. 16 illustrates the physical situation. A linear array of N quantum dots, with on-site energy E_0 , forms a “large” container. The spatial positions of the dots are taken to be uniform from $x_1 = 0$ to $x_N = 100$ in arbitrary units. A tunneling matrix element γ_k couples sites k and $k + 1$. The Hamiltonian for the system is then

$$\hat{H} = \sum_{k=1}^N E_0 |k\rangle\langle k| - \sum_{k=1}^{N-1} \gamma_k \left[|k\rangle\langle k+1| + |k+1\rangle\langle k| \right]. \quad (147)$$

All the coupling elements are identical, $\gamma_k = \gamma$, except those surrounding a segment of length N_{init} dots near the center of the array. These are initially set to $\gamma_{left} = \gamma_{right} = 0$, isolating the N_{init} dots in the small container.

Before $t = 0$, the gas is in the thermal equilibrium state at temperature T and is held in the smaller container of N_{init} sites. We calculate the initial equilibrium density matrix

$$\hat{\rho}^{init} = \frac{e^{-\hat{H}_{init}/k_B T}}{\text{Tr}\left(e^{-\hat{H}_{init}/k_B T}\right)}, \quad (148)$$

where \hat{H}_{init} includes only the dots in the small container.

At $t = 0$ two zero tunnelling matrix elements, γ_{left} and γ_{right} are set to the common value γ , thus opening the ‘‘valves’’ connecting the small container to the larger container. The initial $N_{init} \times N_{init}$ density matrix is embedded in the larger $N \times N$ density matrix describing the whole system and the time development is calculated directly from the von Neumann equation (141). The container is now assumed to be isolated from the heat bath. We calculate the case where $N_{init} = 8$, $N = 64$, and $\gamma/E_0 = 0.1$. The mean value of the energy eigenvalues is E_m and the temperature is chosen to be $T = E_m/(15k_B)$. The smaller container is offset from the center of the larger container slightly to avoid artifacts of symmetry. Time is measured in units of $\tau = \hbar/\gamma$.

The probability of dot occupancy at three snapshots in time are shown in Fig. 17. At $t = 0$ the probability is nonzero only in the smaller container with quantum confinement effects shown by the rounding of the probability at the container edges. At $t = 5\tau$ the gas is expanding freely into the surrounding container. From $t \approx 20\tau$ onward the probability fills the larger container, though because the system has no way to lose energy, quantum oscillations in the probability continue indefinitely. A snapshot at $t = 80\tau$ shows a characteristic distribution.

Fig. 18 shows the probability distribution for each energy eigenstate of the system. Before the expansion ($t = 0^-$) there are $N_{init} = 8$ Hamiltonian eigenstates with the characteristic Boltzmann distribution of probabilities. The red line is the Boltzmann exponential as a continuous function of energy. Just after the valves are opened ($t = 0^+$), the number of eigenstates increases to $N = 64$. This is a far-from-equilibrium situation so the occupancy is no longer thermal. The red line again shows the Boltzmann exponential for the initial temperature, now just for comparison because there is no temperature for the system after $t = 0$. The non-equilibrium distribution in energy is perhaps surprising close to the thermal shape, though less so at low energies. The probabilities for each allowed energy do not change once the valves are open because the time evolution (141) is unitary, preserving the projection onto energy eigenstates, so the ($t = 0^+$) figure is valid for all positive times.

Figure 19 shows two measures of the entropy of the expanding gas. We again note that a thermodynamic entropy cannot be uniquely defined for this non-equilibrium

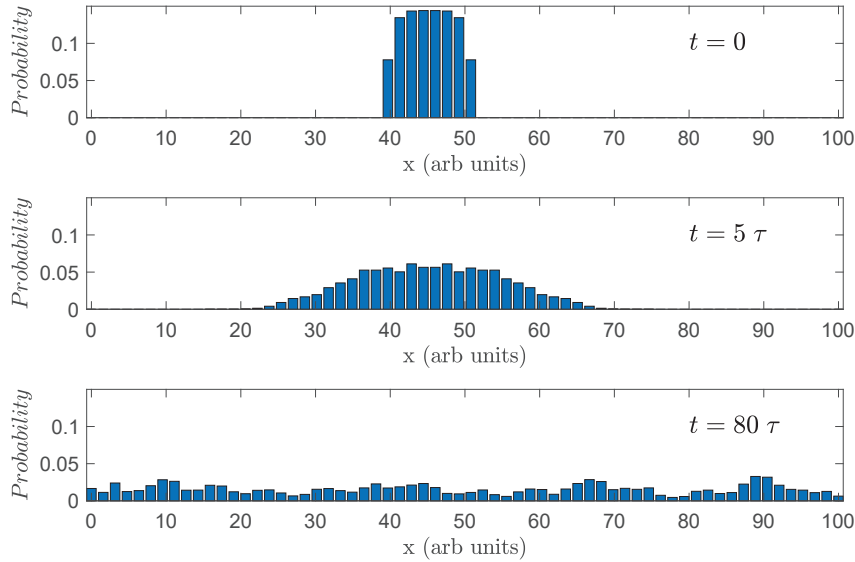


Fig. 17 Free expansion of the quantum gas in 1D. The probability density for the linear chain shown schematically in Fig. 16 is plotted at three snapshots in time during the unitary evolution of the density matrix $\rho(t)$ according to (141). At $t = 0$, the system is confined in the smaller container near the center of the array. The gas is initially at thermal equilibrium and quantum confinement effects are visible in the drop-off at the edges. The single particle gas is isolated from the thermal bath at $t = 0$ and released to expand into the surrounding larger container. Time is measured in units of $\tau = \hbar/\gamma$ where γ is the inter-dot coupling Hamiltonian matrix element (see equation (147)). The middle plot shows the expanding gas at $t = 5\tau$. The lower plot shows the probability at $t = 80\tau$, when it has filled the container. Oscillations persist because there is no energy dissipation mechanism in the model.

situation. The dashed red line shows the von Neumann entropy $S_{vN}(t)$, measured in bits (calculated from (138)). Unsurprisingly, it is constant, 1.81 bits, throughout the expansion precisely because the free expansion is a unitary process. The eigenvalues of the density matrix do not change in time under (141).

The time development of the entropy of outcomes associated with the position operator, $S_x(t)$, calculated from (146) is shown as the solid line in Fig. 19. Initially $S_x \approx 3$, corresponding to the Shannon entropy for $8 = 2^3$ uniformly occupied dots. It is slightly lower because the probability density at $t = 0$ is not quite uniform (Fig. 17). It rises smoothly to near a value of 6, corresponding to the Shannon entropy for $64 = 2^6$ uniformly occupied dots. Again, the residual quantum oscillations visible in Fig. 17 account for the remaining difference.

In this free expansion of an ideal quantum gas, the volume was increased by a factor of 8, which would correspond to a classical increase of the thermodynamic entropy by a factor of $\log_2(V_{final}/V_{init}) = 3$ (see (91)). This is nearly what we see

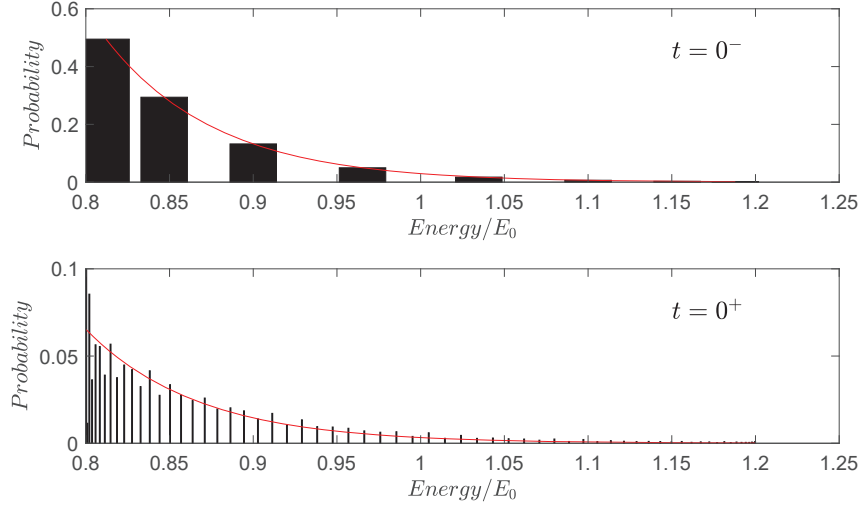


Fig. 18 The probability of each eigen-energy prior to expansion, and just after the expansion “valves” are opened for the single particle quantum gas shown in Figs. 16 and 17. Here $k_B T/E_0 = 0.067$ and $\gamma/E_0 = 0.1$. The initial state is in thermal equilibrium so the probabilities for each eigenenergy of the smaller container, with $N_{init} = 8$, are given by the Boltzmann distribution, equation (142) and indicated by the red line in the top figure. Immediately after the opening, there are more eigenstates because the system is now larger with $N = 64$, as shown in the lower figure. The probability associated with each eigenstate is now only approximately thermal (red line) because the system is no longer in equilibrium. The probabilities in the lower figure are constant after $t = 0^+$ throughout the unitary evolution of the isolated system.

in Fig. 19 for this entirely quantum mechanical calculation using $S_X(t)$. But it is *not* reflected in the von Neumann entropy, which is meant to capture the amount of deviation from a purity in the density matrix. The amount of “mixedness” does not change during isolated free expansion. Using $S_X(t)$ we capture what the classical version captured—the increase in the amount of *position* information that is missing. The entropy of outcomes for energy measurements S_E is constant in time with a value of 4.9 bits.

If the expanded system were to again be put in contact with the bath and allowed to come to thermal equilibrium, then we would have $S_{vN} = 4.78$ bits, $S_E = 4.88$ bits, and $S_X = 5.99$ bits (still reflecting the small quantum edge effects). The von Neumann entropy S_{vN} is the entropy of outcomes S_Q for the case when the relevant operator is the density operator itself, $Q = \rho$, and it will always have the minimum value over the space of all Hermitian operators.

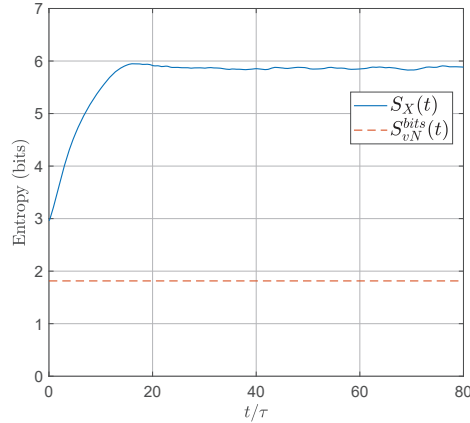


Fig. 19 Two measures of the entropy of the ideal single-particle quantum gas during free expansion. The von Neuman entropy $S_{vN}(t)$, calculated from (138, 139), is constant with a value of 1.81 bits during unitary free expansion (red dashed line). It is a measure of the lack of information about the quantum state of the system in the density matrix—its “mixedness” or deviation from purity. The entropy of outcomes for position, $S_X(t)$, defined in (146) with $\hat{Q} = \hat{X}$, is shown as the solid blue line. S_X is initially 2.96 bits. If there were no quantum size effects in the probability at $t = 0$ (Fig. 17), then we would have $S_X(0) = 3$, corresponding to equal probabilities over $8 = 2^3$ dots. As the expansion proceeds, S_X increases to a value near 6 bits, corresponding to a uniform distribution over $64 = 2^6$ dots. It is S_X that most nearly corresponds to the classical equilibrium result, given by the Sakur-Tetrode equation (91), for the entropy of a gas related to the volume it fills. Using the $S_X(t)$ as a measure of the missing position information offers a natural extension of equilibrium entropy to the non-equilibrium case.

7 Discussion

Discussions of information are sometimes confused by failing to distinguish raw information from encoded information. An encoding scheme adds the mapping between physical states and logical symbols that makes a physical process, which is after all just physics in action, a meaningful logical or arithmetic process. Bits are not part of the ontology of the physical world but rather supervene on physical states.

Logically, the Shannon entropy (SMI), as a measure of missing information in a probability distribution is the most foundational concept for entropy. Probability represents a quantification of incomplete knowledge. Jaynes contribution was built on the insight that the only unbiased way to construct a probability distribution is to find that distribution which maximizes the SMI, a measure of what is unknown, subject to the mathematical constraints that arise from what is known. This is a strict and objective procedure that is the same for all observers who have the same information available to them.

We have seen in Sec. 3 that applying the Jaynes maximum entropy principle yields mathematical results which reproduce standard statistical mechanics. Clas-

sical statistical mechanics is, after all, unnecessary if all the relevant dynamical quantities are known—that would require only mechanics. Statistical mechanics concerns what we can know about a system when the details of its dynamics are unknown. When applied to the problem of a target system in thermal equilibrium with a large heat bath, the Jaynes procedure precisely connects the information theoretic Shannon entropy with the thermodynamic entropy of von Neumann, Gibbs, Boltzmann, and Clausius. The thermodynamic entropy in equilibrium is (within a constant) the Shannon entropy of the unique probability distribution which maximizes the SMI subject to the appropriate constraints. The equilibrium thermodynamic entropy is a special case of the Shannon entropy applied to particular systems that have a well-characterized average energy, number of particles, or other macroscopic constraints.

The Jaynes formulation of physical entropy as a measure of missing information is superior to the often-encountered notion of entropy as a measure of disorder. Disorder is not a quantifiable idea—it is irredeemably subjective, completely in the mind of the beholder. Information (in the information theoretic sense measured in bits) is quantifiable and precise. The information that is missing from a given probability distribution is quantifiable and objective.

The Landauer Principle connects an information theoretic process, bit erasure, with a physical process, heat dissipation. Because it concerns a lower bound for heat dissipation we looked quantitatively at a minimal physical system with a specific encoding scheme. A key step here was to use the Jaynes definition of thermodynamic entropy to describe a memory storage device which is *ipso facto* not in an equilibrium state. This straightforward extension permits a quantitative analysis of the minimal thermodynamic system when a known or unknown bit is erased. We see precisely the expected heat dissipation of $k_B \log(2)$ when an unknown bit is erased, and no lower bound for the heat dissipated when erasing a known bit (with a copy preserved).

In the quantum mechanical case, the von Neumann entropy is the SMI of the eigenvalues of the density matrix. This is a measure of quantum state purity; a pure state has von Neumann entropy of 0. A quantum treatment of the minimal memory system acts essentially the same as the classical system because quantum coherences vanish in thermal equilibrium. It must be emphasized that in quantum mechanics what is unknown includes the fundamental indeterminacy of the physical law. This is now known to be not a peculiarity of the quantum mechanical description, but rather a feature of the nature of the physical world.

Grounding the entropy concept in the Shannon measure also naturally focuses attention on the less well known quantum entropy of outcomes S_Q for measurements of an observable Q . We have seen that in the case of the free expansion of a classical gas, the quantum analogue of the classical entropy was not the von Neumann entropy, but the entropy of outcomes for the position operator, S_X (146). This entropy is not basis-independent—it depends specifically on the observable in which one is interested. Whereas the von Neumann entropy captures the amount of information

missing about the pure state of the system due to entanglement with the environment, S_Q capture the amount of missing information about measurements of the observable Q . It is applicable to both pure and mixed states.

Acknowledgements Thanks to Arieh Ben-Naim for highlighting the Jaynes approach in references [4] and [5], and for helpful conversations. Thanks also to Neal Anderson and Ken Sauer for many stimulating talks on information and related topics.

References

1. W. Porod, R. Grondin, D. Ferry, G. Porod, Phys. Rev. Lett. **52**, 232 (1984)
2. J.D. Norton, Studies In History and Philosophy of Modern Physics **36B**(2), 375 (2005)
3. E. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003)
4. A. Ben-Naim, *Modern Thermodynamics* (World Scientific, 2016)
5. A. Ben-Naim, *A Farewell to Entropy: Statistical Mechanics Based on Information* (World Scientific, 2008)
6. T. Cover, J. Thomas, *Elements of Information Theory*, 2nd edn. (Wiley, 2006)
7. M. Trebus, *Thermostatistics and Thermodynamics* (D. Van Nostrand, 1961)
8. G.M. Jochen Gemmer, M. Michel, *Quantum Thermodynamics: Emergence of Thermodynamic Behavior Within Composite Quantum Systems, Lect. Notes Phys.*, vol. 784 (Springer, Berlin Heidelberg, 2009)
9. I. Amlani, A.O. Orlov, G.L. Snider, C.S. Lent, G.H. Bernstein, Applied Physics Letters **72**(17), 2179 (1998)
10. J. Christie, R. Forrest, S. Corcelli, N. Wasio, R. Quardokus, R. Brown, S. Kandel, C. Lent, Lu, K. Y. Henderson, Angewandte Chemie **127**, 15668 (2015)
11. A.O. Orlov, C.S. Lent, C.C. Thorpe, G.P. Boechler, G.L. Snider, Jpn. J. Appl. Phys. **51**, 06FE10m (2012)
12. R.K. Kumamuru, A.O. Orlov, R. Ramasubramaniam, C.S. Lent, G.H. Bernstein, G.L. Snider, Ieee Transactions On Electron Devices **50**(9), 1906 (2003)
13. L. Szilard, **5**, 840 (1929)
14. G. Paneru, D.Y. Lee, T. Tlustý, H.K. Pak, Phys. Rev. Lett. **120**, 020601 (2018). DOI 10.1103/PhysRevLett.120.020601. URL <https://link.aps.org/doi/10.1103/PhysRevLett.120.020601>
15. M. Giustina, M.A.M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.A. Larsson, C. Abellán, W. Amaya, V. Pruneri, M.W. Mitchell, J. Beyer, T. Gerrits, A.E. Lita, L.K. Shalm, S.W. Nam, T. Scheidl, R. Ursin, B. Wittmann, A. Zeilinger, Phys. Rev. Lett. **115**, 250401 (2015). DOI 10.1103/PhysRevLett.115.250401. URL <http://link.aps.org/doi/10.1103/PhysRevLett.115.250401>
16. J. Handsteiner, A.S. Friedman, D. Rauch, J. Gallicchio, B. Liu, H. Hosp, J. Kofler, D. Bricher, M. Fink, C. Leung, A. Mark, H.T. Nguyen, I. Sanders, F. Steinlechner, R. Ursin, S. Wengerowsky, A.H. Guth, D.I. Kaiser, T. Scheidl, A. Zeilinger, Phys. Rev. Lett. **118**, 060401 (2017). DOI 10.1103/PhysRevLett.118.060401. URL <https://link.aps.org/doi/10.1103/PhysRevLett.118.060401>
17. B. Hensen, H. Bernien, A.E. Dreau, A. Reiserer, N. Kalb, M.S. Blok, J. Ruitenber, R.F.L. Vermeulen, R.N. Schouten, C. Abellan, W. Amaya, V. Pruneri, M.W. Mitchell, M. Markham, D.J. Twitchen, D. Elkouss, S. Wehner, T.H. Taminiau, R. Hanson, Nature **526**, 682 (2015). URL <http://dx.doi.org/10.1038/nature15759>
18. L.K. Shalm, E. Meyer-Scott, B.G. Christensen, P. Bierhorst, M.A. Wayne, M.J. Stevens, T. Gerrits, S. Glancy, D.R. Hamel, M.S. Allman, K.J. Coakley, S.D. Dyer, C. Hodge, A.E.

- Lita, V.B. Verma, C. Lambrocco, E. Tortorici, A.L. Migdall, Y. Zhang, D.R. Kumor, W.H. Farr, F. Marsili, M.D. Shaw, J.A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M.W. Mitchell, P.G. Kwiat, J.C. Bienfang, R.P. Mirin, E. Knill, S.W. Nam, *Phys. Rev. Lett.* **115**, 250402 (2015). DOI 10.1103/PhysRevLett.115.250402. URL <https://link.aps.org/doi/10.1103/PhysRevLett.115.250402>
19. E. Blair, C. Lent, *J. Appl. Phys.* **113**, 124302 (2013)
 20. R.P. Feynman, *Statistical Mechanics: A Set of Lectures* (W.A. Benjamin, Reading, Mass., 1972)
 21. J.P. Timler, Energy dissipation and power gain in quantum-dot cellular automata. Ph.D. thesis, University of Notre Dame (2003)
 22. E.P. Blair, S.A. Corcelli, C.S. Lent, *J. Chem. Phys.* **145**(1), 014307 (2016)