Black Holes, Demons and the Loss of Coherence:

How complex systems get information, and what they do with it.

Seth Lloyd

Ph.D. Thesis

Theoretical Physics

The Rockefeller University

April 1, 1988

Contents

Introduction:

One need only look at the slips of paper on the floor of the New York Stock Exchange at closing to be aware of the extent to which information is gathered, processed, and discarded on the macroscopic level of human affairs. Information is represented physically by the different states of physical systems -- by a modulated frequency of an electromagnetic wave, by neurotransmitters moving across a synapse, by the level of current in a semiconductor. The gathering, processing, and discarding of information is subject to physical laws as is any physical process. Conversely, almost all physical processes involve the exchange and transformation of information. When two electrons interact, their momenta and spins become correlated just as do the the patterns of firing neurons in the brains of two stock brokers talking on the phone. The first two chapters of this thesis use statistical mechanics and information theory to analyze the ways in which physical systems exchange and process information.

Each chapter of this thesis arose out of a paper of the same name. In the first chapter, "A Loophole in the Second Law of Thermodynamics," I examine the constraints that statistical mechanics imposes on devices that gather information and put it to use to decrease entropy. Several theorems on the mechanics of gathering information are proved, and the possibility of violating the second law of thermodynamics by obtaining information is discussed in light of these theorems. Without increasing entropy elsewhere, a Maxwell's demon can lower the entropy of his surroundings by an amount equal to the difference between the maximum entropy of his recording device and its initial entropy. A demon with human-scale recording devices can reduce the entropy of a gas by a negligible amount only; but the proof of the demon's impractibility leaves open the possibility that complex systems in initial states of low entropy, inducing correlations via long-range forces, can reduce the entropy of their surroundings by a substantial amount without increasing entropy elsewhere. In the event that a boundary condition for the universe requires it to be in a state of low entropy when small, the correlations induced between different particle modes during the expansion phase allow the modes to behave like Maxwell's demons during the contracting phase, reducing the entropy of the universe to a low value.

The second chapter, "Complexity as Thermodynamic Depth," came out of a paper written in collaboration with Heinz Pagels. In this chapter, a measure of complexity for the macroscopic states of physical systems is defined. Called thermodynamic depth, the measure is universal: it applies to all physical systems. The form of the measure is uniquely fixed by the requirement that it be a continuous, additive function of the processes that can result in a state, and that it assign states of thermodynamic equilibrium depth zero. The measure satisfies the intuitive requirements that wholly ordered and wholly random systems are not thermodynamically deep, and that a complex object together with a copy are not much deeper than the object alone. A gas at equilibrium is thermodynamically shallow, and so is a salt crystal. One Brahma bull is thermodynamically very deep, but six cloned Brahma bulls are only as deep as the original bull plus the depth of the relatively shallow cloning process.

The thermodynamic depth of a particular macroscopic state of a system is defined to be the difference between the state's coarse- and fine-grained entropy, which in turn is proportional to the amount of information discarded as the system evolved into that state. That the amount of information discarded during some process is a measure of the complexity of its result may at first seem strange. Think of long division: to calculate the quotient of two numbers one must calculate a number of multiples of the divisor as intermediate steps; once the division has been completed, the results of these intermediate multiplications constitute information that is no longer useful -- "junk" information. Animals cull useful information from useless as a simple prerequisite of the complex process of staying alive: light bouncing off a mosquito forms a distinct spot on the retina of a swift, the swift's visual cortex registers the mosquito's position, the wings adjust, the mouth opens, snap, and then the next mosquito comes into view. The position of the mosquito just eaten is no longer relevant to the future survival of the swift, and all the information generated in the process of catching that mosquito, once useful, now beomes useless. To measure the amount of information generated and discarded by a living system is beyond the ability of a physicist; however, one can readily apply the measure of thermodynamic depth to man-made systems, such as computers, where the amount of information processed is simply quantifiable. When one applies thermodynamic depth to physical systems capable of computation, the measure yields a conventional computational measure of complexity as a special case: the computational complexity of a problem is proportional to the number of elementary logical and arithmetical operations that must be performed to solve the problem starting from a given set of initial information. But as the example given above suggests, each logical and arithmetical operation involves discarding a certain amount of information. In the remainder of the chapter, physical and computational examples are given, the relation of thermodynamic depth to previously proposed definitions of complexity is discussed, theorems on the ways that physical systems transfer and create information are proved, and applications to physical, chemical, and mathematical problems are proposed.

The final two chapters of this thesis propose a new version of quantum statistical mechanics. The probabilities inherent in pure quantum states differ markedly from normal statistical mechanical probabilities: because quantum mechanical probabilities come out of underlying probability amplitudes, quantum mechanical systems can exhibit interference phenomena that classical systems cannot. Quantum and classical statistics do not always differ, however. In chapter three, "Pure State Quantum Statistical Mechanics and Black Holes," a number of theorems are proved that imply that the probabilities inherent in the pure quantum states of systems with more and more degrees of freedom differ less and less from the probabilities given by the normal statistical mechanical ensembles. For example, if a gas confined to a volume is in a particular pure quantum state with energy E, then for most measurements made on the gas, the quantum statistics of the pure state predict a distribution of results. (Contrast the quantum case with the classical case: if a classical gas is in a particular microscopic state, then any measurement made on the gas will give a single, well-defined result.) The theorems proved imply that for most measurements, the quantum mechanical distribution of results implied by the pure state has in the thermodynamic limit the same mean value, standard deviation, and higher moments as the distribution implied by the microcanonical ensemble for the gas with energy E. Another way of stating this result is that a system with many degrees of freedom in a pure state behaves with respect to most measurements as if it were in a mixture. This result is applied to black holes, and in chapter 4, "Why Deviations from Wave Function Collapse are Hard to Detect," to the quantum mechanical measurement problem.

Acknowledgements:

I would like to thank my advisor, Heinz Pagels, for his comments, criticism and help with all sections of this thesis. I would like to thank Rolf Landauer for helpful discussions of the work in chapters 1, 2 and 4, E.G.D. Cohen for discussions on chapters 1 and 3, Joel Cohen for discussions on chapters 1 and 2, Jeremy Butterfield for discussions on chapters 1, 3 and 4, and Charles Bennett for discussions on chapters 1 and 2. I would like to thank Nicola Khuri, Tony Sanda, and the Rockefeller Physics department for their welcome support for a sometimes recalcitrant graduate student. I would like to thank Ben, Benta, Cary, Ellen, Julie, Mary, Nancy, Philipp, Robert, Royce, Sadie, Sarah, Simon, Tom, William, The Del Fuegos, my cousins, uncles and aunts, my grandmother, and my parents.

Chapter 1: A loophole in the second law of thermodynamics

Maxwell[1] was the first to note the trade-off between entropy and information: he pointed out that a being who could measure the velocity of individual molecules in a gas could operate a shutter between two containers of the gas, and by shunting fast molecules into one container and slow molecules into the other, create a difference in temperature between the two containers, in apparent contradiction to the second law of thermodynamics. William Thomson, Baron Kelvin, called this being a "demon." Such a demon takes information about the microscopic state of the gas and by acting judiciously, uses that information to reduce the entropy of the gas. Szilard[2] subsequently suggested that the act of acquiring information by its very nature generates entropy; he showed that one bit of information could be used to reduce the entropy of a one-molecule gas by $k_B \ln 2$, and gave an example of a simple measuring device that created at least $k_B \ln 2$ of entropy for each bit of information that it acquired. Brillouin[3] elevated Szilard's result to the status of a "generalized Carnot principle" : the amount by which a demon can reduce entropy by putting his insider information to work is always less than or equal to the amount by which he increases entropy in the act of acquiring information in the first place. More recent authors have dealt with the question of minimum entropy increase in telecommunications[4,5] and in computation.[6,7,8]

The above authors investigate the trade-off between information and entropy in specific devices. In this paper, the mechanics of the general process by which one system obtains information about another are examined. Almost any interaction generates correlation: not only does a drift chamber get information about the trajectories of the particles that pass through it -- an electron can be regarded as getting information about the particles off of which it scatters, as well. Four theorems that exploit the measure-preserving properties of classical and quantum mechanical dynamics are proved. The first two theorems put limits on the amount of information that a given system can gather, and imply that a system such as a drift chamber or an electron that is to obtain information through interaction must either have its entropy reduced during the interaction, or else exist in a state of low entropy beforehand. The third theorem confirms the well-known fact that a reduction of the entropy of one system by normal thermodynamic means requires an increase in entropy elsewhere; the fourth theorem shows that the existence of correlations can mitigate the amount of the required entropy increase. The generalized Carnot principle holds for drift chambers and electrons, photocells and eyes, because in order to gather information repeatedly such systems must repeatedly be placed in states of low entropy, at the cost of a counterbalancing increase in entropy elsewhere. However, the generalized Carnot principle does not necessarily hold for systems in preexisting states of low entropy, such as gravitationally induced clusters of matter, that interact once and for all to acquire information about other systems, such as nearby clusters. Maxwell's demon cannot function to reduce the entropy of a gas, but the existence of states of low entropy at early times together with the presence of long range forces to induce correlations may imply a violation of the second law of thermodynamics for the universe as a whole.

In the event that a boundary condition for the universe requires that it be in a state of high order when small, the second law of thermodynamics must be violated during the recontracting phase. The theorems proved here show how in this event the different particle modes behave as Maxwell's demons, using information acquired during the expanding phase to decrease coarse-grained entropy during the recontracting phase.

Entropy

The Shannon entropy of a system with states i is given by

$$S = -\sum_i p_i \ln p_i \quad ,$$

where p_i is the probability that the system is in its i-th state given macroscopic constraints such as total energy, volume, number of particles, etc.; for a thermodynamic system, the Shannon entropy is equal to the fine-grained entropy divided by Boltzmann's constant. In what follows, the word 'entropy' refers to the Shannon entropy unless otherwise noted. For a discrete system, the entropy is non-negative. For a continuous classical system with coordinates and momenta labelled by a, we have $S = - \int p(a) \ln p(a) da$, where p(a) is now a probability density; S for a continuous system can be negative. For a quantum mechanical system, S can be written $S = -\mathrm{tr}\rho\ln\rho$, where ρ is the density matrix for the system.[9] In each of these expressions for the entropy, S gives a measure of how little we know about the system in question. Note that if $p_i = 1/W$, i=1 to W, so that the system can be in any one of W states with equal probability, the usual expression for the entropy, $S = \ln W$, is recovered.

The relation, $S = \ln W$, does not immediately suggest Carnot's original conception of entropy as a measure of the inability of a system to do work: the closer the entropy of a heat engine together with its reservoirs is to their maximum entropy, the less work the engine can perform. Simply knowing more about a system such as a gas may constrain the number of possible microstates without providing an obvious way to make the gas do work as part of some heat engine; to do work, we must imitate Maxwell's demon and use our knowledge to produce a differential in temperature, pressure, or some other thermodynamic potential. In the cases of an ideal gas, Fermi and Bose gases, spin systems, etc., one can explicitly establish the equality of $k_B\ln W$ and the thermodynamic entropy; and knowledge of a departure from the maximum of the statistical entropy as a function of energy, volume, and particle number gives us the ability to create flows of heat, push pistons, and pump particles. Given the equality of $k_B\ln W$ and thermodynamic entropy in known cases, we assume that the thermodynamic entropy can generally be interpreted as a measure of our lack of knowledge of the actual state of a system. Though not every increase in information about a system such as a gas brings about an obvious increase in our ability to make that system do work, nevertheless, every decrease of ΔS in the thermodynamic entropy of a gas necessarily entails a reduction in the number of possible microscopic states of the system by a factor of $e^{\Delta S}$. In this paper we prove a number of theorems on the mechanics of getting information. As long as the correspondence between thermodynamic entropy and $k_B \ln W$ holds, these theorems proved purely in the context of getting information and reducing the statistical entropy hold for the reduction of the thermodynamic entropy as well.

An immediate problem with identifying the fine-grained entropy with the thermodynamic entropy is that $S = -\sum_i p_i \ln p_i$ has the property of remaining constant under volume-preserving evolution in phase space for classical systems and under unitary evolution for quantum mechanical systems.[10] S = constant is particularly easy to see in the quantum mechanical case. Under unitary evolution U, $\rho \rightarrow \rho' = U\rho U^{\dagger}$, and $S \rightarrow S' = -\mathrm{tr} U\rho U^{\dagger} \ln U\rho U^{\dagger}$. Expanding the logarithm in a power series and using the cyclic property of the trace, we immediately get $S' = S$.

The problem arises as follows: If the probability density $p_0(a)$ or the density matrix ρ_0 reflects accurately our knowledge about the state of a given system at time t=0, then the probability density $p_t(a) = p_0(u_t^{-1}(a))$ or the density matrix $\rho_t = U_t \rho_0 U_t^{\dagger}$ reflects accurately our knowledge at time t only if we know the exact form of the evolution u_t in phase space or of the unitary evolution U_t. But a detailed knowledge of the evolution of a system such as a gas is out of the question. The equations of motion for such a system are rarely integrable, and if we constantly update our probability distributions and density matrices according to experimentally acquired knowledge of the evolution, the corresponding entropies will either remain constant or increase.

If the exact evolution of a system can be determined by integration of the equations of motion, or by experiment, the entropy remains constant. We prove that if the evolution of a system can be determined only inexactly by analytic or experimental means, the fine-grained entropy tends to increase. That is, we give a formal proof of the straightforward idea that if we know something about the state of a system to begin with, but know only approximately how the system evolves in time, then we will know less about the state of the system in the future. We thus prove a version of the second law of thermodynamics for a closed system based on a coarse-graining over the space of the system's possible evolutions, rather than on a coarse-graining over phase or Hilbert space.

Theorem:

(Second law of thermodynamics for isolated systems.)

If the time evolution of an isolated classical or quantum mechanical system is known only inexactly, and if under such an evolution the entropy S goes to S', then $S' \geq S$.

Proof:

Take first the case of a classical, discrete system with n states. A time evolution of such a system is given by a member of the group of permutations of n objects, S_n. Let p_i be the probability that the system is in its ith state. Let $\chi(\sigma)$ be the probability that the actual evolution of the system is σ. As the system evolves, $p_i \to p_i'$, where

$$p_i' = \sum_{\sigma\epsilon S_n} \chi(\sigma) p_{\sigma^{-1}(i)} = \sum_j F_{ij} p_j ,$$

and F_{ij}= the sum over all σ, such that $\sigma^{-1}(i) = j$, of $\chi(\sigma)$. Note that $\sum_i F_{ij} = \sum_j F_{ij} = 1$: the transformation of the probabilities is double stochastic. Under such an inexact evolution, the entropy $S = -\sum_i p_i \ln p_i$ goes to

$$\begin{aligned} S' = -\sum_i p_i' \ln p_i' &= - \sum_{i,j} F_{ij} p_j \ln\Big(\sum_{j'} F_{ij'} p_{j'} \Big) \\ &\geq - \sum_{i,j} F_{ij} p_j \ln p_j \quad , \text{ since xlnx is convex} \\ &= - \sum_j p_j \ln p_j \quad , \text{ since } \sum_i F_{ij} = 1 \\ &= S \quad . \end{aligned}$$

The above proof holds also for any classical continuous system that can be suitably modelled by a discrete system, e.g., any system that we can model on a digital computer.

The proof that $S' \geq S$ for quantum mechanical systems goes as follows: let χ_i be the probability that the actual unitary evolution of the given system is U_i. A particular density matrix ρ then transforms as $\rho \to \rho' = \sum_i \chi_i U_i \rho U_i^\dagger$. Look at a representation in which ρ' is diagonal: $\rho' = \mathrm{diag}(p_1', \ldots, p_n')$. The entropy $S = -\mathrm{tr}\rho\ln\rho$ goes to

$$S' = -\mathrm{tr}\rho'\ln\rho' = -\sum_i p'_i \ln p'_i = -\sum_j\sum_i \chi_i p_i^j \ln\Big(\sum_{i'} \chi_{i'} p_{i'}^j\Big) ,$$

where p_i^j is the jth diagonal element of $U_i \rho\, U_i^\dagger$ in the representation in which ρ' is diagonal. But xlnx is a convex function, and so we have

$$S' \geq -\sum_i \chi_i S_i , \quad \text{where } S_i = -\sum_j p_i^j \ln p_i^j \quad .$$

But if $\rho = \mathrm{diag}(p_1,\ldots,p_n)$, then

$$p_i^j = (\, U_i \rho\, U_i^\dagger\,)^{jj} = \sum_{k,k'} U_i^{jk} \rho_{kk'} \bar{U}_i^{jk'} = \sum_k |U_i^{jk}|^2 p_k \quad ,$$

where the matrix $F^{jk} = |U_i^{jk}|^2$ is double stochastic, $\sum_j F^{jk} = \sum_k F^{jk} = 1$, and so $S_i \geq S$ as in the discrete case. Hence $S' \geq \sum_i \chi_i S_i \geq S$.

We can also define a quantum mechanical entropy with respect to a particular basis: $S^b(A) = -\mathrm{tr}\rho_A^b \ln\rho_A^b$, where $\rho_A^b = \sum_i |b_i\rangle\langle b_i|\, \rho_A |b_i\rangle\langle b_i|$, and $\{|b_i\rangle\}$ is a basis for the Hilbert space of A. The entropy $S = -\mathrm{tr}\rho\ln\rho$ defined above is equal to the entropy defined with respect to a basis in which ρ is diagonal, $\rho = \mathrm{diag}(p_1,\ldots,p_n)$. Note that $S^b = -\sum_i\sum_j |U^{ij}|^2 p_j \ln\big(\sum_{j'} |U^{ij'}|^2 p_{j'}\big)$, where the U^{ij} are the components relating the basis in which ρ is diagonal to $\{|b_i\rangle\}$. Hence $S^b \geq S$, as above: any process, such as measurement, that replaces ρ by ρ^b also tends to increase entropy.

That the statistically defined entropy obeys a version of the second law of thermodynamics confirms the plausibility of identifying the entropy defined in terms of probabilities with the thermodynamic entropy. The proof of the second law given here differs from the proof of Gibbs in that we have introduced a coarse-graining over the space of possible evolutions rather than a coarse-graining over phase or Hilbert space. The form of a particular coarse-graining for a system depends on the interactions that we can arrange between the system and the various devices that can record information about the system's state and evolution. We now turn to the mechanics of getting such information.

Information

The relation, $S = \ln W$, gives an explicit expression to the trade-off between information and entropy. For example, if we compress an ideal gas of N molecules isothermally to half its original volume, we decrease its entropy by $N\ln 2$, and reduce the number of possible microstates of the gas by a factor of 2^N -- we gain information about the actual microscopic state of the gas. Conversely, the more we know about a system, the greater the number of constraints on its microscopic state, the lower its entropy. We can give a numerical measure to information by <u>defining</u> the amount of information obtained about a system during some process to be equal to the amount by which the entropy of the system is reduced during the process:[11]

$$\Delta I = -\Delta S \quad .$$

Two systems are correlated if when we look at one we get information about the other. Given a joint probability distribution $p(a_i b_j)$ for the states $a_i b_j$ of AB, the conditional entropy of A given that B is in the state b_j is $S(A/b_j) = - \sum_i p(a_i/b_j)\ln p(a_i/b_j)$, where $p(a_i/b_j) =$

$p(a_i b_j)/p(b_j)$ is the probability that A is in the state a_i given that B is in the state b_j. The average entropy of A given the state of B is

$$S(A/B) = \sum_j p(b_j)\ S(A/b_j) = S(AB) - S(B)\ .$$

In the quantum mechanical case, the average entropy of A given the state of B depends on which basis $\{|b_i\rangle\}$ we use to describe B: $S^b(A/B) = S^b(AB) - S^b(B)$.

Coarse-grained entropy

Conditional entropies can be used to define thermodynamic, or coarse-grained entropies in terms of interactions with measuring devices. Let A be a system, such as a gas, with n states a_i initially described by a probability distribution $p(a_i) = 1/n$: i.e., nothing is known about the state of A. Let B be a system, initially in state b_0, that interacts with A. If after the interaction B is in the state b_i, then A must have been in a state a_j such that $a_j b_0 \rightarrow a_k b_i$: the transition, $b_0 \rightarrow b_i$ defines a set of states that A could have been in before the interaction. Such a set of states corresponds to a macroscopic state of A defined by the interaction with B. In the continuous case, each transition for B defines a volume in the phase space of A. In the quantum mechanical case, each transition $|b_0\rangle \rightarrow |b_i\rangle$ defines a subspace of the Hilbert space for A. The coarse-grained entropy of the macroscopic state of A given that B went from b_0 to b_i is $\bar{S}(A) = k_B S(A/b_i)$ after the interaction, where k_B is Boltzmann's constant. Boltzmann's constant is brought in to make the connection between information theory and thermodynamics: coarse-grained entropies such as $\bar{S}(A)$ are what is normally meant by "the entropy of a system." For example, if a thermometer B is brought into contact with a gas A confined to a certain volume, then $S(A/b_j)$ is the conditional entropy of the gas given that the thermometer is in a state b_j in which the thermometer's scale registers a certain temperature, $\bar{S}(A)$ is the thermodynamic entropy of the gas at that temperature.

Note that the coarse-grained entropy of a macroscopic state is equal to the fine-grained entropy for a uniform distribution over the microscopic states corresponding to the macroscopic state. The coarse-grained entropy of a state is thus always greater than or equal to the state's fine-grained entropy.

The coarse-grained entropy of a system can always be written as Boltzmann's constant times the conditional entropy of the system relative to the state of a given set of measuring devices after prescribed interactions. The following theorems put restrictions on how much one can reduce the conditional entropy of a system through interaction; since the coarse-grained entropy of a system is Boltzmann's constant times the conditional entropy of the system with respect to certain measuring devices, the theorems proved also apply to how much one can reduce the coarse-grained entropy during interaction. We now use the fact that the fine-grained entropy never decreases to put a limit to the amount by which the conditional entropy can be reduced during any interaction.

Theorem 1:

The amount by which the average entropy of A given the state of B can be reduced during an interaction is limited by the difference between the initial value of B's entropy, and its final value:

$$-\{ S_t(A/B) - S_0(A/B) \} \le S_t(B) - S_0(B) \quad .$$

Proof:

$$S_t(AB) - S_0(AB) \ge 0$$

$$\rightarrow \quad S_t(AB) - S_t(B) - S_0(AB) + S_0(B) \ge S_0(B) - S_t(B)$$

$$\rightarrow \quad -\{ S_t(A/B) - S_0(A/B) \} \le S_t(B) - S_0(B) \quad .$$

In the proof for the quantum mechanical case we must take into account the bases for B with respect to which the average entropy of A is given at time zero and at time t. If at time zero we look at S(A/B) with respect to a basis $\{|b_j\rangle\}$ for B and at time t we look at S(A/B) with respect to another basis, $\{|c_j\rangle\}$, we have $S_t^{bc}(AB) - S_0^{b}(AB) \ge 0$. The rest of the proof continues as after the first line above, and we have

$$-\{ S_t^{bc}(A/B) - S_0^{b}(A/B) \} \le S_t^{bc}(B) - S_0^{b}(B) \quad .$$

According to an outside observer who applies the second law of thermodynamics to AB as an isolated system, a decrease in the entropy of A given B requires an increase in the entropy of B. But for an inside observer who has access to the actual state of B, the entropy of B is always zero. Where the outside observer sees an increase in S(B) in the course of B's interaction with A, the insider sees not an increase in entropy, but an increase in the amount of <u>information</u> that B has about A.

The mutual information, or correlation, between A and B is defined to be the average amount that the entropy of A is reduced given knowledge of the state of B:

I(A,B) = the entropy of A

- the average entropy of A given the state of B

$= S(A) - S(A/B) = S(A) + S(B) - S(AB).$

Two important facts about mutual information: 1) $I(A,B) = I(B,A)$: the average amount by which the entropy of A is reduced if we know the state of B is equal to the average amount by which the entropy of B is reduced if we know the state of A. 2) $I(A,B) \geq 0$: the mutual information is never negative.[11]

The quantum mechanical version of mutual information between two systems A and B with joint density matrix ρ_{AB} can be defined as

$$I(A,B) = S(A) + S(B) - S(AB)$$

$$= -\mathrm{tr}\rho_A \ln\rho_A - \mathrm{tr}\rho_B \ln\rho_B + \mathrm{tr}\rho_{AB} \ln\rho_{AB} \quad ,$$

where ρ_A is the density matrix for A alone, got by taking the trace of ρ_{AB} over B's degrees of freedom, similarly for ρ_B. Once again, $I(A,B) = I(B,A)$ and $I(A,B) \geq 0$. The inherently statistical nature of quantum mechanics implies that $S(A)$, $S(B)$ need not equal zero when $S(AB) = 0$: this makes the quantum mechanical mutual information generally larger than the classical. For example, two electrons in the state $(1/\sqrt{2}) \{ |\uparrow_x\rangle|\downarrow_x\rangle - |\downarrow_x\rangle|\uparrow_x\rangle \}$ exhibit anticorrelation of spins not only along the x-axis, but along any other axis, as well. The mutual information between the electrons is 2ln2, twice the maximum value for the corresponding classical case. We can also define quantum mechanical mutual information with respect to a particular choice of bases $\{|a_i>\},\{|b_j>\}$ for A,B : $I^{ab}(A,B) = S^a(A) + S^b(B) - S^{ab}(AB)$ $= S^a(A) - S^{ab}(A/B)$. Note that $I^{ab}(A,B) \leq I(A,B)$.

Mutual information allows us to give a measure to the amount of information that one system gets about another during any interaction. Suppose A and B are initially described by a classical probability density $p_0(ab)$, where a,b label both coordinates and momenta for A,B respectively. Suppose that A and B undergo a volume-preserving evolution on phase space u_t in which the state ab goes to $u_t(ab)$ at time t; the probability density $p_0(ab)$ then goes to $p_t(ab) = p_0(u_t^{-1}(ab))$. At time t, A and B will in general be correlated: $I_t(A,B) > 0$. The following theorem puts a limit on the amount of information that B can get about A during such an interaction:

Theorem 2:

$I_t(A,B) \leq S_{max}(AB) - S_0(AB)$, where $S_{max}(AB)$ is the maximum entropy for A and B and $S_0(AB) = -\int p_0(ab)\ln p_0(ab)dadb$ is the initial entropy.

Proof:

The mutual information between A and B at time t is $I_t(A,B) = S_t(A) + S_t(B) - S_t(AB)$. But $S_t(AB) \geq S_0(AB)$, as noted above, and the maximum value for $S_t(A) + S_t(B)$ = the maximum value for $S(AB)$ = $\ln V_A + \ln V_B$, where $V_{A,B}$ are the volumes of phase space accessible to A,B respectively. The quantum mechanical version of this proof is identical except that the maximum value for $S(AB)$ is equal to $\ln(n_A) + \ln(n_B)$, where $n_{A,B}$ are the dimensions of the Hilbert spaces for A,B.

Theorem 2 implies that if a system is to function as a measuring device, it together with the system about which it is getting information must be in a state of less than maximum entropy to begin with. If both A and B are in states of maximum entropy at time 0, then $K_t(AB)$ will always be zero.

Theorem 2 puts limits on the amount by which the entropy of a system can be reduced by correlation. The next theorem puts limits on the reduction of entropy by constraint.

Theorem 3:

If A and B are initially uncorrelated, $I_0(A,B) = 0$, then any interaction between the two that decreases A's entropy by ΔS must increase B's by at least ΔS.

Proof:

We have $S_0(A) + S_0(B) = S_0(AB)$ at time zero. At time t we have $S_t(A) + S_t(B) \geq S_t(AB)$. But $S_t(A) = S_0(A) - \Delta S$, and $S_t(AB) \geq S_0(AB)$, so that $S_t(B) \geq S_0(AB) - S_0(A) + \Delta S = S_0(B) + \Delta S$.

In the case of an ideal gas isothermally compressed to half its original volume, with corresponding entropy decrease $\Delta S = N\ln 2$, the heat flowing out of the gas must increase the entropy elsewhere by at least that amount.

We now prove a theorem that puts a limit on the amount by which entropy can be reduced by a combination of correlation and constraint.

Theorem 4:

Any interaction between A and B that decreases A's entropy by ΔS must increase B's entropy by at least $\Delta S - I_0(A,B)$, where $I_0(AB)$ is the correlation between A and B at the start of the interaction.

Proof:

We have $S_0(A) + S_0(B) = S_0(AB) + I_0(A,B)$ at time 0. At time t we have $S_t(A) = S_0(A) - \Delta S$, $S_t(AB) \geq S_0(AB)$, and $S_t(A) + S_t(B) \geq S_t(AB)$. Now

$$\begin{aligned} S_t(B) \geq S_t(AB) - S_t(A) \quad &\rightarrow \quad S_t(B) \geq S_0(AB) - S_0(A) + \Delta S \\ &\rightarrow \quad S_t(B) \geq S_0(A) + S_0(B) - I_0(A,B) - S_0(A) + \Delta S \\ &\rightarrow \quad S_t(B) \geq S_0(B) + \Delta S - I_0(A,B) \quad . \end{aligned}$$

Theorem 4, which includes theorem 3 as a special case, tells us that if A and B are correlated to begin with, a decrease in A's entropy need not be fully compensated for by an increase in the entropy of B: correlation can be "cashed in" to decrease entropy at less than the normal cost. So, when a gas undergoes a Poincaré recurrence cycle the correlations between the molecules conspire to produce a state of abnormally low entropy.

Entropy and Information

The theorems proved above allow us to determine exactly when the collection of information requires an increase in entropy and when it does not. Let us analyze how much information a system B (a photocell, a drift chamber, an electron) can get about another system A (a photon, an electron) by interacting with it, and by how much or how little the entropy of A and B increases during such a process.

Theorem 1 tells us that if the initial entropy of B is less than its maximum value, $S_0(B) < S_{max}(B)$, then the maximum amount of information that B can get about A , i.e., the maximum amount by which the entropy of A can be reduced by becoming correlated with B, is $\Delta I = S_{max}(B) - S_0(B)$. <u>There is no reason why S(AB) must increase during such a process</u>. Nor is there any reason why the coarse-grained entropy must increase. When an electron scatters off of another electron their states become correlated; yet if the electrons begin in a pure state then they end up in a pure state, with entropy zero. Nor need entropy increase when B is macroscopic.[8] Landauer[6] pointed out that a Brownian particle in a potential well that can be continuously modulated between bistability and biased monostability can be used to record one bit of information with vanishingly small increase in entropy as long as the initial configuration of the well is known. Such a system can be realized as a one-domain ferromagnet,[6,7] or as a Josephson junction with a controllable critical current.[12]

Even the prototypical measuring device of Szilard that creates ln2 of entropy for every bit of information acquired can be slightly modified to give as small an entropy increase as desired. Szilard's device consists of a piece of matter that is put in contact with one of two reservoirs at different temperatures, either the first at T_1 or the second at T_2 depending on the position of a pointer. The energy of the matter and the position of the pointer become correlated, but at the cost of an average increase of entropy of at least ln2 as heat flows from matter to reservoir and vice versa. If the initial temperature of the matter is known, however, one can make the increase in entropy as small as one likes simply by supplying reservoirs with incrementally increasing temperatures between T_1 and T_2 , with which the piece of matter comes into equilibrium on its way from one reservoir to the other.

Theorem 1 tells us that the reason that these devices can get information without increasing entropy is that they have been prepared in states of less than maximum entropy: in each case above, the initial state of the device is known at least approximately. Suppose that B has obtained a maximal amount of information about A, so that $S_0(B) = S_{max}(B)$. At this point it is certainly possible to arrange an interaction between B and another system C such that B gets information about C, i.e., $S_t(C/B) - S_0(C/B) < 0$, but during such an interaction the total entropy of A and C given B cannot decrease:

$$S_t(AC/B) - S_0(AC/B) \geq S_0(B) - S_{max}(B) = 0 \quad .$$

That is, once $S(B) = S_{max}(B)$, an interaction with B that decreases the entropy of C by ΔS must increase the entropy of A by at least ΔS. Once B has been saturated as a recording device, any further information that

B gets about any system must be fully compensated for by an increase in entropy elsewhere.

Landauer,[6] and Bennett[8] have pointed out that it is the act of erasure, rather than the act of recording information, that requires an increase in entropy. We apply the theorems proved above to treat the act of erasure in detail. The second law of thermodynamics for isolated systems implies that the demon must arrange an interaction with some external system if he is to decrease the entropy of his recording device, B, when $S(B) = S_{max}(B)$. Note that when $S(B) = S_{max}(B)$, the fine-grained entropy of B equals the coarse-grained entropy, $k_B S(B) = \bar{S}(B)$. But this reduction in the entropy of B cannot take place through correlation: even if B becomes correlated with some other system C, this does the demon no good, since the demon can get no information about C as long as $S(B) = S_{max}(B)$. Hence the demon must arrange to decrease S(B) by constraint. We can now apply theorems 3 and 4: if the demon arranges an interaction between B and a system C that has no initial correlation with B, theorem 3 requires that every decrease of ΔS in B's entropy be compensated for by an increase of ΔS in the entropy of C. In addition, since $\bar{S}(B) = k_B S(B)$, a decrease of $k_B \Delta S$ in $\bar{S}(B)$ requires an increase of $k_B \Delta S$ in the fine-grained entropy of C. If C is already correlated with B, theorem 4 requires that a decrease of ΔS in the entropy of B be matched by an increase in the entropy of C of $\Delta S - I_0(B,C)$, where $I_0(B,C)$ is the initial mutual information between B and C. Theorem 4 thus holds out the hope of decreasing the entropy of B without a fully compensating increase in entropy elsewhere; but by theorem 2, such a decrease can only be accomplished by also reducing the information that B has about C.

Maxwell's demon can function up to a point: he can reduce the entropy of a gas without generating entropy elsewhere by $S_{max}(B)-S_0(B)$, where $S_0(B)$ is the initial entropy of the device that he uses to record information about the gas. He can reduce the coarse-grained entropy of the gas by k_B times this amount. But that is all: any further reductions in entropy can only be accomplished at the expense of generating entropy elsewhere as the demon erases the information that he has already gathered.

Discussion

Maxwell's demon cannot function effectively because in order to decrease the entropy of a gas by a substantial amount without increasing entropy elsewhere, he needs to have a memory bank with at least as many bits as there are molecules in the gas, all at a state of low entropy to begin with. Even if the demon can use all the computers in the world to record his information, with, say, 10^{16} bits of memory, he will run out of memory space before he has reduced the entropy of a gram of gas by a factor of 10^{-10}. In practice, therefore, clever devices that can violate the second law of thermodynamics by any substantial amount are out of the question. Systems in states of low entropy, however, can get a certain amount of information about other systems without increasing entropy overall. One such example is the case of two scattering electrons that are initially in a pure state. Gravitational systems give another example: it was the information encoded in the perturbations of the orbit of Uranus that that allowed the location of Neptune to be pinpointed, yet the interaction between the two planets that produced those perturbations did not produce any increase in entropy (apart from an irrelevant increase due to tidal action). In reducing the uncertainty about the position of Neptune by their calculations, though, Adams and Le Verrier produced more than a compensating amount of uncertainty in the positions and velocities of the molecules in the paper on which they recorded the results of those calculations.

A system in a state of zero entropy to begin with can record an amount of information equal to the logarithm of the number of states accessible to it in the course of interaction. The larger the system, the more information it can record, the more it can reduce the entropy of other systems without increasing entropy in turn. Maxwell's demon operating with human-scale resources can decrease entropy by only the smallest amount. The universe started out in a state of low entropy, however, and a demon that can put to use a substantial fraction of the matter in the universe as a recording device can reduce the entropy of the remaining part by a correspondingly substantial amount. The theorems proved here apply to the mechanics of getting information in general and do not single out specific processes; however, they suggest that by inducing correlations between the positions and velocities of interacting masses the long-range electromagnetic and gravitational forces could accomplish just such a large-scale uncompensated reduction in entropy.

A number of authors[14-18] have pointed out that if there is a boundary condition that requires that the universe must be in a state of high order when small, then entropy must decrease as the universe recollapses. The theorems proved here suggest the mechanism by which such an entropy decrease can take place. The universe starts off with both fine- and coarse-grained entropy low. As the universe expands, the fine-grained entropy remains constant, while the coarse-grained entropy increases. In addition, as the universe expands, the states of different parts of the universe become correlated. As we will show, the mutual information between the various parts of the universe gives a lower bound on the increase in coarse-grained entropy. We will also show that mutual information between the particle modes of the fields in the universe is virtually certain to rise to its maximum value, driving the coarse-grained entropy up to *its* maximum value. As the universe contracts, this mutual information is used to decrease the coarse-grained entropy of the universe as by theorem 4 above: each part of the universe behaves like Maxwell's demon, using the information that it has collected about the other parts of the universe during the expanding phase to decrease their coarse-grained entropy during the contracting phase.

We analyze this process in detail. First, we show that mutual information gives a lower bound to the increase in coarse-grained entropy:

Theorem 5:

Given a macroscopic state b of B at time t determined by the results of inexact measurements on each of the degrees of freedom $l_1,\ldots,l_m$ of B, and given any initial distribution $p_0(l_1\ldots l_m)$ at time t_0, we have

$$\bar{S}_t(b) - \bar{S}_0(l_1\ldots l_n) \geq k_B I_t(l_1,\ldots,l_m) \quad ,$$

where $I_t(l_1,\ldots,l_m) = S_t(l_1) + \ldots + S_t(l_m) - S_t(l_1\ldots l_m)$ is the mutual information shared between $l_1,\ldots,l_m$ at time t.

Proof:

The coarse-grained entropy of b is equal to the sum of the coarse-grained entropies in the degrees of freedom of B partially fixed by the measurent that determines b, $\bar{S}(b) = \bar{S}(l_1) + \ldots + \bar{S}(l_m)$. (For example, a gas with m molecules is confined to a volume V if and only if each of the individual molecules is constrained to lie within V. The entropy of an ideal gas is equal to the sum of the entropies of its individual molecules.) But $\bar{S}(l_i) \geq k_B S_t(l_i)$, where $S_t(l_i)$ is the Shannon entropy for l_i at time t given the underlying distribution, hence

$$\bar{S}(b) \geq k_B\{ S_t(l_1) +\ldots+ S_t(l_m) \} = k_B\{ I_t(l_1,\ldots,l_m) + S_t(l_1\ldots l_m) \},$$

where $I_t(l_1,\ldots,l_m) = S_t(l_1) +\ldots+ S_t(l_m) - S_t(l_1\ldots l_m)$ is the mutual information shared between the degrees of freedom of B at time t. But $S_t(l_1\ldots l_m) = S_0(l_1\ldots l_m) = (1/k_B)\,\bar{S}_0(l_1\ldots l_m)$. Thus,

$$\bar{S}(b) - \bar{S}_0(l_1\ldots l_m) = \bar{S}(b) - k_B S_t(l_1\ldots l_m) \geq k_B I_t(l_1,\ldots,l_m) \ .$$

We can now apply the results of chapter three of this thesis, on the increase of mutual information between the particle modes of quantum fields. The modes of the quantum fields in the universe can be represented as a set of harmonic oscillators. In chapter three we show that, in the thermodynamic limit, an arbitrarily chosen interaction between the oscillators causes them as a group to evolve into a state in which the density matrix for each of the oscillators is exactly that for the oscillator at thermal equilibrium; the evolution drives both the mutual information and the coarse-grained entropy up to their maximum values, equal to the equilibrium thermodynamic entropy.

As the universe expands, the different modes of the quantum fields collect exactly the right amount of information about each other to arrange the required decrease in coarse-grained entropy as the universe contracts. If the boundary condition for the universe requires that entropy be small at the big crunch, the interaction between the different modes causes them to behave like Maxwell's demons; using existing information they decrease the coarse-grained entropy of the remaining modes to its original, low level.

Chapter 2: Complexity as Thermodynamic Depth

Introduction: the intuitive notion of complexity

What is complexity? Before giving a precise definition of complexity it is useful to examine some intuitive properties that a measure of the complexity of physical systems ought to have. These intuitive notions can be compared with the ideas of "algorithmic complexity,"[1-4] "computational complexity,"[5-7] and "logical depth,"[8-10] definitions of complexity that have already appeared in the literature. Following this introductory discussion, we will give a precise, unique definition of the physical complexity of a dynamical system, called thermodynamic depth, which accords with intuitive notions, and show how it is related to computational complexity and logical depth. We give several examples of the application of thermodynamic depth, discuss the implications of this concept of physical complexity for reversible computation and the creation of mutual information, and prove several theorems about the dynamics by which physical systems generate thermodynamical depth.

Dynamical systems range in a continuum from completely ordered, regular systems like the arrangement of carbon atoms in a diamond to completely disordered, chaotic systems like molecules in a gas. The intuitive notion of complexity that we develop here is that complex systems lie somewhere in the continuum between order and chaos. Cells, brains, chickens and chicken DNA are all structurally complex -- they are neither wholly ordered nor wholly disordered. Any reasonable measure of complexity should therefore vanish for the extremes of complete order or disorder and not vanish for the structurally intricate systems between these extremes.

Another intuitive requirement we should expect for a complexity measure is that it be universal -- that it apply to any dynamical system, whether living, non-living, or artificial. While abstract computational definitions of complexity may be useful when applied to specific numerical problems, if a definition of complexity is to be a useful measure for physical systems then it must be defined as a function of physical quantities, which in turn obey physical laws. Furthermore, it is desirable that a physical definition of complexity correspond to the previous definitions of complexity associated with mathematical problem solving if the solutions of these problems are regarded as the outputs of a physical system such as a digital or analog computer.

If we think of complexity as a physical property of an object (such as mass or entropy) then there is a puzzle. Objects can be copied. For example, a bull is a complex object. Are seven bulls seven times as complex as one bull? Can complexity proliferate so cheaply? If complexity is simply an additive property of the parts out of which an object is assembled then it is easy, by copying, to create systems of arbitrarily high complexity.

In order to avoid such indiscriminate proliferation, complexity can not be an additive physical property of a "complex" object. Rather, complexity must be a function of the process -- the assembly routine --

that brought the object into existence. If physical complexity is a measure on the process or set of processes whereby a set of initial states evolves into a final state, then seven bulls need not be very much more complex than one bull. It took billions of years for the earth to evolve one bull; but one bull and a few compliant cows will produce seven bulls relatively speedily.

The definition of complexity that we define below has all of these intuitive properties -- it is a property of the evolution of a state and not of the state itself, it vanishes for ordered and disordered states, it is a universal, physical quantity, and it corresponds to mathematical complexity in problem solving. If, in accord with these intuitive properties, we specify continuity and additivity properties of the complexity measure then the measure is defined uniquely.

Before proceeding it is important to comment on the "algorithmic definition of complexity."[1-4] This definition assigns a complexity measure to computable numbers (or to anything that can be codified as a number) by comparing the length of the numerically codified algorithm for computing the number with the length of the number itself. If the ratio of the two is near unity then the number is said to be complex.

The algorithmic definition of complexity is in our view a misnomer -- it is really a definition of randomness (a profound one) and not complexity. For example, the algorithmic definition of complexity assigns a higher measure of complexity to a random sequence of letters in the Latin alphabet or of nucleotides than to an equal length sequence of The Merchant of Venice or of human DNA. Our intuitive notion of complexity requires that the complexity of random sequences vanishes while the algorithmic definition requires that it be maximal. While useful as a definition of randomness, the algorithmic definition does not provide us with a measure of complexity.

The notion of "computational complexity"[5-7] is well-defined in the field of mathematical problem solving. The complexity of a problem is identified with the amount of computational effort that goes into solving the problem starting from a given formulation. In the travelling salesman problem, for example, one is to find the smallest distance that one must traverse in visiting each of a number of cities, given the distance between each. Each act of multiplication, addition, information transfer, etc., "costs" a certain number of arbitrary units and the complexity of the problem is identified with the total cost of computation. In all known methods of finding the exact solution to the travelling salesman problem, the cost of solving the problem rises exponentially in the number of cities to be visited: the problem rapidly becomes very complex.

The notion of complexity for problems can be extended to a notion of complexity for numbers by regarding a particular number as the solution of a problem -- as the result of a calculation that starts from some initial data. Appealing to the universal properties of Turing machines, Bennett[8-10] proposes a measure of complexity for numbers, called "logical depth": the logical depth of a number is the number of machine cycles that it takes a computer to calculate the number from the shortest possible program. Due to the ability of computers to simulate one other, this definition of complexity is roughly machine independent. Under this measure of complexity, random numbers and ordered but highly regular numbers are logically shallow: the shortest program that can generate a random number N is the program that says "Print N," a fast program, while a number such as 11...1 can be quickly generated by a program that tells the computer to print 1 the requisite number of times. The number made up of the first billion digits of π, however, requires a significant amount of computational effort to generate from its simple representations (e.g., as a continued fraction) and is relatively deep compared to most other billion digit numbers.

Logical depth and the computational definition of complexity are useful and suggestive, but hard to apply to physical systems, which do not come readily encoded as numbers. One might identify the complexity of a physical system with the computational effort required to simulate its evolution, but to simulate arbitrarily accurately a continuous, classical system takes an arbitrarily large amount of computation on a digital computer, and quantum mechanical systems exhibit long-range correlations and statistical behavior that a classical computer cannot simulate. Deutsch[11] has proposed a quantum mechanical computer that can simulate any quantum mechanical system with a finite-dimensional Hilbert space to any desired degree of accuracy,[12] but to simulate accurately a simple continuous process such as a spin-spin interaction between two electrons, or even the evolution of a single electron sitting in a magnetic field, may take the computer a very large number of steps, belying the simplicity of the underlying process. In general, physical systems perform easily processes that are difficult for a computer to simulate, for the simple reason that computers are designed to perform sequences of logical operations, while most systems in nature, though they may abide by logical rules, do not evolve step by step according to Boolean algorithms. A cat following a bird with its eyes need not do floating-point arithmetic to know where to pounce.

A physical measure of complexity

It is desirable, then, to define a measure of complexity for physical systems that does not depend on computer simulations, but is instead a function of physical properties of the systems themselves. Such a measure, in accord with the intuitive notions we developed, should assign low complexity to systems in random states and in ordered but regular states, and, like computational complexity and logical depth, be a measure on the process by which the system evolves from initial to final state. Complexity is then a measure of how hard it is to put something together. If, in addition, one makes the following simple requirements of a measure of complexity for a system that has been determined by experiment to be in a particular macroscopic state:

a) the measure must be a function of the various processes that can result in that state starting from a given set of initial states, and a continuous function of the probabilities p_i that experiment assigns those processes, and

b) the measure must satisfy the requirement of additivity: the measure of increase in complexity of going from the initial set of states to the final state is equal to the average increase in going from the initial set to any intermediate set, plus the increase in going from the intermediate set to the final state,

then the the measure must be proportional to the Shannon entropy[13] of the set of trajectories that lead to that state, $S = -(\sum_i p_i \ln p_i)$, where p_i is the probability of the i-th trajectory, plus a linear combination of terms proportional to the Shannon entropies of other, arbitrary weightings of the trajectories. If we demand that systems in thermodynamic equilibrium have zero complexity, then the form of the measure is completely fixed (up to an overall multiplicative constant which we set equal to one). The part of the measure that depends only on the trajectories is fixed to be the thermodynamic entropy of the state, and to make the connection between information theory and statistical mechanics[14] the constant of proportionality that multiplies the Shannon entropy of the trajectories must be set equal to Boltzmann's constant. The measure of complexity that we define for a state of a system, called the state's <u>thermodynamic depth</u>, is equal to the difference between the coarse-grained entropy of the state (the state's thermodynamic entropy) and the state's fine-grained entropy given the initial distribution of states (Boltzmann's constant times the Shannon entropy of the set of trajectories). Note that if the initial distribution of states is determined by a macroscopic measurement at time t_0, then the fine-grained entropy at that time is set equal to the coarse-grained entropy of the macroscopic state that was the result of the measurement, and the thermodynamic depth of macroscopic states at a later time t is equal to the increase in coarse-grained entropy in going from t_0 to t. Equivalently, since the fine-grained entropy of a closed system remains constant under volume-preserving evolution on phase space or unitary evolution on Hilbert space, the thermodynamic depth identifies a state's complexity with the amount of information pumped into the degrees of freedom of the system other than those fixed by that state, in the process of going from some initial distribution of states to the state in question.

The notion that the amount of entropy generated in arriving at a particular state is a measure of that state's complexity may seem odd at first. Examples will be presented to make clear how the amount of entropy, or junk information, generated by a system during some process is a measure of the total amount of "information processing" that takes place during that process. First, however, the definition of thermodynamic depth will be made formally precise.

Thermodynamic depth

In what follows, the macroscopic states of a system B are defined by prescribed interactions between B and a set of measuring devices; we assume that B and its attendent measuring devices are part of a larger, closed system whose time evolution preserves volume in phase space in the classical case and in Hilbert space in the quantum case. A precise definition of a measure of complexity, called thermodynamic depth, can now be presented:

The <u>thermodynamic</u> <u>depth</u> of a macroscopic state b of a classical system B at time t given an initial distribution $p_0(x)$ over the microscopic states x of B at time t_0 is equal to

$$d_{p_0}(b) = \bar{S}(b) - k_B S_0(b) \quad ,$$

where the arbitrary constant of proportionality has been set to k_B, Boltzmann's constant, $\bar{S}(b)$ is the coarse-grained entropy of b, $= k_B \times \ln($ volume in phase space corresponding to b$)$, and $S_0(b)$ is the Shannon entropy of the set of trajectories that evolve into b, $= -\int p_0'(x) \ln p_0'(x)\, dx$, where $p_0'(x) = p_0(x) \,/\, (\int_{u^{-1}_{t_0 t}(b)} p_0(x)\, dx)$ if $x \in u^{-1}_{t_0 t}(b)$, $p_0'(x) = 0$ otherwise ($u_{t_0 t}$ is the evolution on phase space that maps states at time t_0 to their values at time t).

The quantum mechanical definition of thermodynamic depth is essentially the same, with Hilbert space substituted for phase space:

The thermodynamic depth of a macroscopic state b of a quantum mechanical system B at time t relative to an initial distribution of states described by a density matrix ρ_0 at time t_0 is equal to

$$d_{\rho_0}(b) = \bar{S}(b) - k_B S_0(b) \quad ,$$

where $\bar{S}(b)$ is the coarse-grained entropy of b, equal to $k_B \times \ln($ dimension of the subspace of Hilbert space corresponding to b$)$, and

$S_0(b) = -\mathrm{tr}\ \rho_0' \ln \rho_0'$ is the Shannon entropy of the states that evolve into b, where $\rho_0' = U^\dagger_{t_0 t} P_b U_{t_0 t} \rho_0 U^\dagger_{t_0 t} P_b U_{t_0 t} /$ tr(same) is the density matrix for the states that evolve into b ($U_{t_0 t}$ is the unitary evolution that takes states from t_0 to t; P_b is the projection operator onto the subspace corresponding to b).

We show below that the thermodynamic depth so defined satisfies the properties of continuity and additivity outlined above, and that the form of the measure is unique up to an arbitrary multiplicative constant. It is also a completely physical measure of complexity: the thermodynamic depth of a macroscopic state is equal to the difference between the coarse- and fine-grained entropies of the state given the underlying probability distribution. In addition, the measure satisfies the intuitive requirements on a measure of complexity discussed in the introduction.

One such requirement was that the complexity measure vanish for completely ordered and random systems. Systems at thermodynamic equilibrium can be described completely by a few intensive variables such as pressure, temperature and chemical potential, and by at least one extensive variable such as volume. Simple and well-known relationships hold between these variables: systems at thermodynamic equilibrium are not complex. The definition of thermodynamic depth reflects this fact. The thermodynamic depth of any state of a system relative to the equilibrium distribution (i.e., a uniform distribution over all the accessible phase or Hilbert space) is zero: at equilibrium, the coarse-grained entropy equals the fine-grained entropy, so that $\bar{S} = k_B S_0$ and $d_{p_0} = 0$. Systems in a highly ordered, regular state -- such as a salt crystal -- are also not very complex. The thermodynamic depth of a system in a pure state is zero: $\bar{S} = S_0 = 0$. In addition, systems close to a pure state, with $\bar{S}$ small, have small thermodynamic depth.

Another requirement of the complexity measure was that we not be able to proliferate complexity simply by making copies of a complex object. The thermodynamic depth of a number of identical copies of a complex object that can be simply copied is not much larger than the depth of the original object. The thermodynamic depth of a piece of human DNA and an identical copy is much less than the twice the thermodynamic depth of the piece of DNA alone. The property of additivity implies that the depth of the DNA and its copy is equal to the depth of the DNA (large, for DNA from a living creature) plus the depth of the copying process (small).

The thermodynamic depth of a macroscopic state of a system B depends implicitly on the experiments that have been made on the system to determine by what process the system evolved into that state; these "experiments" correspond to interactions between B and the measuring devices that define B's macroscopic state. Typically, measurements made on a gas close to equilibrium determine very little about the way in which the gas has approached equilibrium; the fine-grained entropy of the gas is on the order of the coarse-grained entropy, and its thermodynamic depth is small. If, however, evidence exists from measurements made at an earlier time that the gas was far from equilibrium -- say, for example, that there was a temperature gradient from one side of the gas to the other -- then the thermodynamic depth of the final, equilibrium configuration of the gas is large. The existence of experimental evidence that the gas started out in a particular state far from equilibrium implies that if one performed detailed enough measurements on the state of the individual molecules of the gas after the temperature has become the same throughout the gas, one could find evidence of the original temperature gradient. At a microscopic level, the motions of the molecules contain subtle and complex correlations that point back to their original, non-equilibrium state. The thermodynamic depth of a final state of the gas depends not only on the characterization of the state itself, but on the experimental evidence of the process by which it got there: the Shannon entropy of the set of trajectories of the final state of a system is equal to the conditional entropy of the system given the state of the measuring devices with which it has interacted. The more precisely the process has been defined, the smaller the fine-grained entropy, and the greater the thermodynamic depth.

If we take a system given by the gas at equilibrium, together with all the measuring devices that register the previous, non-equilibrium state of the gas, then the gas is thermodynamically deep, because evidence exists that points to the gas's detailed causal history. If we take a system given by the gas at equilibrium alone, with no measuring devices that specify the gas's causal history, the gas is thermodynamically shallow. Thermodynamic depth thus satisfies another intuitive property of complexity. Colloquially, a system that appears simple at first glance may be complex on closer inspection. If one has only measured the temperature and pressure of the gas, the gas is thermodynamically shallow. If one has performed measurements to determine more precisely the microscopic state of the gas, and acquired evidence of the gas's detailed causal history, the gas is thermodynamically deep. In contrast, no matter how many measurements one has made on a crystal, the crystal remains thermodynamically shallow, since the amount that one can decrease the fine-grained entropy of the crystal through measurement is limited by the crystal's coarse-grained entropy, which is already small.

Example:

Since the only information that a measuring device outside a black hole can get about the hole is its mass, charge and angular momentum,[15] the fine-grained entropy of the hole is always equal to the coarse-grained entropy of the hole (which is now no longer the equilibrium thermodynamic entropy since black holes have negative specific heat and cannot be in equilibrium). The thermodynamic depth of a black hole is accordingly zero.

Almost any interaction between a system B and another system reduces B's fine-grained entropy: the fact that the other system is in a particular state after the interaction restricts the set of microscopic states accessible to B. We can define the amount that a particular subset of interactions contributes to the thermodynamic depth of a state b of B to be equal to the amount that they reduce B's fine-grained entropy above and beyond the amount that it is reduced by the other interactions that B has undergone. We can thus isolate the amount that different elements of a process that result in a state contribute to the state's thermodynamic depth. For example, when a computer performs a computation, we can separate out the amount that the logical elements contribute to the thermodynamic depth of the result of the computation from the contribution of irrelevantly generated heat by looking at how much the setting of the bits in the input program reduced the fine-grained entropy of the result given the level to which the fine-grained entropy had already been reduced by the other interactions that constructed the computer in the first place.

Additional properties of thermodynamic depth

Suppose that the initial distribution of states is determined by the results of some macroscopic measurement, so that $S_0(b) = \bar{S}(c\ldots d)$, where $c\ldots d$ are the values of the degrees of freedom determined by the measurement and $\bar{S}(c\ldots d)$ is the thermodynamic (coarse-grained) entropy of c...d. The thermodynamic depth of b is then equal to the increase in thermodynamic entropy in going from c...d to b.

Note that the number of bits needed to describe a macroscopic state b in a maximally efficient coding is[13] $I(b) = 1/(k_B \ln 2)\ (\bar{S}_{tot} - \bar{S}(b))$, where $\bar{S}_{tot}$ is k_B times the logarithm of the volume of B's entire phase space. Similarly, the number of bits of information needed to describe the initial distribution $p_0'(x)$ of states that evolve into b is $I_0(b) = (1/\ln 2)\ (\ (1/k_B)\bar{S}_{tot} - S_0(b)\ )$. The thermodynamic depth of a state b is proportional to the difference between the amount of information needed to describe the initial distribution of states, and the amount of information needed to describe b itself:

$$d(b) = \bar{S}(b) - k_B S_0(b) = (k_B \ln 2)\ (\ I_0(b) - I(b)\ ).$$

For any system, such as a Hamiltonian system, whose evolution preserves volume in phase or Hilbert space, the fine-grained (Shannon) entropy of the initial distribution is conserved. If the macroscopic state b is determined by fixing some degrees of freedom of such a system, and if the initial distribution of states that evolve into b has entropy S, then after the system has evolved into b, the entropy of the remaining degrees of freedom must still be S. The different degrees of freedom of a given system can be regarded as interacting subsystems, and all the theorems of the previous chapter concerning the generation and destruction of correlations and the increase and decrease of conditional entropies apply. The thermodynamic depth of a state b is equal to the amount of entropy that the system has pumped from the "relevant" degrees of freedom (those that need to be constrained for the system to evolve into b) into "irrelevant" degrees of freedom (the remaining ones) in the course of constructing the state b. Equivalently, as the relation $d(b) = (k_B \ln 2)\ (\ I_0(b) - I(b)\ )$ suggests, the thermodynamic depth of a state is equal to the amount of information about the processes that resulted in b that now resides in degrees of freedom other than those fixed by b.

We now show that the requirements of continuity and additivity given above imply that the Shannon entropy form for the measure is unique up to a multiplicative constant. We consider separately the form for deterministic and stochastic physical systems.

i) Deterministic case:

The macroscopic states of a deterministic system can be ordered in a hierarchical tree (see figure 1), in which the nodes represent states and the lines between the nodes connect each state with the state below it into which it evolves (this hierarchical tree should in no way be confused with the hierarchical trees of Huberman and Hogg,[16] and of Bachas and Huberman;[17-18] their trees represent the height of potential barriers between, say, the energy minima of a spin glass. The tree given here represents the evolution of a dynamical system). The root of the tree corresponds to the state whose complexity is to be evaluated. The states at the top of the tree correspond to the states of the initial distribution that evolve into that state. The weights of the states at the top are given by experiment, and the weights of the states below are assigned so that the weight of a given state is the sum of the states immediately above it. Normalize the distribution so that the weight of the root is one. Call the root (0). Call the states at the first level (1), (2), ... where (1) is the state at the end of the first branch, (2) is the state at the end of the second branch, etc. Call the states at the second level (1,1), (1,2), ... , (2,1), (2,2), ... where (1,2) is at the end of the second branch off of the state (1), (2,1) is at the end of the first branch off of the state (2), etc. Continue so numbering to the top. Let p(0,...,i) be the weight of the state (0,...,i) at the top of the tree, and let all weights below be assigned so that the weight of a state is equal to the weights of the states directly above it. Then it is a theorem due to Shannon[13] that the only measure $\mu(p(0,\ldots,1), \ldots , p(0,\ldots,n))$ that is a continuous function of the p(0,...,i) and satisfies the previously described additivity requirement that takes the form,

$$\mu(p(0,\ldots,1), \ldots , p(0,\ldots,n)) =$$

$$p(0,1)\mu(p(0,1,\ldots,1)/p(0,1), \ldots , p(0,1,\ldots,n_1)/p(0,1)) +$$

$$\ldots + p(0,m)\mu(p(0,m,\ldots,1)/p(0,m), \ldots , p(0,m,\ldots,n_m)/p(0,m)) + \ldots$$

(where n is the number of initial states; n_1 is the number of initial states that evolve into the state (0,1), . . . ; n_m is the number of initial states that evolve into the state (0,m), etc.) is

$$\mu(p_1, \ldots , p_r) = - k (p_1 \ln p_1 + \ldots + p_r \ln p_r) \quad , \text{ for}$$

arbitrary probabilities $p_1,\ldots, p_r$ such that $p_1+\ldots+p_r=1$; k is an arbitrary constant.

An example of additivity is given in figure 2. There are five possible initial states, four with probability 1/6 and one with probability 1/3. There are two possible intermediate states, each with probability 1/2. There are five different ways of getting to the final state, to which the Shannon entropy assigns the measure $-((4/6)\ln(1/6) + (1/3)\ln(1/3)) = (2/3)\ln 2 + \ln 3$. There are three different ways of getting to the first intermediate state, with Shannon entropy $\ln 3$, and two ways of getting to the second, with Shannon entropy $-(2/3)\ln(2/3) - (1/3)\ln(1/3)$. There are two ways of getting from the intermediate states to the final state, with Shannon entropy $\ln 2$. The property of additivity requires that the Shannon entropy of the total number of ways of getting to the final state be equal to the average Shannon entropy of getting from the initial states to the set of intermediate states, plus the Shannon entropy of getting from the intermediate states to the final state:

$$(2/3)\ln 2 + \ln 3 = (1/2)\ln 3 + (1/2)\{-(2/3)\ln(2/3) - (1/3)\ln(1/3)\} + \ln 2.$$

ii) Stochastic case:

In the deterministic case, the weights of the initial states determine a set of weights for the different possible trajectories that the system may take in going from the distribution of initial states to the final state: each initial state determines a trajectory, and the measure of the complexity of the final state is then just the Shannon entropy of the weighted set of trajectories. When a system evolves probabilistically instead of deterministically, the evolution of its states can no longer be represented immediately by a hierarchical tree. However, the weighting of the initial states together with the form of the stochastic evolution of the system still defines a weighting of the different trajectories that the system can take that result in the state in question (specifically, the form of the evolution determines the probability $p(c_1 c_2 \ldots c_n b)$ that the system follows the sequence of states $c_1 c_2 \ldots c_n b$ given an initial state c_1, and the weighting of the initial states gives us the probability $p(c_1)$ of c_1; the weight of the trajectory $c_1 c_2 \ldots c_n b$ is then $p(c_1)\ p(c_1 c_2 \ldots c_n b)$). In order to satisfy the requirements of continuity and additivity as defined above, the measure of complexity of the process that results in b must once again be the Shannon entropy of the set of trajectories that result in b with total weight normalized to one.

The stochastic system over the given degrees of freedom is equivalent to a deterministic system over the given degrees of freedom and a set of "hidden" degrees of freedom that label the different possible trajectories. The initial distribution over the hidden degrees of freedom is chosen to produce the correct weights for the different trajectories. By the argument of the previous section, the only measure for the deterministic system that is continuous over the probabilities of the various trajectories and additive over intermediate distributions is proportional to the Shannon entropy of the set of trajectories. But the deterministic system over the original plus hidden degrees of freedom is completely equivalent to the stochastic system over the original degrees of freedom. Hence the unique measure for the stochastic system is proportional to the Shannon entropy of the set of trajectories.

The entropy of the set of trajectories that result in a particular state is the only continuous and additive measure of that state's complexity relative to the original distribution of states. This result holds for any dynamical system, stochastic or deterministic, dissipative or conservative.

An additive, continuous measure of complexity for a given macroscopic state b must be proportional to the Shannon entropy of the set of trajectories with probabilities p_i determined by experiment, plus a linear combination of terms proportional to the Shannon entropies of the set of trajectories with other, arbitrary probability distributions. These additional terms do not spoil the continuity of the measure, since they do not depend on the p_i, and obey the additivity property. In particular, the addition of the thermodynamic entropy $\bar{S}(b)$ to the definition of logical depth does not spoil the additive properties of the measure, since $(1/k_B)\bar{S}(b)$ is the Shannon entropy of the trajectories that evolve into b given the uniform measure over the underlying states. If the measure is to assign equilibrium states depth zero, the only terms that can enter are $\bar{S}(b)$ and $S_0(b)$, with coefficients K and $-Kk_B$ respectively, where K is an arbitrary overall multiplicative constant. Setting K equal to one, we arrive at the unique form for the thermodynamic depth, $d(b) = \bar{S}(b) - k_B S_0(b)$.

Generation of entropy during computation

In the course of a calculation a computer performs two basic processes over and over again. 1) It copies information from one place to another (e.g., from the memory to the central processing unit, or CPU, and vice versa). 2) It transforms information using logic circuits (e.g., taking the *and* of two bits, multiplying two floating point numbers).

To copy information is to generate correlation, or mutual information. To transform information using logical operations is to cull the information that one needs from the information that one doesn't need. An *and* gate takes two inputs, each of which can have the values True or False, 0 or 1, and gives only one output, True if both inputs are True, False otherwise. If the output of an *and* gate is False, then one can not tell whether the inputs were True-False, False-True, or False-False.

It is this action of separating information to be used later on in the program from information that is no longer needed that in general generates entropy. If the underlying dynamics of a system are one-to-one, as they are for all known systems at a microscopic level, then two initially distinct states can not evolve into a single state without creating some difference elsewhere. When an *and* gate operates, it discards some information about the inputs: this information must show up elsewhere. Suppose that the *and* gate is realized electronically by means of transistors. If the first input is True and the second input False, then we can not reconstruct the input from the output, False, alone. If, however, we could examine in sufficient detail the motions of the electrons in the transistors after the gate has operated, we would be able to discover some trace of the fact that the inputs were True-False, not False-True or False-False. The information discarded by the <u>and</u> gate has gone into the motions of the electrons at a microscopic level: the <u>and</u> gate, in operating, has generated entropy. Given an initial distribution over the microscopic states of the transistors together with power supply that realize the <u>and</u> gate with inputs True-False, the thermodynamic depth of the output state, False, one machine cycle later, is at least $k_B \ln 3$.

Example:

A program that adds 11011010 to 11001 (binary) first copies the numbers from the program stored in memory to the inputs to the adder in the central processing unit. The CPU then adds the numbers together:

$$\begin{array}{r} 11011010 \\ +\ \underline{00011001} \\ 11110011 \end{array}$$

In general, there are 2^8 pairs of numbers that can be added together to give a particular eight bit number: the addition process begins with two eight bit numbers, finishes with one eight bit number, and has discarded eight bits of information. Given an initial distribution over the microscopic states of the CPU that realizes the two input numbers, the thermodynamic depth of the result of the addition is at least $8k_B \ln 2$.

The two processes, creation of copies and generation of junk information, have analogues in any physical system with more than one degree of freedom. The values of the various degrees of freedom of a system represent information no less than a particular orientation of the magnetic field in one element of a magnetic core memory. Creation of copies corresponds to an increase in mutual information or correlation between different degrees of freedom, while the generation of junk information corresponds to the shunting of information from a "relevant" degree of freedom to another, "irrelevant," degree of freedom. In real physical systems, of course, information is not necessarily processed according to Boolean algorithms. Exactly how the states that represent information transform depends on the Hamiltonian of the system. Regardless of how a given system transforms information locally, however, the amount of information processed globally can be measured in terms of the correlations generated and the unneeded information shunted away.

Example:

There are many ways that a computer can calculate a given result. Since <u>and</u> and <u>or</u>, <u>nand</u> and <u>nor</u>, addition and multiplication are many-to-one operations, one can stop a program at any stage, change the inputs of the previous operation in any way that one likes so long as the operation gives the same result as the original program, and have a way of obtaining the final result that is different (albeit stupidly so) from the way given by the initial program. Any device that can realize the logical functions of a computer, in going from one of a set of inputs that lead to a given final result after a certain amount of time, must generate on average a gross amount of junk information equal to the logarithm of the total number k of inputs that lead to that result in that amount of time; the thermodynamic depth of the result given a distribution over the microscopic states of the computer that realizes the given input is at least $\ln(k)$. In practice, the amount of junk information generated by the "logic" degrees of freedom is roughly equal to $\ln 2$ per elementary logical operation performed. A computer that performs a calculation of d logical steps has thermodynamic depth of at least $\approx k_B d(\ln 2)$. Of course, by profligately generating excess entropy at every turn, a computer can generate large amounts of thermodynamic depth amongst its microscopic degrees of freedom in addition to the depth required by the logical operations. As noted above, we can extract the contribution to the thermodynamic depth due to the operation of the logical elements; even in a profligate computer, that has just completed a 10^6 step calculation while raising the temperature of the room by ten degrees, the contribution of the logical degrees of freedom to the thermodynamic depth is $\approx 10^6 k_B \ln 2$. In addition, it is possible in principle to build a so-called "reversible" computer that does not shunt information generated by logical operations into microscopic degrees of freedom and that generates no extraneous thermodynamic depth: such a computer has thermodynamic depth $\approx k_B d(\ln 2)$ for the result of a d step calculation.

Example:

In an n-piece one dimensional puzzle the edges of the pieces are labeled with patterns of bits. Two pieces fit together if their patterns of bits are complementary (see figure three). Each edge fits onto exactly one other edge. Any process that assembles the puzzle must check pairs of bits to see whether they are equal or not. Any physical system that realizes this checking process must generate junk information, since the inputs 1-0, 0-1 both generate the same output, Fit, while 0-0 and 1-1 both generate the output, No Fit. For each bit checking operation, the system generates junk information ln2. To compare the edges of two pieces to see whether they fit, one need only proceed until one finds a mismatch. Each comparison generates on average junk information $(1/2)\ln 2 + (1/4)2\ln 2 + (1/8)3\ln 2 + \ldots + (1/2^b)b\ln 2$, where b is the total number of bits on each edge. If b is large, the average amount of junk information generated per comparison is approximately equal to $2\ln 2$. In finding the piece that matches the first piece that one picks up, one has to compare $(n-1)/2$ edges, on average. Once the first two are matched, one must compare $(n-2)/2$ edges on average to find the next match. The total amount of junk information generated on average in assembling the puzzle is $d = n(n-1)\ln 2$. In any machine that assembles the puzzle, the elements that perform the comparisons between pieces contribute $k_B d(\ln 2)$ to the total thermodynamic depth of the result. $k_B d(\ln 2)$ is a lower limit on the thermodynamic depth of the puzzle as assembled by any machine.

Example:[19-20]

Newton's method for finding the zeros of an arbitrary polynomial $f(x)$ over $x \epsilon C$ is given by the iteration of the following procedure. Pick a point $x_1 \epsilon C$, construct the tangent to f at x_1, call the point where it hits the x-plane $x_2 = T(x_1)$. Repeat the procedure starting with x_2. The global action of this algorithm gives a map from C into C. Smale has shown that for a slightly modified version of Newton's method, the probability can be made arbitrarily close to 1 that starting with a randomly chosen x_1, the process converges to an approximate zero of f after a finite number of iterations. (An approximate zero of f is a $z_0 \epsilon C$ such that $z_n = T^n(z_0)$ is well-defined for all n, z_n converges to z^* as $n \to \infty$ where $f(z^*)=0$, and $|f(z_n)/f(z_{n+1})| < 1/2$ for all $n=1,2,\ldots$.) For large n the map T^n is thus a compressive map on all the complex numbers except a set of measure zero. If the map T is realized on some finite interval γ by a dynamical system such as an analog computer, the amount of information discarded in n iterations as the system pursues the zeros of f is $\ln\{ \mu(T^n(\gamma))/\mu(\gamma) \}$ (where μ is the usual area measure on the complex plane), which, multiplied by k_B, gives a lower limit to the thermodynamic depth of the result of the first n iterations as performed by any device, digital or analog. When in iterating T the system reaches an approximate zero z_0 of f, the amount of information shunted out becomes greater than ln2 per iteration, and the thermodynamic depth increases by more than $k_B \ln 2$.

Thermodynamic depth and mathematical complexity

Recall that the complexity of a mathematical result is identified with the "cost" of calculating the result from a given formulation.[5-7] If the calculation is performed by a physical device, the measure of thermodynamic depth applies. Any machine that discards d bits of information in obtaining a result must start out with fine-grained entropy at least d(ln2) less than the coarse-grained entropy of the result; a result that is obtained through d elementary logical operations has thermodynamic depth $\geq k_B d(\ln 2)$. If we assign each elementary logical operation a cost proportional to the amount of junk information that it discards, then the mathematical complexity is proportional to the the amount that the physical elements that perform the logical operations contribute to the total thermodynamic depth of the result as calculated by any device. Equivalently, the mathematical complexity is proportional to the thermodynamic depth of the result as calculated on <u>the most efficient device possible</u> -- that is, one that generates no more junk information than that required by the form of the logical operations. The mathematical complexity of a result is equal to the minimum thermodynamic depth of the result "calculated" from the initial conditions by all possible physical systems.

Thermodynamic depth and algorithmic complexity

One can also give an analogue to the shortest program that results in a particular state.[1-2] Given a system B and a probability distribution $p_t(x)$ over the microscopic states of B at time t, the <u>most</u>

plausible cause of a state b of B at t is the state $c_0(b)$ at the most recent time $t_0 \leq t$ that evolves into b given the induced probability distribution $p_0(x)$ at time t_0 and that maximizes the coarse-grained entropy $\overline{S}(c_0)$.

Note that if B is at equilibrium at time t, then the most plausible cause of any state is the state itself (b evolves into itself in no time, and maximizes the coarse-grained entropy among macroscopic states that evolve into b: any state c that evolves into b given an underlying equilibrium distribution has $\overline{S}(c) \leq \overline{S}(b)$). Note also that since $I(c) = (1/k_B \ln 2)\ (\ \overline{S}_{tot} - \overline{S}(c)\ )$ the most plausible cause of b is the state that evolves into b and that requires the fewest number of bits to describe. The algorithmic complexity of a result is the length of the shortest program that can calculate the result. The physical analogue of algorithmic complexity for a macroscopic state b is $I(c_0(b))$ -- the number of bits required to describe b's most plausible cause. The physical analogue of algorithmically random number is a state whose coarse-grained entropy equals its fine-grained entropy: the physical analogue of algorithmic randomness is thermodynamic equilibrium.

Thermodynamic depth and logical depth

The connection between computation and the discarding of information also allows us to relate the thermodynamic depth of a number calculated on a given computer to its logical depth. The logical depth of a number relative to a given computer is equal, roughly, to the number of machine cycles that it takes the computer to calculate the number starting from the shortest possible program.[8-10] The thermodynamic depth of the number is approximately $k_B \ln 2$ times its logical depth, given the most efficient computer preprogrammed with the shortest program to calculate the number.

More precisely, a result has logical depth d with significance b if any program that can calculate the result in fewer than d steps can itself be expressed as the output of a program b bits shorter than the original program. This refinement in the notion of logical depth makes the definition robust and reasonably machine independent. We will connect thermodynamic depth to a closely analogous alternative definition of logical depth: a result has logical depth d with significance b if the programs that calculate the result in fewer than d steps contribute a factor of less than 2^{-b} to the results total algorithmic probability.

To connect the notion of logical depth to thermodynamic depth, suppose that we have some system whose evolution realizes the operation of a Turing machine. The initial distribution of microscopic states of the system is such as to allow the machine to perform the required data manipulations (moving the tape, copying and erasing data, etc.), and to give a uniform distribution over all one-sided sequences of bits on the input tape of the machine. The initial weight of a sequence in which the first n bits are fixed at particular values and the remaining bits can vary is then 2^{-n}.

The logical depth of a result is measured in machine cycles. As seen above, a digital computer that derives a result through d elementary logical operations discards about d(ln2) worth of information on average, and has thermodynamic depth $\geq k_B d(\ln 2)$. Thermodynamic depth gives a measure of the complexity of constructing a particular result at a particular time, starting from a given distribution, while logical depth gives a measure of the complexity of constructing the result at an indeterminate time, starting from a uniform distribution over the space of input sequences. Thermodynamic depth can straightforwardly encompass a definition analogous to logical depth by defining a macroscopic state (determined by fixing the first m bits of the output tape) to have thermodynamic depth d with significance b if the set of initial macroscopic states with respect to which the result has thermodynamic depth of less than d (given the initial distribution restricted to any one of those states) contributes a factor of less than 2^{-b} of the total probability that the final state arises in the first place. With this definition a result's logical depth becomes proportional to its thermodynamic depth relative to any device that is set up to calculate in a maximally efficient fashion, creating only the minimum amount of junk information.

Reversible computers

In the previous chapter a number of theorems were proved on the generation of mutual information shared between two systems, and on the transfer of information from one system to another. In particular, these theorems imply that one can not decrease the entropy of one system without increasing the entropy of another unless the two systems are correlated to begin with. That is, one can not erase junk information except by creating junk information elsewhere, or by arranging an interaction between the degrees of freedom that register the junk information and the degrees of freedom that register the useful information.

In principle it is possible[21-28] (using Fredkin gates, for example) to build computers that do not relegate the information that is not needed to the microscopic level, but record it instead. If one copies the useful information to memory and then runs the computer backwards, the junk information can be erased reversibly, foregoing an increase in entropy. It might seem that the in principle existence of reversible computers that can perform complex calculations without increasing entropy provide a counterexample to the definition of complexity given. This is not the case.

For a reversible computer there exists exactly one program that starts with code on part of the input tape and blanks on the rest, and that results after d steps in a specified output on the first n bits of the output tape and blanks on the rest. However, in general there exist many programs that result after d steps in the specified output on the first n bits of the output tape and other, junk, information on the remainder. In fact, since the logical operations on a reversible digital computer are the same as on a conventional digital computer, the number of programs that result in a specified output on the first n bits of the output tape and unspecified junk on the rest must be roughly the same for both. A reversible computer simply saves the information that is no longer needed to complete the calculation while a conventional computer discards it. Since the definition of thermodynamic depth depends only on the coarse- and fine-grained entropies of the specified output and not on the exact state of the other degrees of freedom of the computer -- in particular, it does not depend on whether the remaining bits on the output tape are all zero or random -- the thermodynamic depth of a particular computational result on a reversible computer is about the same as the thermodynamic depth of the result on an optimal conventional computer, that discards no more information than it has to to derive the result.

The thermodynamic depth of a result on a reversible computer depends on all trajectories that lead to that result, not merely the unique trajectory that cleans up after itself. The trajectory that reversibly erases the junk information created has special properties: in following such a trajectory the computer must systematically destroy mutual information, and as will be shown, natural processes that decrease mutual information substantially are few and far between. Mutual information behaves like (the negative of) Boltzmann's H-function: if mutual information starts at a low level, for almost all interactions it rises to a high level, and stays high for a long time with small downward fluctuations. In the next section, two theorems will be exhibited that imply that for systems with many degrees of freedom, the amount of time that one has to wait for mutual information to be erased is of the order of the Poincare recurrence time.

The reversible erasure of junk information is not generally found in nature (or in conventional computation) for the simple reason that it is not necessary to erase the junk information to obtain the result. Since the junk information is by definition that which is not needed to obtain the end result, whether it is erased or not is irrelevant to obtaining the result. The possibility of reversible computation is interesting for systems that have strong constraints on how much entropy they can generate, however; when close to equilibrium and subject to strong constraints on how much negentropy is available, complex systems must be ecologically minded.

Thermodynamic depth and mutual information

The amount of computational effort that it takes to obtain a result can be identified with the number of bits of junk information created, which by simple counting is equal to the number of bits of information copied (= the total amount of mutual information created) minus the difference between the length of the final result and the input program (= the amount of useful information generated). For a computer, then, the thermodynamic depth of a result can be identified either with the amount of junk information created, or with the amount of information copied. The analogue of this result for physical systems in general can be derived from theorem 5 of chapter one, which we restate here:

Theorem:

Given a macroscopic state b of B at time t determined by the results of inexact measurements on each of the degrees of freedom $l_1,\dots,l_m$ of B, and given any initial distribution $p_0(l_1\dots l_m)$ at time t_0, we have

$$d_{p_0}(b) \geq k_B I_t(l_1,\dots,l_m) \quad ,$$

where $I_t(l_1,\dots,l_m) = S_t(l_1) + \dots + S_t(l_m) - S_t(l_1\dots l_m)$ is the mutual information shared between $l_1,\dots,l_m$ at time t.

The thermodynamic depth of b relative to any initial distribution is bounded below by Boltzmann's constant times the mutual information between the degrees of freedom of B not completely fixed by b. The mutual information is maximized for $k_B S_t(l_i) = \bar{S}(l_i)$, at which values $d(b) = k_B I$. In the next section theorems are proved that imply that for most interactions between degrees of freedom, their mutual information tends to rise to near its maximum value and stay there to within fluctuations. The previous theorem implies that as the mutual information rises, the thermodynamic depth tends to rise as well, although increase in correlation need not accompany an increase in thermodynamic depth.[29]

Bennett has suggested the following counterexample to the idea that mutual information is a good measure of complexity: Take a glass and smash it with a hammer. Keep on smashing, and you will in short order create a mass of shards that contains an amount of mutual information greater than that between all the DNA in the human body. Since a bowl of glass dust ought not to be more complex than human genes, Bennett argues, mutual information ought not to be considered a measure of complexity. The problem with this counterexample is that the complexity of the genes in the body of a given human being is far greater than just the mutual information that they possess <u>amongst</u> <u>themselves</u>. By virtue of the extensive process of evolution by which these genes came into existence, sequences of nucleotides in the DNA of the human genome are highly correlated not only with sequences in the DNA of other human beings, but with sequences in the genome of any other living thing on the planet Earth. The structure of the DNA of living things is in turn highly correlated with the features of the environment. In short, in the course of the process that resulted in a particular piece of DNA in a human cell, a whole lot of mutual information was produced, accompanied by a correspondingly large production of thermodynamic depth. This is not to say that a smashed up piece of glass is not complex (try reassambling it as a puzzle): it simply cannot compare in thermodynamic depth to a sequence of human DNA.

<u>Transfer of information and creation of mutual information</u>

Thermodynamic depth identifies the complexity of a state of a physical system with the amount of information processed in the course of constructing that state. So far, we have concentrated on how information is processed in computers. In this section we derive results on how information is exchanged between the parts of physical systems in general, not simply systems that evolve according to Boolean algorithms. Specifically, we show that almost all interactions between two or more initially uncorrelated degrees of freedom tend to transfer information from one to the other; in addition, almost all interactions tend to produce mutual information between degrees of freedom, and that mutual information, once created, tends to stay at a high level until the Poincare recurrence time of the system in question.

i) Discrete classical systems:

Given a system AB with degrees of freedom A and B and with possible states (a_i,b_j), $1 \le i \le m$, $1 \le j \le n$, given initial probabilities $p(ij) = p(a_i,b_j)$, $\sum_{i,j} p(ij) = 1$, $p(i) = p(a_i)$, $p(j) = \sum_i p(ij)$, (the index i will always refer to a's, the index j to b's), and given a one-to-one evolution $t: (a_i,b_j) \rightarrow (a_{i'},b_{j'})$, where t is a member of the group of permutations of mn objects, $t \epsilon S^{mn}$, $t(ij) = i'j'$, then there is a natural action of t on the probability distribution $p(ij)$: $t: p(ij) \rightarrow p_t(ij)$, where $p_t(ij) = p(t^{-1}(ij))$, with a corresponding action on p_i and p_j: $p_t(i) = \sum_j p_t(ij) = \sum_j p(t^{-1}(ij))$, $p_t(j) = \sum_i p_t(ij) = \sum_i p(t^{-1}(ij))$.

The average amount of information transferred from B to A in the course of the interaction is got by looking at the initial probability distribution $p(ij) = (1/n)\delta_{ii_0}$ for some initial state of A i_0: i.e., A is fixed at a given value i_0, while B can be anything. The amount of information that B transfers to A in the course of the interaction is equal to $S_t(A)$ after the interaction, since any variation in the value of A after the interaction is due to the initial variation of B. $S_t(A)$

depends both on the value of i_0 and on the form of the joint evolution t; however, we will now show that for any i_0 and for most $t \epsilon S^{mn}$, $S_t(A)$ is close to its maximum possible value, which is equal to ln(m) and arises when $p_t(i) = 1/m$ when $m \leq n$ and which is equal to ln(n) when $m \geq n$. This result follows from the following theorems:

Theorem 1:

Given a system determined by degrees of freedom A and B initially characterized by a probability distribution over their joint values $p(a_i, b_j) = p(ij)$, $1 \leq i \leq m$, $1 \leq j \leq n$, as above, the average value of $p_t(i)$ as t ranges over S^{mn} is 1/m, and the standard deviation of $p_t(i)$ from this value is

$$\left\{ \frac{n(m-1)}{m(mn-1)} \left(\sum_{i,j} (p(ij))^2 - \frac{1}{mn} \right) \right\}^{\frac{1}{2}} .$$

(Proof in Appendix A.)

A stronger result can be derived if the initial probability distribution is one in which the value of A is fixed at a particular value i_0 and the value of B is arbitrary:

Theorem 2:

Given degrees of freedom AB with possible values $(a_i b_j)$, $1 \leq i \leq m$, $1 \leq j \leq n$, $n >> m$, and initial probability distribution $p(a_i b_j) = (1/n)\delta_{i_0 i}$, and given an interaction between A and B, corresponding to a joint evolution $t \epsilon S^{mn}$ selected at random, then the probability that $p_t(a_i)$

take on a particular set of values $p_t(a_i) = p_i$ after the interaction is proportional to e^{nS}, where $S = -\sum_i p_i \ln p_i$.

Proof:

t selected at random permutes the mn probabilities $p(a_i b_j)$ at random, and adds at random the n initial probabilities with values 1/n to the m "bins" corresponding to the m possible values of a_i. The value for $p_t(a_i)$ after the interaction is just n_i/n, where n_i is the number of non-zero probabilities that have been added to the i-th bin. The probability that $p_t(a_1),\ldots,p_t(a_m)$ take on the values $p_1=n_1/n,\ldots,p_m=n_m/m$ after the interaction is proportional to $n!/(n_1!\ldots n_m!)$, as in the Ehrenfest urn model.

$$\begin{aligned} \text{But } \ln(n!/(n_1!\ldots n_m!)) &\approx n\ln(n) - n - n_1\ln(n_1) + n_1 - \ldots - n_m\ln(n_m) + n_m \\ &= -n\sum_i (n_i/n)\ln(n_i/n) \\ &= nS. \end{aligned}$$

The probability that $p_t(a_i)$ take on the values p_i is proportional to e^{nS}.

If the initial probability distribution $p(ij) = (1/n)\delta_{ii_0}$, as above, then the first theorem above implies that the standard deviation of $p_t(i)$ from 1/m is

$$(m-1)/\{m(mn-1)\}^{\frac{1}{2}} \approx 1/\sqrt{n} \quad \text{for } m,n>>1.$$

For $n>>m^2$, this result implies that $p_t(i) \approx 1/m$ and that $S_t(A)$ is close to its maximum value, ln(m). The second theorem implies that for n>>m, the amount of information transferred from B to A behaves exactly like the negative of Boltzmann's H-function: if it starts out at zero it is very likely to rise to close to its maximum value and stay there for a long time, occasionally making small downward fluctuations. For m>>n, $p_t(i)$ is a distribution in which almost all of the $p_t(i)$ that are not zero are equal to 1/n: $S_t(A)$ is again close to its maximum possible value given the initial distribution, this time ln(n).

Example:

Suppose that A has two possible values, labeled 1 and 2, and B has n, n>>2. Suppose that A is initially fixed at 1, uncorrelated with the value of B, about which nothing is known. I.e., the initial probability distribution for AB is p(1j) = 1/n, p(2j) = 0, for all j. After interaction, we expect $p_t(1)$, $p_t(2)$ to be (inserting m=2, n=n in the formula above) $1/2 + \{ \frac{1}{4(2n-1)} \}^{\frac{1}{2}} \approx 1/2 + (1/8n)^{\frac{1}{2}}$, $1/2 - \{ \frac{1}{4(2n-1)} \}^{\frac{1}{2}} \approx 1/2 - (1/8n)^{\frac{1}{2}}$. The value of A after the interaction is likely to be almost completely uncertain: A has gotten close to the maximum possible information about the initial state of B. As the interaction continues, the expectation values and standard deviations for p(1) and p(2) remain the same: the information that B has transferred to A remains close to its maximum value. If n is large, the chance that p(1) and p(2) will deviate by a substantial amount from 1/2 is small; we will have to wait for a long time for A to get rid of the information that B has saddled it with. The only way to be sure that A and B become uncorrelated again is to wait for the Poincare recurrence time -- on the order of 2n iterations for a typical $t \epsilon S^{mn}$.

Example:

Given k interacting binary systems, with states 0...00, 0...01, 0...10, etc., take an initial distribution in which all of the systems are initially in the state 0 except for the last, which has a fifty-fifty chance of being in either 0 or 1. I.e., p(0...00)=1/2, p(0...01)=1/2, all other probabilities zero. After the systems have interacted, we expect probabilities for any of the systems:

$$p(0) = 1 - p(1) = 1/2 \pm \{\ 2^{k-2}/(2^{k}-1)\ (1/2 - 1/2^{k})\ \} \approx 1/2 \pm 1/2\sqrt{2}\ ,$$

for $k \gg 1$. The fact that the standard deviation is less than 1/2 indicates that a substantial fraction of the systems are now correlated with the initial state of the kth system, and <u>ipso</u> <u>facto</u> with eachother. The following argument shows that roughly half are so correlated.

Under a transformation $t \epsilon S^{2^{k}}$, the states 0...00 and 0...01 go to t(0...00) and t(0...01). If we look at the binary representations of t(0...00) and t(0...01) and compare them bit by bit then the number of systems that have p(0) = p(1) = 1/2 after the interaction is equal to the number of bits of t(0...00) and t(0...01) that don't match. If t is selected at random from $S^{2^{k}}$, the number of bits that don't match is k/2 on average, and about k/2 of the systems will become correlated with the initial state of the last system and with eachother; by transferring a bit of information to each of k/2 other systems, the interaction has created mutual information $(k/2 - 1)\ln 2$ between them. The probability that none of the first k-1 systems will become correlated with the initial state of the last system is $2^{-(k-1)}$. Once again the average amount of time that we have to wait for the information created to be destroyed is on the order of the Poincare recurrence time, $2^{(k-1)}$ iterations.

Of course, if k is small, the Poincare recurrence time is short.

Example: Fredkin gates

Take k=3 in the previous example, and let t be the permutation on three bits that does not change the value of the third bit, and that switches the values of the first and second bit if the third bit is 1 and leaves them alone otherwise. t represents the operation of a Fredkin gate,[26] a gate that can realize various logical operations reversibly, without shunting unneeded information to the microscopic level. If the first two bits are 01 initially, and if the third bit can be either 0 or 1 with probability 1/2, then t generates maximal mutual information between each of the first two bits and the third. Since $t^2 = 1$, operating the Fredkin gate a second time brings one back to the initial probability distribution and zero mutual information.

The ability of Fredkin gates to destroy mutual information without consigning it to the microscopic level allows the possibility of reversible computers that can calculate without increasing entropy in principle. In practice, however, almost any interaction between two or more degrees of freedom generates hard-to-destroy mutual information, and macroscopic information devices generate entropy whether they want to or not by engendering correlation between the computing degrees of freedom and microscopic, inaccessible degrees of freedom. In practice, reversible computers are hard to build.

ii) Quantum mechanical information transfer:

Quantum systems differ from classical systems in that pure quantum mechanical states have intrinsically statistical properties. For example, if two electrons are in the state $(1/\sqrt{2})\{|\uparrow\rangle_1|\downarrow\rangle_2 - |\downarrow\rangle_1|\uparrow\rangle_2\}$, then each electron taken on its own is described by the density matrix $\rho = (1/2)\{|\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|\}$, a statistical mixture. In general, if the Hilbert space H for a quantum mechanical system can be decomposed into a tensor product space $H = H_A \otimes H_B$, as it can if there exist two commuting operators A and B that together form a Casimir set for H, then any $|\psi\rangle \epsilon H$ can be written $|\psi\rangle = \sum_{i,j=1}^{m,n} c_{ij}|a_i\rangle|b_j\rangle$, where m,n are the dimensions of $H_{A,B}$ and $\{|a_i\rangle\}$, $\{|b_j\rangle\}$ are bases for $H_{A,B}$, and $\sum_{i.j}|c_{ij}|^2 = 1$. One of m,n is less than or equal to the other, suppose it is m; then by a suitable change of bases for $H_{A,B}$, $|\psi\rangle$ can be written in the form $|\psi\rangle = \sum_{i=1}^{m} \gamma_i|a_i'\rangle|b_i'\rangle$, where $\sum_i |\gamma_i|^2 = 1$. The entropy of A, S(A) = $-\mathrm{tr}\rho(A)\ln\rho(A) = \sum_i |\gamma_i|^2 \ln|\gamma_i|^2 = S(B)$, and the mutual information between A and B is then $S(A) + S(B) - S(AB) = -2\sum_i |\gamma_i|^2\ln|\gamma_i|^2$. A typical pure quantum mechanical state for two degrees of freedom of a system exhibits mutual information between them.

If in the initial pure state, S(A) = 0, i.e., A has no information about B, S(A) will increase for most interactions between A and B: for almost any interaction, A gets information about B. The average value of S(A) after the interaction is determined by the following theorem.

Theorem:

If $|\psi_0\rangle = |a_0\rangle|b_0\rangle$ is taken to a state $|\psi\rangle = \sum_{i,j=1}^{m,n} c_{ij}|a_i\rangle|b_j\rangle$ by a unitary evolution U_t selected at random from U^{mn}, $m\leq n$, then the probability that $|\psi\rangle = \sum_{i=1}^{m} \gamma_i|a_i'\rangle|b_i'\rangle$, for a particular set of $|\gamma_i|$ is proportional to $(|\gamma_1|\ldots|\gamma_m|)^{2(n-m)+1} \prod_{1\leq i<j\leq m} (|\gamma_i|^2 - |\gamma_j|^2)^2$.

(Proof in Appendix B.)

The distribution for the $|\gamma_i|$ that maximizes $S_t(A) = S_t(B)$ after the interaction is $|\gamma_i| = 1/m$. The second factor in the expression above insures that the probability that $|\gamma_i|=|\gamma_j|=(1/m)$ is zero, but the first factor insures that if $n>>m$, most of the $|\gamma_i|$ are likely to be very close to 1/m. For example, if $m=2$ and $n>>2$, the most likely values for $|\gamma_1|$, $|\gamma_2|=(1-|\gamma_1|^2)^{\frac{1}{2}}$ are

$$|\gamma_1|^2 = (1/2)\ \{\ 1 + 1/(2n-2)^{\frac{1}{2}}\},\quad |\gamma_2|^2 = (1/2)\ \{\ 1 - 1/(2n-2)^{\frac{1}{2}}\}.$$

After the interaction, $S_t(A)$ and $S_t(B)$ are likely to be close to their maximum value of ln2.

If A and B start out in a mixture instead of a pure state, $S_t(A)$ is still likely to be close to its maximum value.

If $\rho_t(AB) = \sum_k d_k|\psi_k\rangle\langle\psi_k|$, where $\sum_k d_k = 1$, then $\rho_t(A) = \sum_k d_k\rho_k(A)$, where $\rho_k(A)$ is the density matrix for A when AB is in the state $|\psi_k\rangle$. But $S_t(A) = -\mathrm{tr}\rho_t(A)\ln\rho_t(A) \geq \sum_k d_k\mathrm{tr}\rho_k(A)\ln\rho_k(A) = \sum_k d_k S_k(A)$, and each of the $S_k(A)$ is likely to be close to its maximum value; hence $S_t(A)$ is likely to be close to its maximum value.

Example:

A system of n interacting spins, initially in a state that exhibits no mutual information, tend to evolve into a state that exhibits a great deal of mutual information. The states of such a system can be labeled $|\psi\rangle = \sum_{ij\ldots k} c_{ij\ldots k}|ij\ldots k\rangle$, where each index i,j,...,k can be ↑ or ↓. If the spins do not interact, each of the states $|ij\ldots k\rangle$ are degenerate and no increase in mutual information takes place. Given a small perturbing interaction, H_{int}, whose eigenstates are randomly selected superpositions of the $|ij\ldots k\rangle$, the amount of time that it takes an initially uncorrelated state such as $|\psi_0\rangle = |\uparrow\downarrow\ldots\uparrow\rangle$ to evolve into a highly correlated state is $t = h/(2\pi\Delta E)$, where ΔE is the average energy level spacing of H_{int}. The amount of time that it takes the system to return to a state that exhibits close to zero mutual information is on the order of $2^n h/(2\pi\Delta E)$ -- the Poincare recurrence time.

Conclusion

Thermodynamic depth is a universal measure of complexity: any macroscopic state of a physical system can be assigned a thermodynamic depth purely as a function of physical quantities. Thermodynamic depth is the unique (up to a multiplicative constant) measure of a state's complexity that is a continuous function of the probabilities of the various processes that can lead to that state, that is additive over time, and that assigns measure zero to systems in thermodynamic equilibrium. Thermodynamic depth satisfies intuitive requirements for a measure of complexity: it assigns low complexity to random and to highly ordered but regular states. The thermodynamic depth of a complex system and a copy is not equal to twice the depth of the system but to the depth of the system plus the depth of the copying process. Thermodynamic depth encompasses mathematical definitions of complexity: the computational complexity of a problem is proportional to the thermodynamic depth of the most efficient machine that solves that problem. The thermodynamic depth of a result calculated from a minimal program on a reversible computer is proportional to the logical depth of the result.

Thermodynamic depth is equal to the difference between the coarse- and fine-grained entropy, or equivalently, to $1/k_B$ times the amount of information shunted between the various parts of a system in the process of constructing a particular state. The processes that lead to thermodynamically deep states are also universal: almost any interaction between two or more degrees of freedom of a system transfers information from each degree of freedom to the others and generates mutual information between them.

Applications:

Since thermodynamic depth applies universally to all physical systems, systems can be classified according to how thermodynamically deep they are. For example, during the formation of macromolecules in the early terrestrial environment, enzyme A is thermodynamically deeper than enzyme B if more entropy had to be generated in the process of constructing A than constructing B. The additive property of thermodynamic depth implies that complexity increases as new depth-producing chemical reactions come into play using the products of previous reactions.[30-32]

If one extracts the genetic contribution to the overall thermodynamic depth of an organism, one arrives at Kuhn's notion of "Knowledge":[33] the genetic complexity of an organism is proportional to the amount of genetic information tried out and discarded by the process of natural selection on the ancestors of the organism. If one could map out the evolutionary path by which simple organisms came into being, then one could order organisms on the basis of their thermodynamic depth. In addition, one could classify the various functions that organisms perform by estimating, say, the depth of the process whereby a cell assembles a particular molecule.

The thermodynamic depth of a state of any dissipative dynamical system can be got by regarding it as a subsystem of a larger, conservative system and calculating how much information the system must exchange and discard in the process of arriving at the given state. That is, one evaluates the thermodynamic depth of a dissipative system by calculating how much the degrees of freedom of the dissipative system contribute to the thermodynamic depth of the conservative system in which it is embedded. For example, one could estimate the complexity of an interlocking set of markets by modelling the exchange of goods as a dynamical system and calculating that system's thermodynamic depth. Thermodynamically deep systems are necessarily far from equilibrium: are complex systems unstable? Complexity may come at a cost.

Finally, whenever a number is represented as the state of a physical system, it has a thermodynamic depth. The relationship between thermodynamic depth and logical depth implies that when a system represents a logically deep number, that state is thermodynamically deep. Since only a finite amount of thermodynamic depth has been generated within our event horizon since the big bang, and since there are an infinite number of arbitrarily logically deep numbers, there are an infinity of numbers that are too deep to be represented physically at this time in the universe. For example, a random integer greater than e^{S/k_B}, where S is the entropy of the universe within our event horizon, is unlikely to have any representation at this time since there are an insufficient number of degrees of freedom available in the universe to 'count' that high. Of course, there are many numbers greater than e^{S/k_B}

that can be represented concisely at this time; $10^{10^{10^{10}}}$ is such a number.

The limit on the possible thermodynamic depth of any physical system at this stage in the evolution of the universe suggests the existence of a further class of numbers. These are numbers that have a concise description, but the problem of arriving at this description is so complex that its thermodynamic depth exceeds that allowed for the universe at this time. Such numbers are in principle indistinguishible from random numbers at the moment; but as the universe expands and the limits on the possible thermodynamic depth of the description goes up, such numbers cease to be random. As defined by thermodynamic depth, the notion of algorithmic randomness is time-dependent. Conversely, in the very early universe, numbers such as π that are far from algorithmically random now, were random.

Chapter 3: Pure State Quantum Statistical Mechanics and Black Holes

If one flips a coin twenty times, and gets heads nine times and tails eleven, one ascribes the variation in the results to differences in how hard one flipped the coin, how it hit the ground, etc. If one prepares twenty electrons in the state spin x up, then makes a measurement of spin z on each electron and gets spin z up nine times and spin z down eleven, then (if one does not believe hidden variable theories) one ascribes the variation to the statistical properties of pure quantum mechanical states.

In quantum statistical mechanics, the inherently statistical nature of quantum mechanical pure states adds an additional element of chance to the already chancy results of measurements made on systems with many degrees of freedom, not all of which have been fixed by experiment. The sort of probabilities that come out of quantum mechanical pure states differ significantly from those that come out of classical probability distributions. In this chapter we show, however, that the pure states of quantum mechanical systems with many degrees of freedom reproduce the statistics of the normal statistical mechanical ensembles such as the microcanonical and canonical ensemble to a high degree of accuracy: quantum mechanical systems with many degrees of freedom in pure states behave like statistical mixtures, with respect to most measurements. In the thermodynamic limit, as the number of degrees of freedom of the system in question goes to infinity, the difference between the statistics implied by the pure states of the system and the statistical mechanical probabilities becomes impossible to detect.

Suppose, for example, that one has a gas composed of n particles, confined to a box with sides of length L. Classically, if one fixes the state of the gas by fixing the position and momentum of each particle, then the results of any measurement that one makes on the gas are fixed as well. Quantum mechanically, if one puts the gas as a whole in a pure state by putting each of the particles in a pure state that is an eigenstate of the energy and momentum for a particle in a box of the given size, then only the results of measurements that correspond to operators that commute with the particles' momenta are fixed: if one makes a measurement on the position of a particle with energy and momentum high compared with $h/(2\pi L)$, the particle can turn up in any part of the box with equal probability. This uniform distribution for the position of the particle in the box arises from the form of the pure state of the gas as a whole, but it is exactly the same as the distribution predicted by the normal statistical mechanical ensemble for the positions of the particles of the gas with total energy and momentum fixed.

If a quantum mechanical system is in a pure state, then only measurements that correspond to operators of which that state is an eigenstate will give the same result each time. Other measurements will give a statistical distribution of results when performed on an ensemble of systems all prepared in that state. We will prove that for a given measurement, if many copies of a system with many degrees of freedom have all been prepared in the same pure state chosen at random from a subspace of Hilbert space (such as the subspace $H_{E,E+dE}$ composed of all states with energy between E and E+dE), the results of the measurement made on these systems in this state lie in a statistical distribution whose mean, standard deviation, and higher moments differ from the mean, standard deviation, and higher moments predicted for the results of the measurement by a conventional uniform distribution (the microcanonical ensemble, for the subspace $H_{E,E+dE}$) over all states in that subspace by a factor of $1/\sqrt{n}$, on average, where n is the dimension of the subspace. For a given measurement, most pure states of a system with many degrees of freedom give statistical distributions of results that differ by only a small amount from those predicted by the ensemble average. Note that this result has no classical analogue: many copies of a classical system all prepared in the same state all give the same result for any measurement made upon them -- they imply no distribution whatsoever.

As a corollary, we prove that for many copies of a quantum mechanical system with many degrees of freedom prepared in a given state (for example, a state with energy between E and E+dE), the results of most measurements restricted to a given subspace ($H_{E,E+dE}$) fall in distribution that have expectation values, standard deviations, and higher moments that differ by only a small amount ($1/(\dim H_{E,E+dE})^{\frac{1}{2}}$) from the expectation values, standard deviations, and higher moments predicted by the ensemble average over that subspace (the microcanonical ensemble for states with energy between E and E+dE). For a system with many degrees of freedom, the quantum mechanical probabilities inherent in the actual state of the system are likely to mimic the normal statistical mechanical probabilities. In the thermodynamic limit, $n\to\infty$, the correspondence is exact.

Another way of phrasing this result is that systems with many degrees of freedom in pure states behave as if they were in statistical mixtures with regard to most measurements. We apply this result to the foundations of statistical mechanics, to black holes, and in chapter 4 to the quantum measurement problem.

Our results follow directly from

Theorem:[1,2]

Let $H \cong C^n$ be a subspace of the Hilbert space for a quantum mechanical system **A**. Let F be an Hermitian operator corresponding to the measurement of some quantity on **A**. Then

$$\{ \text{ the average over all } |\psi\rangle \epsilon H \text{ of } (\langle\psi|F|\psi\rangle - (1/n)\mathrm{tr}F)^2 \}^{\frac{1}{2}}$$
$$= (1/(n+1)^{\frac{1}{2}}) \{ \mathrm{tr}(F^2)/n - (\mathrm{tr}F)^2/n^2 \}^{\frac{1}{2}}$$

(proof in appendix C).

In the case that H represents the Hilbert space of states compatible with the results of macroscopic measurements that have been performed on the system, we may paraphrase this theorem as follows: If the dimension of the Hilbert space of states compatible with our macroscopic knowledge of the system is large, then the amount by which the expectation value of an operator F on a typical state $|\psi\rangle$ differs from its average expectation value over all compatible states is likely to be small. Indeed, since

$$\{ \mathrm{tr}(F^2)/n - (\mathrm{tr}F)^2/n^2 \}^{\frac{1}{2}} \leq \max |f_i| ,$$

where the f_i are the possible results of the measurement, the theorem above implies that the amount by which the expectation value of an operator on a typical state differs from its expectation value over all states is likely to be less than $1/\sqrt{n}$ × (the maximum magnitude of the result).

For example, if a number of systems identical to A have all been prepared in the same pure state $|\psi\rangle \in H_{E,E+dE}$, the theorem above tells us that if $n = \dim H_E$ is large, the distribution of results for a measurement corresponding to a Hermitian operator F are likely to have the same expectation value as that implied for the measurement by the microcanonical ensemble. Of course, if $|\psi\rangle$ is an eigenstate of F, or a superposition of states dominated by a single eigenstate of F, the distribution of the different results, f_i, of the measurement implied by $|\psi\rangle$ differs markedly from the statistics given by the microcanonical ensemble. If we make a measurement corresponding to F on a number of systems all of which have been prepared in some eigenstate of F, we will obtain the same result for each system. The microcanonical ensemble, in contrast, predicts a range of results. The theorem above says that if n is large, the probability that a state selected at random will be a superposition dominated by a single eigenstate of F is small.

The proof of the following corollary follows immediately from the proof of the theorem above:

Corollary:

Given a state $|\psi\rangle \in H \cong C^n$, and a set of real numbers $\{f_i\}$, the average over all bases $\{\ |e_i\rangle\ \}$ for H of

$$(\langle\psi|F|\psi\rangle - (1/n)\mathrm{tr}F)^2$$

is equal to

$$1/(n+1)\ \{\ \mathrm{tr}F^2/n - (\mathrm{tr}F)^2/n^2\ \}\ ,$$

where $F = \sum_i f_i|e_i\rangle\langle e_i|$.

This corollary has the consequence that almost all measurements that we can make on a number of systems all prepared in a particular state whose energy lies between E and E+dE will have expectation values very close to those predicted by the microcanonical ensemble. Even if some clever fellow has prepared our system in a very particular state, most of the measurements that we perform on the system will have results that follow a microcanonical distribution. To put the same point in a different way: if we do not know in what pure state a system has been prepared, the chances of our making a measurement that has that state as one of its eigenstates are slim.

We have the following results: 1) The expectation value of a particular operator over most quantum states of a system with n degrees of freedom, and with energy in the interval [E, E+dE], is likely to be the same as the operator's expectation value over the microcanonical ensemble to within a factor of $1/\sqrt{n}$. 2) The expectation values of most operators over a particular quantum state with energy in the interval $[E, E+dE]$ are likely to be equal to those operators' expectation values over the microcanonical ensemble to within a factor of $1/\sqrt{n}$.

We can go further. Not only do the pure quantum states of a system with many degrees of freedom give expectation values for measurements that are very close to the expectation values predicted by the microcanonical ensemble, these pure states also give standard deviations from those expectation values that are very close to the standard deviations predicted by the microcanonical ensemble.

For a particular measurement, F, the microcanonical ensemble over $H_{E,E+dE}$ predicts a variance from the mean value of F of $(1/n)\mathrm{tr}F^2 - \{(1/n)\mathrm{tr}F\}^2$, where the traces are taken over $H_{E,E+dE}$. For an ensemble of systems prepared in a pure state $|\psi\rangle$, the variance of the results of a measurement of F from $(1/n)\mathrm{tr}F$ is $\langle\psi|(F - (1/n)\mathrm{tr}F)^2|\psi\rangle$. Let us look at the average amount by which the variance of F over $|\psi\rangle$ differs from F's variance over the microcanonical ensemble, i.e., let us look at

$$[\ \text{the average of}\ \underset{\text{over } |\psi\rangle\ \epsilon\ H_{E,E+dE}}{}\ \{ \langle\psi|(F - (1/n)\mathrm{tr}F)^2|\psi\rangle - (\mathrm{tr}F^2/n - (\mathrm{tr}F/n)^2) \}^2\ ]^{\frac{1}{2}} .$$

By the theorem above, this quantity is equal to

$$(1/(n+1)^{\frac{1}{2}})\ \{ (1/n)\mathrm{tr}B^2 - ((1/n)\mathrm{tr}B)^2 \}^{\frac{1}{2}}, \text{ where } B = (\ F - (1/n)\mathrm{tr}F\ )^2 .$$

The quantity $\{ (1/n)\mathrm{tr}B^2 - ((1/n)\mathrm{tr}B)^2 \}^{\frac{1}{2}}$ is on the order of the average magnitude of an eigenvalue of F^2 and is independent of n; the factor of $1/(n+1)^{\frac{1}{2}}$ out front then insures that as n gets large, the amount by which the variance given by the pure state quantum statistics differs from the variance implied by the microcanonical ensemble tends to become small.

One can apply the same argument to the higher moments of the distribution. Since the average over all $|\psi\rangle$ of

$$(\langle\psi|F^m|\psi\rangle - (1/n)\mathrm{tr}F^m)^2 = (1/(n+1)) \{ (1/n)\mathrm{tr}F^{2m} - ((1/n)\mathrm{tr}F^m)^2 \} ,$$

as n gets large the moments given by the pure quantum states converge on the microcanonical moments.

For quantum systems with more and more degrees of freedom, not only do the expectation values implied by the pure quantum states of the system tend to mimic more and more exactly the expectation values of statistical mechanics -- the quantum deviations away from those expectation values get closer and closer to the deviations predicted by statistical mechanics, as well. In the thermodynamic limit, $n\to\infty$, the probability that a pure state selected at random from $H_{E,E+dE}$ gives a distribution for the results of a given measurement that differs from the distribution implied by the microcanonical ensemble, is zero.

We now apply the results derived above to the canonical and grand canonical ensembles.

The Canonical Ensemble

In this section we show that the exact state for a system in contact with a thermostat at temperature T is likely to be a mixture that has the same form as the canonical ensemble for the system.

Suppose we have a joint system **AB**. If **A** and **B** are weakly interacting, then their Hilbert space can be decomposed into the tensor product space: $H_{AB} = H_A \otimes H_B$. To lowest order in perturbation theory, the subspace H^E_{AB} of $H_A \otimes H_B$ corresponding to energy E is spanned by a basis of vectors of the form

$|E_i\rangle^k_A \otimes |E_j\rangle^l_B$, where $E_i + E_j = E$, and the $|E_i\rangle^k_A$, $k = 1$ to $d_A(E_i)$ ($d_A(E_i)$ is the degeneracy of E_i in H_A), span the subspace of H_A corresponding to energy E_i; similarly for the $|E_j\rangle^l_B$, $l = 1$ to $d_B(E_j)$.

A typical state $|\psi\rangle \in H^E_{AB}$ can be written

$$|\psi\rangle = \sum_i \sum_{k=1}^{d_A(E_i)} \sum_{l=1}^{d_B(E-E_i)} \alpha^i_{kl} \, |E_i\rangle^k_A \quad |E-E_i\rangle^l_B \quad .$$

The corresponding density matrix is

$$\rho = |\psi\rangle\langle\psi| = \sum_{i,i'} \sum_{k,k'=1}^{d_A(E_i),d_A(E_{i'})} \sum_{l,l'=1}^{d_B(E-E_i),d_B(E-E_{i'})}$$

$$\left[\alpha^i_{kl} \bar{\alpha}^{i'}_{k'l'} \, |E_i\rangle^k_A \, {}^{k'}\langle E_{i'}| \otimes |E-E_i\rangle^l_B \, {}^{l'}\langle E-E_{i'}| \right] ,$$

and the density matrix for **A** alone is got by taking the trace of ρ over the degrees of freedom of **B** :

$$\rho_A = \sum_i \sum_{l=1}^{d_B(E-E_i)} \sum_{k,k'=1}^{d_A(E_i)} \alpha^i_{kl} \bar{\alpha}^i_{k'l} \, |E_i\rangle^k_A \, {}^{k'}\langle E_i| \quad .$$

Note that ρ_A has no off-diagonal terms between states of different energy. In general, for a particular pure state $|\psi\rangle$ with energy E, **A** is in a mixture of states with different energies E_i, each correlated with a state of B with energy $E-E_i$.

We can write $\rho_A = \sum_i p(E_i)\ \rho_{E_i}$, where $\rho_{E_i} \epsilon\ \overline{H}_A^{E_i} \otimes H_A^{E_i}$ is a density matrix for a system with energy E_i, tr $\rho_{E_i}=1$: the $p(E_i)$ are the probabilities that a measurement of energy on **A** will find the value E_i. Tr $\rho_A=$ tr $\rho_{E_i} = 1$ implies that $\sum_i p(E_i) = 1$. We can now ask, if we select $|\psi> \epsilon\ H_{AB}^E$ at random, what are the most probable values for the $p(E_i)$, and how much deviation do we expect from those values?

Applying theorem 1 above, we know that the most likely value for $p(E_i)$ is given by its value over the ensemble average. For a particular ρ_{AB}, $p(E_i) = \text{tr}\ \rho_{AB}\ P_A^{E_i} \otimes P_B^{E-E_i}$, where $P_A^{E_i}, P_B^{E-E_i}$ are the projection operators onto the E_i eigenspace of H_A and the $E-E_i$ eigenspace of H_B. Over the ensemble, $p(E_i)$ averages out to n_i/n, where $n_i= d_A(E_i)d_B(E-E_i)$; i.e., over the ensemble, the probability of finding a certain energy E_i for **A** is just proportional to the degeneracy of such states. Theorem 2 now tells us that $p(E_i)$ deviates from its most likely value by

$$(1/(n+1)^{\frac{1}{2}})\ \{\ (n_i/n)\ -\ (n_i/n)^2\ \}^{\frac{1}{2}} \ \approx\ (\sqrt{n_i})/n \quad \text{, on average.}$$

So we have

$$\rho_A = (1/n) \sum_i d_A(E_i)d_B(E-E_i)\ (\ 1 \pm 1/\sqrt{n_i}\ )\ \rho_{E_i} \quad .$$

But $d_A(E_i) = e^{S_A(E_i)}$, and $d_B(E-E_i) = e^{S_B(E-E_i)}$, where S_A and S_B are the entropies of **A** and B. If B has many more degrees of freedom than **A**, then terms with E_i near zero dominate the sum and we can approximate $S_B(E-E_i) = S_B(E) - E_i/T$, where $1/T = \partial S_B/\partial E$, and we have

$$\rho_A = (1/N) \sum_i \{ e^{-(E_i - TS_A(E_i))/T} (1 \pm 1/\sqrt{n_i}) \} \rho_{E_i} ,$$

where $1/N = e^{S_B(E)}/n$, and $\mathrm{tr}\, \rho_{E_i} = 1$.

We see that for a system **A** in contact with a thermostat **B** at temperature T, for almost all pure states $|\psi> \epsilon H^E_{AB}$, the <u>exact</u> <u>state</u> of **A** is likely to give (to within a small fluctuation) the same density matrix as the canonical ensemble for **A** . In the thermodynamic limit, $n_i \to \infty$, and the probability that the density matrix for **A** differs from a thermal density matrix becomes zero.

<u>Non-equilibrium pure state statistical mechanics</u>

The results derived so far imply that if the initial state for a system with many degrees of freedom is chosen at random, the distributions of results for measurements that that state implies do not differ from the predictions of the normal statistical mechanical ensembles in the thermodynamic limit. We now derive a non-equilibrium result. We show that if a system **A** has an arbitrary weak interaction with a larger system **B**, then even if **A** and **B** start out in a pure state far from equilibrium, their joint state evolves into one in which the density matrix for **A** has a thermal form. Suppose that **A** and **B** are initially noninteracting with total energy E, as in the previous section, and that one perturbs the system by adding a small interaction Hamiltonian to the original Hamiltonian: $H' = H + H_{int}$. Suppose that the initial state for **A**B is $|\chi_0\rangle$, $H|\chi_0\rangle = E|\chi_0\rangle$. According to first order degenerate perturbation theory, the eigenvectors of H_{int} confined to the subspace of energy E are, to lowest order, arbitrary orthogonal linear combinations of the unperturbed states of the subspace. If we wait an amount of time $t=h/(2\pi\Delta E)$, where ΔE is the average energy level spacing of H_{int}, the evolution of the system will take $|\chi_0\rangle$ to a state that is an arbitrary superposition of the original states $|E_i\rangle^k_{\mathbf{A}}|E-E_i\rangle^l_{\mathbf{B}}$ with total energy E. Applying the results of the section on the canonical ensemble above, the density matrix for **A** is then

$$\rho_A = (1/N) \sum_i \{ e^{-(E_i - TS_{\mathbf{A}}(E_i))/T} (1 \pm 1/\sqrt{n_i}) \} \rho_{E_i} \quad ,$$

as above. Once again $\rho_{\mathbf{A}}$ takes a thermal form. In the thermodynamic limit, an arbitrarily chosen H_{int} causes **A** and **B** to evolve into a state in which the density matrix for **A** is exactly that for **A** at thermal equilibrium.

Grand canonical ensembles

In the above analysis, the only requirement on E was that it be an additively conserved quantity. We can repeat the steps above including other additively conserved quantities. For example, if the total electric charge of **A** and **B** is Q_{tot}, then we find that the most likely exact state for **A** is

$$\rho_{\mathbf{A}} = (1/N') \sum_{i,Q} \rho_{E_iQ} \{ e^{-(E_i - \Phi Q - TS_{\mathbf{A}}(E_i,Q))/T} (1 \pm 1/\sqrt{n_{iQ}}) \} \quad ,$$

where $\rho_{E_iQ} \in \underline{H}_{\mathbf{A}}^{E_iQ} \times H_{\mathbf{B}}^{E_i}$ has trace 1, $n_{iQ} = d_{\mathbf{A}}(E_i,Q)d_{\mathbf{B}}(E-E_i,Q_{tot}-Q)$, and

$\Phi = -T(\partial S_B/\partial Q)\big|_{Q=Q_{tot}}$ is the electric potential of **B**.

The exact state of a system **A** in contact with reservoirs of heat, charge, particle #, etc., is likely to give to within a small fluctuation the same density matrix as the various grand canonical ensembles for **A**. In the limit that the size of the reservoirs goes to infinity, the probability that the density matrix for A takes on an exact grand canonical form goes to one.

Applications of pure state quantum statistical mechanics

We apply the results derived above in three areas: 1) the interpretation of quantum statistical mechanics, 2) horizon radiation from black holes and cosmological particle production, and 3) the quantum measurement problem.

The interpretation of probability in statistical mechanics

Historically, there are two approaches to interpreting the probability distribution function in classical statistical mechanics. In the ensemble approach, the probabilities that describe a complex system are taken to refer to an imaginary ensemble of systems identical to the one of interest, whose states are spread out uniformly over the set of microscopic states consistent with our macroscopic information. In the ergodic approach, the probability distribution is regarded as giving the relative frequencies of occurrence of the microscopic states of the actual system of interest, averaged over time. For example, if all we know about a system is that it has energy between E and E+dE, then we describe the system by the microcanonical ensemble, characterized by a probability distribution that assigns equal probability to all states with energy in the interval [E, E+dE]. In the ensemble approach, this distribution is taken to refer to an imaginary ensemble of identical systems whose states are spread out evenly over the hypersurface in phase space made up of points representing states with energy between E and E+dE. In the ergodic approach the microcanonical ensemble is held to describe the actual system of interest in accordance with the ergodic hypothesis -- the trajectory of the representative point of the system is hypothesized to fill up the hypersurface of energy E uniformly, spending an equal amount of time in equal volumes of phase space.

When Boltzmann proposed the first version of the ergodic hypothesis in 1871,[3] one of his goals was to provide a purely mechanical description of the approach to equilibrium, independent of statistical considerations. In introducing his famous H-theorem in 1872,[4] he claimed to have proved the second law of thermodynamics on purely mechanical grounds: any initial distribution of kinetic energy, Boltzmann asserted, would eventually approach Maxwell's distribution. It was not until several years later, prodded by Loschmidt, that he acknowledged the statistical nature of his proof, inherent in the assumption of molecular chaos.

The results on the statistics of pure quantum states of systems with many degrees of freedom derived above suggest a purely mechanical interpretation of results normally derived statistically by taking ensemble averages -- or rather, a purely quantum mechanical interpretation. If we prepare a system with many degrees of freedom in any state with energy between E and E+dE, then the great majority of measurements performed on that system will give results that follow closely the statistical predictions of the microcanonical ensemble. The statistical nature of the outcome of such measurements arises not because we do not know what state the system is in (we prepared it in a pure state), nor because we are making measurements over a long period of time (we can make the measurements over as short a period of time as we like as long as we do not conflict with the uncertainty principle), but because quantum mechanics is inherently statistical. For a system with many degrees of freedom, the quantum mechanical statistics inherent in the pure states of the system converge on the statistics implied by the microcanonical ensemble.

Horizon radiation and cosmological particle production

i) Horizon radiation

It is well established[5-9] that a black hole of mass M emits radiation with a thermal spectrum at temperature $T = (8\pi M)^{-1}$, where we have set G=h=c=1. Hawking uses this result to argue that a black hole behaves like a black body (albeit a strange black body, with negative specific heat) with entropy $S = 4\pi M^2$. In the normal derivation of black-hole radiance, however, one treats the hole as a feature of the background spacetime -- a geometric, not a thermodynamic object. The thermal form of the radiation arises from the periodicity of the Schwarzschild metric in imaginary time, not from overtly statistical considerations. T.D. Lee[9] has pointed out that while black-body radiation is incoherent, the radiation coming from the horizon of a hole is fundamentally coherent over spacetime as a whole; Lee argues that the coherence of the horizon radiation implies black holes are not black bodies.

Using methods of field theory in curved spacetime[8-9] one can show that in the presence of a black hole of mass M the incoming vacuum state $|0>_{in}$ evolves into a state of the form

$$|0>_{out} = (1/N) \sum_{i} \sum_{k}^{d(E_i)} e^{-4\pi M E_i} |-E_i>^k_{inside} \, |E_i>^k_{outside}$$

where the $|E_i>^k_{outside}$ are states with total energy E_i, outside the hole (the energy is determined by an observer in the asymptotically flat area of spacetime; $d(E_i)$ = the degeneracy of E_i), and the $|-E_i>^k_{inside}$ are states with total energy $-E_i$ inside the hole. The total energy and charge of this state are still zero, but the state of the fields outside the hole is a thermal mixture at temperature $T = (8\pi M)^{-1}$.

To see the thermal form of the radiation, note that the density matrix for the fields over the whole of spacetime is

$$\rho_{\text{total}} = (1/|N|^2) \sum_{i,i'} \sum_{k,k'=1}^{d(E_i)} e^{-4\pi(E_i+E_{i'})} |-E_i>^{k}_{\text{inside}}{}^{k'}\!\!<-E_{i'}| \otimes |E_i>^{k}_{\text{outside}}{}^{k'}\!\!<E_{i'}| \quad .$$

A measurement made outside the horizon corresponds to an operator of the form $I_{\text{inside}} \otimes A_{\text{outside}}$, where I_{inside} is the identity operator on the Hilbert space of states inside the hole, and A_{outside} is an Hermitian operator on the space of states outside the hole. We have

$$\begin{aligned} < I_{\text{inside}} \otimes A_{\text{outside}} > &= \operatorname{tr} \rho_{\text{total}}(I_{\text{inside}} \otimes A_{\text{outside}}) \\ &= \operatorname{tr} \rho_{\text{outside}} A_{\text{outside}} \text{ , where} \end{aligned}$$

ρ_{outside} = the trace over the internal degrees of freedom of the hole of ρ_{total},

$$\begin{aligned} &= (1/|N|^2) \sum_{\substack{i,i' \\ k,k'}} e^{-4\pi M(E_i+E_{i'})} |E_i>^{k}_{\text{outside}}{}^{k'}\!\!<E_{i'}| \; \delta_{ii'} \delta_{kk'} \\ &= (1/|N|^2) \sum_{i,k} e^{-8\pi M E_i} |E_i>^{k}_{\text{outside}}{}^{k}\!\!<E_i| \quad . \end{aligned}$$

Note that even though the state of the fields over the whole of spacetime is pure, the fields outside the hole are in a mixture. In fact, the form of ρ_{outside} is exactly that of a thermal mixture at temperature $T = (8\pi M)^{-1}$. The hole radiates with a black-body spectrum.

T.D. Lee has argued that the coherence of the quantum fields over the whole of spacetime implies that black holes should not be regarded as black bodies, the radiation from which is incoherent. Our results from the previous section imply the opposite: <u>the form of the coherent fields in the presence of a black hole is exactly what we expect of the coherent fields in contact with a black body in the thermodynamic limit</u>. The results of the previous sections on the canonical ensemble imply that the pure state of the quantum fields interacting with a black body of temperature T is very likely to take the form of a thermal mixture at temperature T. The results on non-equilibrium pure state statistical mechanics derived above imply that if the quantum fields around the hole are not initially in a thermal form, an arbitrary interaction between the field modes and the degrees of freedom of the gravitational field will induce the state of the fields to attain a thermal form, at least locally (since black holes have negative specific heat, the fields can not be in global stable equilibrium with the hole). If a black hole is a black body with temperature $T = (8\pi M)^{-1}$, then the factor by which the form of the coherent state of the fields interacting with the hole deviates from the form of a thermal mixture is on the order of $e^{-2\pi M^2} = e^{-(2\pi \times 10^{76})}$, for $M = M_\odot$. Black holes may not be black bodies, but if they aren't, it is very difficult to tell the difference.

ii) Increase of mutual information and cosmological particle production

The theorems on Hilbert space proved above allow us to carry further the arguments of Hu and Kandrup,[10] who treat the problem of cosmological particle generation and entropy production in terms of increase of the mutual information shared between the oscillators that represent the modes of quantum fields in Fock space. Kandrup and Hu show that increase in the mutual information between the oscillators is closely linked to cosmological partical production. (As we have seen, a black hole creates particles with induced entropy $-\mathrm{tr}\rho_{\text{outside}} \ln\rho_{\text{outside}}$ outside the hole, even though the fields over all of spacetime are in a pure state.) They point out that the mutual information is never negative, and if the mutual information is initially zero, any interaction will cause it to increase, at least initially.

We use the non-equilibrium results derived above to show that the mutual information shared between the particle modes of the quantum fields in the universe is very likely to increase to its maximum value.

The modes of the quantum fields can be represented as a collection of harmonic oscillators: we first examine the case of n operators without interactions, then introduce interactions and take the thermodynamic limit $n \to \infty$. n noninteracting harmonic oscillators are described by a Hamiltonian $H = H_1 + H_2 + \ldots + H_n$, where $H_1 = \sum_{i=0}^{\infty} i(h/2\pi)\omega_1 |i\rangle_1 \langle i|$, where i_1 is the i-th excited state of the 1-th oscillator, ω_1 is its fundamental frequency, and we have set the zero point energy to zero. Suppose that the oscillators start out in the state $|\chi_0\rangle = |i\rangle_1 |j_2\rangle \ldots |k\rangle_n$, with total energy $E = (h/2\pi)\ (i\omega_1 + j\omega_2 + \ldots + k\omega_n)$. If there is no interaction between the oscillators, then they will stay in this state forever. Suppose now that one perturbs the system by adding a small interaction Hamiltonian to the original Hamiltonian: $H' = H + H_{int}$. According to first order perturbation theory, the eigenvectors of H_{int} confined to the subspace of energy E are, to lowest order, arbitrary orthogonal linear combinations of the unperturbed states of the subspace. If we wait an amount of time $t = h/(2\pi\Delta E)$, where ΔE is the average energy level spacing of H_{int}, the evolution of the system will take $|\chi_0>$ to a state that is an arbitrary superposition of the original states $|i>|j>\dots|k>$ with total energy E. Applying the results of the section on the canonical ensemble above, the density matrix for the first oscillator is then

$$\rho_1 = (1/N) \sum_{i=0}^{\infty} |i>_1<i|\ d_{2\dots n}(E-h\omega_1 i)\ (\ 1 \pm 1/(d(E-\hbar\omega_1 i)^{\frac{1}{2}}\ )$$

where $d_{2\dots n}(E-\hbar\omega_1 i)$ is the dimension of the space of states of the oscillators 2...n with total energy $E-h\omega_1 i$, N is a normalization constant, and the $\pm 1/(\dots)^{\frac{1}{2}}$ expresses the uncertainty in the result due to the lack of knowledge of the exact form of H_{int}. If E is large and $n >> 1$, we can write

$$d_{2\dots n}(E-\hbar\omega_1 i) = e^{S_{2\dots n}(E-\hbar\omega_1 i)} = e^{S_{2\dots n}(E) - (\partial S_{2\dots n}/\partial E)\hbar\omega_1 i}.$$

Taking the thermodynamic limit $n\to\infty$, we obtain

$$\rho_1 = (1/N') \sum_{i=0}^{\infty} |i>_1<i|\ e^{-\hbar\omega_1 i/T}, \text{ where } 1/T = \partial S_{2\dots n}(E)/\partial E.$$

The expressions for $\rho_2,\dots,\rho_n$ are identical, with appropriate re-indexing. In the thermodynamic limit, an arbitrarily chosen H_{int} causes the oscillators as a group to evolve into a state in which the density matrix for each of the oscillators is exactly that for the oscillators at thermal equilibrium; the evolution drives both the mutual information and the coarse-grained entropy up to their maximum values, equal to the equilibrium thermodynamic entropy.

Our results imply that correlations exhibited by interacting quantum systems in a pure state are not only likely to increase from an initial zero value, but that they are likely to tend to continue rising and attain an equilibrium value given by the normal statistical mechanical entropy for the systems.

Chapter 4: Why Deviations from Wave Function Collapse are Hard to Detect

In this chapter, we use the example of a Stern-Gerlach apparatus to show that the quantum mechanical measurement problem ceases to be a physical problem for a measuring device with many degrees of freedom in the sense that deviations from wave function collapse become virtually impossible to detect.

The quantum mechanical measurement problem

The quantum mechanical measurement problem stems from the following curious fact. When a quantum mechanical system evolves and interacts normally its state vector changes in a continuous and deterministic fashion, governed by the Schroedinger equation; yet if it undergoes an interaction that constitutes a 'measurement,' the system jumps stochastically into an eigenstate of the observable measured, a process -- called collapse of the wave function -- that is incompatible with Schroedinger evolution. A number of papers have tried to reconcile the determinism of quantum dynamics with the inherent chanciness of measurement.[1,2] In this chapter, we give a treatment of the measurement problem along the lines of Peres[3] and Zurek;[4] in particular, we show generally that if our measuring devices are systems with many degrees of freedom, and if we are confined to making a reasonable number of measurements on these devices, then the measurement problem ceases to be problematic in the sense that deviations from wave function collapse become virtually impossible to detect.

We use Dirac notation to give a quantum mechanical description of an object system, on which a certain quantity is to be measured, coupled to a measuring device. If we label the eigenstates of the operator to which the quantity to be measured corresponds as $|i\rangle$, a measurement interaction will be effected by the following joint evolution for the object and measuring device:

$$|i\rangle_{\text{object}}|0\rangle_{\text{device}} \quad \rightarrow \quad |i\rangle_{\text{object}}|a_i\rangle_{\text{device}} \quad ,$$

where $|0\rangle$ is the initial state of the measuring device, and $|a_i\rangle$ is a state that functions as a pointer, to tell us that the object system is in the ith eigenstate. Note that the object system remains undisturbed by the measurement; our argument does not require this idealization, but is made less ponderous by it. If the object is not in a single eigenstate but in a superposition of different eigenstates, we have the following Schroedinger evolution of the state of the composite system:

$$\sum_i d_i|i\rangle|0\rangle \quad \rightarrow \quad \sum_i d_i|i\rangle|a_i\rangle \quad .$$

Now the problem comes in. When we perform a real experiment in which the measuring device has been constructed to give the sort of evolution above, we find that the final state of the object-device system is of the form $|i\rangle|a_i\rangle$: that is, the wave function of the object system has 'collapsed'--jumped to a single eigenstate, in spite of our setting up the measurement so as not to disturb the object system. If we perform the experiment a number of times, preparing the object system in the state $\sum_i d_i|i\rangle$ and the measuring device in the state $|0\rangle$ each time, we find that the system jumps into the eigenstate $|i\rangle$ of the superposition with probability $|d_i|^2$. Hence we have von Neumann's projection postulate[5]: during measurement the state of a system is projected onto (collapses stochastically to) one of the eigenstates of the operator measured, i.e.,

$$\sum_i d_i|i\rangle|0\rangle \quad \rightarrow \quad |i\rangle|a_i\rangle \quad , \qquad \text{with probability } |d_i|^2 .$$

This evolution is clearly different from the evolution given by the Schroedinger equation for the combined object-device system. Von Neumann showed that we can not explain away this difference by ascribing it to insufficient knowledge either of the initial state of the object or of the apparatus.[6] The uncertainty of the final experimental result is inherent in the quantum mechanical measurement process. Such a stochastic evolution is incompatible with the deterministic evolution given by the Schroedinger equation, and is irreversible if we run our measurement process backwards in time: we can not be sure that our present state will jump backward into the state it just came from.

Given the apparently irreducibly statistical nature of quantum mechanics, one might still hope to give an unambiguous interpretation to the final superposition in terms of probabilities: that is, one might claim that the state $\sum_i d_i|i\rangle|a_i\rangle$ represents a situation in which the object and measuring device are either in the state $|1\rangle|a_1\rangle$, <u>or</u> in the state $|2\rangle|a_2\rangle$, <u>or</u> in the state $|3\rangle|a_3\rangle$, etc. . Unfortunately, the different components of the superposition can still interfere with each other after the measurement interaction has taken place, and if we have not looked at the results of the measurement yet, we cannot ignore the possibility of interference between states, only one of which we would like to imagine as present.

To make this point mathematically explicit, imagine that we have isolated our object system and apparatus in a box, so that we know only that the measurement interaction has taken place after some time, but not what the result was. We know the probability is $|d_i|^2$ that the state of the object and apparatus will be found to be $|i\rangle|a_i\rangle$ when the box is opened. Our probabilistic description of the object-apparatus system must be given in terms of a weighted average of states, conveniently expressed in density matrix notation as

$$\bar{\rho} = \sum_i |d_i|^2 |i\rangle\langle i| \otimes |a_i\rangle\langle a_i| \quad .$$

Note that $\bar{\rho}^2$ does not in general equal $\bar{\rho}$; the density matrix that corresponds to our description of the object-apparatus system is that of a system not in a pure state, but in a mixture. However, the object system and apparatus began in a pure state:

$$\rho = |\psi\rangle\langle\psi| \ , \ |\psi\rangle = \sum_i d_i |i\rangle|0\rangle, \ \rho^2 = |\psi\rangle\langle\psi|\psi\rangle\langle\psi| = |\psi\rangle\langle\psi| = \rho,$$

and pure states are carried to pure states by Schroedinger evolution. Pure states cannot evolve into mixtures.

We can restate the measurement problem as follows: Why does the state of an object-apparatus system after measurement appear to be characterized by the density matrix

$$\bar{\rho} = \sum_i |d_i|^2 |i\rangle\langle i| \otimes |a_i\rangle\langle a_i| \quad ,$$

a mixture, when Schroedinger evolution suggests that the proper density matrix is

$$\rho = \sum_{i,j} d_i \bar{d}_j |i\rangle\langle j| \otimes |a_i\rangle\langle a_j| = \bar{\rho} + \sum_{i\neq j} d_i \bar{d}_j |i\rangle\langle j| \otimes |a_i\rangle\langle a_j| \quad ,$$

a pure state? (There is, of course, the further question of how the system goes from a mixture of eigenstates to a single one when we open the box and look in, but if the object and apparatus can indeed be described by $\bar{\rho}$, this change is no stranger than the probability of a flipped coin being heads going from 1/2 to 0 or 1 when we look at it. That is, we can give an ignorance interpretation to the mixture where we cannot to the pure state.[7])

To see that the measurement problem is really a problem, suppose that we 'measure' the spin of one electron by correlating it with the spin of another electron in the following interaction:

$$(1/\sqrt{2})\ \{\ |\uparrow_x\rangle - |\downarrow_x\rangle\ \}\ |\uparrow_z\rangle \quad \rightarrow \quad (1/\sqrt{2})\ \{\ |\uparrow_x\rangle|\downarrow_x\rangle - |\downarrow_x\rangle|\uparrow_x\rangle\ \} \quad .$$

The final state is just as in the Bohm incarnation of the Einstein-Podolsky-Rosen thought experiment.[8] Here we obviously cannot replace ρ, the density matrix for the pure state, by the mixture,

$$\bar{\rho} = (1/2)|\uparrow_x\rangle\langle\uparrow_x| \otimes |\downarrow_x\rangle\langle\downarrow_x| + (1/2)|\downarrow_x\rangle\langle\downarrow_x| \otimes |\uparrow_x\rangle\langle\uparrow_x| \quad ,$$

since $\bar{\rho}$ does not exhibit correlation between the two particles with respect to spin along the z-axis, a correlation that we know exists.

To paraphrase our statement of the measurement problem: after a measurement interaction has taken place, quantum mechanics tells us that our object system and apparatus are in a pure state, while observation indicates that they are in a mixture. The measurement problem is only a physical problem if we can make a further measurement on both object system and apparatus that distinguishes between the pure state ρ and the mixture $\bar{\rho}$. We have just seen that if our measuring device is an electron, the difference between ρ and $\bar{\rho}$ is simple to detect. But an electron is hardly a typical measuring device. We now show that if we correlate the state of a quantum system to the state of a macroscopic object with many degrees of freedom, the difference between ρ and $\bar{\rho}$ is possible to detect in principle but very hard to detect in practice.

If the equations of motion for the measuring device are integrable, then we can follow the method of Peres[3] to exhibit an operator B for the object system and measuring apparatus that will reveal large deviations from wave function collapse. Since, in general, the equations of motion for a system with many degrees of freedom are not integrable, one cannot in general use the method of Peres to exhibit operators that give large deviations from wave function collapse. (A measurement corresponding to such an operator might be very difficult to make in any case.) However, as shown below, almost any measurement exhibits deviations from wave function collapse, as long as the operator to which it corresponds does not commute with the operator that defines the macroscopic states of the measuring device to which the states of the object system has become correlated. Nonetheless, to detect such a deviation from wave function collapse, for a measuring device with many degrees of freedom, we show that we must perform an exponentially large number of measurements on the measuring device and quantum system. (In fact, the number of measurements that we must make on average is equal to the dimension of the subspace of Hilbert space corresponding to the original macroscopic measurement that defines the macroscopic states of the measuring device that became correlated with the states of the object system in the first place.) The measurement problem is not a problem if we confine ourselves to making a reasonable number of measurements with the normal sort of instruments that one finds in a laboratory.

For the sake of concreteness, we examine what happens when we couple the spin of an electron to the state of scintillators in a Stern-Gerlach apparatus; the result derived depends only on the many degrees of freedom of the scintillators, and our treatment is easily generalized to any quantum system coupled to an arbitrary macroscopic measuring device. Suppose that we have a device that uses a magnetic field with a non-zero gradient in the z-direction to split an atomic beam into two components, in one of which the electronic spins have spin z up and pass through scintillator 1, in the other of which the electrons have spin z down and pass through scintillator 2. Let us assume that the scintillators and magnet are initially in the pure state $|0>$; an apparatus that is not initially in a pure state can only make it harder to detect a difference between ρ and $\bar{\rho}$.

If we pass an electron in the state $a|\uparrow_z\rangle + b|\downarrow_z\rangle$ through the apparatus, the evolution of the combined electron-apparatus system is as follows:

$$(a|\uparrow_z\rangle + b|\downarrow_z\rangle)|0\rangle \quad \rightarrow \quad |\psi\rangle = a|\uparrow_z\rangle|\mathrm{up}\rangle + b|\downarrow_z\rangle|\mathrm{down}\rangle \quad ,$$

where $|\mathrm{up}\rangle$ is a state in which only scintillator 1 has fired, and $|\mathrm{down}\rangle$ is a state in which only scintillator 2 has fired.

The difference between using a scintillator to measure the spin of the electron, and using another electron, is that the scintillator is an object with many degrees of freedom: the equations of motion for the scintillator are in general not integrable, and our knowledge about the state of the scintillator is given by the results of a set of macroscopic measurements made on the apparatus itself. Even if the system is in a pure state, the measurements that determine the macroscopic state of the system only locate that state to within a large-dimensional subspace of Hilbert space. Since the scintillator has many degrees of freedom, what these measurements reveal about about the state $|\mathrm{up}\rangle$ is that it belongs to some subspace H_{up}, where the dimension of H_{up} is large (on the order of the exponential of Avogadro's number). Similarly, what the measurements reveal about $|\mathrm{down}\rangle$ is that $|\mathrm{down}\rangle \in H_{\mathrm{down}}$, where H_{down} is a subspace orthogonal to H_{up}; the dimension of H_{down} is also large.

If the measurement problem is to be of physical importance for the measurement described, then after the initial measurement interaction we must be able to make a further measurement on the electron and Stern-Gerlach apparatus that distinguishes between ρ and $\bar{\rho}$, where
$\rho = |\psi\rangle\langle\psi|$, and

$$\bar{\rho} = |a|^2 |\uparrow_z\rangle\langle\uparrow_z| \otimes |\mathrm{up}\rangle\langle\mathrm{up}| + |b|^2 |\downarrow_z\rangle\langle\downarrow_z| \otimes |\mathrm{down}\rangle\langle\mathrm{down}| \quad .$$

Such a measurement corresponds to an Hermitian operator, B, such that

$$\mathrm{tr}\rho B \neq \mathrm{tr}\bar{\rho}B \quad .$$

That is, we require B such that

$$\langle\psi| B - P_{\mathrm{up}} B P_{\mathrm{up}} - P_{\mathrm{down}} B P_{\mathrm{down}} |\psi\rangle \neq 0 \quad ,$$

where P_{up} , P_{down} are the projection operators onto H_{up}, H_{down}, respectively.

One need not exhibit exotic operators to find such a B: virtually any operator that does not commute with P_{up} and P_{down} will satisfy this inequality. But we can now show that if our measuring device has many degrees of freedom, then for any B, even the most advantageously chosen, the difference between $\mathrm{tr}\rho B$ and $\mathrm{tr}\bar{\rho}B$ is likely to be very small and hard to detect. (One might also imagine that one could detect a difference between ρ and $\bar{\rho}$ by comparing the standard deviations that they give from $\mathrm{tr}\rho B$, $\mathrm{tr}\bar{\rho}B$. The results of chapter three imply that the differences between the standard deviations and between the higher moments are just as hard to detect as the difference between $\mathrm{tr}\rho B$ and $\mathrm{tr}\bar{\rho}B$.)

A given initial state of the electron, $a|\uparrow_z\rangle + b|\downarrow_z\rangle$, gives a final state $|\psi\rangle = a|\uparrow_z\rangle|up\rangle + b|\downarrow_z\rangle|down\rangle$ for the object and Stern-Gerlach apparatus, where $|up\rangle$ is an arbitrary state in the Hilbert space H_{up}, and $|down\rangle$ is an arbitrary state in H_{down}. The difference between $tr\rho B$ and $tr\bar{\rho}B$ is impossible to calculate in the absence of exact knowledge of the states $|up\rangle$, $|down\rangle$; we can, however, give an expectation value to the difference betweent $tr\rho B$ and $tr\bar{\rho}B$ by calculating the average amount by which by which $tr\rho B$ differs from $tr\bar{\rho}B$ as $|up\rangle$, $|down\rangle$ range over H_{up}, H_{down} respectively. In fact, the average over $|up\rangle\epsilon H_{up}$, $|down\rangle\epsilon H_{down}$ of ($tr\rho B - tr\bar{\rho}B$) is zero. A better measure of deviation from wave function collapse is the root mean square average amount that the expectation value of B over ρ differs from its expectation value over $\bar{\rho}$: i.e., our measure of deviation from wave function collapse is

$$\Delta = \{ \text{ average over } |\chi\rangle\epsilon H_{up}, |\xi\rangle\epsilon H_{down} \text{ of } (tr\rho B - tr\bar{\rho}B)^2 \}^{\frac{1}{2}}$$

$$= \{ \text{ average over } |\chi\rangle\epsilon H_{up}, |\xi\rangle\epsilon H_{down} \text{ of }$$

$$(\langle\psi| \ B - P_{up}B \ P_{up} - P_{down}B \ P_{down}|\psi\rangle)^2 \ \}^{\frac{1}{2}}$$

$$\text{where } |\psi\rangle = a|\uparrow_z\rangle|\chi\rangle + b|\downarrow_z\rangle|\xi\rangle \quad ,$$

$$= \{ \text{ average over } |\chi\rangle\epsilon H_{up}, |\xi\rangle\epsilon H_{down} \text{ of }$$

$$(\bar{a}b\langle\chi|\langle\uparrow_z| \ B \ |\downarrow_z\rangle|\xi\rangle \ + \ \bar{b}a\langle\xi|\langle\downarrow_z| \ B \ |\uparrow_z\rangle|\chi\rangle)^2 \ \}^{\frac{1}{2}}$$

$$= \left\{ \left(\frac{2|a|^2|b|^2}{k(n-k)}\right) \operatorname{tr} (P_{up}B\, P_{down})(P_{up}B\, P_{down})^{\dagger} \right\}^{\frac{1}{2}} ,$$

where $k = \dim H_{up}$, $n-k = \dim H_{down}$,

$$= \left\{ \frac{2n|a|^2|b|^2}{k(n-k)} \right\}^{\frac{1}{2}} \left\{(\operatorname{tr} P_{up}B\, P_{down}B)/n \right\}^{\frac{1}{2}} .$$

To detect deviations from wave function collapse, we must measure B to within an accuracy of Δ.

To get a scale-free measure of the deviation from wave function collapse, we divide Δ by the root mean square average value of the eigenvalues of B over the states in question:

$$\Delta/\bar{B} = \left\{ \frac{2n|a|^2|b|^2}{k(n-k)} \right\}^{\frac{1}{2}} \left\{(\operatorname{tr} P_{up}B\, P_{down}B)/n \right\}^{\frac{1}{2}} / \bar{B} \quad , \text{where}$$

$$\bar{B} = \left\{ (1/n) \operatorname{tr} \left[(P_{up}+ P_{down}) B (P_{up}+ P_{down})\right]^2 \right\}^{\frac{1}{2}} .$$

$\Delta/\bar{B}$ is essentially the ratio between the amount by which deviations from wave function collapse disturb the the result of a measurement of B, and the magnitude of the result; in order to detect deviations from wave function collapse, we must measure B to within a fractional accuracy of better than $\Delta/\bar{B}$.

Note that $\left\{ \frac{2n|a|^2|b|^2}{k(n-k)} \right\}^{\frac{1}{2}} \leq \left(2/\min(k,n-k) \right)^{\frac{1}{2}}$, and that even if we choose B so as to maximize Δ, we have $(\operatorname{tr} P_{up}B\, P_{down}B)/n \leq \bar{B}^2$. Hence

$$\Delta/\bar{B} \leq \left\{ 2/\min(k,n-k) \right\}^{\frac{1}{2}} .$$

As the number of degrees of freedom of the measuring device gets greater and greater, $\Delta/\bar{B}$ is suppressed by as factor of 1/(dimension of the space of accessible states)$^{\frac{1}{2}}$, and the difference between ρ and $\bar{\rho}$ becomes harder and harder to detect. In fact, since in order to detect a factor of $1/\sqrt{m}$ difference in expectation values it is necessary to make at least m repetitions of the experiment, to detect a difference between $\mathrm{tr}\rho B$ and $\mathrm{tr}\bar{\rho}B$ we must make a number of measurements equal to the dimension of the space of states accessible to the measuring device, a number on the order of the exponential of the number of degrees of freedom of the device. It is possible, of course, that the particular state $|\psi>$ in which the electron and measuring device end up gives a large value for $\mathrm{tr}\rho B - \mathrm{tr}\bar{\rho}B$, but if our apparatus has very many degrees of freedom such a large value is very unlikely. In the thermodynamic limit $\Delta/\bar{B} \rightarrow 0$ and deviation from wave function collapse becomes impossible to detect. That is to say, the measurement problem is not a problem for measuring devices with many degrees of freedom if after the initial measurement interaction we confine ourselves to making a reasonable number of measurements.

Discussion

In principle, it is always possible to detect deviations from wave function collapse. Now matter how complex our measuring apparatus, if we know its exact state, can integrate its equations of motion, and can arrange a measurement corresponding to any Hermitian operator that we like, then deviations from wave function collapse can be detected with only a few measurements. In practice, when our measuring apparatus has many degrees of freedom, so that we can not integrate its equations of motion, and so that precise knowledge of its state requires a very large number of measurements to be made on the apparatus itself, then even if we can make measurements corresponding to any Hermitian operator that we want, such deviations are virtually impossible to detect. We have shown this result here for a Stern-Gerlach apparatus making measurements on electrons, but the result is easily generalized to any apparatus with many degrees of freedom, making measurements on any quantum system. In the thermodynamic limit, as the number of degrees of freedom of the apparatus goes to infinity, the quantum measurement problem ceases to be a problem.

Appendices

Appendix A:

Proof of theorem 1, chapter 2.

First, the mean values of $p_t(i)$ and $p_t(j)$ as t ranges over S^{mn} are $1/m$, $1/n$ respectively: $\langle p_t(i)\rangle = 1/(mn)! \sum_{t \in S^{mn}} p_t(i)$, and $\langle p_t(i)\rangle = \langle p_t(i')\rangle$, so $\langle p_t(i)\rangle = 1/(mn)! \sum_{t \in S^{mn}} (1/m) \sum_{i=1}^{m} p_t(i) = 1/(mn)! \times (mn)! \times (1/m) \times 1 = 1/m$. Similarly, $\langle p_t(j)\rangle = 1/n$.

The variance of $p_t(i)$ from its mean is $\langle (p_t(i) - \langle p_t(i)\rangle)^2 \rangle = \langle p_t(i)^2 \rangle - \langle p_t(i)\rangle^2 = 1/(mn)! \sum_{t \in S^{mn}} p_t(i)^2 - 1/m^2$

= 1/(mn)! × (the sum over all partitions of the p(ij) into m ordered sets of n numbers of (the sum of numbers in the i-th set)^2) - $1/m^2$

$= 1/(mn)! \sum_{t \in S^{mn}} \{ \sum_{j} p(t^{-1}(ij) \} \{ \sum_{j'} p(t^{-1}(ij')) \} - 1/m^2$.

Pick one of the p(ij); call it p. If one were to write out all the terms in the sum over permutations, p would appear (mn)!/m times multiplied by itself, and (mn)!/m times multiplied by 2×(the sum of n-1 of the remaining p(ij)). p appears with each of the mn-1 remaining p(ij) an equal number of times in the sum, and so the variance

$= 1/(mn)! \{ (mn)!/m [\sum_{i,j=1}^{m,n} p(ij)^2 + (n-1)/(mn-1) \sum_{ij \neq i'j'} p(ij)p(i'j')] \}$.

But since $\sum_{i,j} p(ij) = 1$, $\sum_{ij \neq i'j'} p(ij)p(i'j') = \sum_{i,j} p(ij) \{ 1 - p(ij) \}$, and the expression for the variance, after a little algebra, is

$$\langle p_t(i)^2 \rangle - \langle p_t(i) \rangle^2 = \frac{n(m-1)}{m(mn-1)} \{ \sum_{i,j} p(ij)^2 - 1/(mn) \}.$$

The standard deviation stated in the theorem is the square root of the variance.

Appendix B:

Proof of theorem 3, chapter 2.

Every state $|\psi\rangle = \sum_{i,j=1}^{m,n} c_{ij}|a_i\rangle|b_j\rangle$, where $m \leq n$ and $\sum_{i,j} |c_{ij}|^2 = 1$, an be written in the form $|\psi\rangle = \sum_{i=1}^{m} \gamma_i |a_i'\rangle|b_i'\rangle$, where $\sum_i |\gamma_i|^2 = 1$. The robability that the $|\gamma_i|$ take on a particular set of values if the c_{ij} re chosen at random subject to the normalization constraint $\Sigma|c_{ij}|^2=1$ s proportional to the volume on the unit sphere in C^{mn} taken up by tates of the form $|\chi\rangle = \sum_{i=1}^{m} \gamma_i |a_i\rangle|b_i\rangle$ for the particular set of $|\gamma_i|$ ut for any normalized, orthogonal $|a_i\rangle$, $|b_i\rangle \in H_A, H_B$. We calculate his volume as follows.

The set whose volume is to be calculated is equal to the orbit of ny state of the form $|\psi^0\rangle = \sum_{i=1}^{m} \gamma_i |a_i^0\rangle|b_i^0\rangle$, for the given $|\gamma_i|$, under he action of the tensor product unitary group $U(m) \otimes U(n)$. We can arametrize this orbit in terms of a set of the group parameters. The nit sphere in C^{mn} is isomorphic to the unit sphere in R^{2mn}. The set of ormalized states corresponds to a 2mn-1 dimensional manifold in R^{2mn}. ixing the values of $|\gamma_i|$ subject to the requirement $\sum_i |\gamma_i|^2 = 1$ adds m-1 further independent constraints: the set whose volume we are to calculate is a 2mn-1 - (m-1) = 2mn-m dimensional manifold in R^{2mn}.

Any $U_1 \epsilon U(m)$ can be written $U_1 = e^{-i \sum_{i,j} a_{ij} A_{ij}} = e^{-i\, \vec{a} \cdot \vec{A}}$, where i,j range from 1 to m, a_{ij} are real numbers, and A_{ij} is an m×m Hermitian matrix written with respect to the basis $\{|a_i^0>\}$ with entries $(A_{ij})^{kl} = \delta_{ik}\delta_{jl}$ if i=j, $= (1/\sqrt{2})\ (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$ if i<j, $= (i/\sqrt{2})\ (\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk})$ if i>j. The A_{ij} form a basis for the Lie algebra of U(m) under the trace norm: tr $A_{ij}A_{kl} = \delta_{ik}\delta_{jl}$. Similarly, any $U_2 \epsilon U(n)$ can be written $U_2 = e^{-i \sum_{i,j} b_{ij} B_{ij}} = e^{-i\, \vec{b} \cdot \vec{B}}$, where i,j range from one to n, the b_{ij} are real, and B_{ij} are defined in the same way as A_{ij}, with respect to a basis the first m of whose members are $\{|b_i^0>\}$.

Different members A_{ij}, B_{kl} of the algebra for $U(m) \otimes U(n)$ generate different infinitesimal translations in the neighbourhood of $|\psi^0>$. For each translation, we can define a one-form on the space of parameters a_{ij}, b_{kl}:

$$D_{ij} = e^{-i\ da_{ij}A_{ij}} \otimes I_{n\times n}\ |\psi^0> \quad - \quad |\psi^0>$$

$$= \ -i\ (\ da_{ij}A_{ij} \otimes I_{n\times n}\ ) \sum_{i=1}^{m} \gamma_i |a_i^0>|b_i^0>.$$

Similarly, define $E_{kl} = I_{m\times m} \otimes e^{-i\ db_{kl}B_{kl}}|\psi^0> \ - \ |\psi^0>$. There are $m^2 + n^2$ such infinitesimal translations, but only 2mn-m of them are linearly independent. Note that $E_{kl} = 0$ if both k and l are greater than m, and that D_{ii} is proportional to E_{ii} for $1 \leq i \leq m$. The one forms D_{ij},

$1 \le i,j \le m$, and E_{kl} , $k \ne l$, $1 \le k,l \le n$, either k or $l \le m$, correspond to a set of $m^2 + n^2 - (n-m)^2 - m = 2mn - m$ linearly independent infinitesimal translations about the point $|\psi^0>$ in C^{mn}. Let $\bar{D}_{ij}$ and $\bar{E}_{kl}$ be the one forms corresponding linearly independent infinitesimal transformations about the point that corresponds to $|\psi^0>$ in R^{2mn}.

The volume 2mn-m form on the manifold whose volume we wish to calculate is then equal to the wedge product,

$$\bar{D}_{11} \wedge \bar{D}_{12} \wedge \bar{D}_{21} \wedge \ldots \wedge \bar{D}_{mm} \wedge \bar{E}_{12} \wedge \bar{E}_{21} \wedge \ldots \wedge \bar{E}_{mn} \wedge \bar{E}_{nm} .$$

Writing out the explicit form of the terms in this product, and doing the algebra, one finds that the dependence of this 2mn-m form on the $|\gamma_i|$ is

$$(\ |\gamma_1| \ \ldots \ |\gamma_m| \ )^{2(n-m)+1} \prod_{1 \le i<j \le m} (\ |\gamma_i|^2 - |\gamma_j|^2) \ .$$

The volume of the manifold of states that can be written in the form $|\psi> = \sum_i \gamma_i |a_i>|b_i>$ is equal to the integral of the volume 2mn-m form over this manifold. The total volume of this set of states is also proportional to $(\ |\gamma_1| \ \ldots \ |\gamma_m| \ )^{2(n-m)+1} \prod_{1 \le i<j \le m} (\ |\gamma_i|^2 - |\gamma_j|^2)$.

Appendix C:

Proof of theorem 1, chapter 3.

We want to prove the following: Given an Hermitian operator F on $H \cong C^n$, then

$$\{\text{ the average over all } |\psi\rangle \epsilon H \text{ of } (\langle\psi|F|\psi\rangle - (1/n)\mathrm{tr}F)^2 \}^{\frac{1}{2}}$$

$$= (1/(n+1)^{\frac{1}{2}}) \{ (\mathrm{tr}F^2)/n - (\mathrm{tr}F)^2/n^2 \}^{\frac{1}{2}} .$$

Proof:

Pick a basis $\{|e_i\rangle\}$ for H in which F is diagonal:

$F = \sum_i f_i |e_i\rangle\langle e_i|$. The average over all $|\psi\rangle = \sum_i \alpha_i |e_i\rangle \epsilon H$ of $(\langle\psi|F|\psi\rangle - (1/n)\mathrm{tr}F)^2$ is equal to the average over $(\alpha_1,\ldots,\alpha_n) \epsilon$ the unit sphere in C^n of

$$(\sum_{i=1}^{n} f_i(|\alpha_i|^2 - 1/n))^2 ,$$

which is equal to the average over the unit sphere of

$$\sum_{i,j=1}^{n} f_i f_j \{ |\alpha_i|^2|\alpha_j|^2 - (1/n)(|\alpha_i|^2 + |\alpha_j|^2) + 1/n^2 \} .$$

But the average over the unit sphere of $|\alpha_i|^2$ is 1/n, the average of $|\alpha_i|^2|\alpha_j|^2$ is 2/n(n+1) if i=j, and is 1/n(n+1) if i≠j. The average over all $|\psi\rangle \epsilon H$ of $(\langle\psi|F|\psi\rangle - (1/n)\mathrm{tr}F)^2$ is equal to

$$\sum_{i,j} f_i f_j \{ 2\delta_{ij}/n(n+1) + (1-\delta_{ij})/n(n+1) - 2/n^2 + 1/n^2 \}$$

$$= \quad 1/(n+1) \{ (\mathrm{tr}F^2)/n - (\mathrm{tr}F)^2/n^2 \} .$$

Taking the square root of both sides completes the proof.

References

Chapter 1

1. J.C. Maxwell, Theory of Heat, London (1871).
2. L. Szilard, Z.Phys. 53, 840 (1929).
3. L. Brillouin, Science and Information Theory, 2d.ed., New York (1962).
4. D.A. Bell, Am.Sci. 40, 682 (1952).
5. R.C. Raymond, Am.Jour.Phys. 19, 109 (1951).
6. R. Landauer, IBM Journal of Research and Development 5, 183 (1961).
7. R.W. Keyes and R. Landauer, Ibid. 14, 152 (1970).
8. C.H. Bennett, Ibid. 17, 525 (1973); Int.J.Th.P. 21, 905 (1982); Sci.Am. 257, 108 (1987).
9. J. von Neumann, Mathematical Foundations of Quantum Mechanics, Princeton (1955).
10. R.C. Tolman, The Principles of Statistical Mechanics, Oxford (1938).
11. For a more extensive approach to information theory, the classic text is C.E. Shannon and W. Weaver, The Mathematical Theory of Communication, Urbana (1949).
12. K.K. Likharev, Int.J.Theor.Phys. 21, 311 (1982).
13. T. Gold, Am.J.Phys. 30, 403 (1962).
14. Y. Ne'eman, Int.J.Theor.Phys. 3, 1 (1970).
15. W.J. Cocke, Phys.Rev. 160, 1165 (1967).

16. H. Schmidt, J.Math.Phys. 7, 495 (1966).

17. S.W. Hawking, Phys.Rev.D 32, 2489 (1985).

18. D. Page, Phys.Rev.D 32, 2496 (1985).

Chapter 2

1. G. Chaitin, IBM J.Res.Dev. 21, 350 (1977).

2. G. Chaitin, J.Ass.Comp.Mach. 22, 329 (1975).

3. R.J. Solomonoff, Inf.and Cont. 7, 1 (1964).

4. A.N. Kolmogoroff, Probl.Perde.Inf. 1/1 (1965).

5. J.F. Traub, G.W. Wasilkowski, H. Wozniakowski, Information, Uncertainty, Complexity, Reading (1983).

6. E.W. Packel, H. Wozniakowski, Bull.Am.Math.Soc. 17, 9 (1987).

7. In addition, there is an entire journal associated with this approach, the Journal of Complexity.

8. C.H. Bennett, Found.Phys. 16, 585 (1986).

9. C.H. Bennett, in Emerging Syntheses in Science, D. Pines ed., Santa Fe, (1985).

10. C.H. Bennett, in preparation.

11. D. Deutsch, Proc.R.Soc.Lond.A 400, 97 (1985).

12. See also R. Feynman, Int.J.Th.Phys. 21, 467 (1982).

13. C.E. Shannon, W. Weaver, The Mathematical Theory of Communication, Urbana, (1949).

14. L. Brillouin, Science and Information Theory, 2nd ed., London (1962).
15. R.H. Price, Phys.Rev.D 5, 2419,2439 (1972).
16. B.A. Huberman, T. Hogg, Physica 22D, 376 (1986).
17. C.P. Bachas, B.A. Huberman, Phys.Rev.Lett. 57, 1965 (1986).
18. C.P. Bachas, B.A. Huberman, J.Phys.A 20, 4995 (1987).
19. S. Smale, Bull.Am.Math.Soc. 4, 1 (1981).
20. S. Smale, Bull.Am.Math.Soc. 13, 87 (1985).
21. R. Landauer, IBM J.Res.Dev. 5, 183 (1961).
22. R. Landauer, IBM J.Res.Dev. 14, 152 (1970).
23. R. Landauer, Phys.Rev.Lett. 53, 1205 (1984).
24. R. Landauer, Physica Scripta 35, 88 (1988).
25. C.H. Bennett, Int.J.Theor.Phys. 21, 905 (1982).
26. E. Fredkin, T. Toffoli, Int.J.Theor.Phys. 21, 219 (1982).
27. P.A. Benioff, Phys.Rev.Lett. 48, 1581 (1982).
28. P.A. Benioff, Int.J.Theor.Phys. 21, 177 (1982).
29. J. Rothstein suggested mutual information as a measure of self-organization in The Maximum Entropy Formalism, R.D. Levine, M.Tribus ed., Cambridge (1979).
30. F. Dyson, J.Mol.Evol. 18, 344 (1982).
31. P. Anderson, Proc.Nat.Acad.Sci. USA 80, 3386 (1983).
32. L. Demetrius, Proc.Nat.Acad.Sci. USA 81, 6068 (1984).
33. H. Kuhn, J. Waser in Biophysics, ed. W. Hoppe, W. Lohmann, H. Markte, H. Ziegler., New York (1983). Kuhn also notes that computation involves discarding information.

Chapter 3

1. This theorem is implied by the proof of the first part of Jancel's "first statistical ergodic theorem,' in R. Jancel, Foundations of Classical and Quantum Statistical Mechanics, Oxford (1969). Our theorem is not an ergodic theorem, however: it applies to the state of a system at a given time.
2. P. Bocchieri, G.M. Prosperi, in Statistical Mechanics, ed. T.A. Bak, (1967).
3. L. Boltzmann, Wiener Berichte 63, 391 (1871).
4. L. Boltzmann, Wiener Berichte 66, 275 (1872).
5. S.W. Hawking, Nature 248, 30 (1974).
6. S.W. Hawking, Commun.Math.Phys. 43, 199 (1975).
7. J.B. Hartle and S.W. Hawking, Phys.Rev.D 13, 870 (1976).
8. W.G. Unruh, Phys.Rev.D 14, 870 (1976).
9. T.D. Lee, Nucl.Phys.B 264, 437 (1985).
10. B.L. Hu, H.E. Kandrup, Phys.Rev.D 35, 1776 (1987).

Chapter 4

1. See, for example, the papers collected in J. Wheeler and W.H. Zurek, Quantum Theory and Mesurement, Princeton (1983).

2. The proceedings of the New York Academy of Sciences conference, 'New Techniques and Ideas in Quantum Measurement Theory,' N.Y. Acad. Sc., **480** (1986).
3. A. Peres, Phys. Rev. D**22**, 879 (1980).
4. W.H. Zurek, in 'New Techniques and Ideas in Quantum Measurement Theory,' *op.cit.*
5. J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, Princeton (1955).
6. Von Neumann, *op.cit.* His demonstration has a loophole that allows for certain types of hidden variables.
7. One can still ask, 'What happened to the other components of the superposition?' But as long as one cannot tell the difference between ρ and $\bar{\rho}$, this is not a question of physics but of interpretation of quantum mechanical formalism. Whether one regards the components of the superposition other than result of the measurement as being still present in some sense or as having disappeared depends on whether one subscribes to the many-worlds interpretation of quantum mechanics or not.
8. A. Einstein, B. Podolsky, N. Rosen, Phys. Rev. **47**, 777, (1935).

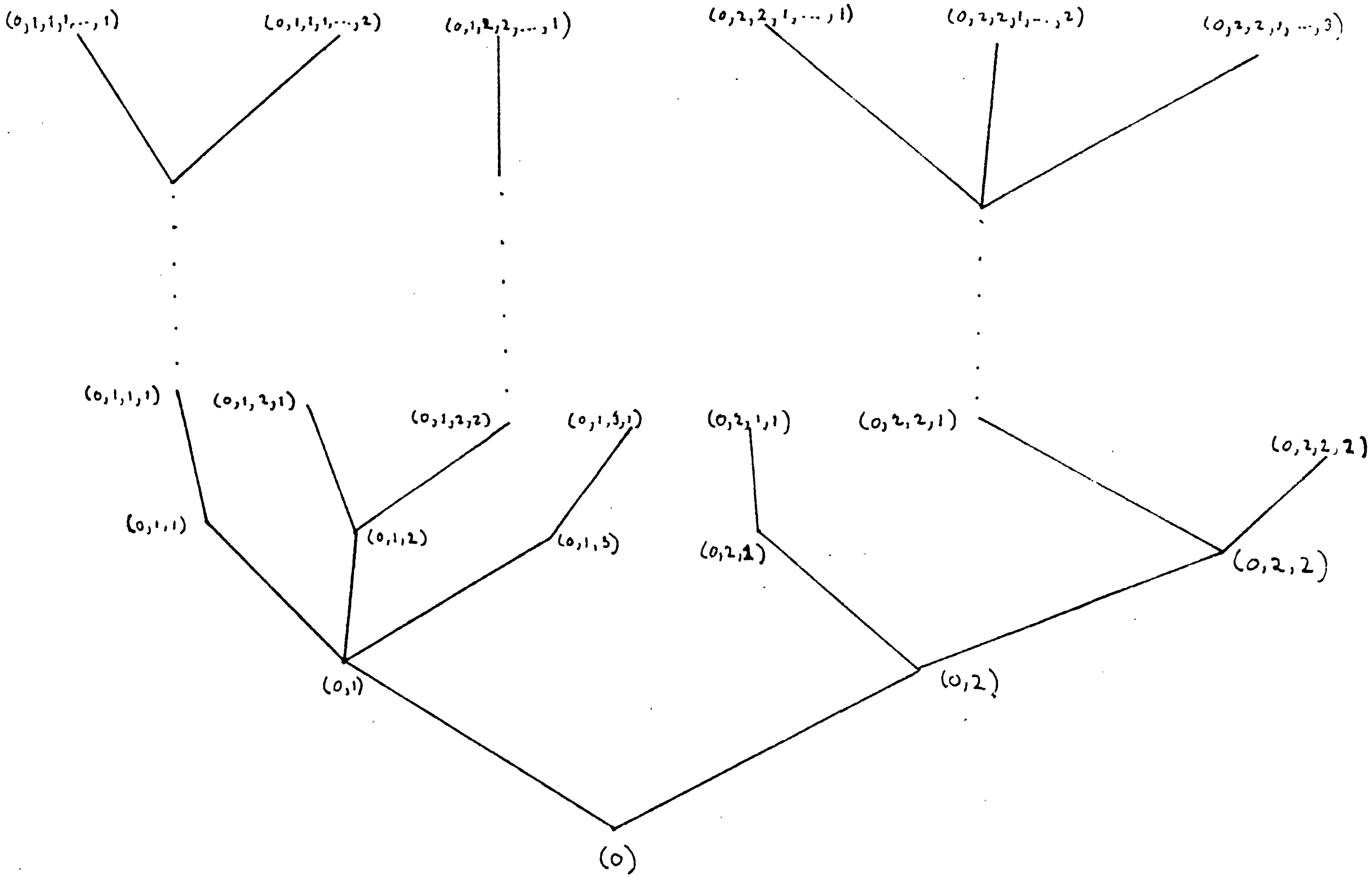

figure 1

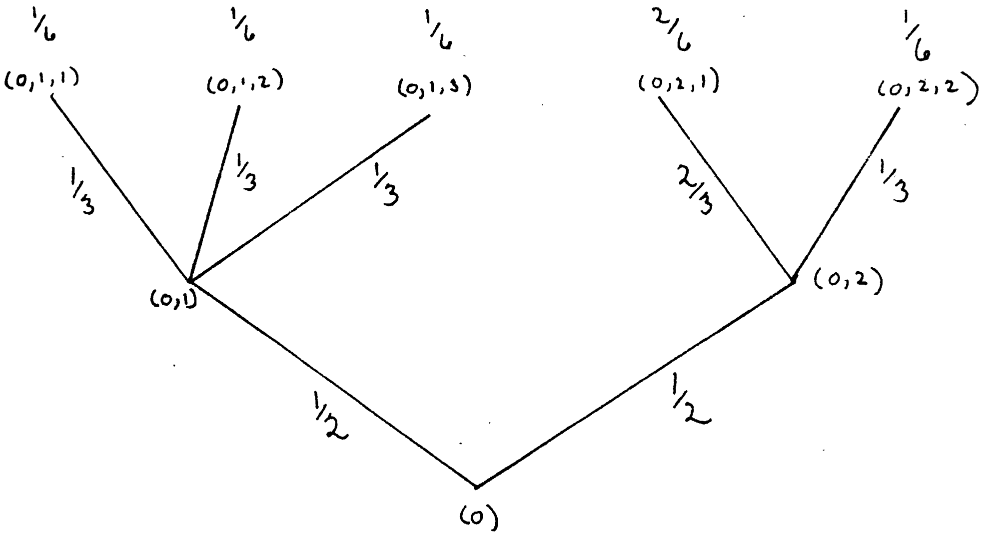

figure 2

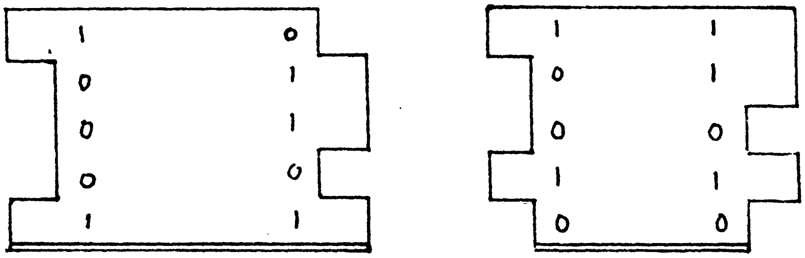

figure 3.