PHYSICAL REVIEW D VOLUME 20, NUMBER 6 15 SEPTEMBER 1979

Entropy and black-hole thermodynamics

Robert M. Wald
Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637

(Received 10 April 1979)

The concept of entropy is examined with an eye toward gaining insight into the nature of black-hole thermodynamics. Definitions of entropy are given for ordinary classical and quantum-mechanical systems which lead to plausibility arguments for the ordinary laws of thermodynamics. The treatment of entropy for a classical system is in the spirit of the information-theory viewpoint, but by explicitly incorporating the coarse-grained observable into the definition of entropy, we eliminate any nonobjective features. The definition of entropy for a quantum-mechanical system is new, but directly parallels the classical treatment. We then apply these ideas to a self-gravitating quantum system which contains a black hole. Under some assumptions—which, although nontrivial, are by no means exotic— about the nature of such a system, it is seen that the same plausibility arguments which lead to the ordinary laws of thermodynamics for ordinary systems now lead to the laws of black-hole mechanics, including the generalized second law of thermodynamics. Thus, it appears perfectly plausible that black-hole thermodynamics is nothing more than ordinary thermodynamics applied to a self-gravitating quantum system.

I. INTRODUCTION

The concept of entropy and its role in the laws of thermodynamics has undoubtedly been one of the most widely discussed issues in physics in the past century. While there are nearly as many points of view on this subject as there are authors, there is basic agreement—at least in the domain of classical mechanics—as to what entropy is and as to the statistical nature of the laws of thermodynamics. However, some remarkable recent developments in the theory of black holes have given cause for reexamination of some of these ideas.

First, a complete mathematical analogy was found between the ordinary laws of thermodynamics and the laws of black-hole mechanics in general relativity.[1] Most striking is the analogy between the law of entropy increase in thermodynamics and the law of area increase[2] for black holes. Since the laws of thermodynamics are statistical in nature—valid only with high probability—while the laws of black-hole mechanics of classical general relativity are rigorous theorems in differential geometry, most theorists at first viewed this analogy as merely a mathematical curiosity. However, this situation changed dramatically after Hawking's discovery[3] of the quantum particle-creation process near black holes which results in an exactly thermal emission of particles to large distances.[4] The temperature of this radiation equals (up to a numerical factor) the surface gravity of the black hole, which is the quantity playing the role analogous to temperature in the laws of black-hole mechanics. Furthermore, this result made entirely viable the generalized second law of thermodynamics which had earlier been proposed by Bekenstein.[5]

The idea here is the following: If black holes are present it is unpleasant to have to include the entropy of matter that has fallen into a black hole in counting up total entropy, since neither the entropy nor any other property of this matter can in principle be measured by an observer outside the black hole. But the total entropy of matter outside black holes can of course, decrease by simply allowing matter to fall into a black hole. Bekenstein defined a new quantity, called the generalized entropy S', to be the sum of the total entropy of matter outside black holes (S_m) plus a multiple of the total area of the black holes (S_{bh}). He proposed that the generalized entropy can never decrease, since when matter falls into a black hole (thereby decreasing S_m), it also increases S_{bh}. This generalized second law does not quite work with a purely classical picture of black holes because, by sending in low enough temperature radiation, one can decrease S_m by a large amount while keeping the increase in S_{bh} small. But the quantum particle-creation process saves one from this potential violation of nondecrease of S' by making the black hole emit more than it would absorb in this situation. Furthermore, although the particle-creation process violates the area-increase theorem of classical black-hole physics—black-holes area decreases as a result of it—it does so at the expense of emission of thermal radiation with high S_m and thus no decrease of S'. Thus while S_m and S_{bh} may individually decrease, it appears to be true that S' never decreases.

The above results and ideas have left little doubt that the connection between black hole and thermodynamics is considerably more than a mathematical analogy between certain formulas. But the full nature of this fundamental connection remains to be explored. Indeed, it is widely believed that the notion of entropy in quantum gravitational physics may be much more fundamental than it is in ordinary physics (where it has the clear status of a derived concept, dependent on coarse graining, etc.) and that this may somehow be related to black-hole thermodynamics.

The main purpose of this paper is to gain more insight into this issue. We shall argue in Sec. VI that the laws of black-hole mechanics—including the generalized second law of thermodynamics—are nothing more than the ordinary laws of thermodynamics applied to a self-gravitating quantum system containing a black hole. In particular, there appears to be no need to postulate a more fundamental role of entropy for a quantum gravitational system than for an ordinary system.

Before discussing self-gravitating quantum systems with black holes, it will be necessary to give a more precise treatment of entropy for ordinary dynamical systems. Therefore, we begin in Sec. II with a discussion of entropy in classical mechanics. We give a precise definition of entropy which explicitly incorporates the dependence of entropy on the choice of coarse-grained observable. This definition is in the spirit of information theory but clears up some of the nonobjective aspects of that viewpoint. Here the entropy of a system depends on the observable but not on the observer or how much he knows about the system. The thermodynamic entropy of a subsystem—defined as the maximum of the entropy over all states—is, for a wide class of observables, independent of observable and is given by the standard microcanonical ensemble formula, so standard thermodynamic formulas may be recovered. The treatment given here also provides a simple resolution of the Gibb's paradox concerning the entropy of mixing.

The definition of entropy in ordinary Schrödinger quantum mechanics is the subject of Sec. III. As in classical mechanics the precise definition of entropy depends on what observable is being measured. Entropy itself will be defined to be an observable, i.e., an operator. The definition given in Sec. III will *not* reduce to the commonly given $\mathrm{tr}(\rho \ln \rho)$ formula for density matrices ρ. In fact, the entropy of pure states can be high, the entropy of density matrices can be low, and the entropy of both can, and usually will, vary with time (as it must if it is to reduce to the classical notion of entropy in the classical limit). However, for a wide class of observables, the thermodynamic entropy will again be independent of observable and be given by the standard microcanonical ensemble formula, so the usual thermodynamic formulas may again be recovered.

The utility of the notion of entropy stems from the fact that, as argued in Secs. II and III, it is highly plausible that for most (complicated) systems the total entropy of a given state is the logarithm of the fraction of time the system spends in that state or others which have the same value of the coarse-grained observable. In other words, it measures the fraction of time the system "looks like" it is in the given state. In classical mechanics the main argument leading to this conclusion is the standard ergodic argument usually employed to justify the microcanonical ensemble. Similar arguments in the quantum-mechanical case are given in Sec. III.

Thus if a system is in a state of low entropy, it is likely to quickly change its appearance (with respect to the given observable) and go to a state of higher entropy, since it spends very little of its time in states of low entropy. On the other hand, if the system is in a state of maximum entropy, it is likely to "just sit there" in a state of apparent equilibrium (with respect to the given observable). In other words, we have a plausibility argument for the second law of thermodynamics, derived in a manifestly time-reversal-invariant manner. We also have an important entropy criterion for the (apparent) equilibrium of a system. The potential uses and pitfalls of entropy in stability analysis are discussed in Sec. IV and illustrated there by the example of a Newtonian self-gravitating gas.

Plausibility arguments for the laws of thermodynamics are given in Sec. V. These arguments are quite simple and general, but of course are far from rigorous. To show that thermodynamics rigorously applies to a given system would require a detailed study of that system, and such a study is not attempted here. Our aim in Sec. V is not to prove the laws of thermodynamics but to gain a basic understanding of why they work in the cases where they do apply.

Finally, in Sec. VI we apply these ideas to a self-gravitating quantum system with a black hole. A satisfactory quantum theory of gravitation does not yet exist, so the ideas of this section are speculative. The main assumption that will be made is that a quantum self-gravitating system behaves basically like an ordinary quantum system. This is a much stronger assumption than it might at first appear to be, since a classical self-gravitating system, either Newtonian or general relativistic, does not behave like an "ordinary" classical system. However, as explained in Sec. VI, the quantum particle-creation results lend plausibility to the validity of this assumption. Assuming further that the degrees of freedom of the system can be (approximately) divided between "black-hole states" and "other states" and assuming a specific formula for black-hole entropy with respect to a naturally chosen observable, we argue for the main conclusion of this paper that black-hole thermodynamics is, plausibly, just ordinary thermodynamics applied to a quantum self-gravitating system.

II. ENTROPY IN CLASSICAL MECHANICS

The state of a system in classical mechanics is described by a point in its phase space $\mathfrak{M}$. (For a gas consisting of N particles, $\mathfrak{M}$ is $6N$-dimensional.) If we knew the exact state of the system (as well as the exact equations of motion) we could predict the exact future behavior of the system. However, the observables which we can measure in practice do not come close to telling us the exact state of the system and force us into a statistical description of its future behavior. In addition, since we observe only very limited properties of the system, the system may appear to be in equilibrium (i.e., the observables may not change with time over long periods) even though the system is actually undergoing very complicated dynamical motion. It should be emphasized that our notion of equilibrium of such systems depends entirely on what observables are being measured. Indeed, an alien civilization with eyesight good enough to keep track of the motion of individual particles in a classical gas would undoubtedly be surprised that we would ever attribute any equilibrium properties to it.

An isolated system is one whose interaction with other systems can be neglected. It can be characterized by it total energy E as well as other "state parameters" which will be collectively labeled as α. Restriction to given values of E and α will define a submanifold M of the original phase space $\mathfrak{M}$. We shall assume that this "reduced phase space" M is compact. The state parameters α are supposed to be chosen so that the dynamical trajectories on M are (at least, approximately) ergodic. Thus for example, for a gas confined to a box, α ordinarily might consist only of parameters describing the size and shape of the box; however, if the box were exactly spherical, it would also have to include the angular momentum of the gas since it would now be conserved. The issue of determining the correct state parameters for a given system is a nontrivial one. We attempt no general answer here, but merely assume that the correct state parameters have been identified.

An observable on an isolated system is simply a map from M into another set (typically $\mathbb{R}$ or $\mathbb{R}^n$). We define a *coarse-grained observable* $\mathcal{O}$ to be an observable which is piecewise constant, i.e., such that M can be divided into a finite number of cells with $\mathcal{O}$ constant on each cell. (By "cell" we mean a set whose interior is connected smooth manifold and whose boundary is a C^0 manifold of one lower dimension.) Because of finite measurement accuracy, physically measurable quantities correspond much more closely to coarse-grained observables than continuous observables.

Let $\mathcal{O}$ be a coarse-grained observable. We define the total entropy with respect to $\mathcal{O}$ of the state $x \in M$ by

$$S(x) = \ln[\operatorname{vol}\{y \in M \mid \mathcal{O}(y) = \mathcal{O}(x)\}], \tag{1}$$

where the natural (Liouville) measure on M is used to compute the volume, and we use units where Boltzmann's constant k is set equal to 1. In other words, $S(x)$ is the logarithm of the volume of the cell to which x belongs together with the other cells for which $\mathcal{O}$ takes the same value.

As an example, consider a gas of N particles. The most natural observable to consider for this system is the distribution function, measured to finite accuracy; this represents the most we could reasonably expect to measure without keeping track of the individual particles. To define $\mathcal{O}$, we break up the 1-particle phase space into cells of uniform size and take $\mathcal{O}(x)$ to be the collection $\{n_i\}$, where n_i is the number of particles in the ith cell of 1-particle phase space when the system is in state $x \in M$. The cells in M will then be of approximately[6] uniform size so, up to an additive constant, the entropy of x will just be the logarithm of the number of cells which have the same value of $\mathcal{O}$ as x. Hence we recover the familiar formula

$$S(x) = \ln \frac{N!}{\prod_i n_i!} + \text{const} \tag{2}$$

which, upon use of Stirling's approximation, leads to the standard formula

$$S(x) = -\sum_i n_i \ln n_i + \text{const}. \tag{3}$$

However, if we had chosen a totally different coarse-grained observable $\mathcal{O}$ (e.g., one which kept track of the individual particles) we would have obtained a quite different formula for S.

The importance and utility of S stems from the idea that $S(x)$ measures the fraction of time the system spends during its dynamical evolution in states which are indistinguishable (by the observable $\mathcal{O}$) from x. The main arguments for this are the well-known arguments used to justify the microcanonical ensemble, which we shall not repeat in detail here (see, e.g., Ref. 7). If M is metrically indecomposable (i.e., if M cannot be written as the disjoint union of two sets of positive measure each of which is taken into itself by the dynamical evolution) it follows from Birkhoff's theorem[7] that for almost all initial states, the fraction of time spent in each cell equals its volume of that cell divided by the volume of M. This means that for almost all states $S(x)$ is (up to an additive constant) the logarithm of the fraction of time $\mathcal{O}$ has the same value as it does at x. The assumption of metric indecomposability is unrealistically strong and in any case the time scale required to the system to manifest this ergodic behavior is enormous compared with normal observation time scales. Nevertheless, if the cells of $\mathcal{O}$ are large, it is plausible (and one can argue, borne out experimentally by the validity of thermodynamics) that for most systems the above interpretation of S will be valid to high accuracy. Furthermore, even if the system—when perfectly isolated—were to display violations of ergodicity (e.g., many trajectories avoid large regions of M), it is likely that small outside disturbances would help give "effective ergodicity" to the system. Estimates of the effects of the tiniest disturbances on systems such as a hard-sphere gas show that they magnify enormously in a few collision times and thus have a very large influence on dynamical trajectories in M. If gross violations of ergodicity still occur, then, of course, new state parameters must be identified to further reduce M to the submanifolds where ergodic behavior is achieved.

Consider now the case where the total phase space $\mathcal{M}$ of the isolated system can be expressed as a product of phase spaces of subsystems,

$$\mathcal{M}=\mathcal{M}_1\times\cdots\times\mathcal{M}_n\,, \tag{4}$$

and the coupling between these subsystems is sufficiently weak that, at least over a limited period of time, we may treat each of them as an (approximately) isolated system in its own right. Then we may assign state parameters E_i, α_i to the ith subsystem defining its reduced phase space M_i. If we have a coarse-grained observable $\mathcal{O}_i$ on M_i, we may define the entropy of a state x_i of the ith subsystem in exactly the same manner as for the original system, namely,

$$S_i(x_i)=\ln[\mathrm{vol}\{y_i\in M_i \mid \mathcal{O}_i(y_i)=\mathcal{O}_i(x_i)\}]\,. \tag{5}$$

The reason for making two separate definitions Eqs. (1) and (5) is logical clarity with regard to the following point. For the total isolated system we do not contemplate every varying its state parameters E, α, whereas for the subsystem we do. Hence for the total system we have the above ergodic argument and interpretation of S as measuring fractions of time (and thus, as discussed in Sec. V, a plausibility argument for the second law of thermodynamics) but no meaningful dependence of S on state parameters, whereas for a subsystem in general we have no ergodic argument but can meaningfully vary E_i and α_i.

Suppose now, as ordinarily is the case, that for each subsystem the coarse-grained observable $\mathcal{O}_i$ arises from a coarse-grained observable $\mathcal{Q}_i$ defined on all of $\mathcal{M}_i$. That is, suppose the cells of $\mathcal{O}_i$ arise from intersecting the cells of $\mathcal{Q}_i$ in $\mathcal{M}_i$ with the submanifold M_i. The collection $\mathcal{Q}=(\mathcal{Q}_1,\ldots,\mathcal{Q}_n)$ defines a coarse-grained observable on the total phase space $\mathcal{M}$. Let $\mathcal{O}$ be the coarse-grained observable obtained from $\mathcal{Q}$ by intersection with M. Each cell of $\mathcal{Q}$ in $\mathcal{M}$ is a product of cells of $\mathcal{Q}_i$ in $\mathcal{M}_i$. Thus the volume of each $\mathcal{Q}$ cell is a product of the volumes of $\mathcal{Q}_i$ cells. If the relative sizes (i.e., ratio of volumes) of the $\mathcal{O}_i$ cells in M_i induced by these $\mathcal{Q}_i$ cells do not vary strongly with the state parameters E_i, α_i, then the volumes of the $\mathcal{O}_i$ cells will be approximately proportional to the volumes of the corresponding $\mathcal{Q}_i$ cells. Similarly, if the relative sizes of the $\mathcal{O}$ cells in M do not vary strongly with E, α, the volumes of the $\mathcal{O}$ cells will be approximately proportional to the volumes of the corresponding $\mathcal{Q}$ cells. Hence the volume of each $\mathcal{O}$ cell will be approximately proportional to the product of the volumes of the $\mathcal{O}_i$ cells. Taking logarithms, we find that the total entropy of a state $x=(x_1,\ldots,x_n)$ in M (where each $x_i\in M_i$) is, to a good approximation, given by

$$S(x)=\sum_{i=1}^{n} S_i(x_i)\,. \tag{6}$$

However, if the observable $\mathcal{O}$ is not chosen in the above manner, the additivity of entropy need not hold.

The above result is best illustrated by the following example: Suppose the isolated system consist of two noninteracting gases confined to the same box. Suppose we take as observables $\mathcal{O}_1$ and $\mathcal{O}_2$ the (coarse-grained) distribution functions of the two gases. The entropy of each gas will then be given by Eq. (3). If we take $\mathcal{O}$ to be the pair of distribution functions (i.e., if the two gases are observably distinct and we measure both distribution functions), then Eq. (6) applies and the total entropy S will be the sum of S_1 and S_2. On the other hand, if $\mathcal{O}$ is say, the sum of the two distribution functions (i.e., the gases are observably indistinct and we measure only their cumulative effect), then S will *not* be $S_1 + S_2$. This resolves the Gibb's paradox on the entropy of mixing. Some textbook treatments[8] of this point claim that the Gibbs paradox cannot be resolved in the context of classical mechanics.

We define the *thermodynamic entropy* $\mathcal{S}_i$ of the ith subsystem of an isolated system to be the maximum of the entropy over all states x_i corresponding to state parameters E_i, α_i:

$$\mathcal{S}_i = \max_{x_i \in M_i} S_i(x_i)\,. \tag{7}$$

Thus, $\mathcal{S}_i$ is a function of the state parameters E_i α_i, not the individual states x_i. Of course, $\mathcal{S}_i$ also depends on the choice of observable $\mathcal{O}_i$. However, for observables which we can in practice measure for most complicated systems, a single group of cells with the same value of $\mathcal{O}_i$ will dominate the reduced phase space M_i. For example, for a gas this will be true if $\mathcal{O}_i$ is taken to be the coarse-grained distribution function (with the phase-space cells not chosen too small and with additional coarse-graining defined on the $\{n_i\}$). In practice we measure considerably less than this about a gas, and the domination by the largest group of cells will be even greater. Thus we can estimate $\mathcal{S}_i$ as

$$\mathcal{S}_i(E_i, \alpha_i) \approx \ln(\mathrm{vol} M_i)\,. \tag{8}$$

Equation (8) is just the standard microcanonical ensemble formula for entropy. From it we can recover all the usual (observable, independent) formulas for the thermodynamic entropy of classical systems.

III. ENTROPY IN QUANTUM MECHANICS

In this section we shall define entropy for an ordinary quantum-mechanical system in close parallel with the treatment for a classical system given in Sec. II. For a density matrix ρ, our definition will *not* reduce to the familiar formula[9] $S = -\mathrm{tr}(\rho \ln \rho)$. This latter expression has the properties that (a) it remains constant with time for a perfectly isolated system and (b) it is not measurable in practice. For these and other reasons, it clearly does not correspond to the classical notion of entropy. Our formula for the thermodynamic entropy of a subsystem will, however, agree with the standard quantum microcanonical ensemble formula. The treatment given here is close in spirit to a discussion of Tolman.[20]

In quantum mechanics, states of a system are described by vectors in a Hilbert space $\mathcal{H}$. However, unlike classical mechanics, an isolated system in quantum mechanics need not have a definite energy. Indeed, it is well beyond our capacity to prepare (or measure) a complicated system to be in an exact energy eigenstate; if we could, then—since the exact energy eigenstates are probably nondegenerate—we could determine the state exactly and would have no need for thermodynamics. However, we can prepare an isolated system so that its probability of being in a state with energy differing from E by more than a small interval ΔE is negligible. The remaining state parameters α for such an isolated system will consist of parameters defining the nature of the system (e.g., for a gas in a box, parameters characterizing the box) together with observables which we measure and which commute with the Hamiltonian (e.g., for a gas in a spherical box, the angular momentum). Thus we shall take the *reduced Hilbert space* H of the isolated system to be the subspace spanned by the energy eigenstates with eigenvalues within ΔE of E, intersected with the similar subspaces of other measured observables (if any) which commute with the Hamiltonian. We shall assume that the reduced Hilbert space H is finite dimensional. This is completely analogous to the assumption that the reduced phase space M in the classical case is compact.

In quantum mechanics an observable is described by a self-adjoint operator acting on the Hilbert space of states $\mathcal{H}$. A measurement of a quantum-mechanical system affects the system, and if we are observing an isolated system it is reasonable to require that we choose an observable which does not disturb the system too greatly. Thus for example, the measurement of the position of particles in a gas to a very high accuracy is not an acceptable observable. We shall take as the criterion for $\mathcal{O}$ to be an acceptable observable that it leave the reduced Hilbert space invariant, i.e., $\mathcal{O}$ maps H into H. However, we do *not* assume that $\mathcal{O}$ commutes with the Hamiltonian; indeed, if $\mathcal{O}$ does, it should be included as a state parameter.

Since in quantum mechanics all information about simultaneously measurable (i.e., commuting) observables can be encoded into a single observable, there is no loss of generality in considering only a single operator $\mathcal{O}$. There is no need in quantum mechanics to make a separate mathematical definition of a coarse-grained observable as, in effect, the quantum theory itself has already done the coarse graining for us. A coarse-grained observable in quantum mechanics is merely an ordinary observable whose degeneracy subspaces are "large." Because of our very limited measurement abilities for complicated systems, the observables which we can measure, in practice, are coarse grained in this sense.

Let $\mathcal{O}: H \rightarrow H$ be an observable for our isolated system. Since $\mathcal{O}$ is self-adjoint, and H is finite dimensional, we can write $\mathcal{O}$ as

$$\mathcal{O} = \sum_j \lambda_j \mathcal{P}_j \tag{9}$$

where the λ_j are the (distinct) eigenvalues of $\mathcal{O}$ and the $\mathcal{P}_j$ are the projection operators onto the eigensubspaces. We define the *entropy operator* S with respect to the observable $\mathcal{O}$ by

$$S = \sum_j (\ln d_j) \mathcal{P}_j \tag{10}$$

where $d_j = \operatorname{tr} \mathcal{P}_j$ is the dimension of the jth eigensubspace. Thus entropy is an observable which, as in the classical case, can be measured by measuring $\mathcal{O}$. The expected entropy of a system described by density matrix ρ is

$$\langle S \rangle = \operatorname{tr}(S\rho) . \tag{11}$$

Unlike the commonly seen $\operatorname{tr}(\rho \ln \rho)$ formula,[9] $\langle S \rangle$ given by Eq. (11) can change with time. Furthermore, the entropy of a pure state lying in a subspace of large dimension d_j is high, whereas the entropy of a density matrix constructed from states lying in a subspace of small d_j is low.

In the classical case discussed in Sec. II we argued that most dynamical trajectories on M spent equal times in equal volumes so that the entropy of a state is the logarithm of the fraction of time the system spends in a state indistinguishable (as far as $\mathcal{O}$ is concerned) from the given one. We shall now argue that in the quantum-mechanical case, most systems spend equal times in subspaces of equal dimension in H and thus that the same interpretation of entropy applies.

Suppose for simplicity that the isolated system is initially described by a pure state $v \in H$ rather than a density matrix. We may ask what is the probability that at time t the observable $\mathcal{O}$ will have value λ_j. This probability is

$$\begin{aligned} p_j(t) &= (\mathcal{P}_j v(t), \mathcal{P}_j v(t)) \\ &= \sum_{k,l} (\mathcal{P}_j \mathcal{E}_k v(0), \mathcal{P}_j \mathcal{E}_l v(0)) e^{i(E_k - E_l)t} , \end{aligned} \tag{12}$$

where $\{\mathcal{P}_j\}$ are the projection operators (spectral family) of $\mathcal{O}$ and $\{\mathcal{E}_k\}$ are the projection operators of the Hamiltonian. The average $\bar{p}_j$ of $p_j(t)$ over time is the closest analog in quantum mechanics to the fraction of time the system spends with the given value λ_j of the observable. We find

$$\begin{aligned} \bar{p}_j &= \lim_{T\to\infty} \frac{1}{T} \int_0^\infty p_j(t) dt \\ &= \sum_k (\mathcal{P}_j \mathcal{E}_k v(0), \mathcal{P}_j \mathcal{E}_k v(0)) \\ &= \sum_k \| \mathcal{P}_j \mathcal{E}_k v \|^2 . \end{aligned} \tag{13}$$

We can estimate the right-hand side of Eq. (13) if we assume that the relation between the invariant subspaces of $\mathcal{O}$ and the Hamiltonian is a "random" one, and that v has been chosen in a random manner with respect to the energy eigenspaces. Projection of the unit vector v onto the kth energy subspace should produce a vector of squared norm l_k/n, where l_k is the dimension of this subspace. Further projection onto the jth subspace of $\mathcal{O}$ should further reduce the squared norm by the factor d_j/n. Thus we estimate

$$\bar{p}_j = \sum_k \frac{d_j}{n} \frac{l_k}{n} = \frac{d_j}{n} , \tag{14}$$

which gives the desired result that the fraction of time the system spends in the jth subspace is proportional to d_j. Note that this estimate applies even if the original vector v is an eigenstate of $\mathcal{O}$, provided, of course, that the relation between $\mathcal{O}$ and the Hamiltonian is random. Even if the relation between $\mathcal{O}$ and the Hamiltonian is not sufficiently random to give the ergodic behavior Eq. (14) for nearly all state vectors for a perfectly isolated system, the inability to perfectly isolate a system from small disturbances should, as in the classical case, contribute toward giving "effectively ergodic" behavior in H. (The extreme sensitivity of the system to tiny disturbances manifests itself in the quantum-mechanical case by the energy difference denominator in the formula for induced transition amplitudes.) If the relation between $\mathcal{O}$ and the Hamiltonian is grossly nonrandom (e.g., in the extreme case, if $\mathcal{O}$ commutes with the Hamiltonian), and if sufficient isolation can be achieved, then of course $\mathcal{O}$ should be included as a state parameter.

Consider now the case where the full Hilbert space of system can be expressed as the tensor product of Hilbert spaces of subsystems,

$$\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n , \tag{15}$$

and the interaction between the subsystems is sufficiently small that we can (over limited periods of time) treat each subsystem as an (approximately) isolated system. Then for each subsystem we can introduce state parameters E_i, α_i and define a reduced Hilbert space H_i. If we have an acceptable observable $\mathcal{O}_i$ for the ith subsystem, we can define the entropy operator $S_i : H_i \rightarrow H_i$ in exactly the same manner as for the original isolated system.

The reduced Hilbert space H of the total system will be contained in the direct sum K of tensor products of the reduced Hilbert spaces H_i.

$$H \subset K = \bigoplus_{\{E_i,\alpha_i\}} \left[\bigotimes_{i=1}^{n} H_i(E_i, \alpha_i) \right], \tag{16}$$

where the sum is taken over all collection of state parameters $\{E_i, \alpha_i\}$ which are compatible with the original state parameters E, α, of the total system. Define the observable $\mathcal{Q}$ on K to be the observation of which reduced Hilbert space H_i each subsystem belongs to together with the observation of $\mathcal{O}_i$ on each subsystem. In other words, the eigenspaces of $\mathcal{Q}$ are precisely the subspaces left invariant by all the projection operators onto the H_i and all the operators $\mathcal{O}_i$. Thus each eigensubspace of $\mathcal{Q}$ is a tensor product over i of invariant subspaces of the $\mathcal{O}_i$ in H_i. Hence, the dimension of each eigensubspace of $\mathcal{Q}$ is the product of the dimensions of the corresponding eigensubspaces of the $\mathcal{O}_i$. Suppose we take our observable $\mathcal{O}$ on H to be the projection of $\mathcal{Q}$ onto H, i.e., $\mathcal{O} = \mathcal{L}\mathcal{Q}\mathcal{L}$, where $\mathcal{L}$ is the projection operator onto H. Assuming that the eigensubspaces of $\mathcal{O}$ are in essence just the projections onto H of those of $\mathcal{Q}$, and assuming that the dimensions of these eigensubspaces of $\mathcal{O}$ are (approximately) proportional to the corresponding ones of $\mathcal{Q}$, (i.e., assuming that $\mathcal{Q}$ itself is essentially an acceptable observable on H), we find that if $\psi_i \in H_i$ are entropy eigenstates with eigenvalues s_i and $\psi = \psi_i \otimes \cdots \otimes \psi_n$ lies in H, then at least to a good approximation, ψ is an eigenstate of total entropy with eigenvalue s given by

$$s = \sum_{i=1}^{n} s_i \,. \tag{17}$$

Since this is (approximately) valid for all eigenvectors $\psi_i \in H_i$, we actually have the operator identity

$$S = \sum_{i=1}^{n} S_i \tag{18}$$

holding to a good approximation on the intersection of $\otimes_{i=1}^{n} H_i$ with H. Thus we recover the additivity of entropy (for suitable choice of $\mathcal{O}$) in a manner analogous to the classical case.[10]

The *thermodynamic entropy* $\mathcal{S}_i$ of the *i*th subsystem is defined as

$$\mathcal{S}_i = \max_{\psi_i \in H_i} (\psi_i, S\psi_i) = \ln D_i \,, \tag{19}$$

where D_i is the dimension of the largest invariant subspace in H_i of $\mathcal{O}_i$. Thus $\mathcal{S}_i$ is a number (not an operator) which depends on the state parameters E_i, α_i. $\mathcal{S}_i$ of course also depends on the choice of observable $\mathcal{O}_i$. However, for observables which we can in practice measure, a single invariant subspace will dominate H_i. Thus we may estimate $\mathcal{S}_i$ as

$$\mathcal{S}_i = \ln D_i \approx \ln(\dim H_i) \,. \tag{20}$$

Equation (20) is the standard (observable independent) formula for the entropy of the microcanonical ensemble; it *is* equal to $-\mathrm{tr}(\rho \ln \rho)$, with ρ the microcanonical density matrix. (If E is the only observable in the state parameters, then $\dim H_i$ is just the density of energy levels at energy E times ΔE.) From Eq. (20) we can recover the usual formulas for the thermodynamic entropy of quantum systems.

IV. TOTAL ENTROPY AND STABILITY

Most systems with many degrees of freedom, such as a gas of particles, are never in equilibrium in a true sense; the system is always undergoing complicated dynamical motion. However, if we measure only the (coarse-grained) observable $\mathcal{O}$, it is possible that the value of $\mathcal{O}$ may remain unchanged over a long period of time, i.e., the system may be in *apparent stable equilibrium* with respect to $\mathcal{O}$. Using the ideas of the previous two sections (for respectively, the classical and quantum-mechanical cases) it is easy to obtain a criterion for apparent stable equilibrium.

By the arguments of Secs. II and III, the entropy of a given state is (up to an additive constant) the logarithm of the fraction of time the system spends in states where $\mathcal{O}$ has the same value as in the given state. Therefore if the entropy of a given state is low, $\mathcal{O}$ should soon change its value as the dynamical evolution proceeds. On the other hand, for a state of maximum entropy (with a choice of $\mathcal{O}$ where classically the largest group of cells fills almost all of M, or quantum mechanically where the largest invariant subspace of $\mathcal{O}$ fills almost all of H) the value of $\mathcal{O}$ is likely to remain unchanged over a long period of time as the dynamical evolution proceeds. Thus a criterion for apparent stable equilibrium with respect to $\mathcal{O}$ is that the system be in a state of maximum entropy.

Thus in principle we should be able to determine the (apparent) stability of a given configuration of a complicated system (such as a self-gravitating star cluster) by (1) identifying the observable we are measuring (e.g., the coarse-grained distribution function for a star cluster), (2) computing the entropy of the configuration whose stability we are testing [using Eq. (3) in this example], and (3) comparing this value with the value of entropy for other configurations with the same state parameters to determine if the entropy of the original configuration is a maximum. In practice however, this approach is not very reliable mainly for the following reason: Although we have an estimate of the relative amounts of time the system spends in the various macroscopic configurations during its dynamical evolution, we obtain no estimate of how rapidly it evolves from one configuration to the next. Thus if we find a configuration of higher entropy than the given one, the given configuration is indeed unstable, but the time scale required to manifest this instability may be so long as to make the instability totally uninteresting; the given configuration may in practice be perfectly stable. On the other hand, if we fail to find a configuration of higher entropy, this may be due to a poor choice of trial configurations rather than the stability of the given configuration.

The example of a classical Newtonian self-gravitating gas of point particles provides a good illustration of both the potential utility and pitfalls of this approach as well as providing a useful background for the discussion of Sec. VI. The major added difficulty that arises in this case is that, even for a system confined to a finite box, the reduced phase space M is not compact. Normally the finite total energy E bounds the momenta of the individual particles, thus confining M to a bounded region of the full phase space. However, in the Newtonian self-gravitating case, one can have aribtrarily large negative binding energies of groups of particles, so no such bound on the individual momenta of particles exists. This means that a Newtonian self-gravitating gas does not conform to the framework of ideas of Sec. II, since there is no reasonable possibility of ergodic behavior over all of M if M is noncompact. This difficulty by itself need not lead to disastrous consequences for equilibrium. Although all Newtonian self-gravitating systems would eventually evolve to configurations where groups of particles become more and more tightly bound with the gas becoming hotter and hotter (gravothermal instability[11]), the time scales required to achieve such tightly bound systems may be enormous compared with time scales of interest. Therefore we may be justified in excluding such configurations from consideration, in which case the remaining section of M will be compact, and we can expect the ideas of Sec. II to apply with at least some degree of validity. If the Newtonian self-gravitating gas is confined to a sufficiently small box, there will be a local maximum of the entropy (defined with respect to the coarse-grained distribution function) which, if we exclude tightly bound configurations, should become a global maximum and thus correspond physically to a stable equilibrium configuration. However, if the box is large, then, because of the long-range forces and the absence of screening, there is not even a local maximum of the entropy,[11] so physically no true equilibrium exists in this case, and even more caution must be used when employing this approach.

Nevertheless, some analytic calculations illustrate that the above approach for determining stable equilibrium does give correct results in simple cases. If we ask when a uniform distribution of Newtonian gas has maximum entropy compared with a nearby configuration with a sinusoidally perturbed density distribution, we reproduce the usual Jeans' length criterion.[12] If we maximize the entropy of a gas with state parameters E and angular momentum J with respect to scaling the size of the configuration, we obtain the virial theorem.[12] If we further constrain (artificially) the gas to be a uniform density spheroid, we find the maximum entropy is achieved when its eccentricity is that of a Maclaurin spheroid.[12]

These results given encouragement to try the approach in some interesting cases where alternative methods are extremely difficult to apply. One example is to estimate the "final" spectrum of masses of a box of Newtonian gas which is initially Jeans unstable. It is hopelessly difficult to follow the nonlinear dynamics of such a system, but it is an extremely simple matter to compute entropies of final trial configurations. A second example concerns the potential instability of rotating systems to nonaxisymmetric perturbations and the obtaining of estimates of the degree of nonaxisymmetry in the final configuration when such an instability exists. Some simple numerical calculations[13] showed that the presence of bars increases the entropy of configurations with large angular momentum J. However, for the reasons discussed above, the reliability of these results is questionable. To obtain completely reliable results, one would need a knowledge of the nature of the dominant potential instability (so as to pick a good trial configuration) and, if a configuration of larger entropy is found, one would have to know that the time scale for achieving this instability is within the range of interest. Of course to achieve this knowledge, one would in effect have to solve the problem by other means. Nevertheless, the calculation of entropies of configurations is so simple that it has the potential to play an extremely valuable role in preliminary stability analyses of systems.

V. THE LAWS OF THERMODYNAMICS

In this section we shall give some crude plausibility arguments for the validity of the laws of thermodynamics for both classical and quantum systems. To obtain a precise, rigorous proof that thermodynamics applies to a given system certainly requires further detailed assumptions about the system as well as further idealizations such as infinite size of the system. Thus our aim is not to give a general proof of the validity of the laws of thermodynamics but rather to give a simple explanation of why they hold in the cases where they are valid. As we shall see in Sec. VI, these same plausibility arguments plus some further assumptions concerning the nature of a quantum self-gravitating system will yield the laws of black-hole mechanics.

The argument for the validity of the second law of thermodynamics—the nondecrease of total entropy—is basically the one already given in Sec. IV. Namely, since the entropy of a state measures the fraction of time the system spends in a state observably indistinguishable from the given one, the condition for (apparent) stable equilibrium—i.e., no change of the value of $\mathcal{O}$ for a long period of time—is that S be maximum. If S is not maximum, then the system is likely to change its "macroscopic state," i.e., the value of $\mathcal{O}$ will change, in a relatively short period of time. If the system is in a state of low entropy it is extremely unlikely to evolve to a state of still lower entropy, as the system spends even less time in such states. Thus, *with high (and, in practice for most systems, overwhelming) probability, a system in a state of low entropy will evolve to states of higher entropy, while a system in a state of maximum entropy will remain unchanged in appearance for a long period of time.* Of course, this law holds only with high probability, and in fact, by clever arrangements of the system—such as in the "spin echo" effect—we can produce brief violations of it.

The remaining laws of thermodynamics apply to the situation where the total system can be divided into subsystems (as described in Secs. II and III) such that each subsystem is in "thermal equilibrium," i.e., in a state of maximum entropy for the values of its state parameters. Physically it is reasonable to expect this to be the case if each subsystem self-interacts sufficiently rapidly that it has time to evolve to a state of maximum entropy, but the subsystems interact sufficiently weakly with each other that the total system need not be in a state of maximum total entropy.

We define the temperature T_i of the ith subsystems by

$$\frac{1}{T_i}=\left(\frac{\partial \mathcal{S}_i}{\partial E_i}\right)_{\alpha_i}, \tag{21}$$

where $\mathcal{S}_i$ is the thermodynamic entropy given by Eq. (8) [or Eq. (20)]. By Eq. (6) [or Eq. (18)] plus the assumption that each subsystem is in a state of maximum entropy for its state parameters E_i, α_i, the total entropy is given by

$$S=\sum_{i=1}^{n} \mathcal{S}_i \,. \tag{22}$$

As discussed above, for the total system to be in thermodynamic equilibrium, S must be maximum. In particular, it must be maximum with respect to arbitrary interchanges of energy among the subsystems. Using Eq. (21), we thus obtain the 0th law of thermodynamics: *A necessary condition for the thermodynamic equilibrium of the total system is that the temperatures T_i of the subsystems all be equal.*

The first law of thermodynamics (or, really, the combined first and second laws) is entirely trivial in this formulation. The thermodynamic entropy $\mathcal{S}_i$ is a function of only the state parameters E_i and α_i. Assuming it varies smoothly with these parameters, we have

$$d\mathcal{S}_i=\left(\frac{\partial \mathcal{S}_i}{\partial E_i}\right)_{\alpha_i} dE_i+\left(\frac{\partial \mathcal{S}_i}{\partial \alpha_i}\right)_{E_i} d\alpha_i\,, \tag{23}$$

where, if α_i represents more than one parameter, the second term on the right-hand side of Eq. (23) is understood as a sum over these parameters. Using the definition of T_i, we have

$$dE_i=T_i d\mathcal{S}_i-T_i\left(\frac{\partial \mathcal{S}_i}{\partial \alpha_i}\right)_{E_i} d\alpha_i\,, \tag{24}$$

which is the first law.

Finally, we give a particularly crude plausibility argument for the third law of thermodynamics in the form expressing the impossibility of achieving absolute zero temperature. This law is considerably less rigorous than the other laws and really expresses the difficulty, in practice, of achieving and maintaining $T_i=0$. By Eq. (21) at $T_i=0$ the entropy $\mathcal{S}_i$ has infinite derivative with respect to E_i (with the other state parameters held fixed). Thus the total entropy of a system which has a subsystem at $T_i=0$ will enormously lower than the total entropy would be if a tiny bit of energy were transferred to the ith subsystem. Hence unless perfect energy isolation can be achieved, the subsystem should quickly evolve away from absolute zero temperature.

An alternate version of the third law which is sometimes given for quantum systems is that the entropy $\mathcal{S}_i$ at absolute zero temperature vanishes.

The rationale behind this version is that, for most quantum systems, the only regime where $\mathcal{S}_i$ has infinite derivative with respect to E_i—i.e., the only regime where the density of energy levels changes enormously with energy—is at the ground state. Thus for most systems being at absolute zero temperature is equivalent to being in the ground state. But for most systems the density of energy levels at the ground state is very small (at least compared with other energies), i.e., $\dim H_i \sim 1$ at the ground state. Thus $\mathcal{S}_i \approx 0$ at $T_i = 0$. However, there is no general principle of quantum mechanics which requires that the ground state be the only state at which $T_i = 0$. Indeed, if (as we shall argue in the next section) black holes fit into the general framework described here, they would provide an example of a system which can have $T_i = 0$ at high entropy and thus would violate this formulation of the third law (though not the first formulation above).

VI. THE THERMODYNAMICS OF BLACK HOLES

In this section we shall apply the ideas of the previous sections to a quantum self-gravitating system containing a black hole. This is a relatively bold application in view of the fact that a satisfactory quantum theory of gravity does not yet exist. Furthermore, the Schrödinger picture framework for the quantum-mechanical notion of entropy given in Sec. III is not very natural in the space-time viewpoint of general relativity. In particular, it is not obvious how to properly reformulate concepts such as "the fraction of time the system spends in a given state," etc. However, we shall ignore these difficult issues in the discussion below.

Consider an isolated, self-gravitating system confined to a box. As already discussed in Sec. IV, a classical Newtonian self-gravitating gas of point particles is *not* really a system for which ordinary thermodynamics is applicable because the reduced phase space M is noncompact, so ergodic behavior on M cannot be expected, and there is no equilibrium state of maximum entropy. As previously discussed, the cause of this problem is the lack of a bound on the magnitude of the binding energy of groups of particles. However, this difficulty is resolved when one passes to a (classical) general relativistic description of the system. According to the positive energy conjecture,[14] the total energy of a bound system in general relativity can never become negative, i.e., roughly speaking, the binding energy of a group of particles can never exceed their rest mass plus kinetic energy. If one attempts to bind particles so tightly that, in the Newtonian description, their binding energy would exceed this limit, one finds in general relativity that gravitational collapse to a black hole will occur.

However, this cure of general relativity brings with it an equally bad disease. Assuming the validity of the cosmic censor hypothesis, in classical general relativity black holes are completely stable; they can form and grow in size but they can never bifurcate or disappear.[15] Thus for a general relativistic self-gravitating system one does not get ergodic behavior either. Given enough time, a closed self-gravitating system will form a black hole; given enough time, this black hole will swallow everything else up and then remain in this state forever.

Remarkably, this second problem appears to be resolved by quantum theory. When quantum effects are taken into account, black holes no longer are stable objects; they can decrease in size and can "evaporate" within a finite time.[3] Thus, there is no obvious contradiction to ergodic behavior of a closed, self-gravitating quantum system. *It is perfectly plausible that a quantum self-gravitating system behaves in essence like an ordinary quantum-mechanical system as described in Sec. III.* Note that it is not necessary for the validity of this idea that—as Hawking[16] has proposed—the process of black-hole evaporation be the time reverse of the process of black-hole formation. As Penrose[17] has argued, this latter idea appears implausible.

To proceed further, we make the assumption that the states of the self-gravitating quantum system can be decomposed (at least approximately) into black-hole states and "other" states, so that as in Sec. III the total Hilbert space $\mathcal{H}$ can be expressed as

$$\mathcal{H} = \mathcal{H}_{\mathrm{bh}} \otimes \mathcal{H}_{\mathrm{oth}}, \tag{25}$$

where $\mathcal{H}_{\mathrm{bh}}$ is the space of black-hole states. The Hilbert space $\mathcal{H}_{\mathrm{oth}}$ describes the states of "ordinary matter outside black holes" as well as possibly other more exotic states (e.g., perhaps white-hole states). We further assume that the interaction between the black hole (if present) and other states is sufficiently weak that we can treat each as a subsystem in the sense of Sec. III above. This is not unlike decomposing the states of water molecules in a box into "liquid water states" and "water vapor states" and treating each phase as a thermodynamic subsystem.

Next we assume that the correct state parameters characterizing the black-hole subsystem are simply its energy E (i.e., its mass) and its angular momentum J. This is strongly suggested by the black-hole uniqueness theorems[18] of classical general relativity which state that stationary vacuum black holes are uniquely characterized by E and J. (For simplicity, we ignore the possibility that the black hole may be charged; otherwise the total charge Q should also be included as a state parameter.)

To define the entropy of a black hole, we need to introduce an observable $\mathcal{O}$. As emphasized in Secs. II and III above, the notion of entropy (as opposed to thermodynamic entropy) depends critically on the choice of coarse-grained observable. However, for a black-hole subsystem, there is a completely natural choice of coarse-grained observable $\mathcal{O}$, namely, the properties of the "instantaneous" black-hole state which can be simultaneously measured by observers *outside* the event horizon of the black hole. For a subsystem without a horizon, such a complete set of observables would not be coarse grained, i.e., it would distinguish between distinct states and thus lead to an assignment of zero entropy to each state. However, because of the possibly large number of internal configurations of the black hole which produce the same external configurations, this observable $\mathcal{O}$ here should be coarse grained, i.e., in the quantum-mechanical case, have large degeneracy subspaces. It is worth emphasizing that the notion of entropy thus obtained for black holes attains a more fundamental and absolute status than elsewhere. For other systems we measure only coarse-grained observables because our measurement ability is limited by practical considerations; in principle we could determine the state of the system precisely. However, for black holes we cannot in principle measure the internal configurations unless we go inside the black hole, in which case we still would not be able to report our results to observers remaining outside the black hole. Thus for black holes the limit to our measurement abilities implied by $\mathcal{O}$ is a more fundamental one.

Finally, we postulate the formula for the entropy of a black hole defined with respect to our observable $\mathcal{O}$

$$S=\tfrac{1}{4}A \tag{26}$$

where A is the area of the event horizon (expressed in Planck units $G=c=\hbar=1$). If Eq. (26) is correct, it remains a mystery—presumably to be resolved only when a full quantum theory of gravitation is available—why such a remarkably simple formula should hold.

The thermodynamic entropy $\mathcal{S}$ of a black hole is the maximum of S over all configurations with fixed state parameters E and J,

$$\mathcal{S}=\tfrac{1}{4}\max A(E,J)\equiv\tfrac{1}{4}\,\mathcal{A}(E,J)\,, \tag{27}$$

where $\mathcal{A}(E,J)$ is the maximum possible area. The uniqueness theorems[18] for stationary black holes together with the area-increase theorem[2] strongly indicate that the maximum possible value of A is achieved by the Kerr black hole, with

$$\mathcal{A}(E,J)=8\pi[E^2+(E^4-J^2)^{1/2}]\,. \tag{28}$$

Assuming that the states of maximum entropy fill up most of the reduced Hilbert space H_{bh}, Eq. (27) is equivalent to the formula for the dimension N of H_{bh}, i.e., the number of black-hole states with energy and angular momentum within ΔE and ΔJ of given values,

$$N\propto\exp[\tfrac{1}{4}\mathcal{A}(E,J)]\,. \tag{29}$$

This formula previously has been postulated by Hawking[16] and (with a slightly different interpretation) by Bekenstein.[19]

Thus our picture of a closed self-gravitating quantum system with a black hole is the following: The total system behaves essentially like an ordinary quantum system and displays ergodic motion (in the sense of Sec. III) through its reduced Hilbert space. The black-hole states and the remaining states can be treated as thermodyanmic subsystems. The state parameters of the black-hole subsystem are E and J. Furthermore, the formula (26) holds for the entropy defined with respect to the observable measuring all properties of the black hole which can in principle, be determined by observers outside the black hole. While all these assumptions are certainly nontrivial, none of them can be considered in any sense "exotic."

Our aim now is to argue that under the above assumptions, the ordinary laws or thermodynamics yield black-hole thermodynamics when applied to this system. We choose as the observable for the total system the properties of the black-hole subsystem which can be measured outside the black hole together with a suitable coarse-grained observable for the non-black-hole subsystem. By the general argument of Sec. III, the total entropy will be additive

$$S_{\text{tot}}=S_{\text{bh}}+S_{\text{oth}}\,. \tag{30}$$

Using Eq. (26) for S_{bh}, we see that the ordinary second law of thermodynamics yields in this case the *generalized second law*: The sum of "ordinary entropy of matter outside black holes" plus black hole area never decreases. If S_{oth} is negligible, the generalized second law reduces to the area theorem.[2]

The remaining laws of thermodynamics apply to the case when the two subsystems (i.e., the black-hole states and the other states) are each in a state of maximum entropy, i.e., each is in thermodynamic equilibrium. A black-hole equilibrium configuration is uniquely characterized by its state parameters E and J. It temperature is

$$T_{\mathrm{bh}}^{-1} = \left(\frac{\partial \mathcal{S}_{\mathrm{bh}}}{\partial E}\right)_J = 4\pi[E + E^3(E^4 - J^2)^{-1/2}] \tag{31}$$

in agreement with the Hawking formula.[3] The 0th law of thermodynamics, applied in this case, states that a necessary condition for equilibrium between the black hole and other states is that $T_{\mathrm{bh}} = T_{\mathrm{oth}}$. That this equilibrium can actually be achieved is of course essentially the Hawking particle-creation result.[3,4]

The first law of thermodynamics, Eq. (24), applied to the black-hole subsystem, yields

$$dE = T_{\mathrm{bh}} dS_{\mathrm{bh}} + \Omega_{\mathrm{bh}} dJ, \tag{32}$$

where

$$\Omega_{\mathrm{bh}} = -T_{\mathrm{bh}}\left(\frac{\partial \mathcal{S}_{\mathrm{bh}}}{\partial J}\right)_E = \frac{J}{2E[E + (E^4 - J^2)^{1/2}]}. \tag{33}$$

Equation (32) is just the first law of black-hole mechanics.[1]

Finally, the first formulation of the third law of thermodynamics given in Sec. V, now yields the statement that it should be very difficult in practice to achieve $T_{\mathrm{bh}} = 0$. This is of course just the third law of black-hole mechanics.[1] The alternate formulation of the third law does not apply to black holes, presumably for the reason given at the end of Sec. V.

Thus it appears perfectly plausible that the connection between black holes and thermodynamics is extremely simple and direct: Black-hole thermodynamics is just ordinary thermodynamics applied to a quantum self-gravitating system. Many further aspects of this connection remain to be explored, however, such as the derivation of Eq. (26) and the relation of the thermal particle spectrum seen by an accelerated observer in flat spacetime to black-hole thermodynamics.[21]

ACKNOWLEDGMENTS

I wish to thank Rafael Sorkin for many valuable discussions and arguments over the notion of entropy. This work was supported in part by NSF Grant PHY 78-24275 to the University of Chicago and by the Alfred P. Sloan Foundation.

---

[1] J. M. Bardeen, B. Carter, and S. W. Hawking, Commun. Math. Phys. 31, 161 (1973).

[2] S. W. Hawking, Phys. Rev. Lett. 26, 1344 (1971).

[3] S. W. Hawking, Commun. Math. Phys. 43, 199 (1975).

[4] R. M. Wald, Commun. Math. Phys. 45, 9 (1975); S. W. Hawking, Phys. Rev. D 14, 2460 (1976); L. Parker, *ibid.* 12, 1519 (1975); P. Panangaden and R. M. Wald, *ibid.* 16, 929 (1977).

[5] J. D. Bekenstein, Phys. Rev. D 7, 2333 (1973).

[6] The cells in $\mathfrak{M}$ will be identical in size; the cells induced in M by intersection with these cells will be only approximately uniform in size.

[7] A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics* (Dover, New York, 1949).

[8] See, e.g., K. Huang, *Statistical Mechanics* (Wiley, New York, 1963).

[9] See, e.g., A. Wehrl, Rev. Mod. Phys. 50, 221 (1978).

[10] The argument for additivity of entropy in the classical case could be made more manifestly analogous to that given here for the quantum case by slightly "thickening" the reduced phase spaces.

[11] D. Lynden-Bell and R. Wood, Mon. Not. R. Astron. Soc. 138, 495 (1968).

[12] R. M. Wald, Maryland Report No. 74-031 (unpublished).

[13] R. M. Wald (unpublished).

[14] For the most recent results on the positive-energy conjecture, see R. Schoen and S. T. Yau, Proc. Nat. Acad. Sci. U. S. A. 75, 2567 (1978).

[15] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge University Press, Cambridge, England, 1973).

[16] S. W. Hawking, Phys. Rev. D 13, 191 (1976).

[17] R. Penrose, in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1979).

[18] See, e.g., the contribution by B. Carter in Ref. 17.

[19] J. D. Bekenstein, Phys. Rev. D 12, 3077 (1975).

[20] R. C. Tolman, *Principles of Statistical Mechanics* (Clarendon, Oxford, 1938), Chap. XII.

[21] W. G. Unruh, Phys. Rev. D 14, 870 (1976); D. Sciama (unpublished).