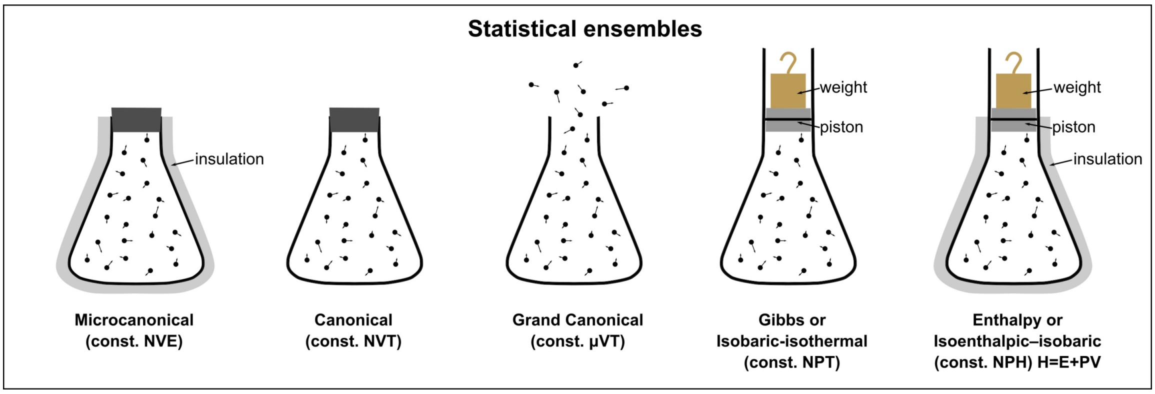
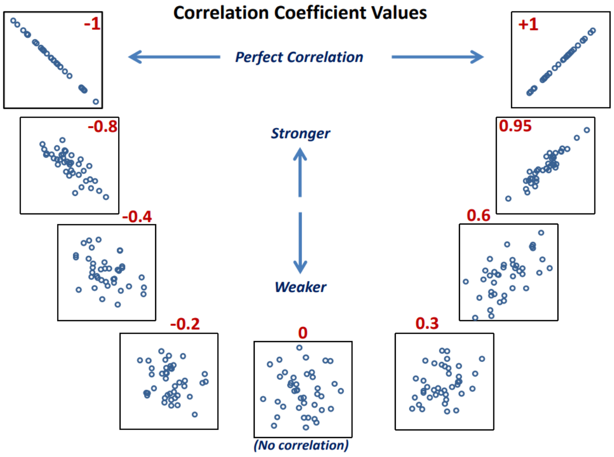
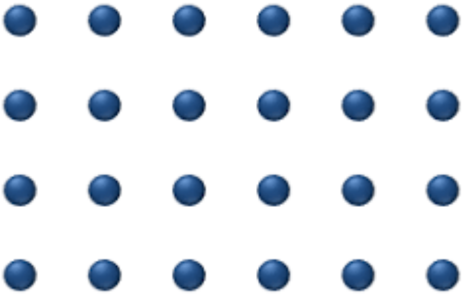
Statistical Mechanics



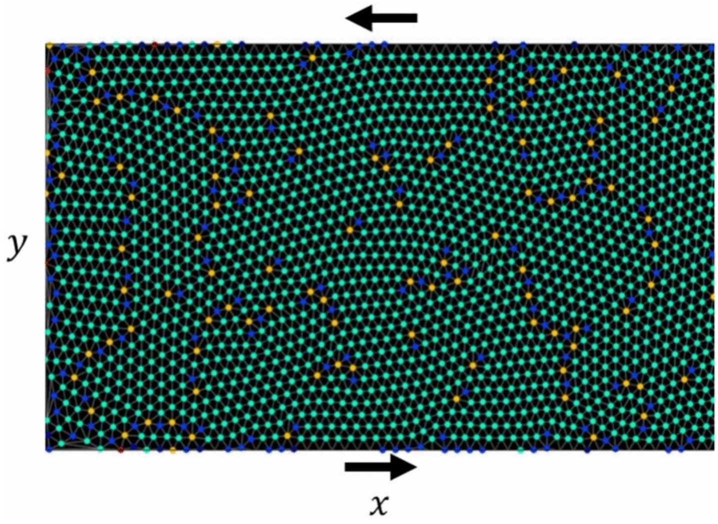
Correlations



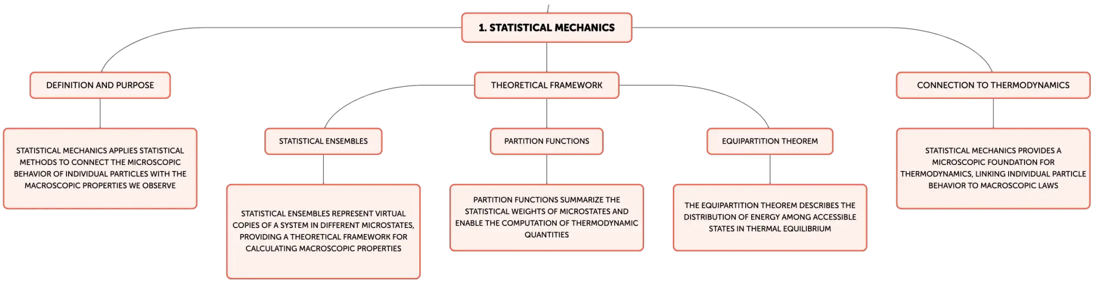
Fluctuations



Strongly interacting particles



Microscopic origin of thermodynamic laws



1. STATISTICAL MECHANICS

DEFINITION AND PURPOSE

STATISTICAL MECHANICS APPLIES STATISTICAL METHODS TO CONNECT THE MICROSCOPIC BEHAVIOR OF INDIVIDUAL PARTICLES WITH THE MACROSCOPIC PROPERTIES WE OBSERVE

THEORETICAL FRAMEWORK

STATISTICAL ENSEMBLES

STATISTICAL ENSEMBLES REPRESENT VIRTUAL COPIES OF A SYSTEM IN DIFFERENT MICROSTATES, PROVIDING A THEORETICAL FRAMEWORK FOR CALCULATING MACROSCOPIC PROPERTIES

PARTITION FUNCTIONS

PARTITION FUNCTIONS SUMMARIZE THE STATISTICAL WEIGHTS OF MICROSTATES AND ENABLE THE COMPUTATION OF THERMODYNAMIC QUANTITIES

EQUIPARTITION THEOREM

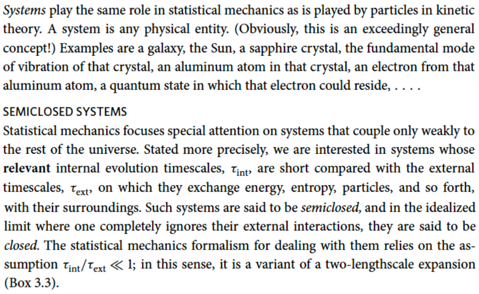
THE EQUIPARTITION THEOREM DESCRIBES THE DISTRIBUTION OF ENERGY AMONG ACCESSIBLE STATES IN THERMAL EQUILIBRIUM

CONNECTION TO THERMODYNAMICS

STATISTICAL MECHANICS PROVIDES A MICROSCOPIC FOUNDATION FOR THERMODYNAMICS, LINKING INDIVIDUAL PARTICLE BEHAVIOR TO MACROSCOPIC LAWS

The founding fathers

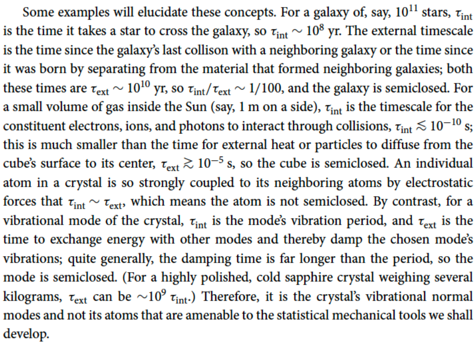
Systems: closed and semi-closed



Systems play the same role in statistical mechanics as is played by particles in kinetic theory. A system is any physical entity. (Obviously, this is an exceedingly general concept!) Examples are a galaxy, the Sun, a sapphire crystal, the fundamental mode of vibration of that crystal, an aluminum atom in that crystal, an electron from that aluminum atom, a quantum state in which that electron could reside, ...

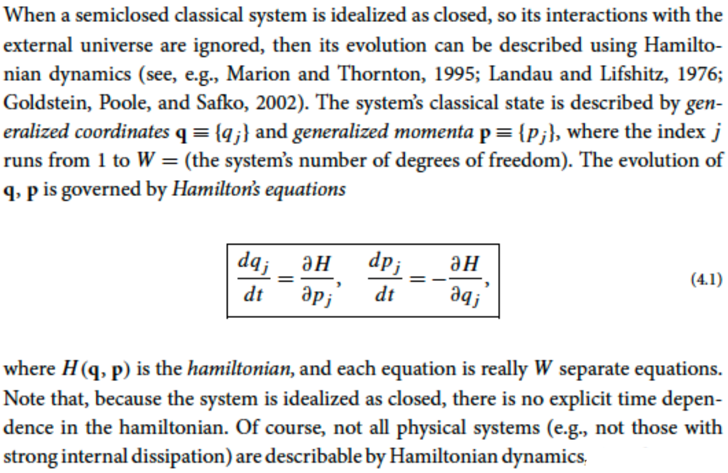
SEMICLOSED SYSTEMS

Statistical mechanics focuses special attention on systems that couple only weakly to the rest of the universe. Stated more precisely, we are interested in systems whose **relevant** internal evolution timescales, τ_{int}, are short compared with the external timescales, τ_{ext}, on which they exchange energy, entropy, particles, and so forth, with their surroundings. Such systems are said to be *semiclosed*, and in the idealized limit where one completely ignores their external interactions, they are said to be *closed*. The statistical mechanics formalism for dealing with them relies on the assumption $\tau_{\text{int}}/\tau_{\text{ext}} \ll 1$; in this sense, it is a variant of a two-lengthscale expansion (Box 3.3).



Some examples will elucidate these concepts. For a galaxy of, say, 10^{11} stars, τ_{int} is the time it takes a star to cross the galaxy, so $\tau_{\text{int}} \sim 10^8$ yr. The external timescale is the time since the galaxy's last collison with a neighboring galaxy or the time since it was born by separating from the material that formed neighboring galaxies; both these times are $\tau_{\text{ext}} \sim 10^{10}$ yr, so $\tau_{\text{int}}/\tau_{\text{ext}} \sim 1/100$, and the galaxy is semiclosed. For a small volume of gas inside the Sun (say, 1 m on a side), τ_{int} is the timescale for the constituent electrons, ions, and photons to interact through collisions, $\tau_{\text{int}} \lesssim 10^{-10}$ s; this is much smaller than the time for external heat or particles to diffuse from the cube's surface to its center, $\tau_{\text{ext}} \gtrsim 10^{-5}$ s, so the cube is semiclosed. An individual atom in a crystal is so strongly coupled to its neighboring atoms by electrostatic forces that $\tau_{\text{int}} \sim \tau_{\text{ext}}$, which means the atom is not semiclosed. By contrast, for a vibrational mode of the crystal, τ_{int} is the mode's vibration period, and τ_{ext} is the time to exchange energy with other modes and thereby damp the chosen mode's vibrations; quite generally, the damping time is far longer than the period, so the mode is semiclosed. (For a highly polished, cold sapphire crystal weighing several kilograms, τ_{ext} can be $\sim 10^9\, \tau_{\text{int}}$.) Therefore, it is the crystal's vibrational normal modes and not its atoms that are amenable to the statistical mechanical tools we shall develop.

Hamiltonian dynamics for closed systems



When a semiclosed classical system is idealized as closed, so its interactions with the external universe are ignored, then its evolution can be described using Hamiltonian dynamics (see, e.g., Marion and Thornton, 1995; Landau and Lifshitz, 1976; Goldstein, Poole, and Safko, 2002). The system's classical state is described by *generalized coordinates* $\mathbf{q} \equiv \{q_j\}$ and *generalized momenta* $\mathbf{p} \equiv \{p_j\}$, where the index j runs from 1 to $W =$ (the system's number of degrees of freedom). The evolution of $\mathbf{q}, \mathbf{p}$ is governed by *Hamilton's equations*

$$\boxed{\frac{dq_j}{dt} = \frac{\partial H}{\partial p_j}, \quad \frac{dp_j}{dt} = -\frac{\partial H}{\partial q_j},} \tag{4.1}$$

where $H(\mathbf{q}, \mathbf{p})$ is the *hamiltonian*, and each equation is really W separate equations. Note that, because the system is idealized as closed, there is no explicit time dependence in the hamiltonian. Of course, not all physical systems (e.g., not those with strong internal dissipation) are describable by Hamiltonian dynamics.

Ensembles

In kinetic theory, we study statistically a collection of a huge number of particles. Similarly, in statistical mechanics, we study statistically a collection or *ensemble* of a huge number of systems. This ensemble is actually only a conceptual device, a foundation for statistical arguments that take the form of thought experiments. As we shall see, there are many different ways that one can imagine forming an ensemble, and this freedom can be used to solve many different types of problems.

In some applications, we require that all the systems in the ensemble be closed and be identical in the sense that they all have the same number of degrees of freedom, W; are governed by hamiltonians with the same functional forms $H(\mathbf{q}, \mathbf{p})$; and have the same volume V and total internal energy E (or $\mathcal{E}$, including rest masses). However, the values of the generalized coordinates and momenta at a specific time t, $\{\mathbf{q}(t), \mathbf{p}(t)\}$, need not be the same (i.e., the systems need not be in the same state at time t). If such a conceptual ensemble of identical closed systems (first studied by Boltzmann) evolves until it reaches statistical equilibrium (Sec. 4.5), it then is called *microcanonical;* see Table 4.1.

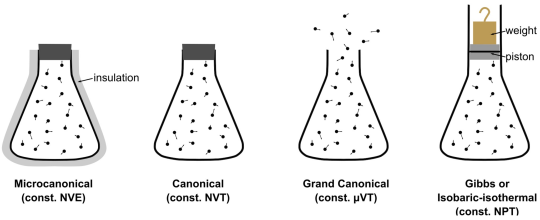


TABLE 4.1: Statistical-equilibrium ensembles used in this chapter

Ensemble	Quantities exchanged with surroundings
Microcanonical	Nothing
Canonical	Energy E
Gibbs	Energy E and volume V
Grand canonical	Energy E and number of particles N_I of various species I

Preliminaries

η = mean number of systems that occupy a quantum state at location {q, p} in an ensemble's 2W-dimensional phase space, at time t

1. This generalized η is proportional to the number of systems $N_{\rm sys}$ in our ensemble. (If we double $N_{\rm sys}$, then η will double.) Because our ensemble is only a conceptual device, we don't really care how many systems it contains, so we divide η by $N_{\rm sys}$ to get a renormalized, $N_{\rm sys}$-independent distribution function, $\rho = \eta/N_{\rm sys}$, whose physical interpretation is

$$\rho(t; \mathbf{q}, \mathbf{p}) = \left(\begin{array}{l}\text{probability that a system, drawn randomly}\\ \text{from our ensemble, will be in a quantum state}\\ \text{at location } (\mathbf{q}, \mathbf{p}) \text{ in phase space at time } t\end{array}\right). \qquad (4.4)$$

2. If the systems of our ensemble can exchange particles with the external universe (as is the case, for example, in the grand canonical ensemble of Table 4.1), then their number of degrees of freedom, W, can change, so ρ may depend on W as well as on location in the $2W$-dimensional phase space: $\rho(t; W, \mathbf{q}, \mathbf{p})$.

Distribution function

In the sector of the system's phase space with W degrees of freedom, denote the number density of quantum states by

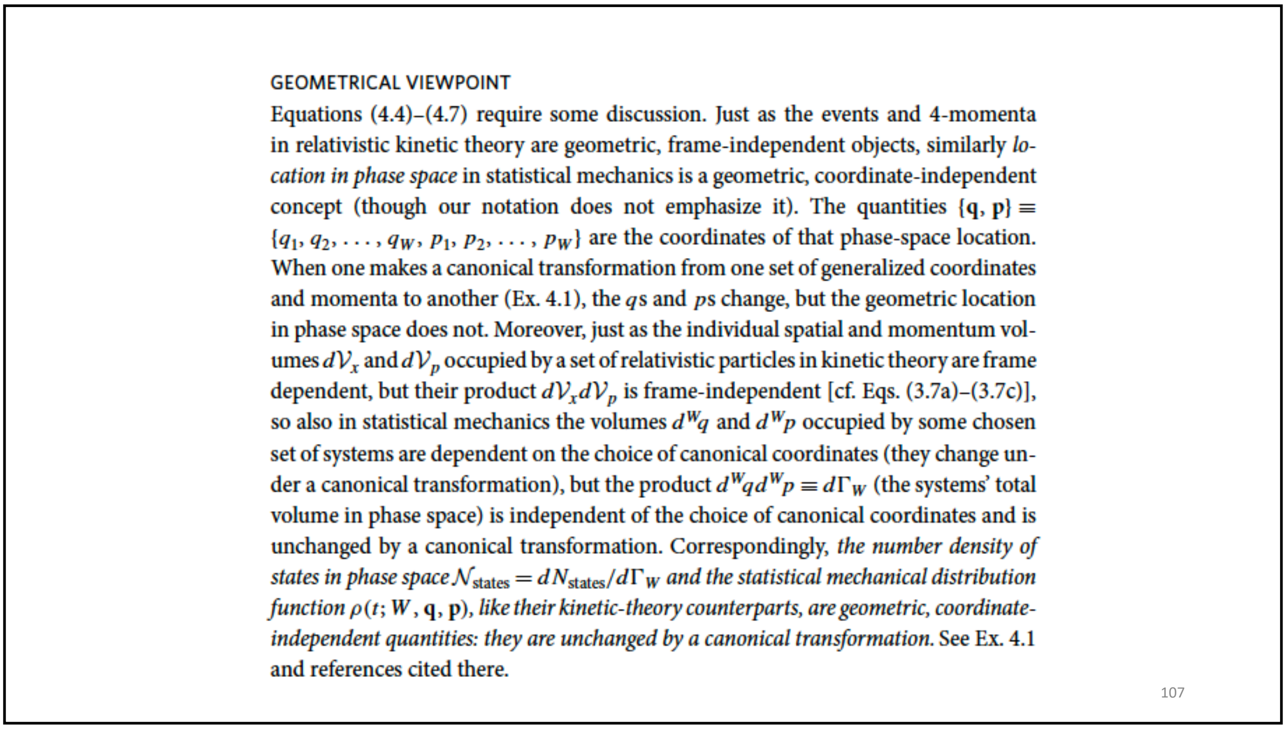
$$\mathcal{N}_{\rm states}(W, \mathbf{q}, \mathbf{p}) = \frac{dN_{\rm states}}{d^W q\, d^W p} \equiv \frac{dN_{\rm states}}{d\Gamma_W}. \qquad (4.5)$$

Here we have used

$$d^W q \equiv dq_1 dq_2 \cdots dq_W, \quad d^W p \equiv dp_1 dp_2 \cdots dp_W, \quad d\Gamma_W \equiv d^W q\, d^W p. \qquad (4.6)$$

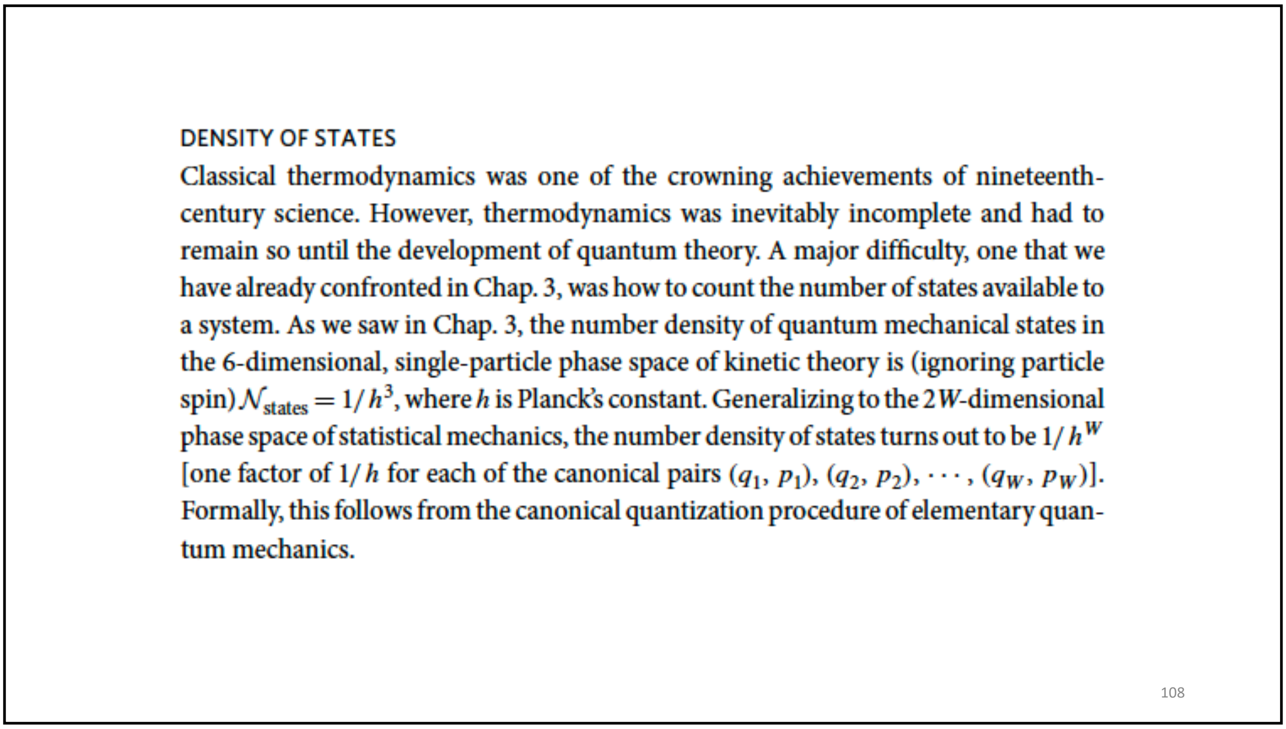
Then the sum of the occupation probability ρ over all quantum states, which must (by the meaning of probability) be unity, takes the form

$$\sum_n \rho_n = \sum_W \int \rho \mathcal{N}_{\rm states} d\Gamma_W = 1. \qquad (4.7)$$



GEOMETRICAL VIEWPOINT

Equations (4.4)–(4.7) require some discussion. Just as the events and 4-momenta in relativistic kinetic theory are geometric, frame-independent objects, similarly *location in phase space* in statistical mechanics is a geometric, coordinate-independent concept (though our notation does not emphasize it). The quantities $\{\mathbf{q}, \mathbf{p}\} \equiv \{q_1, q_2, \ldots, q_W, p_1, p_2, \ldots, p_W\}$ are the coordinates of that phase-space location. When one makes a canonical transformation from one set of generalized coordinates and momenta to another (Ex. 4.1), the qs and ps change, but the geometric location in phase space does not. Moreover, just as the individual spatial and momentum volumes $d\mathcal{V}_x$ and $d\mathcal{V}_p$ occupied by a set of relativistic particles in kinetic theory are frame dependent, but their product $d\mathcal{V}_x d\mathcal{V}_p$ is frame-independent [cf. Eqs. (3.7a)–(3.7c)], so also in statistical mechanics the volumes $d^W q$ and $d^W p$ occupied by some chosen set of systems are dependent on the choice of canonical coordinates (they change under a canonical transformation), but the product $d^W q\, d^W p \equiv d\Gamma_W$ (the systems' total volume in phase space) is independent of the choice of canonical coordinates and is unchanged by a canonical transformation. Correspondingly, *the number density of states in phase space* $\mathcal{N}_{\text{states}} = dN_{\text{states}}/d\Gamma_W$ *and the statistical mechanical distribution function* $\rho(t; W, \mathbf{q}, \mathbf{p})$*, like their kinetic-theory counterparts, are geometric, coordinate-independent quantities: they are unchanged by a canonical transformation.* See Ex. 4.1 and references cited there.



DENSITY OF STATES

Classical thermodynamics was one of the crowning achievements of nineteenth-century science. However, thermodynamics was inevitably incomplete and had to remain so until the development of quantum theory. A major difficulty, one that we have already confronted in Chap. 3, was how to count the number of states available to a system. As we saw in Chap. 3, the number density of quantum mechanical states in the 6-dimensional, single-particle phase space of kinetic theory is (ignoring particle spin) $\mathcal{N}_{\text{states}} = 1/h^3$, where h is Planck's constant. Generalizing to the $2W$-dimensional phase space of statistical mechanics, the number density of states turns out to be $1/h^W$ [one factor of $1/h$ for each of the canonical pairs $(q_1, p_1), (q_2, p_2), \cdots, (q_W, p_W)$]. Formally, this follows from the canonical quantization procedure of elementary quantum mechanics.

When the laws of quantum mechanics were developed, it became clear that all identical particles are indistinguishable, so having particle 1 at location $\mathcal{A}$ in phase space and an identical particle 2 at location $\mathcal{B}$ must be counted as the same state as particle 1 at $\mathcal{B}$ and particle 2 at $\mathcal{A}$. Correspondingly, if we attribute half the quantum state to the classical phase-space location {1 at $\mathcal{A}$, 2 at $\mathcal{B}$} and the other half to {1 at $\mathcal{B}$, 2 at $\mathcal{A}$}, then the classical number density of states per unit volume of phase space must be reduced by a factor of 2—and more generally by some *multiplicity factor* $\mathcal{M}$. In general, therefore, we can write the actual number density of states in phase space as

$$\boxed{\mathcal{N}_{\text{states}} = \frac{dN_{\text{states}}}{d\Gamma_W} = \frac{1}{\mathcal{M}h^W},} \qquad (4.8a)$$

and correspondingly, we can rewrite the normalization condition (4.7) for our probabilistic distribution function as

$$\boxed{\sum_n \rho_n \equiv \sum_W \int \rho \mathcal{N}_{\text{states}} d\Gamma_W = \sum_W \int \rho \frac{d\Gamma_W}{\mathcal{M}h^W} = 1.} \qquad (4.8b)$$

This equation can be regarded, in the classical domain, as defining the meaning of the sum over states n. We shall make extensive use of such sums over states.

For N identical and indistinguishable particles with zero spin, it is not hard to see that $\mathcal{M} = N!$. If we include the effects of quantum mechanical spin (and the spin states can be regarded as degenerate), then there are g_s [Eq. (3.16)] more states present in the phase space of each particle than we thought, so an individual state's multiplicity $\mathcal{M}$ (the number of different phase-space locations to be attributed to the state) is reduced to

$$\boxed{\mathcal{M} = \frac{N!}{{g_s}^N} \quad \text{for a system of N identical particles with spin } s.} \qquad (4.8c)$$

This is the quantity that appears in the denominator of the sum over states [Eq. (4.8b)].

ENSEMBLE AVERAGE

Each system in an ensemble is endowed with a total energy that is equal to its hamiltonian, $E = H(\mathbf{q}, \mathbf{p})$ [or relativistically, $\mathcal{E} = H(\mathbf{q}, \mathbf{p})$]. Because different systems reside at different locations $(\mathbf{q}, \mathbf{p})$ in phase space, they typically will have different energies. A quantity of much interest is the *ensemble-averaged energy*, which is the average value of E over all systems in the ensemble:

$$\langle E \rangle = \sum_n \rho_n E_n = \sum_W \int \rho\, E \mathcal{N}_{\text{states}} d\Gamma_W = \sum_W \int \rho\, E \frac{d\Gamma_W}{\mathcal{M}h^W}. \qquad (4.10a)$$

For any other function $A(\mathbf{q}, \mathbf{p})$ defined on the phase space of a system (e.g., the linear momentum or the angular momentum), one can compute an ensemble average by the obvious analog of Eq. (4.10a):

$$\boxed{\langle A \rangle = \sum_n \rho_n A_n.} \qquad (4.10b)$$

Our probabilistic distribution function $\rho_n = \rho(t; W, \mathbf{q}, \mathbf{p})$ has deeper connections to quantum theory than the above discussion reveals. In the quantum domain, even if we start with a system whose wave function ψ is in a *pure* state (ordinary, everyday type of quantum state), the system may evolve into a *mixed state* as a result of (i) interaction with the rest of the universe and (ii) our choice not to keep track of correlations between the universe and the system (Box 4.2 and Sec. 4.7.2). The system's initial, pure state can be described in geometric, basis-independent quantum language by a state vector ("ket") $|\psi\rangle$; but its final, mixed state requires a different kind of quantum description: a *density operator* $\hat{\rho}$. In the classical limit, the quantum mechanical density operator $\hat{\rho}$ becomes our classical probabilistic distribution function $\rho(t, W, \mathbf{q}, \mathbf{p})$; see Box 4.2 for some details.

Conservation law for systems & Liouville equation

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial q_j}\left(\rho \frac{dq_j}{dt}\right) + \frac{\partial}{\partial p_j}\left(\rho \frac{dp_j}{dt}\right) = 0. \qquad (4.13)$$

Equation (4.13) has an implicit sum, from 1 to W, over the repeated index j (recall the Einstein summation convention, Sec. 1.5). Using Hamilton's equations, we can rewrite this as

$$\begin{aligned} 0 &= \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial q_j}\left(\rho \frac{\partial H}{\partial p_j}\right) - \frac{\partial}{\partial p_j}\left(\rho \frac{\partial H}{\partial q_j}\right) \\ &= \frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial q_j}\frac{\partial H}{\partial p_j} - \frac{\partial \rho}{\partial p_j}\frac{\partial H}{\partial q_j} = \frac{\partial \rho}{\partial t} + [\rho, H]_{\mathbf{q},\mathbf{p}}, \end{aligned} \qquad (4.14)$$

where $[\rho, H]_{\mathbf{q},\mathbf{p}}$ is the *Poisson bracket* (e.g., Landau and Lifshitz, 1976; Marion and Thornton, 1995; Goldstein, Poole, and Safko, 2002). By using Hamilton's equations once again in the second expression, we discover that this is the time derivative of ρ moving with a fiducial system through the $2W$-dimensional phase space:

$$\boxed{\left(\frac{d\rho}{dt}\right)_{\substack{\text{moving with a} \\ \text{fiducial system}}} \equiv \frac{\partial \rho}{\partial t} + \frac{dq_j}{dt}\frac{\partial \rho}{\partial q_j} + \frac{dp_j}{dt}\frac{\partial \rho}{\partial p_j} = \frac{\partial \rho}{\partial t} + [\rho, H]_{\mathbf{q},\mathbf{p}} = 0.}$$

Therefore, the probability ρ is constant along the system's phase space trajectory, as was to be proved.

We call Eq. (4.15), which embodies this Liouville theorem, the statistical mechanical *Liouville equation* or *collisionless Boltzmann equation*.

Liouville's theorem

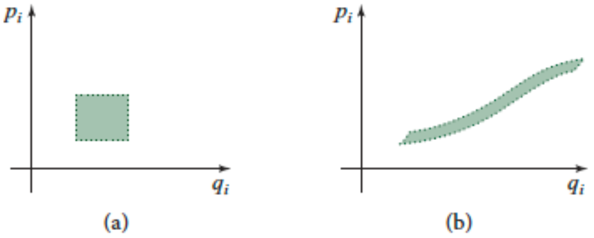


FIGURE 4.1 Liouville's theorem. (a) The region in the q_i-p_i part of phase space (with i fixed) occupied by a set $\mathcal{S}$ of identical, closed systems at time $t = 0$. (b) The region occupied by the same set of systems a short time later, $t > 0$. The hamiltonian-generated evolution of the individual systems has moved them in such a manner as to skew the region they occupy, but the volume $\int dp_i dq_i$ is unchanged.

Boltzmann transport equation

What happens if the systems being studied interact weakly with their surroundings? We must then include an interaction term on the right-hand side of Eq. (4.15), thereby converting it into the statistical mechanical version of the Boltzmann transport equation:

$$\boxed{\left(\frac{d\rho}{dt}\right)_{\substack{\text{moving with a}\\ \text{fiducial system}}} = \left(\frac{d\rho}{dt}\right)_{\text{interactions}}.} \qquad (4.16)$$

The time derivative on the left is now taken moving through phase space with a fiducial system that does not interact with the external universe.

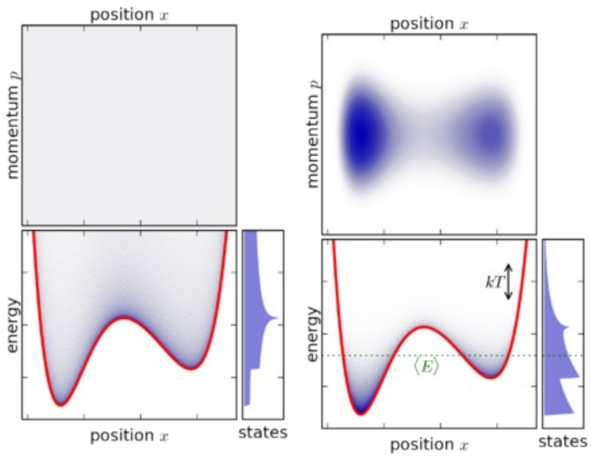
STATISTICAL EQUILIBRIUM AND JEANS' THEOREM

Consider an ensemble of identical systems, all of which have the same huge number of degrees of freedom (dimensionality $W \gg 1$). Put all the systems initially in the same state, and then let them exchange heat (but not particles, volume, or anything else) with an external *thermal bath* that has a huge heat capacity and is in thermodynamic equilibrium at some temperature T. (For example, the systems might be impermeable cubes of gas 1 km on a side near the center of the Sun, and the thermal bath might be all the surrounding gas near the Sun's center; or the systems might be identical sapphire crystals inside a huge cryostat, and the thermal bath might be the cryostat's huge store of liquid helium.) After a sufficiently long time, $t \gg \tau_{\text{ext}}$, the ensemble will settle down into equilibrium with the bath (i.e., it will become the canonical ensemble mentioned in Table 4.1 above). In this final, canonical equilibrium state, the probability $\rho(t, \mathbf{q}, \mathbf{p})$ is independent of time t, and it no longer is affected by interactions with the external environment. In other words, the interaction terms in the evolution equation (4.16) have ceased to have any net effect: on average, for each interaction event that feeds energy into a system, there is an interaction event that takes away an equal amount of energy. The distribution function, therefore, satisfies the interaction-free, collisionless Boltzmann equation (4.15) with the time derivative $\partial\rho/\partial t$ removed:

$$[\rho, H]_{\mathbf{q},\mathbf{p}} \equiv \frac{\partial\rho}{\partial q_j}\frac{\partial H}{\partial p_j} - \frac{\partial\rho}{\partial p_j}\frac{\partial H}{\partial q_j} = 0. \qquad (4.17)$$

We use the phrase *statistical equilibrium* to refer to any ensemble whose distribution function has attained such a state and thus satisfies Eq. (4.17).

Canonical distribution



To summarize, an ensemble of identical systems with many degrees of freedom $W \gg 1$, which have reached statistical equilibrium by exchanging energy but nothing else with a huge thermal bath, has the following canonical distribution function:

$$\boxed{\rho_{\text{canonical}} = C\,\exp(-\mathcal{E}/k_BT)}, \quad \boxed{\rho_{\text{canonical}} = C'\,\exp(-E/k_BT) \text{ nonrelativistically.}} \tag{4.20}$$

Here $\mathcal{E}(\mathbf{q}, \mathbf{p})$ is the energy of a system at location $\{\mathbf{q}, \mathbf{p}\}$ in phase space, k_B is Boltzmann's constant, T is the temperature of the heat bath, and C is whatever normalization constant is required to guarantee that $\sum_n \rho_n = 1$. The nonrelativistic expression is obtained by removing all the particle rest masses from the total energy $\mathcal{E}$ and then taking the low-temperature, low-thermal-velocities limit.

GENERAL EQUILIBRIUM ENSEMBLE

We can easily generalize the canonical distribution to an ensemble of systems that exchange other additive conserved quantities (extensive variables) $K_1, K_2, \ldots$, in addition to energy $\mathcal{E}$, with a huge, thermalized bath. By an obvious generalization of the argument in Sec. 4.4.1, the resulting statistical equilibrium distribution function must have the form

$$\boxed{\rho = C\,\exp\left(-\beta\mathcal{E} - \sum_A \beta_A K_A\right).} \tag{4.21}$$

When the extensive variables K_A that are exchanged with the bath (and thus appear explicitly in the distribution function ρ) are energy $\mathcal{E}$, momentum $\mathbf{P}$, angular momentum $\mathbf{J}$, the number N_I of the species I of conserved particles, volume V, or any combination of these quantities, it is conventional to rename the multiplicative factors β and β_A so that ρ takes on the form

$$\boxed{\rho = C\,\exp\left[\frac{-\mathcal{E} + \mathbf{U}\cdot\mathbf{P} + \boldsymbol{\Omega}\cdot\mathbf{J} + \sum_I \tilde{\mu}_I N_I - PV}{k_BT}\right].} \tag{4.22}$$

Here T, $\mathbf{U}$, $\boldsymbol{\Omega}$, $\tilde{\mu}_I$, and P are constants (called *intensive* variables) that are the same for all systems and subsystems (i.e., that characterize the full ensemble and all its subensembles and therefore must have been acquired from the bath); *any extensive variable that is not exchanged with the bath must be omitted from the exponential and be replaced by an implicit delta function.*

SPECIAL EQUILIBRIUM ENSEMBLES

Henceforth (except in Sec. 4.10.2, when discussing black-hole atmospheres), we restrict our baths always to be at rest in our chosen reference frame and to be nonrotating with respect to inertial frames, so that $\mathbf{U} = \boldsymbol{\Omega} = 0$. The distribution function ρ can then either be a delta function in the system momentum $\mathbf{P}$ and angular momentum $\mathbf{J}$ (if momentum and angular momentum are not exchanged with the bath), or it can involve no explicit dependence on $\mathbf{P}$ and $\mathbf{J}$ (if momentum and angular momentum are exchanged with the bath; cf. Eq. (4.22) with $\mathbf{U} = \boldsymbol{\Omega} = 0$). In either case, if energy is the only other quantity exchanged with the bath, then the distribution function is the canonical one [Eq. (4.20)]:

$$\boxed{\rho_{\text{canonical}} = C \exp\left[\frac{-\mathcal{E}}{k_B T}\right] = C' \exp\left[\frac{-E}{k_B T}\right],} \tag{4.25a}$$

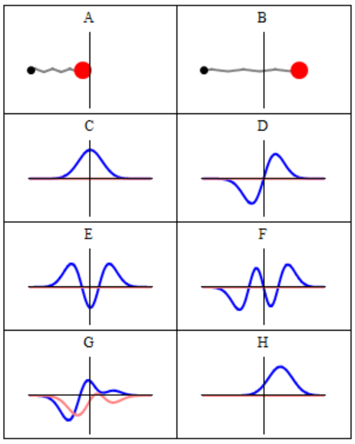
where (obviously) the constants C and C' are related by

$$C' = C \exp\left[-\sum_I N_I m_I / k_B T\right].$$

$$\boxed{\rho_{\text{Gibbs}} = C \exp\left[\frac{-(\mathcal{E} + PV)}{k_B T}\right] = C' \exp\left[\frac{-(E + PV)}{k_B T}\right]} \tag{4.25b}$$

(with an implicit delta function in N_I and possibly in $\mathbf{J}$ and $\mathbf{P}$). The combination $\mathcal{E} + PV$ is known as the *enthalpy* H. If the exchanged quantities are energy and particles but not volume (e.g., if the systems are 1-m cubes inside the Sun with totally imaginary walls through which particles and heat can flow), then the equilibrium is the *grand canonical ensemble*, with

$$\boxed{\rho_{\text{grand canonical}} = C \exp\left[\frac{-\mathcal{E} + \sum_I \tilde{\mu}_I N_I}{k_B T}\right] = C \exp\left[\frac{-E + \sum_I \mu_I N_I}{k_B T}\right]} \tag{4.25c}$$



SINGLE-PARTICLE QUANTUM STATES (MODES)

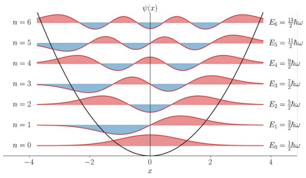
As an important example, let each system be a single-particle quantum state of some field. These quantum states can exchange particles (quanta) with one another. As we shall see, in this case the above considerations imply that, in statistical equilibrium at temperature T, the mean number of particles in a state, whose individual particle energies are $\mathcal{E}$, is given by the Fermi-Dirac formula (for fermions) $\eta = 1/(e^{(\mathcal{E}-\tilde{\mu})/(k_BT)} + 1)$ and Bose-Einstein formula (for bosons) $\eta = 1/(e^{(\mathcal{E}-\tilde{\mu})/(k_BT)} - 1)$, which we used in our kinetic-theory studies in the last chapter [Eqs. (3.22a), (3.22b)]. Our derivation of these mean occupation numbers will

Choose one specific mode $\mathcal{S}$ [e.g., a nonrelativistic electron plane-wave mode in a box of side L with spin up and momentum $\mathbf{p} = (5, 3, 17)h/L$]. There is one such mode $\mathcal{S}$ in each of the systems in our ensemble, and these modes (all identical in their properties) form a subensemble of our original ensemble. Our derivation focuses on this subensemble of identical modes $\mathcal{S}$. Because each of these modes can exchange energy and particles with all the other modes in its system, the subensemble is grand canonically distributed.

The *(many-particle) quantum states* allowed for mode $\mathcal{S}$ are states in which $\mathcal{S}$ contains a finite number n of particles (quanta). Denote by $\mathcal{E}_{\mathcal{S}}$ the energy of one particle residing in mode $\mathcal{S}$. Then the mode's total energy when it is in the state $|n\rangle$ (when it contains n quanta) is $\mathcal{E}_n = n\mathcal{E}_{\mathcal{S}}$. [For a freely traveling, relativistic electron mode, $\mathcal{E}_{\mathcal{S}} = \sqrt{m^2 + \mathbf{p}^2}$, Eq. (1.40), where $\mathbf{p}$ is the mode's momentum, $p_x = jh/L$ for some integer j and similarly for p_y and p_z; for a phonon mode with angular eigenfrequency of vibration ω, $\mathcal{E}_{\mathcal{S}} = \hbar\omega$.] Since the distribution of the ensemble's modes among the allowed quantum states is grand canonical, the probability ρ_n of being in state $|n\rangle$ is [Eq. (4.25c)]

$$\rho_n = C\ \exp\left(\frac{-\mathcal{E}_n + \tilde{\mu} n}{k_B T}\right) = C\ \exp\left(\frac{n(\tilde{\mu} - \mathcal{E}_{\mathcal{S}})}{k_B T}\right), \qquad (4.26)$$

where $\tilde{\mu}$ and T are the chemical potential and temperature of the bath of other modes, with which the mode $\mathcal{S}$ interacts.[8]



FERMION MODES: FERMI-DIRAC DISTRIBUTION

Suppose that $\mathcal{S}$ is a fermion mode (i.e., its particles have half-integral spin). Then the Pauli exclusion principle dictates that $\mathcal{S}$ cannot contain more than one particle: n can take on only the values 0 and 1. In this case, the normalization constant in the distribution function (4.26) is determined by $\rho_0 + \rho_1 = 1$, which implies that

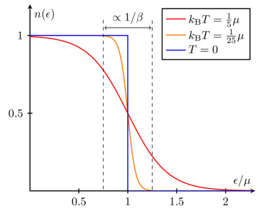
$$\rho_0 = \frac{1}{1 + \exp[(\tilde{\mu} - \mathcal{E}_{\mathcal{S}})/(k_B T)]}, \quad \rho_1 = \frac{\exp[(\tilde{\mu} - \mathcal{E}_{\mathcal{S}})/(k_B T)]}{1 + \exp[(\tilde{\mu} - \mathcal{E}_{\mathcal{S}})/(k_B T)]}. \qquad (4.27a)$$

This is the explicit form of the grand canonical distribution for a fermion mode. For many purposes (including all those in Chap. 3), this full probability distribution is more than one needs. Quite sufficient instead is the mode's mean occupation number

$$\boxed{\eta_{\mathcal{S}} \equiv \langle n \rangle = \sum_{n=0}^{1} n\rho_n = \frac{1}{\exp[(\mathcal{E}_{\mathcal{S}} - \tilde{\mu})/(k_B T)] + 1} = \frac{1}{\exp[(E_{\mathcal{S}} - \mu)/(k_B T)] + 1}.} \qquad (4.27b)$$

Here $E_{\mathcal{S}} = \mathcal{E}_{\mathcal{S}} - m$ is the energy of a particle in the mode with rest mass removed, and $\mu = \tilde{\mu} - m$ is the chemical potential with rest mass removed—the quantities used in the nonrelativistic (Newtonian) regime.

Equation (4.27b) is the *Fermi-Dirac mean occupation number* asserted in Chap. 3 [Eq. (3.22a)] and studied there for the special case of a gas of freely moving, noninteracting fermions. Because our derivation is completely general, we conclude that this mean occupation number and the underlying grand canonical distribution (4.27a) are valid for any mode of a fermion field—for example, the modes for an electron trapped in an external potential well or a magnetic bottle, and the (single-particle) quantum states of an electron in a hydrogen atom.



BOSON MODES: BOSE-EINSTEIN DISTRIBUTION

Suppose that $\mathcal{S}$ is a boson mode (i.e., its particles have integral spin), so it can contain any nonnegative number of quanta; that is, n can assume the values $0, 1, 2, 3, \ldots$. Then the normalization condition $\sum_{n=0}^{\infty} \rho_n = 1$ fixes the constant in the grand canonical distribution (4.26), resulting in

$$\rho_n = \left[1 - \exp\left(\frac{\tilde{\mu} - \mathcal{E}_{\mathcal{S}}}{k_B T}\right)\right] \exp\left(\frac{n(\tilde{\mu} - \mathcal{E}_{\mathcal{S}})}{k_B T}\right). \tag{4.28a}$$

From this grand canonical distribution we can deduce the mean number of bosons in mode $\mathcal{S}$:

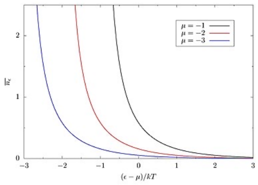
$$\boxed{\eta_{\mathcal{S}} \equiv \langle n \rangle = \sum_{n=1}^{\infty} n\rho_n = \frac{1}{\exp[(\mathcal{E}_{\mathcal{S}} - \tilde{\mu})/(k_B T)] - 1} = \frac{1}{\exp[(E_{\mathcal{S}} - \mu)/(k_B T)] - 1},} \tag{4.28b}$$

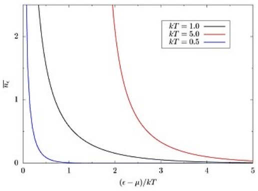
in accord with Eq. (3.22b). As for fermions, this *Bose-Einstein mean occupation number* and underlying grand canonical distribution (4.28a) are valid generally, and not solely for the freely moving bosons of Chap. 3.

When the mean occupation number is small, $\eta_{\mathcal{S}} \ll 1$, both the bosonic and the fermionic distribution functions are well approximated by the classical *Boltzmann mean occupation number*

$$\boxed{\eta_{\mathcal{S}} = \exp[-(\mathcal{E}_{\mathcal{S}} - \tilde{\mu})/(k_B T)].} \tag{4.29}$$

In Sec. 4.9 we explore an important modern application of the Bose-Einstein mean occupation number (4.28b): *Bose-Einstein condensation* of bosonic atoms in a magnetic trap.





4.4.4 Equipartition Theorem for Quadratic, Classical Degrees of Freedom

As a second example of statistical equilibrium distribution functions, we derive the classical equipartition theorem using statistical methods.

To motivate this theorem, consider a diatomic molecule of nitrogen, N_2. To a good approximation, its energy (its hamiltonian) can be written as

$$E = \frac{p_x^2}{2M} + \frac{p_y^2}{2M} + \frac{p_z^2}{2M} + \frac{P_\ell^2}{2M_\ell} + \frac{1}{2} M_\ell \, \omega_v^2 \, \ell^2 + \frac{J_x^2}{2I} + \frac{J_y^2}{2I}. \tag{4.30}$$

Here M is the molecule's mass; p_x, p_y, and p_z are the components of its translational momentum; and the first three terms are the molecule's kinetic energy of translation. The next two terms are the molecule's longitudinal vibration energy, with ℓ the change of the molecule's length (change of the separation of its two nuclei) from equilibrium, P_ℓ the generalized momentum conjugate to that length change, ω_v the vibration frequency, and M_ℓ the generalized mass associated with that vibration. The last two terms are the molecule's energy of end-over-end rotation, with J_x and J_y the components of angular momentum associated with this two-dimensional rotator and I its moment of inertia.

Notice that every term in this hamiltonian is quadratic in a generalized coordinate or generalized momentum! Moreover, each of these coordinates and momenta appears only in its single quadratic term and nowhere else, and the density of states is independent of the value of that coordinate or momentum. We refer to such a coordinate or momentum as a *quadratic degree of freedom*.

In some cases (e.g., the vibrations and rotations but not the translations), the energy $E_\xi = \alpha\xi^2$ of a quadratic degree of freedom ξ is quantized, with some energy separation ε_0 between the ground state and first excited state (and with energy separations to higher states that are $\lesssim \varepsilon_0$). If (and only if) the thermal energy $k_B T$ is significantly larger than ε_0, then the quadratic degree of freedom ξ will be excited far above its ground state and will behave classically. The equipartition theorem applies only at these high temperatures. For diatomic nitrogen, the rotational degrees of freedom J_x and J_y have $\varepsilon_0 \sim 10^{-4}$ eV and $\varepsilon_0/k_B \sim 1$ K, so temperatures big compared to 1 K are required for J_x and J_y to behave classically. By contrast, the vibrational degrees of freedom ℓ and P_ℓ have $\varepsilon_0 \sim 0.1$ eV and $\varepsilon_0/k_B \sim 1{,}000$ K, so temperatures of a few thousand Kelvins are required for them to behave classically. Above $\sim 10^4$ K, the hamiltonian (4.30) fails: electrons around the nuclei are driven into excited states, and the molecule breaks apart (dissociates into two free atoms of nitrogen).

The equipartition theorem holds for any classical, quadratic degree of freedom [i.e., at temperatures somewhat higher than $T_o = \varepsilon_o/(k_B T)$]. We derive this theorem using the canonical distribution (4.25a). We write the molecule's total energy as $E = \alpha\xi^2 + E'$, where E' does not involve ξ. Then the mean energy associated with ξ is

$$\langle E_\xi \rangle = \frac{\int \alpha\xi^2 \, e^{-\beta(\alpha\xi^2 + E')} d\xi \; d(\text{other degrees of freedom})}{\int e^{-\beta(\alpha\xi^2 + E')} d\xi \; d(\text{other degrees of freedom})}. \tag{4.31}$$

Here the exponential is that of the canonical distribution function (4.25a), the denominator is the normalizing factor, and we have set $\beta \equiv 1/(k_B T)$. Because ξ does not appear in the portion E' of the energy, its integral separates out from the others in both numerator and denominator, and the integrals over E' in numerator and denominator cancel. Rewriting $\int \alpha\xi^2 \exp(-\beta\alpha\xi^2) \, d\xi$ as $-d/d\beta[\int \exp(-\beta\alpha\xi^2) \, d\xi]$, Eq. (4.31) becomes

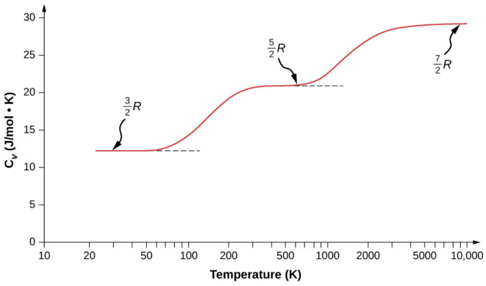
$$\begin{aligned} \langle E_\xi \rangle &= -\frac{d}{d\beta} \ln \left[\int \exp(-\beta\alpha\xi^2) \, d\xi \right] \\ &= -\frac{d}{d\beta} \ln \left[\frac{1}{\sqrt{\beta\alpha}} \int du e^{-u^2} du \right] = \frac{1}{2\beta} = \frac{1}{2} k_B T. \end{aligned} \tag{4.32}$$

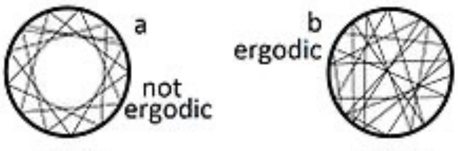
Therefore, *in statistical equilibrium, the mean energy associated with any classical, quadratic degree of freedom is $\frac{1}{2} k_B T$*. This is the equipartition theorem. Note that the factor $\frac{1}{2}$ follows from the quadratic nature of the degrees of freedom.

For our diatomic molecule, at room temperature there are three translational and two rotational classical, quadratic degrees of freedom $(p_x, p_y, p_z, J_x, J_y)$, so the mean total energy of the molecule is $\frac{5}{2} k_B T$. At a temperature of several thousand Kelvins, the two vibrational degrees of freedom, ℓ and P_ℓ, become classical and the molecule's mean total energy is $\frac{7}{2} k_B T$. Above $\sim 10^4$ K the molecule dissociates, and its two parts (the two nitrogen atoms) have only translational quadratic degrees of freedom, so the mean energy per atom is $\frac{3}{2} k_B T$.

Translational, rotational and vibrational degrees of freedom

Motion	Structure	Degrees of freedom	Average energy
Translation	All	3	$\frac{3}{2}kT$
Rotation	Linear	2	kT
	Non-linear	3	$\frac{3}{2}kT$
Vibration	Linear	$3N-(3+2)=3N-5$	$2(3N-5)\frac{1}{2}kT$
	Non-linear	$3N-(3+3)=3N-6$	$2(3N-6)\frac{1}{2}kT$
Total	Linear	$3N$	$\left(3N-\frac{5}{2}\right)kT$
	Non-linear	$3N$	$3(N-1)kT$





4.6 The Ergodic Hypothesis

The ensembles we have been studying are almost always just conceptual ones that do not exist in the real universe. We have introduced them and paid so much attention to them not for their own sakes, but because, in the case of statistical-equilibrium ensembles, they can be powerful tools for studying the properties of a single, individual system that really does exist in the universe or in our laboratory.

This power comes about because a sequence of snapshots of the single system, taken at times separated by sufficiently large intervals Δt, has a probability distribution ρ (for the snapshots' instantaneous locations $\{\mathbf{q}, \mathbf{p}\}$ in phase space) that is the same as the distribution function ρ of some conceptual, statistical-equilibrium ensemble. If the single system is closed, so its evolution is driven solely by its own hamiltonian, then the time between snapshots should be $\Delta t \gg \tau_{\rm int}$, and its snapshots will be (very nearly) microcanonically distributed. If the single system exchanges energy, and only energy, with a thermal bath on a timescale $\tau_{\rm ext}$, then the time between snapshots should be $\Delta t \gg \tau_{\rm ext}$, and its snapshots will be canonically distributed; similarly for the other types of bath interactions. This property of snapshots is equivalent to the statement that *for the individual system, the long-term time average*[10] *of any function of the system's location in phase space is equal to the statistical-equilibrium ensemble average:*

$$\boxed{\bar{A} \equiv \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{+T/2} A(\mathbf{q}(t), \mathbf{p}(t)) = \langle A\rangle \equiv \sum_n A_n \rho_n.} \qquad (4.33)$$

This property comes about because of *ergodicity:* the individual system, as it evolves, visits each accessible quantum state n for a fraction of the time that is equal to the equilibrium ensemble's probability ρ_n.

Entropy

For an ensemble

$$\boxed{S \equiv -k_B \sum_n \rho_n \ln \rho_n.}$$

Microcanonical

$$\boxed{S = k_B \ln N_{\text{states}}.}$$

The entropy, so defined, has some important properties. One is that when the ensemble can be broken up into statistically independent subensembles of subsystems (as is generally the case for big systems in statistical equilibrium), so that $\rho = \prod_a \rho_a$, then the entropy is additive: $S = \sum_a S_a$ (see Ex. 4.3). This permits us to regard the entropy, like the systems' additive constants of motion, as an extensive variable.

A second very important property is that, as an ensemble of systems evolves, its entropy cannot decrease, and it generally tends to increase. This is the statistical mechanical version of the second law of thermodynamics.

Entropy is maximized when ρ is constant (microcanonical ensemble)

Consider the class of all distribution functions ρ that: (i) vanish unless the constants of motion have the prescribed values $\mathcal{E}$ (in the tiny range $\delta\mathcal{E}$) and K_A; (ii) can be nonzero anywhere in the region of phase space, which we call $\mathcal{Y}_o$, where the prescribed values $\mathcal{E}$, K_A are taken; and (iii) are correctly normalized so that

$$\sum_n \rho_n \equiv \int_{\mathcal{Y}_o} \rho \mathcal{N}_{\text{states}} d\Gamma = 1 \qquad (4.36a)$$

[Eq. (4.8b)]. We ask which ρ in this class gives the largest entropy

$$S = -k_B \sum_n \rho_n \ln \rho_n.$$

$$\delta S = \delta \int_{\mathcal{Y}_o} (-k_B \rho \ln \rho - \Lambda\rho) \mathcal{N}_{\text{states}} d\Gamma = 0. \qquad (4.36b)$$

Here Λ is a Lagrange multiplier that enforces the normalization (4.36a). Performing the variation, we find that

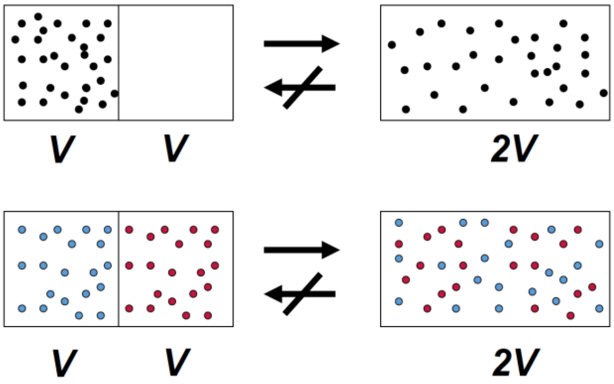
$$\int_{\mathcal{Y}_o} (-k_B \ln \rho - k_B - \Lambda) \delta\rho \mathcal{N}_{\text{states}} d\Gamma = 0, \qquad (4.36c)$$

which is satisfied if and only if ρ is a constant, $\rho = e^{-1-\Lambda/k_B}$, independent of location in the allowed region $\mathcal{Y}_o$ of phase space (i.e., if and only if ρ is that of the microcanonical ensemble). This calculation actually only shows that the microcanonical ensemble has stationary entropy. To show it is a maximum, one must perform the second variation (i.e., compute the second-order contribution of $\delta\rho$ to $\delta S = \delta \int (-k_B \rho \ln \rho) \mathcal{N}_{\text{states}} d\Gamma$). That second-order contribution is easily seen to be

$$\delta^2 S = \int_{\mathcal{Y}_o} \left(-k_B \frac{(\delta\rho)^2}{2\rho} \right) \mathcal{N}_{\text{states}} d\Gamma < 0. \qquad (4.36d)$$

Thus, the microcanonical distribution does maximize the entropy, as claimed.

Examples



Paradox ?

There is an apparent paradox at the heart of statistical mechanics, and, at various stages in the development of the subject it has led to confusion and even despair.

It still creates controversy (see, e.g., Hawking and Penrose, 2010; Penrose, 1999).

Its simplest and most direct expression is to ask: how can the time-reversible, microscopic laws, encoded in a time-independent hamiltonian, lead to the remorseless increase of entropy?

Entropy increase

Assume, for simplicity, that at time t = 0 all the systems are concentrated in a small but finite region of phase space with volume $\Delta\Gamma$ as shown in Fig. 4.2a, with $\rho = 1/(\mathcal{N}_{\text{states}}\Delta\Gamma)$ in the occupied region and ρ = 0 everywhere else.

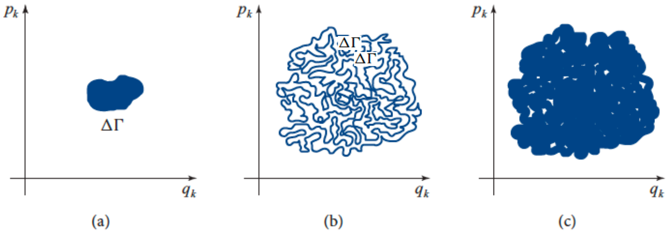


FIGURE 4.2 Evolution of a classical ensemble at $t = 0$ (a) toward statistical equilibrium by means of phase mixing (b) (cf. Fig. 4.1) followed by coarse-graining of one's viewpoint (c).

As time passes each system evolves under the action of the systems' common hamiltonian. As depicted in Fig. 4.2b, this evolution distorts the occupied region of phase space; but Liouville's theorem dictates that the occupied region's volume remain unchanged and, correspondingly, that the ensemble's entropy remains unchanged.

$$S = -k_B \int (\rho \ln \rho)\mathcal{N}_{\text{states}} d\Gamma = k_B \ln(\mathcal{N}_{\text{states}}\Delta\Gamma)$$

How can this be so? The ensemble is supposed to evolve into statistical equilibrium, with its distribution function uniformly spread out over that entire portion of phase space allowed by the hamiltonian's constants of motion—a portion of phase space far, far larger —and in the process the entropy is supposed to increase.

Coarsening

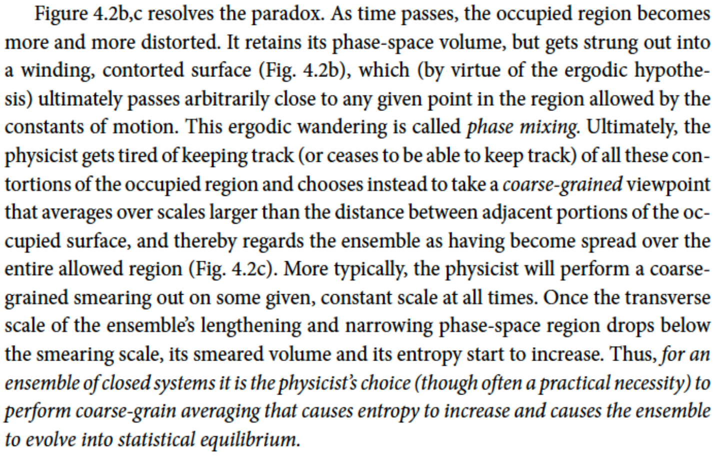
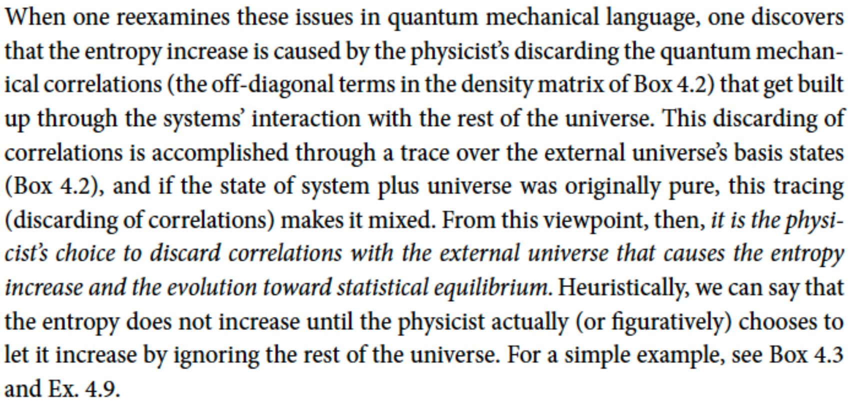


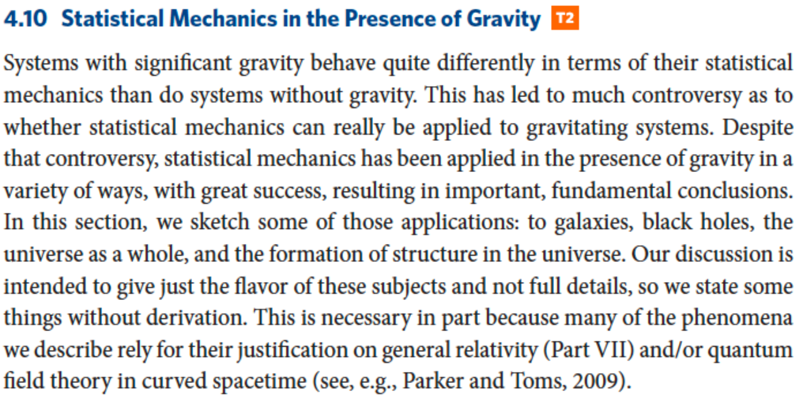
Figure 4.2b,c resolves the paradox. As time passes, the occupied region becomes more and more distorted. It retains its phase-space volume, but gets strung out into a winding, contorted surface (Fig. 4.2b), which (by virtue of the ergodic hypothesis) ultimately passes arbitrarily close to any given point in the region allowed by the constants of motion. This ergodic wandering is called *phase mixing*. Ultimately, the physicist gets tired of keeping track (or ceases to be able to keep track) of all these contortions of the occupied region and chooses instead to take a *coarse-grained* viewpoint that averages over scales larger than the distance between adjacent portions of the occupied surface, and thereby regards the ensemble as having become spread over the entire allowed region (Fig. 4.2c). More typically, the physicist will perform a coarse-grained smearing out on some given, constant scale at all times. Once the transverse scale of the ensemble's lengthening and narrowing phase-space region drops below the smearing scale, its smeared volume and its entropy start to increase. Thus, *for an ensemble of closed systems it is the physicist's choice (though often a practical necessity) to perform coarse-grain averaging that causes entropy to increase and causes the ensemble to evolve into statistical equilibrium.*

Discarding correlations



When one reexamines these issues in quantum mechanical language, one discovers that the entropy increase is caused by the physicist's discarding the quantum mechanical correlations (the off-diagonal terms in the density matrix of Box 4.2) that get built up through the systems' interaction with the rest of the universe. This discarding of correlations is accomplished through a trace over the external universe's basis states (Box 4.2), and if the state of system plus universe was originally pure, this tracing (discarding of correlations) makes it mixed. From this viewpoint, then, *it is the physicist's choice to discard correlations with the external universe that causes the entropy increase and the evolution toward statistical equilibrium.* Heuristically, we can say that the entropy does not increase until the physicist actually (or figuratively) chooses to let it increase by ignoring the rest of the universe. For a simple example, see Box 4.3 and Ex. 4.9.

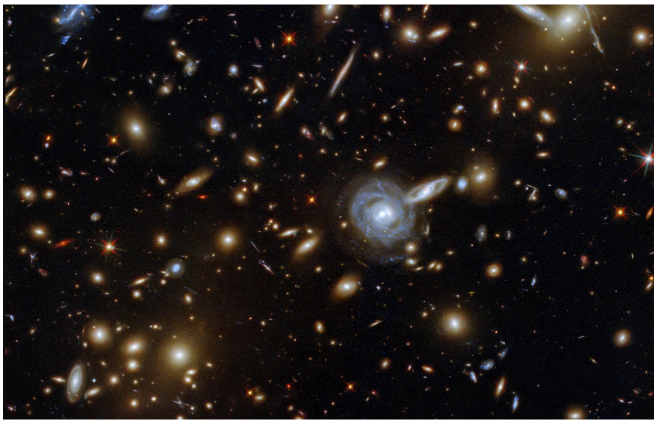
Gravity



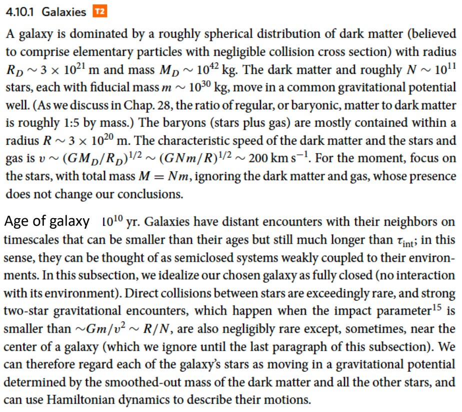
4.10 Statistical Mechanics in the Presence of Gravity T2

Systems with significant gravity behave quite differently in terms of their statistical mechanics than do systems without gravity. This has led to much controversy as to whether statistical mechanics can really be applied to gravitating systems. Despite that controversy, statistical mechanics has been applied in the presence of gravity in a variety of ways, with great success, resulting in important, fundamental conclusions. In this section, we sketch some of those applications: to galaxies, black holes, the universe as a whole, and the formation of structure in the universe. Our discussion is intended to give just the flavor of these subjects and not full details, so we state some things without derivation. This is necessary in part because many of the phenomena we describe rely for their justification on general relativity (Part VII) and/or quantum field theory in curved spacetime (see, e.g., Parker and Toms, 2009).

Galaxies as closed systems of non-interacting stars



$$\tau_{\text{int}} \sim 2R/v \sim 10^8 \text{ yr}.$$



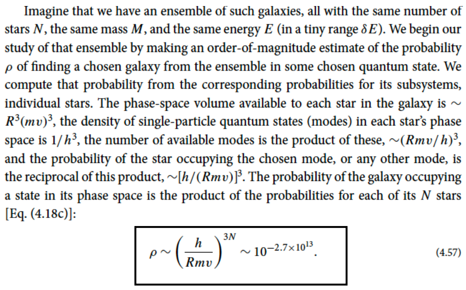
4.10.1 Galaxies T2

A galaxy is dominated by a roughly spherical distribution of dark matter (believed to comprise elementary particles with negligible collision cross section) with radius $R_D \sim 3 \times 10^{21}$ m and mass $M_D \sim 10^{42}$ kg. The dark matter and roughly $N \sim 10^{11}$ stars, each with fiducial mass $m \sim 10^{30}$ kg, move in a common gravitational potential well. (As we discuss in Chap. 28, the ratio of regular, or baryonic, matter to dark matter is roughly 1:5 by mass.) The baryons (stars plus gas) are mostly contained within a radius $R \sim 3 \times 10^{20}$ m. The characteristic speed of the dark matter and the stars and gas is $v \sim (GM_D/R_D)^{1/2} \sim (GNm/R)^{1/2} \sim 200 \text{ km s}^{-1}$. For the moment, focus on the stars, with total mass $M = Nm$, ignoring the dark matter and gas, whose presence does not change our conclusions.

Age of galaxy 10^{10} yr. Galaxies have distant encounters with their neighbors on timescales that can be smaller than their ages but still much longer than τ_{int}; in this sense, they can be thought of as semiclosed systems weakly coupled to their environments. In this subsection, we idealize our chosen galaxy as fully closed (no interaction with its environment). Direct collisions between stars are exceedingly rare, and strong two-star gravitational encounters, which happen when the impact parameter[15] is smaller than $\sim Gm/v^2 \sim R/N$, are also negligibly rare except, sometimes, near the center of a galaxy (which we ignore until the last paragraph of this subsection). We can therefore regard each of the galaxy's stars as moving in a gravitational potential determined by the smoothed-out mass of the dark matter and all the other stars, and can use Hamiltonian dynamics to describe their motions.

Entropy of a galaxy

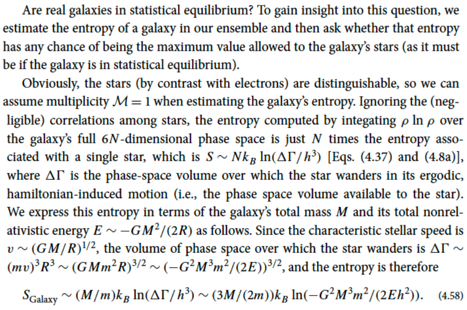
Probability density of N stars



Imagine that we have an ensemble of such galaxies, all with the same number of stars N, the same mass M, and the same energy E (in a tiny range δE). We begin our study of that ensemble by making an order-of-magnitude estimate of the probability ρ of finding a chosen galaxy from the ensemble in some chosen quantum state. We compute that probability from the corresponding probabilities for its subsystems, individual stars. The phase-space volume available to each star in the galaxy is $\sim R^3(mv)^3$, the density of single-particle quantum states (modes) in each star's phase space is $1/h^3$, the number of available modes is the product of these, $\sim(Rmv/h)^3$, and the probability of the star occupying the chosen mode, or any other mode, is the reciprocal of this product, $\sim[h/(Rmv)]^3$. The probability of the galaxy occupying a state in its phase space is the product of the probabilities for each of its N stars [Eq. (4.18c)]:

$$\boxed{\rho \sim \left(\frac{h}{Rmv}\right)^{3N} \sim 10^{-2.7\times 10^{13}}.} \qquad (4.57)$$

Are galaxies at equilibrium ?

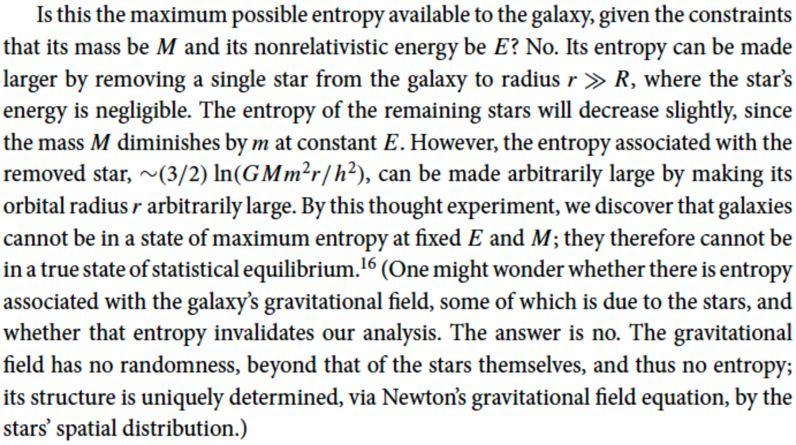


Are real galaxies in statistical equilibrium? To gain insight into this question, we estimate the entropy of a galaxy in our ensemble and then ask whether that entropy has any chance of being the maximum value allowed to the galaxy's stars (as it must be if the galaxy is in statistical equilibrium).

Obviously, the stars (by contrast with electrons) are distinguishable, so we can assume multiplicity $\mathcal{M} = 1$ when estimating the galaxy's entropy. Ignoring the (negligible) correlations among stars, the entropy computed by integating $\rho \ln \rho$ over the galaxy's full $6N$-dimensional phase space is just N times the entropy associated with a single star, which is $S \sim Nk_B \ln(\Delta\Gamma/h^3)$ [Eqs. (4.37) and (4.8a)], where $\Delta\Gamma$ is the phase-space volume over which the star wanders in its ergodic, hamiltonian-induced motion (i.e., the phase space volume available to the star). We express this entropy in terms of the galaxy's total mass M and its total nonrelativistic energy $E \sim -GM^2/(2R)$ as follows. Since the characteristic stellar speed is $v \sim (GM/R)^{1/2}$, the volume of phase space over which the star wanders is $\Delta\Gamma \sim (mv)^3R^3 \sim (GMm^2R)^{3/2} \sim (-G^2M^3m^2/(2E))^{3/2}$, and the entropy is therefore

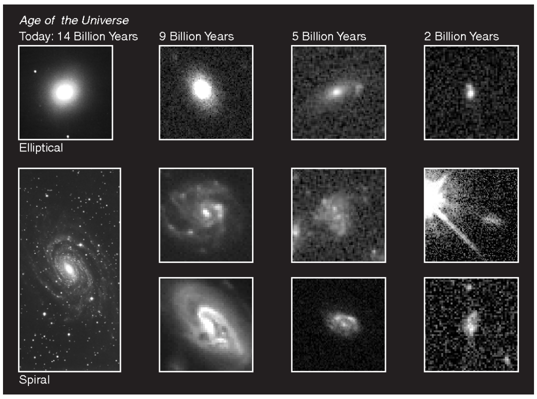
$$S_{\text{Galaxy}} \sim (M/m)k_B \ln(\Delta\Gamma/h^3) \sim (3M/(2m))k_B \ln(-G^2M^3m^2/(2Eh^2)). \qquad (4.58)$$

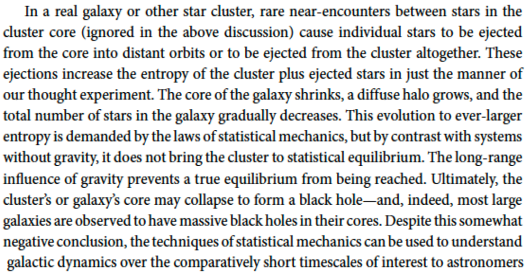
The galaxy entropy increases



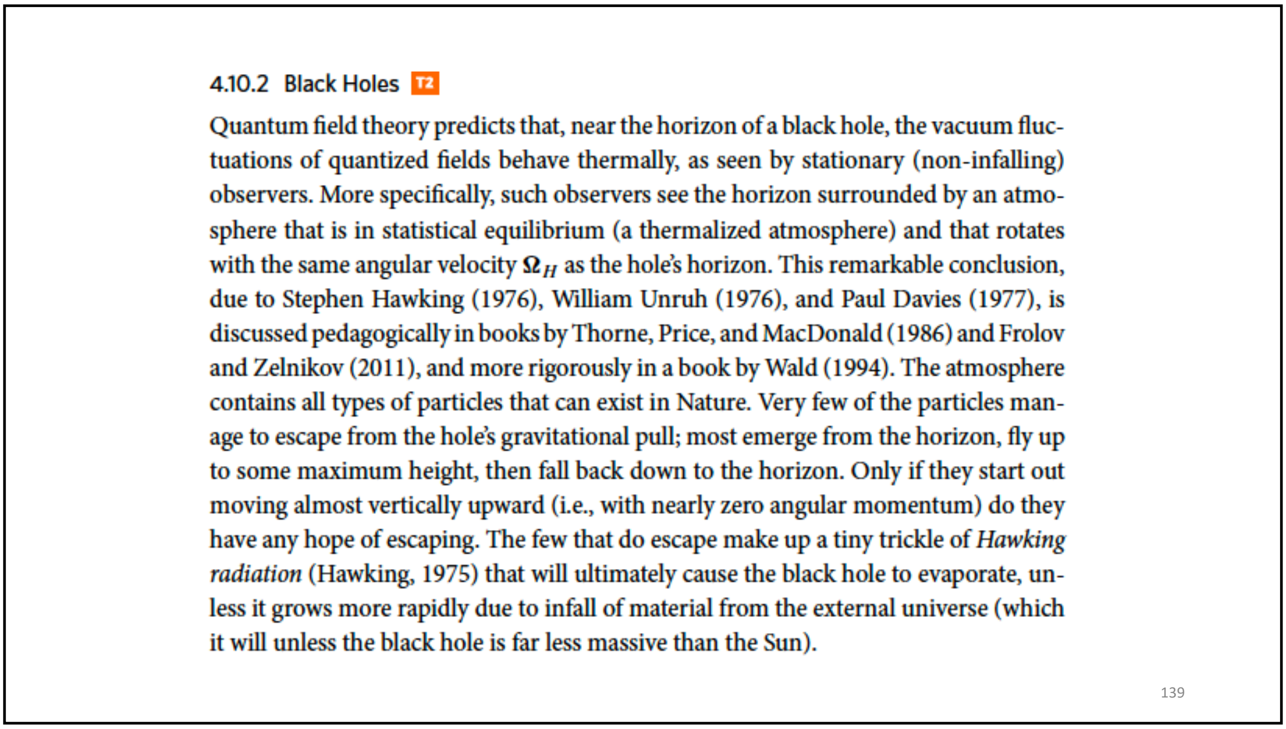
Is this the maximum possible entropy available to the galaxy, given the constraints that its mass be M and its nonrelativistic energy be E? No. Its entropy can be made larger by removing a single star from the galaxy to radius $r \gg R$, where the star's energy is negligible. The entropy of the remaining stars will decrease slightly, since the mass M diminishes by m at constant E. However, the entropy associated with the removed star, $\sim(3/2)\ln(GMm^2r/h^2)$, can be made arbitrarily large by making its orbital radius r arbitrarily large. By this thought experiment, we discover that galaxies cannot be in a state of maximum entropy at fixed E and M; they therefore cannot be in a true state of statistical equilibrium.[16] (One might wonder whether there is entropy associated with the galaxy's gravitational field, some of which is due to the stars, and whether that entropy invalidates our analysis. The answer is no. The gravitational field has no randomness, beyond that of the stars themselves, and thus no entropy; its structure is uniquely determined, via Newton's gravitational field equation, by the stars' spatial distribution.)

Galaxies are not in statistical equilibrium



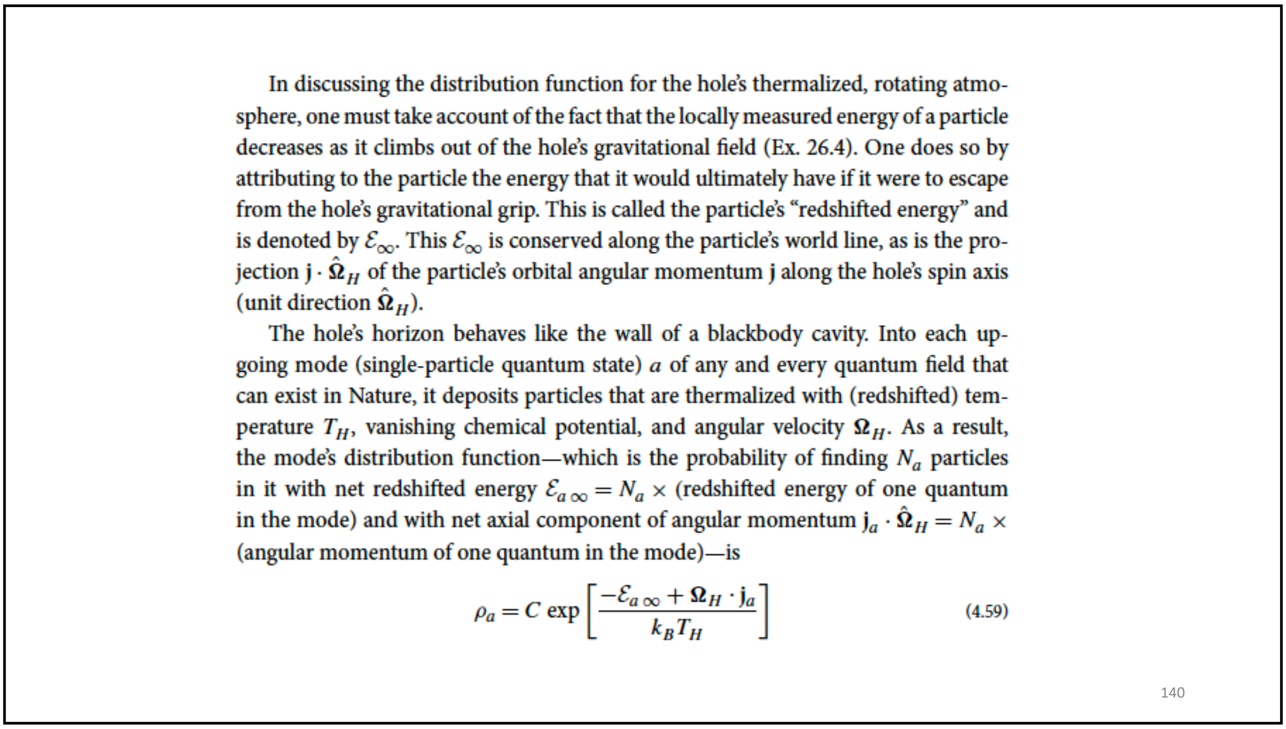


In a real galaxy or other star cluster, rare near-encounters between stars in the cluster core (ignored in the above discussion) cause individual stars to be ejected from the core into distant orbits or to be ejected from the cluster altogether. These ejections increase the entropy of the cluster plus ejected stars in just the manner of our thought experiment. The core of the galaxy shrinks, a diffuse halo grows, and the total number of stars in the galaxy gradually decreases. This evolution to ever-larger entropy is demanded by the laws of statistical mechanics, but by contrast with systems without gravity, it does not bring the cluster to statistical equilibrium. The long-range influence of gravity prevents a true equilibrium from being reached. Ultimately, the cluster's or galaxy's core may collapse to form a black hole—and, indeed, most large galaxies are observed to have massive black holes in their cores. Despite this somewhat negative conclusion, the techniques of statistical mechanics can be used to understand galactic dynamics over the comparatively short timescales of interest to astronomers



4.10.2 Black Holes T2

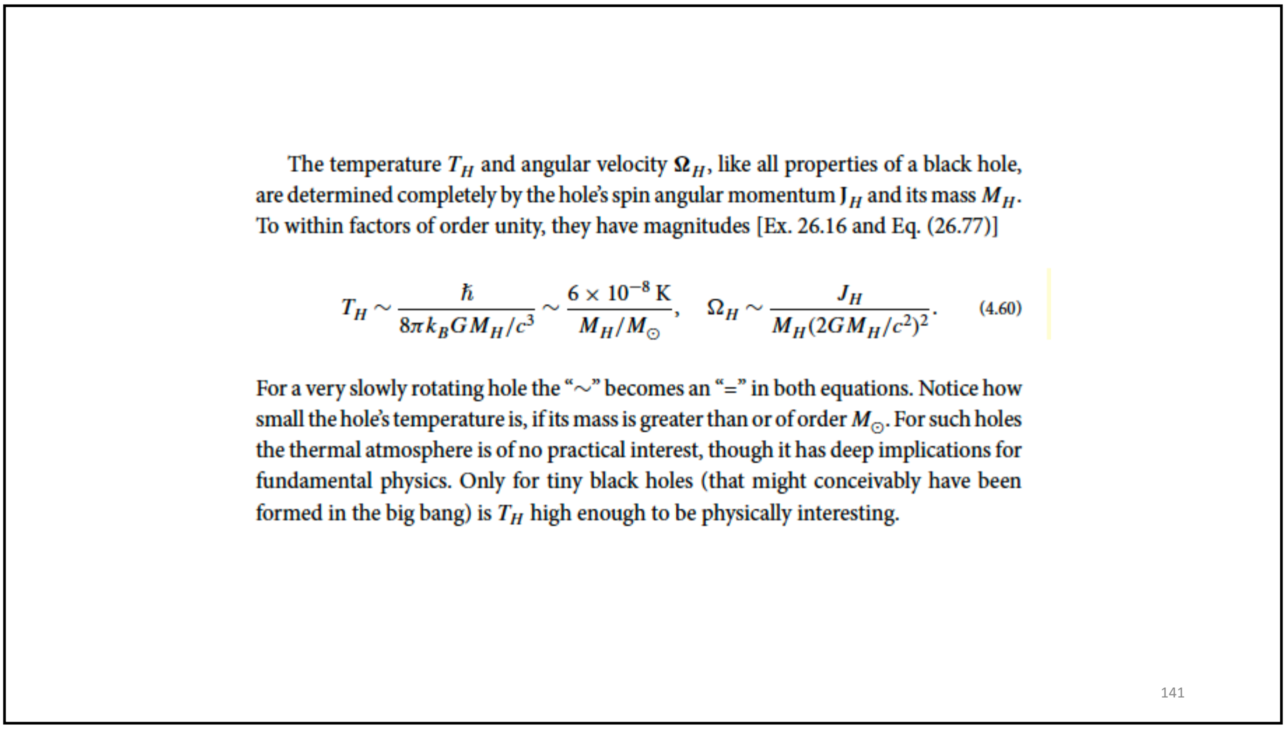
Quantum field theory predicts that, near the horizon of a black hole, the vacuum fluctuations of quantized fields behave thermally, as seen by stationary (non-infalling) observers. More specifically, such observers see the horizon surrounded by an atmosphere that is in statistical equilibrium (a thermalized atmosphere) and that rotates with the same angular velocity $\boldsymbol{\Omega}_H$ as the hole's horizon. This remarkable conclusion, due to Stephen Hawking (1976), William Unruh (1976), and Paul Davies (1977), is discussed pedagogically in books by Thorne, Price, and MacDonald (1986) and Frolov and Zelnikov (2011), and more rigorously in a book by Wald (1994). The atmosphere contains all types of particles that can exist in Nature. Very few of the particles manage to escape from the hole's gravitational pull; most emerge from the horizon, fly up to some maximum height, then fall back down to the horizon. Only if they start out moving almost vertically upward (i.e., with nearly zero angular momentum) do they have any hope of escaping. The few that do escape make up a tiny trickle of *Hawking radiation* (Hawking, 1975) that will ultimately cause the black hole to evaporate, unless it grows more rapidly due to infall of material from the external universe (which it will unless the black hole is far less massive than the Sun).



In discussing the distribution function for the hole's thermalized, rotating atmosphere, one must take account of the fact that the locally measured energy of a particle decreases as it climbs out of the hole's gravitational field (Ex. 26.4). One does so by attributing to the particle the energy that it would ultimately have if it were to escape from the hole's gravitational grip. This is called the particle's "redshifted energy" and is denoted by $\mathcal{E}_\infty$. This $\mathcal{E}_\infty$ is conserved along the particle's world line, as is the projection $\mathbf{j} \cdot \hat{\boldsymbol{\Omega}}_H$ of the particle's orbital angular momentum $\mathbf{j}$ along the hole's spin axis (unit direction $\hat{\boldsymbol{\Omega}}_H$).

The hole's horizon behaves like the wall of a blackbody cavity. Into each up-going mode (single-particle quantum state) a of any and every quantum field that can exist in Nature, it deposits particles that are thermalized with (redshifted) temperature T_H, vanishing chemical potential, and angular velocity $\boldsymbol{\Omega}_H$. As a result, the mode's distribution function—which is the probability of finding N_a particles in it with net redshifted energy $\mathcal{E}_{a\,\infty} = N_a \times$ (redshifted energy of one quantum in the mode) and with net axial component of angular momentum $\mathbf{j}_a \cdot \hat{\boldsymbol{\Omega}}_H = N_a \times$ (angular momentum of one quantum in the mode)—is

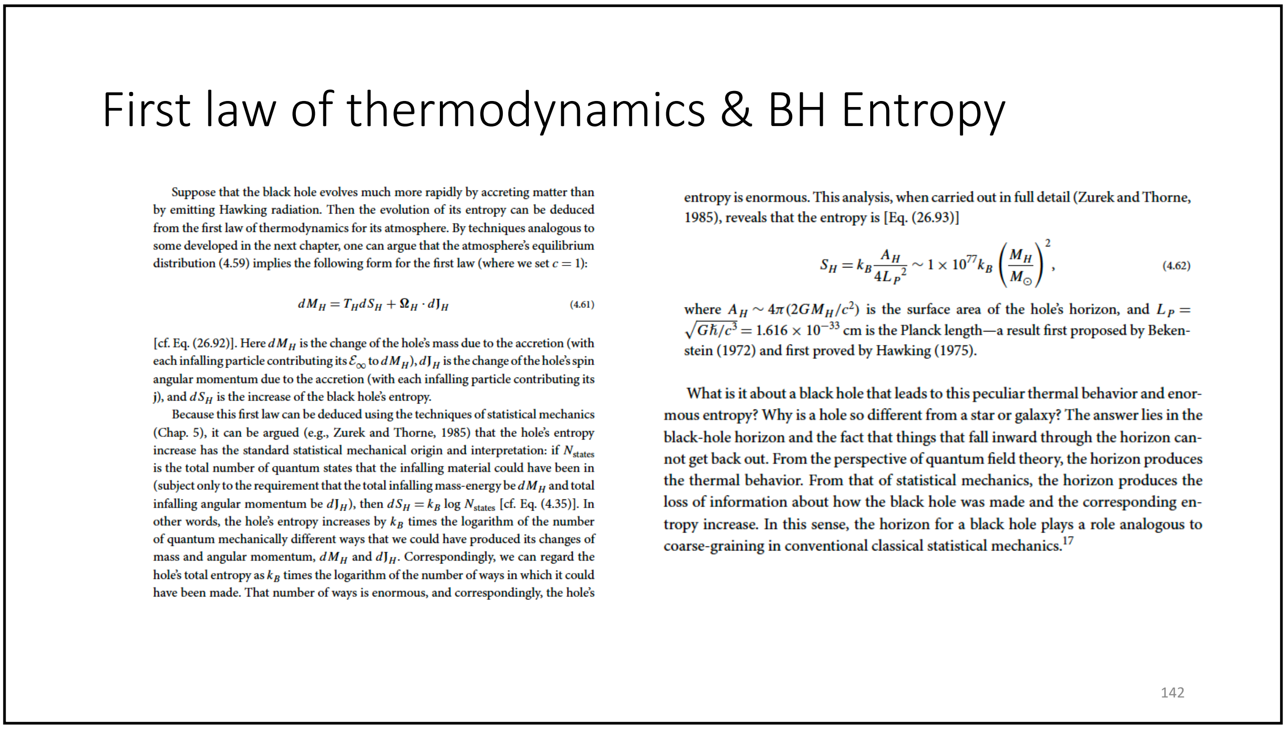
$$\rho_a = C \exp\left[\frac{-\mathcal{E}_{a\,\infty} + \boldsymbol{\Omega}_H \cdot \mathbf{j}_a}{k_B T_H}\right] \qquad (4.59)$$



The temperature T_H and angular velocity $\boldsymbol{\Omega}_H$, like all properties of a black hole, are determined completely by the hole's spin angular momentum $\mathbf{J}_H$ and its mass M_H. To within factors of order unity, they have magnitudes [Ex. 26.16 and Eq. (26.77)]

$$T_H \sim \frac{\hbar}{8\pi k_B G M_H/c^3} \sim \frac{6\times 10^{-8}\,\mathrm{K}}{M_H/M_\odot}, \qquad \Omega_H \sim \frac{J_H}{M_H(2GM_H/c^2)^2}. \tag{4.60}$$

For a very slowly rotating hole the "$\sim$" becomes an "=" in both equations. Notice how small the hole's temperature is, if its mass is greater than or of order $M_\odot$. For such holes the thermal atmosphere is of no practical interest, though it has deep implications for fundamental physics. Only for tiny black holes (that might conceivably have been formed in the big bang) is T_H high enough to be physically interesting.



First law of thermodynamics & BH Entropy

Suppose that the black hole evolves much more rapidly by accreting matter than by emitting Hawking radiation. Then the evolution of its entropy can be deduced from the first law of thermodynamics for its atmosphere. By techniques analogous to some developed in the next chapter, one can argue that the atmosphere's equilibrium distribution (4.59) implies the following form for the first law (where we set $c = 1$):

$$dM_H = T_H dS_H + \boldsymbol{\Omega}_H \cdot d\mathbf{J}_H \tag{4.61}$$

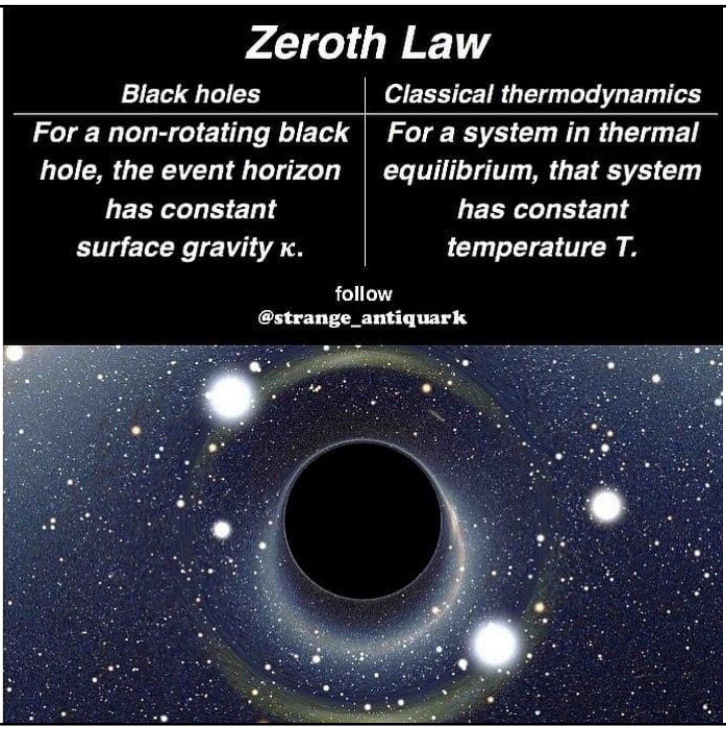
[cf. Eq. (26.92)]. Here dM_H is the change of the hole's mass due to the accretion (with each infalling particle contributing its $\mathcal{E}_\infty$ to dM_H), $d\mathbf{J}_H$ is the change of the hole's spin angular momentum due to the accretion (with each infalling particle contributing its j), and dS_H is the increase of the black hole's entropy.

Because this first law can be deduced using the techniques of statistical mechanics (Chap. 5), it can be argued (e.g., Zurek and Thorne, 1985) that the hole's entropy increase has the standard statistical mechanical origin and interpretation: if N_{states} is the total number of quantum states that the infalling material could have been in (subject only to the requirement that the total infalling mass-energy be dM_H and total infalling angular momentum be $d\mathbf{J}_H$), then $dS_H = k_B \log N_{\text{states}}$ [cf. Eq. (4.35)]. In other words, the hole's entropy increases by k_B times the logarithm of the number of quantum mechanically different ways that we could have produced its changes of mass and angular momentum, dM_H and $d\mathbf{J}_H$. Correspondingly, we can regard the hole's total entropy as k_B times the logarithm of the number of ways in which it could have been made. That number of ways is enormous, and correspondingly, the hole's entropy is enormous. This analysis, when carried out in full detail (Zurek and Thorne, 1985), reveals that the entropy is [Eq. (26.93)]

$$S_H = k_B \frac{A_H}{4{L_P}^2} \sim 1\times 10^{77} k_B \left(\frac{M_H}{M_\odot}\right)^2, \tag{4.62}$$

where $A_H \sim 4\pi(2GM_H/c^2)$ is the surface area of the hole's horizon, and $L_P = \sqrt{G\hbar/c^3} = 1.616\times 10^{-33}$ cm is the Planck length—a result first proposed by Bekenstein (1972) and first proved by Hawking (1975).

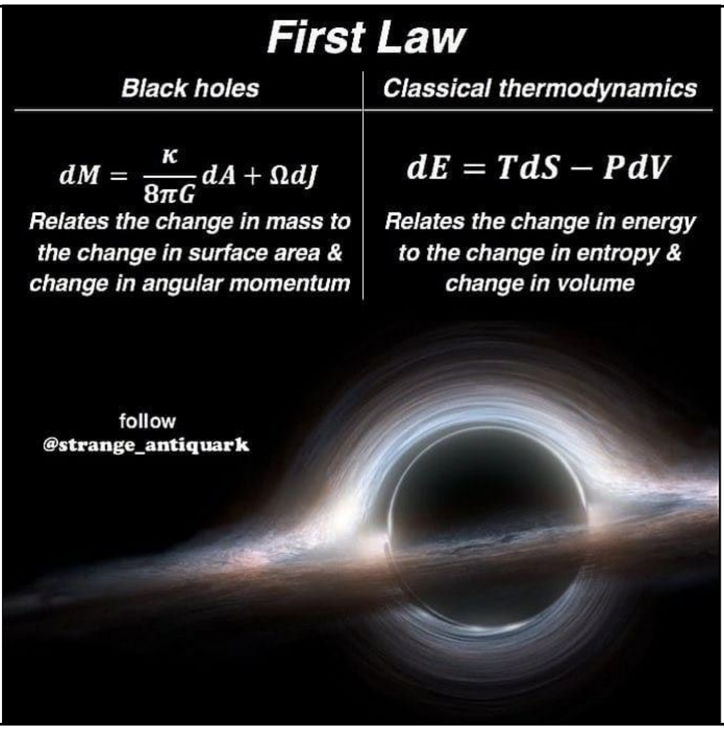
What is it about a black hole that leads to this peculiar thermal behavior and enormous entropy? Why is a hole so different from a star or galaxy? The answer lies in the black-hole horizon and the fact that things that fall inward through the horizon cannot get back out. From the perspective of quantum field theory, the horizon produces the thermal behavior. From that of statistical mechanics, the horizon produces the loss of information about how the black hole was made and the corresponding entropy increase. In this sense, the horizon for a black hole plays a role analogous to coarse-graining in conventional classical statistical mechanics.[17]



Zeroth Law

Black holes	*Classical thermodynamics*
For a non-rotating black hole, the event horizon has constant surface gravity κ.	*For a system in thermal equilibrium, that system has constant temperature T.*

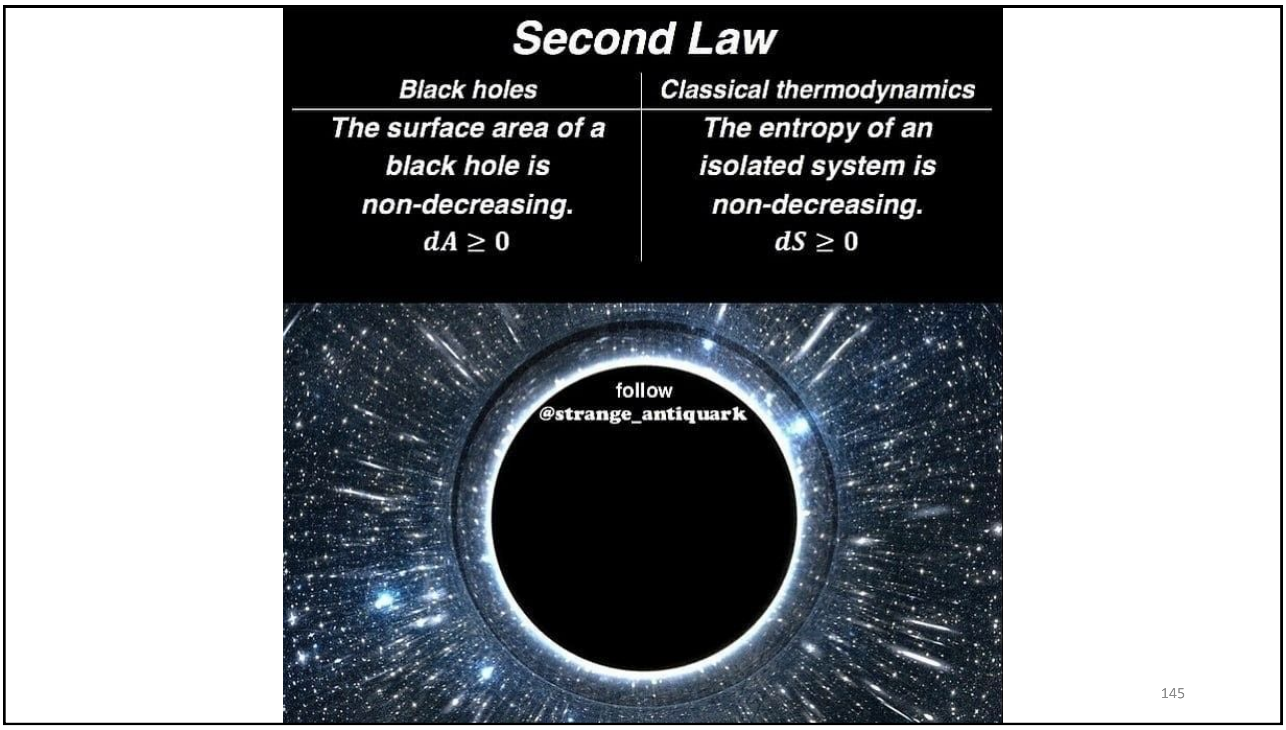
follow
@strange_antiquark



First Law

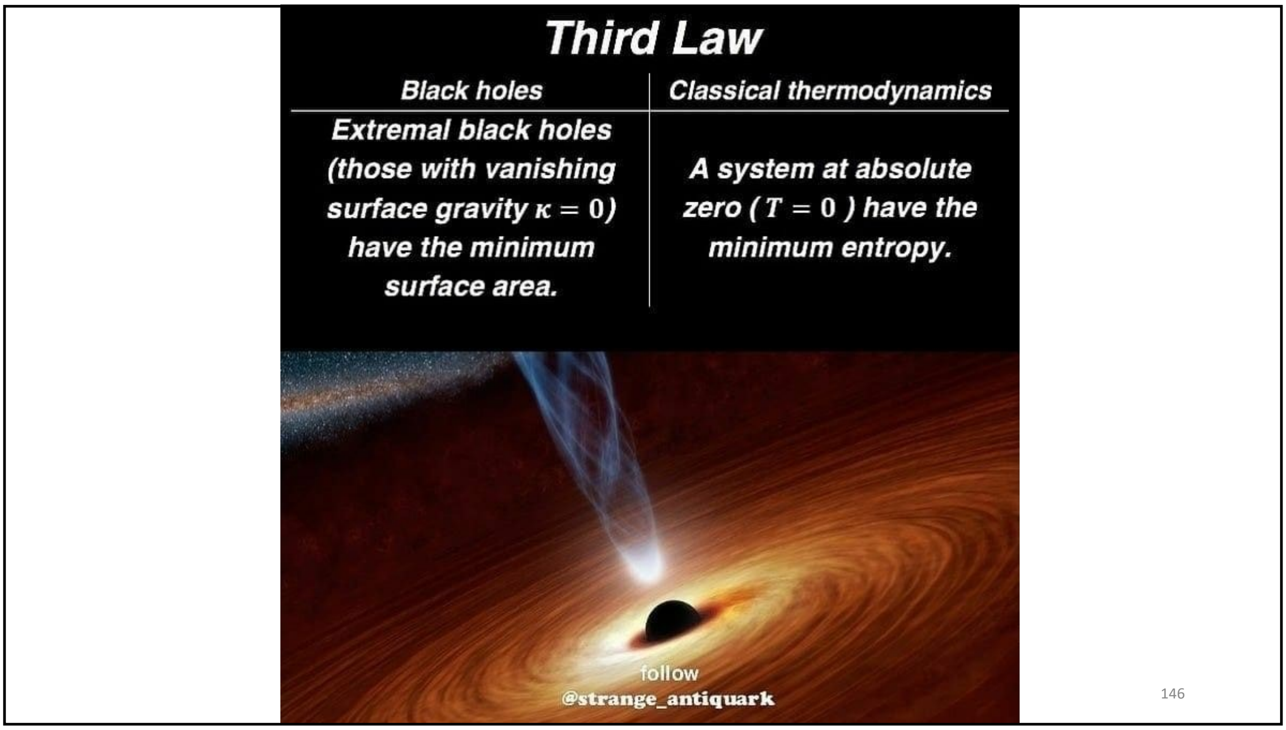
Black holes	*Classical thermodynamics*
$dM = \frac{\kappa}{8\pi G} dA + \Omega dJ$ *Relates the change in mass to the change in surface area & change in angular momentum*	$\boldsymbol{dE = TdS - PdV}$ *Relates the change in energy to the change in entropy & change in volume*

follow
@strange_antiquark



Second Law

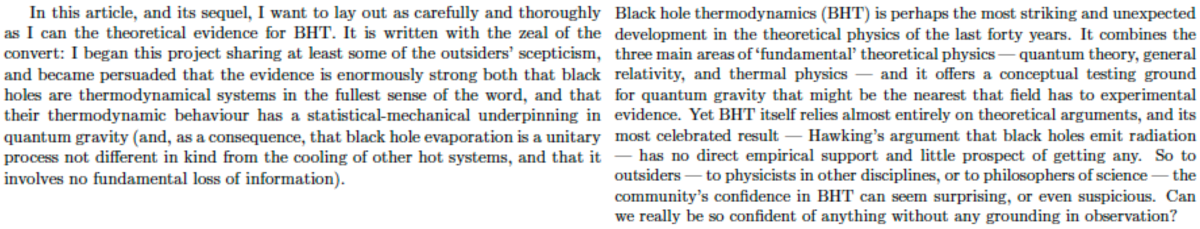
Black holes	Classical thermodynamics
The surface area of a black hole is non-decreasing. $dA \geq 0$	The entropy of an isolated system is non-decreasing. $dS \geq 0$



Third Law

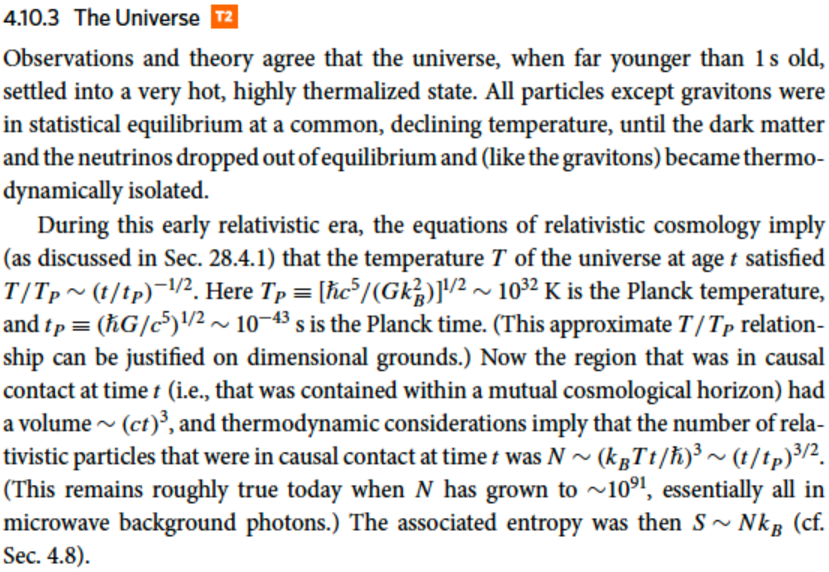
Black holes	Classical thermodynamics
Extremal black holes (those with vanishing surface gravity $\kappa = 0$) have the minimum surface area.	A system at absolute zero ($T = 0$) have the minimum entropy.

More on Black Hole Thermodynamics & SM (D. Wallace)



In this article, and its sequel, I want to lay out as carefully and thoroughly as I can the theoretical evidence for BHT. It is written with the zeal of the convert: I began this project sharing at least some of the outsiders' scepticism, and became persuaded that the evidence is enormously strong both that black holes are thermodynamical systems in the fullest sense of the word, and that their thermodynamic behaviour has a statistical-mechanical underpinning in quantum gravity (and, as a consequence, that black hole evaporation is a unitary process not different in kind from the cooling of other hot systems, and that it involves no fundamental loss of information).

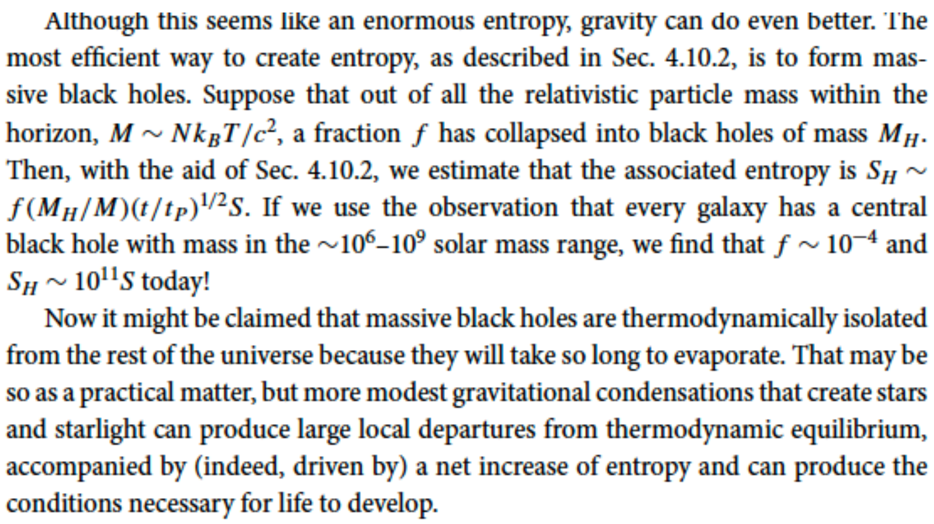
Black hole thermodynamics (BHT) is perhaps the most striking and unexpected development in the theoretical physics of the last forty years. It combines the three main areas of 'fundamental' theoretical physics — quantum theory, general relativity, and thermal physics — and it offers a conceptual testing ground for quantum gravity that might be the nearest that field has to experimental evidence. Yet BHT itself relies almost entirely on theoretical arguments, and its most celebrated result — Hawking's argument that black holes emit radiation — has no direct empirical support and little prospect of getting any. So to outsiders — to physicists in other disciplines, or to philosophers of science — the community's confidence in BHT can seem surprising, or even suspicious. Can we really be so confident of anything without any grounding in observation?



4.10.3 The Universe T2

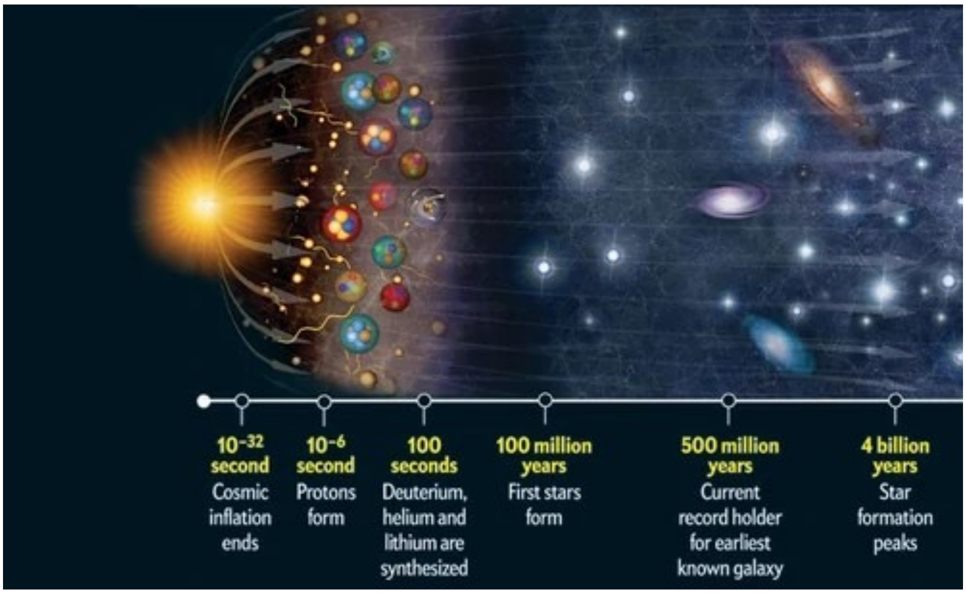
Observations and theory agree that the universe, when far younger than 1 s old, settled into a very hot, highly thermalized state. All particles except gravitons were in statistical equilibrium at a common, declining temperature, until the dark matter and the neutrinos dropped out of equilibrium and (like the gravitons) became thermodynamically isolated.

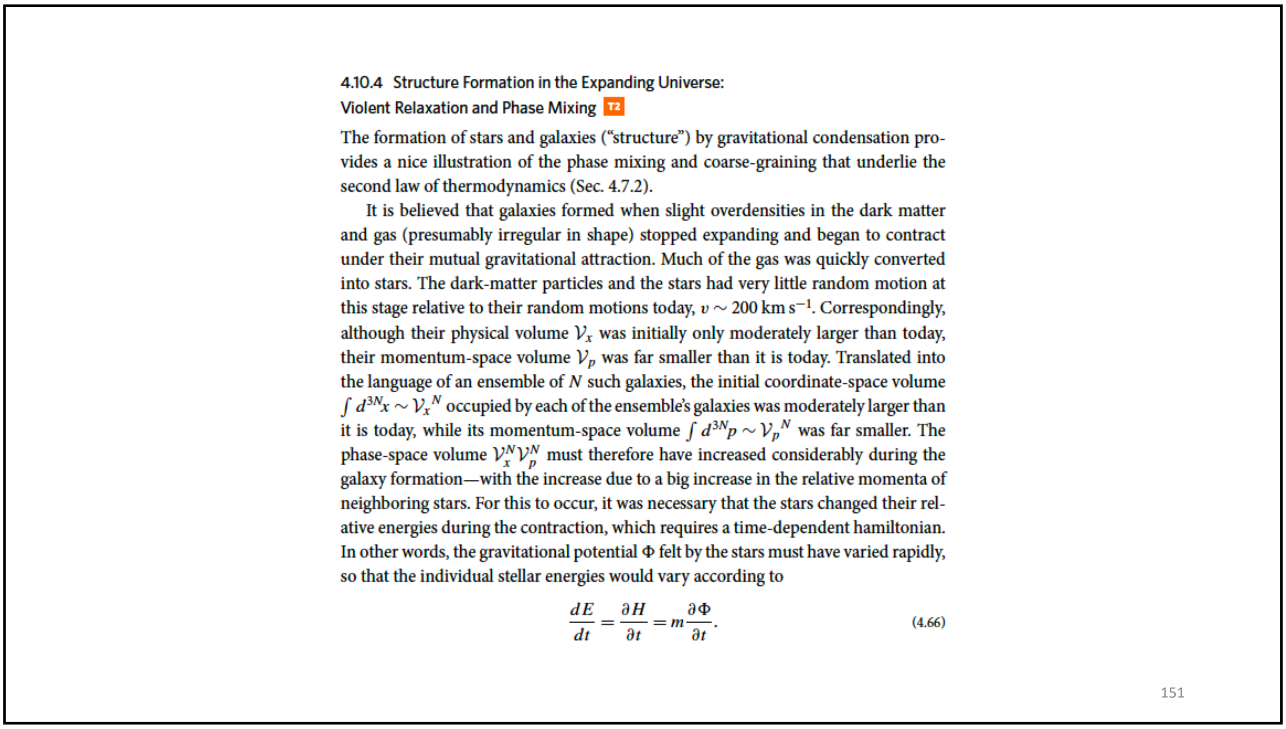
During this early relativistic era, the equations of relativistic cosmology imply (as discussed in Sec. 28.4.1) that the temperature T of the universe at age t satisfied $T/T_P \sim (t/t_P)^{-1/2}$. Here $T_P \equiv [\hbar c^5/(Gk_B^2)]^{1/2} \sim 10^{32}$ K is the Planck temperature, and $t_P \equiv (\hbar G/c^5)^{1/2} \sim 10^{-43}$ s is the Planck time. (This approximate T/T_P relationship can be justified on dimensional grounds.) Now the region that was in causal contact at time t (i.e., that was contained within a mutual cosmological horizon) had a volume $\sim (ct)^3$, and thermodynamic considerations imply that the number of relativistic particles that were in causal contact at time t was $N \sim (k_B T t/\hbar)^3 \sim (t/t_P)^{3/2}$. (This remains roughly true today when N has grown to $\sim 10^{91}$, essentially all in microwave background photons.) The associated entropy was then $S \sim N k_B$ (cf. Sec. 4.8).



Although this seems like an enormous entropy, gravity can do even better. The most efficient way to create entropy, as described in Sec. 4.10.2, is to form massive black holes. Suppose that out of all the relativistic particle mass within the horizon, $M \sim Nk_BT/c^2$, a fraction f has collapsed into black holes of mass M_H. Then, with the aid of Sec. 4.10.2, we estimate that the associated entropy is $S_H \sim f(M_H/M)(t/t_P)^{1/2}S$. If we use the observation that every galaxy has a central black hole with mass in the $\sim 10^6$–10^9 solar mass range, we find that $f \sim 10^{-4}$ and $S_H \sim 10^{11}S$ today!

Now it might be claimed that massive black holes are thermodynamically isolated from the rest of the universe because they will take so long to evaporate. That may be so as a practical matter, but more modest gravitational condensations that create stars and starlight can produce large local departures from thermodynamic equilibrium, accompanied by (indeed, driven by) a net increase of entropy and can produce the conditions necessary for life to develop.





4.10.4 Structure Formation in the Expanding Universe: Violent Relaxation and Phase Mixing T2

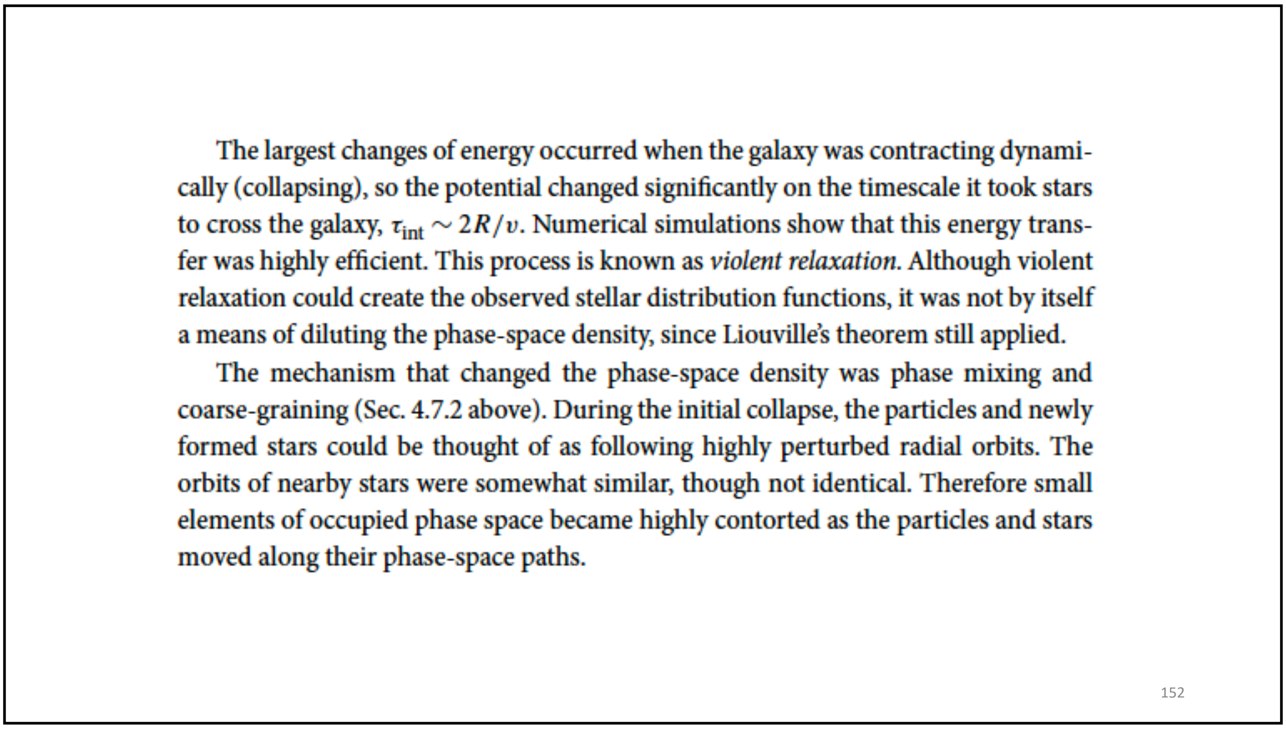
The formation of stars and galaxies ("structure") by gravitational condensation provides a nice illustration of the phase mixing and coarse-graining that underlie the second law of thermodynamics (Sec. 4.7.2).

It is believed that galaxies formed when slight overdensities in the dark matter and gas (presumably irregular in shape) stopped expanding and began to contract under their mutual gravitational attraction. Much of the gas was quickly converted into stars. The dark-matter particles and the stars had very little random motion at this stage relative to their random motions today, $v \sim 200\ \mathrm{km\ s^{-1}}$. Correspondingly, although their physical volume $\mathcal{V}_x$ was initially only moderately larger than today, their momentum-space volume $\mathcal{V}_p$ was far smaller than it is today. Translated into the language of an ensemble of N such galaxies, the initial coordinate-space volume $\int d^{3N}x \sim \mathcal{V}_x{}^N$ occupied by each of the ensemble's galaxies was moderately larger than it is today, while its momentum-space volume $\int d^{3N}p \sim \mathcal{V}_p{}^N$ was far smaller. The phase-space volume $\mathcal{V}_x^N \mathcal{V}_p^N$ must therefore have increased considerably during the galaxy formation—with the increase due to a big increase in the relative momenta of neighboring stars. For this to occur, it was necessary that the stars changed their relative energies during the contraction, which requires a time-dependent hamiltonian. In other words, the gravitational potential Φ felt by the stars must have varied rapidly, so that the individual stellar energies would vary according to

$$\frac{dE}{dt} = \frac{\partial H}{\partial t} = m\frac{\partial \Phi}{\partial t}. \qquad (4.66)$$

151

151

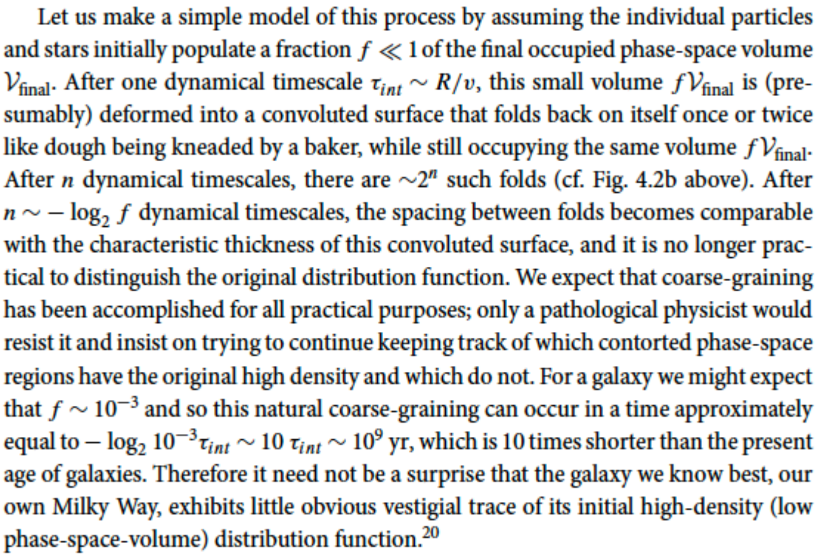


The largest changes of energy occurred when the galaxy was contracting dynamically (collapsing), so the potential changed significantly on the timescale it took stars to cross the galaxy, $\tau_{\mathrm{int}} \sim 2R/v$. Numerical simulations show that this energy transfer was highly efficient. This process is known as *violent relaxation*. Although violent relaxation could create the observed stellar distribution functions, it was not by itself a means of diluting the phase-space density, since Liouville's theorem still applied.

The mechanism that changed the phase-space density was phase mixing and coarse-graining (Sec. 4.7.2 above). During the initial collapse, the particles and newly formed stars could be thought of as following highly perturbed radial orbits. The orbits of nearby stars were somewhat similar, though not identical. Therefore small elements of occupied phase space became highly contorted as the particles and stars moved along their phase-space paths.

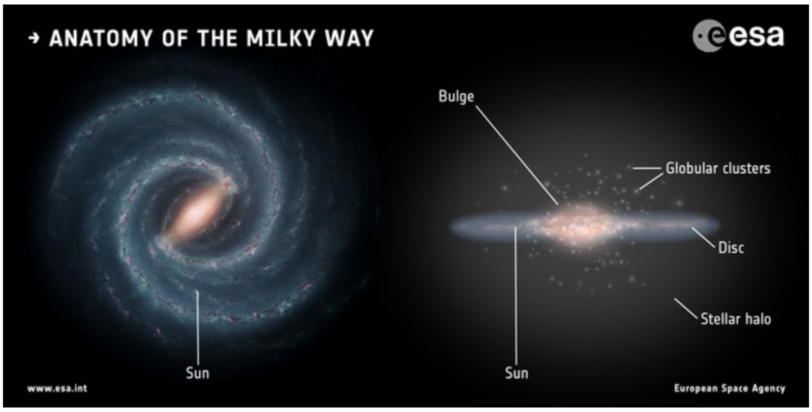
152

152

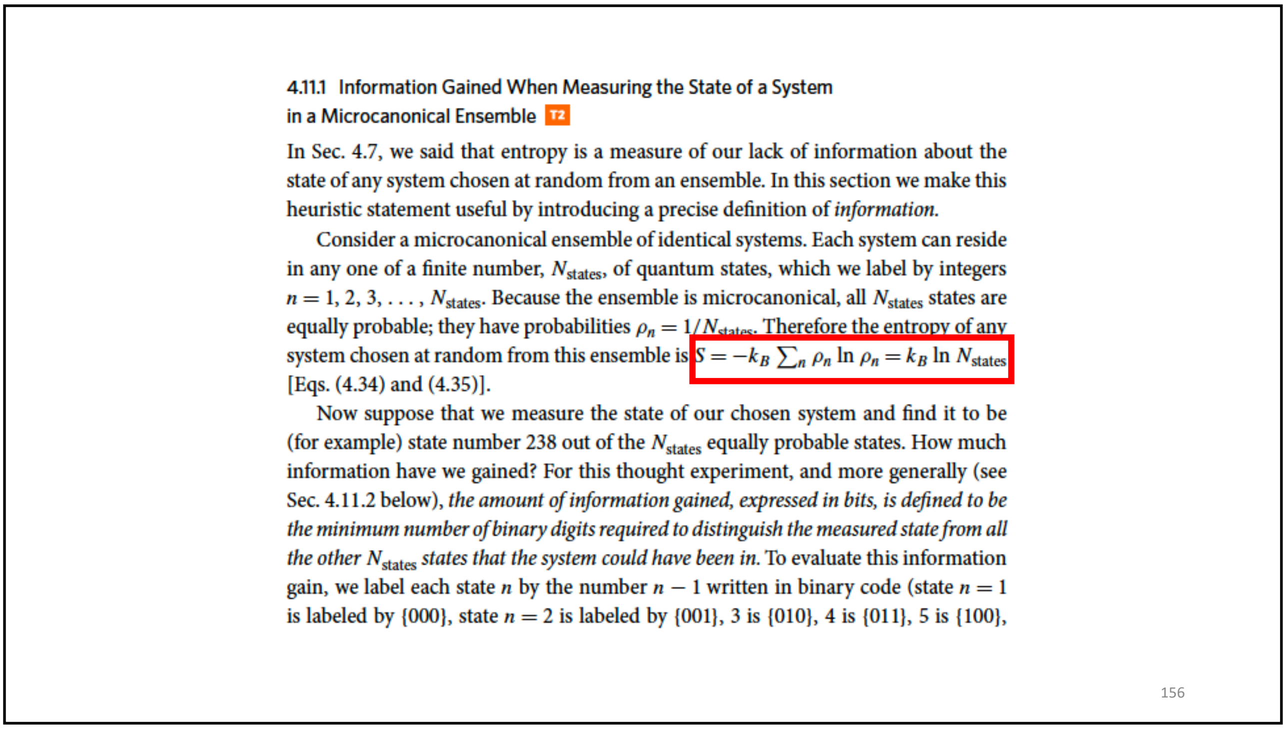


Let us make a simple model of this process by assuming the individual particles and stars initially populate a fraction $f \ll 1$ of the final occupied phase-space volume $\mathcal{V}_{\text{final}}$. After one dynamical timescale $\tau_{int} \sim R/v$, this small volume $f\mathcal{V}_{\text{final}}$ is (presumably) deformed into a convoluted surface that folds back on itself once or twice like dough being kneaded by a baker, while still occupying the same volume $f\mathcal{V}_{\text{final}}$. After n dynamical timescales, there are $\sim 2^n$ such folds (cf. Fig. 4.2b above). After $n \sim -\log_2 f$ dynamical timescales, the spacing between folds becomes comparable with the characteristic thickness of this convoluted surface, and it is no longer practical to distinguish the original distribution function. We expect that coarse-graining has been accomplished for all practical purposes; only a pathological physicist would resist it and insist on trying to continue keeping track of which contorted phase-space regions have the original high density and which do not. For a galaxy we might expect that $f \sim 10^{-3}$ and so this natural coarse-graining can occur in a time approximately equal to $-\log_2 10^{-3}\tau_{int} \sim 10\ \tau_{int} \sim 10^9$ yr, which is 10 times shorter than the present age of galaxies. Therefore it need not be a surprise that the galaxy we know best, our own Milky Way, exhibits little obvious vestigial trace of its initial high-density (low phase-space-volume) distribution function.[20]

Structure of the milky way



Entropy and information

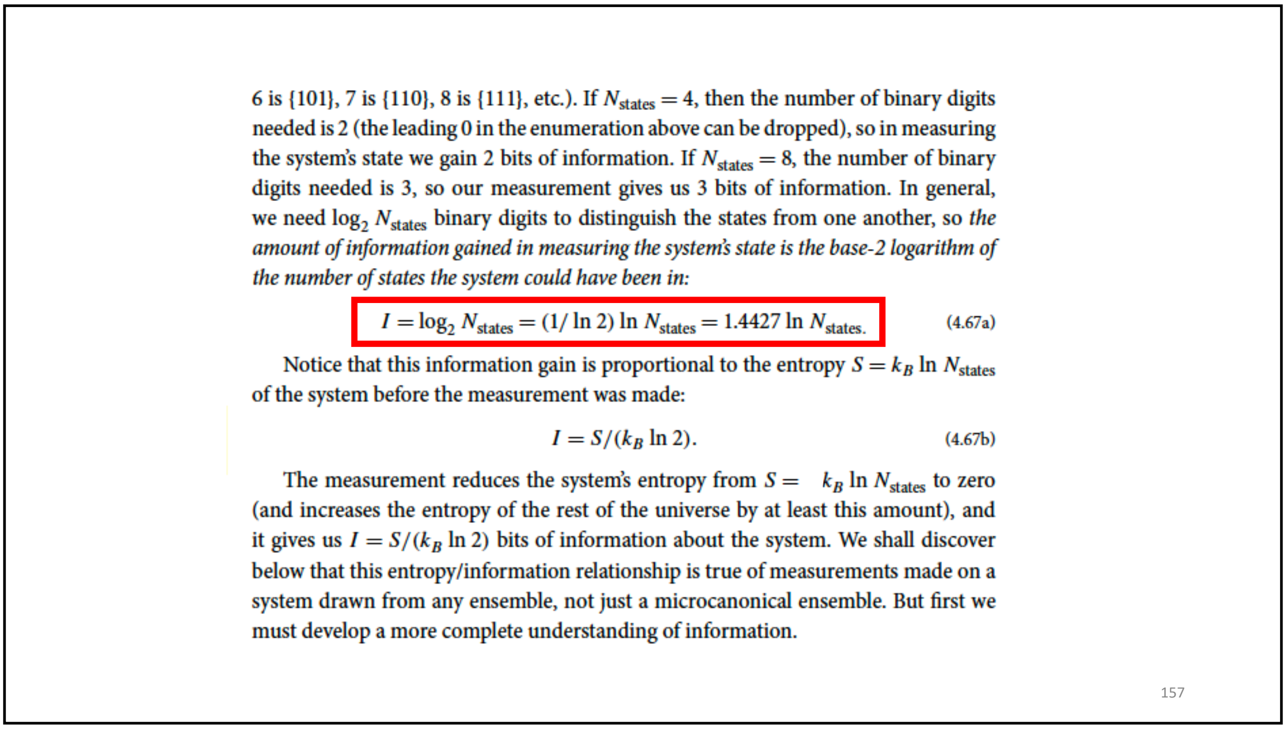


4.11.1 Information Gained When Measuring the State of a System in a Microcanonical Ensemble T2

In Sec. 4.7, we said that entropy is a measure of our lack of information about the state of any system chosen at random from an ensemble. In this section we make this heuristic statement useful by introducing a precise definition of *information*.

Consider a microcanonical ensemble of identical systems. Each system can reside in any one of a finite number, N_{states}, of quantum states, which we label by integers $n = 1, 2, 3, \ldots, N_{\text{states}}$. Because the ensemble is microcanonical, all N_{states} states are equally probable; they have probabilities $\rho_n = 1/N_{\text{states}}$. Therefore the entropy of any system chosen at random from this ensemble is $\boxed{S = -k_B \sum_n \rho_n \ln \rho_n = k_B \ln N_{\text{states}}}$ [Eqs. (4.34) and (4.35)].

Now suppose that we measure the state of our chosen system and find it to be (for example) state number 238 out of the N_{states} equally probable states. How much information have we gained? For this thought experiment, and more generally (see Sec. 4.11.2 below), *the amount of information gained, expressed in bits, is defined to be the minimum number of binary digits required to distinguish the measured state from all the other* N_{states} *states that the system could have been in*. To evaluate this information gain, we label each state n by the number $n - 1$ written in binary code (state $n = 1$ is labeled by {000}, state $n = 2$ is labeled by {001}, 3 is {010}, 4 is {011}, 5 is {100},



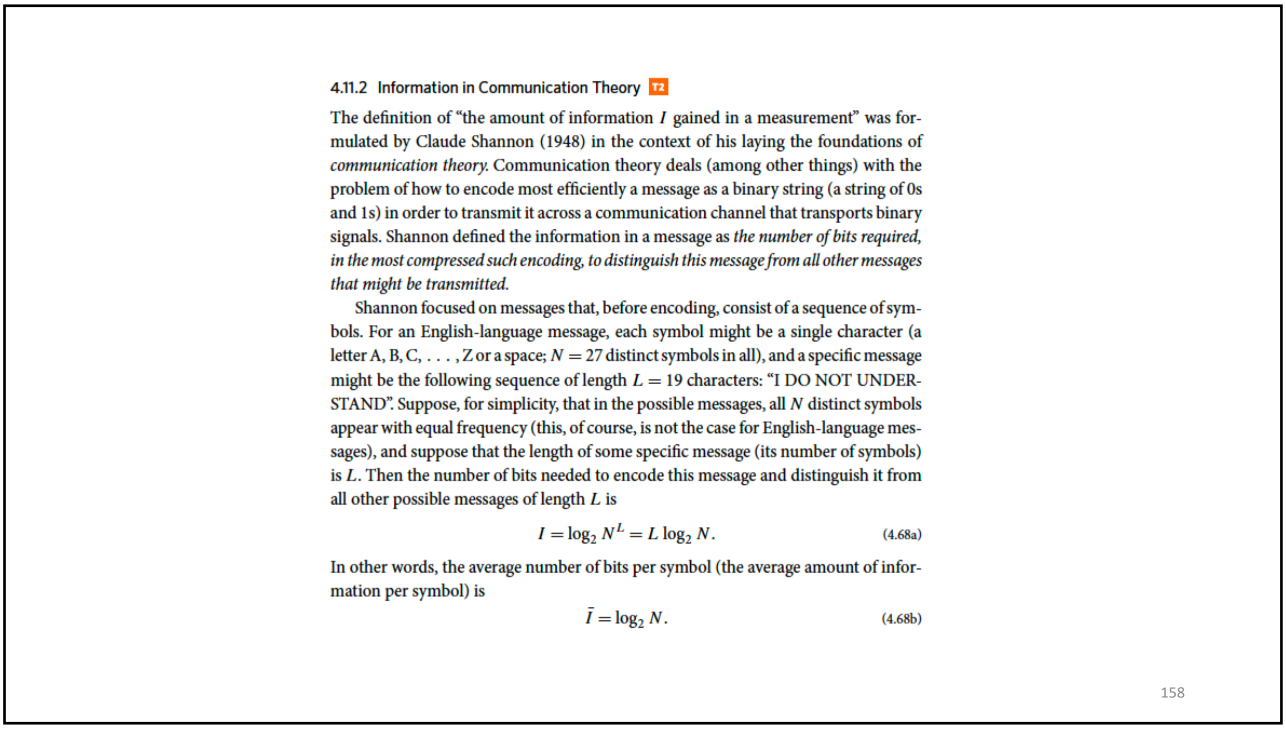
6 is {101}, 7 is {110}, 8 is {111}, etc.). If $N_{\text{states}} = 4$, then the number of binary digits needed is 2 (the leading 0 in the enumeration above can be dropped), so in measuring the system's state we gain 2 bits of information. If $N_{\text{states}} = 8$, the number of binary digits needed is 3, so our measurement gives us 3 bits of information. In general, we need $\log_2 N_{\text{states}}$ binary digits to distinguish the states from one another, so *the amount of information gained in measuring the system's state is the base-2 logarithm of the number of states the system could have been in:*

$$\boxed{I = \log_2 N_{\text{states}} = (1/\ln 2) \ln N_{\text{states}} = 1.4427 \ln N_{\text{states}}.} \quad (4.67a)$$

Notice that this information gain is proportional to the entropy $S = k_B \ln N_{\text{states}}$ of the system before the measurement was made:

$$I = S/(k_B \ln 2). \quad (4.67b)$$

The measurement reduces the system's entropy from $S = k_B \ln N_{\text{states}}$ to zero (and increases the entropy of the rest of the universe by at least this amount), and it gives us $I = S/(k_B \ln 2)$ bits of information about the system. We shall discover below that this entropy/information relationship is true of measurements made on a system drawn from any ensemble, not just a microcanonical ensemble. But first we must develop a more complete understanding of information.



4.11.2 Information in Communication Theory T2

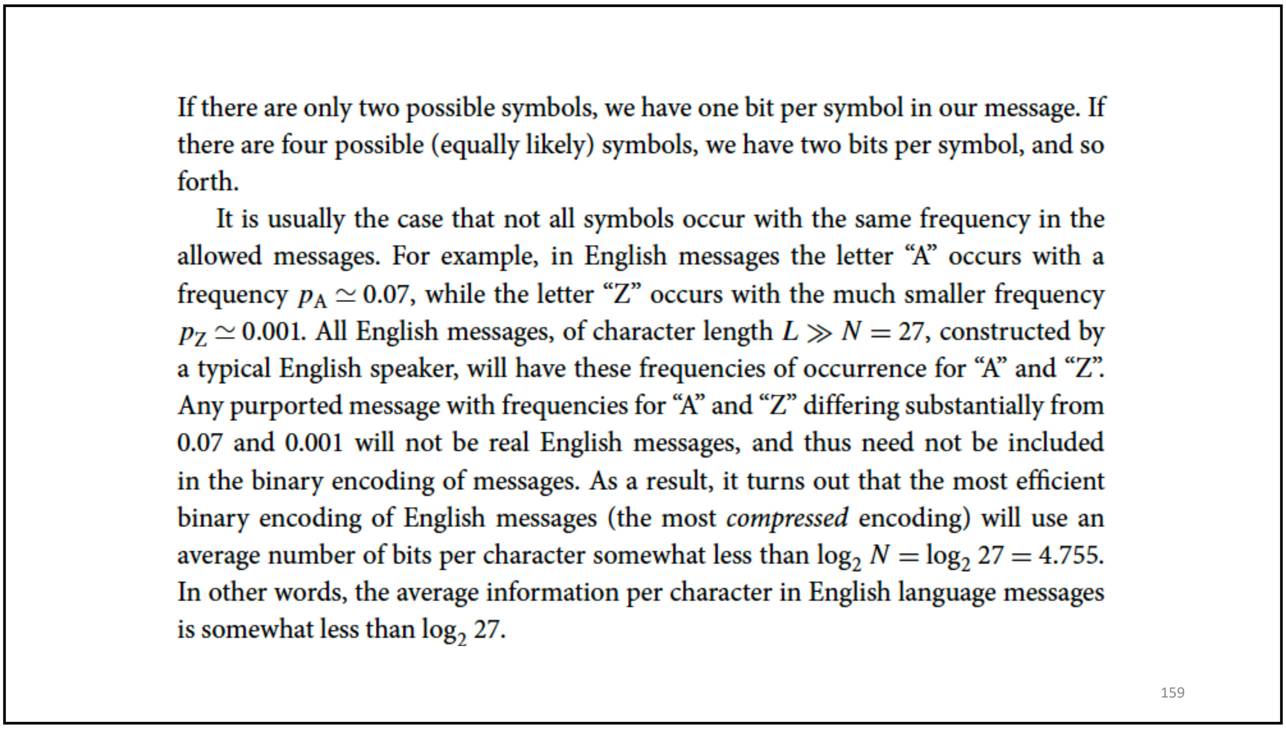
The definition of "the amount of information I gained in a measurement" was formulated by Claude Shannon (1948) in the context of his laying the foundations of *communication theory.* Communication theory deals (among other things) with the problem of how to encode most efficiently a message as a binary string (a string of 0s and 1s) in order to transmit it across a communication channel that transports binary signals. Shannon defined the information in a message as *the number of bits required, in the most compressed such encoding, to distinguish this message from all other messages that might be transmitted.*

Shannon focused on messages that, before encoding, consist of a sequence of symbols. For an English-language message, each symbol might be a single character (a letter A, B, C, . . . , Z or a space; $N = 27$ distinct symbols in all), and a specific message might be the following sequence of length $L = 19$ characters: "I DO NOT UNDERSTAND". Suppose, for simplicity, that in the possible messages, all N distinct symbols appear with equal frequency (this, of course, is not the case for English-language messages), and suppose that the length of some specific message (its number of symbols) is L. Then the number of bits needed to encode this message and distinguish it from all other possible messages of length L is

$$I = \log_2 N^L = L \log_2 N. \quad (4.68a)$$

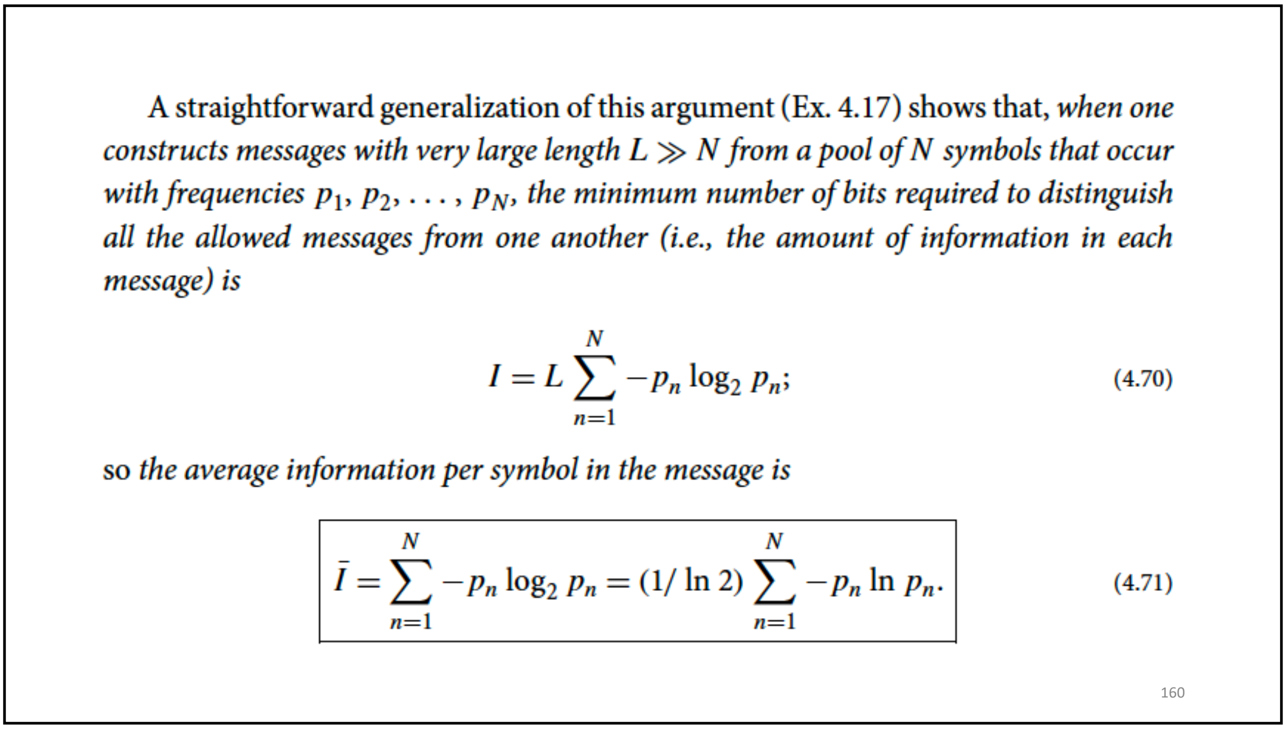
In other words, the average number of bits per symbol (the average amount of information per symbol) is

$$\bar{I} = \log_2 N. \quad (4.68b)$$



If there are only two possible symbols, we have one bit per symbol in our message. If there are four possible (equally likely) symbols, we have two bits per symbol, and so forth.

It is usually the case that not all symbols occur with the same frequency in the allowed messages. For example, in English messages the letter “A” occurs with a frequency $p_A \simeq 0.07$, while the letter “Z” occurs with the much smaller frequency $p_Z \simeq 0.001$. All English messages, of character length $L \gg N = 27$, constructed by a typical English speaker, will have these frequencies of occurrence for “A” and “Z”. Any purported message with frequencies for “A” and “Z” differing substantially from 0.07 and 0.001 will not be real English messages, and thus need not be included in the binary encoding of messages. As a result, it turns out that the most efficient binary encoding of English messages (the most *compressed* encoding) will use an average number of bits per character somewhat less than $\log_2 N = \log_2 27 = 4.755$. In other words, the average information per character in English language messages is somewhat less than $\log_2 27$.

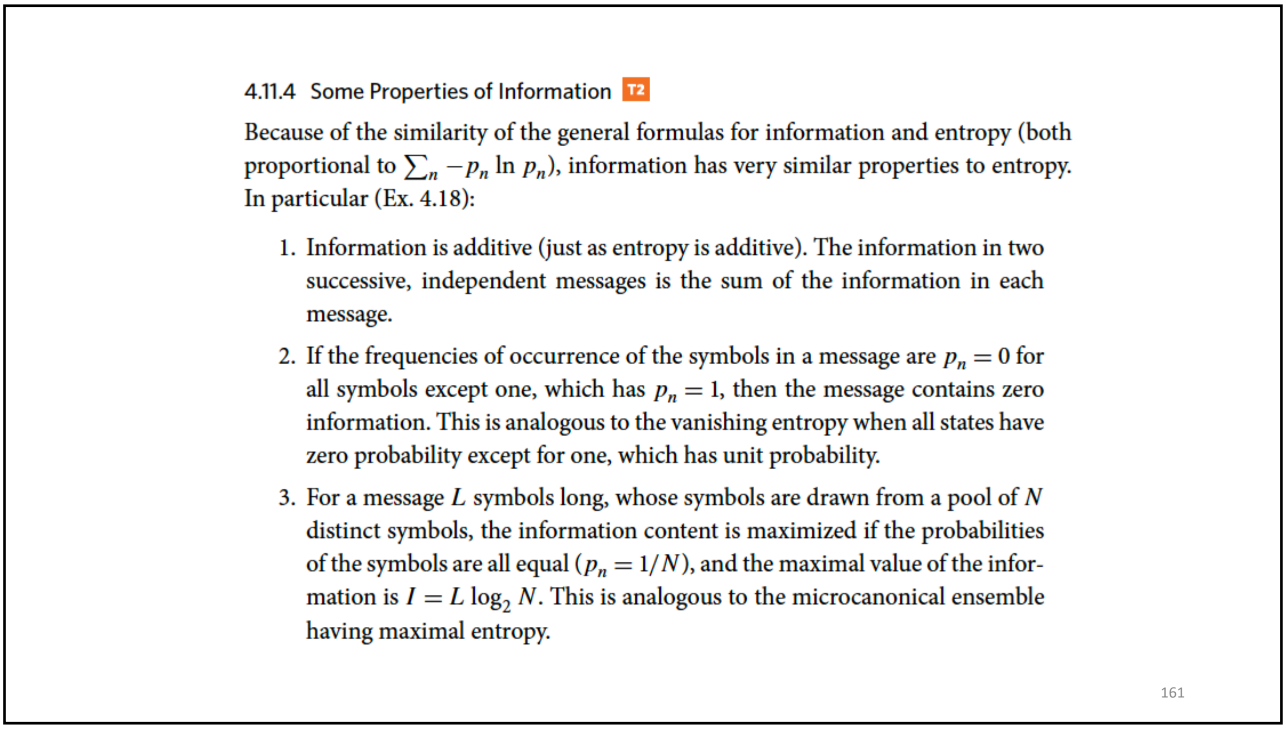


A straightforward generalization of this argument (Ex. 4.17) shows that, *when one constructs messages with very large length $L \gg N$ from a pool of N symbols that occur with frequencies $p_1, p_2, \ldots, p_N$, the minimum number of bits required to distinguish all the allowed messages from one another (i.e., the amount of information in each message) is*

$$I = L \sum_{n=1}^{N} -p_n \log_2 p_n; \tag{4.70}$$

so *the average information per symbol in the message is*

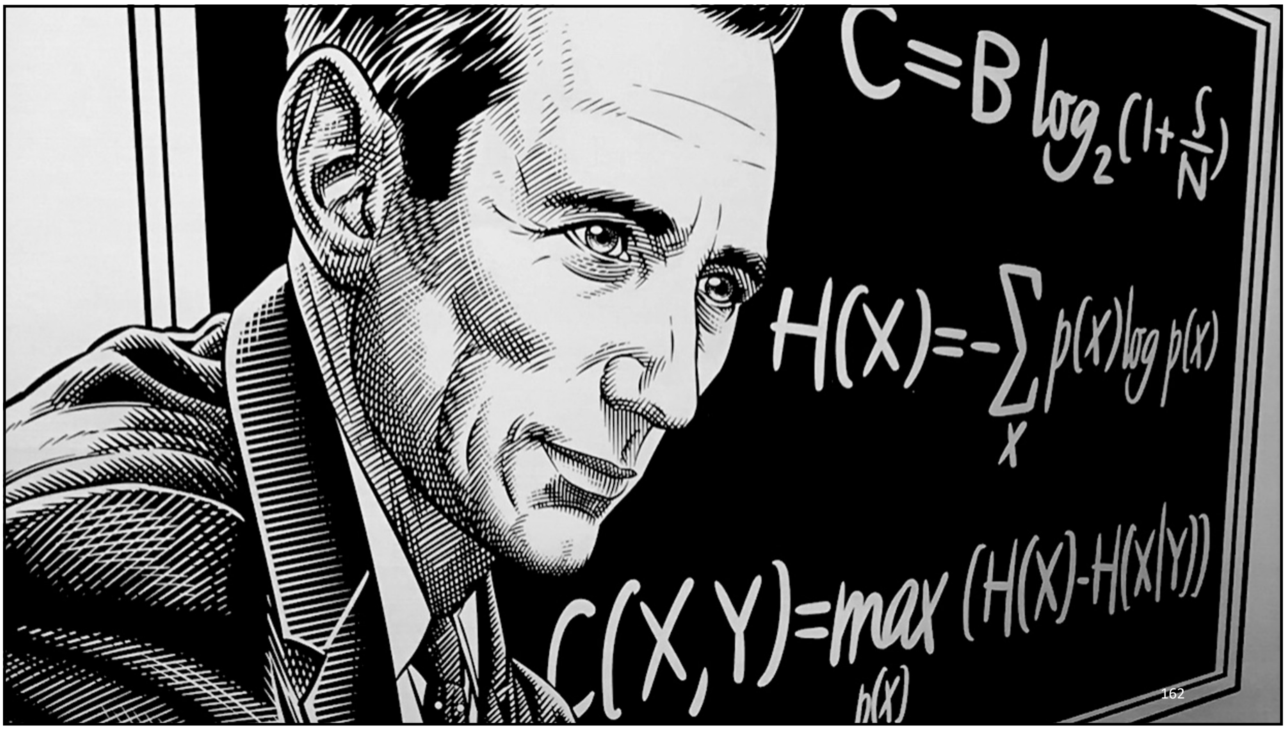
$$\boxed{\bar{I} = \sum_{n=1}^{N} -p_n \log_2 p_n = (1/\ln 2) \sum_{n=1}^{N} -p_n \ln p_n.} \tag{4.71}$$

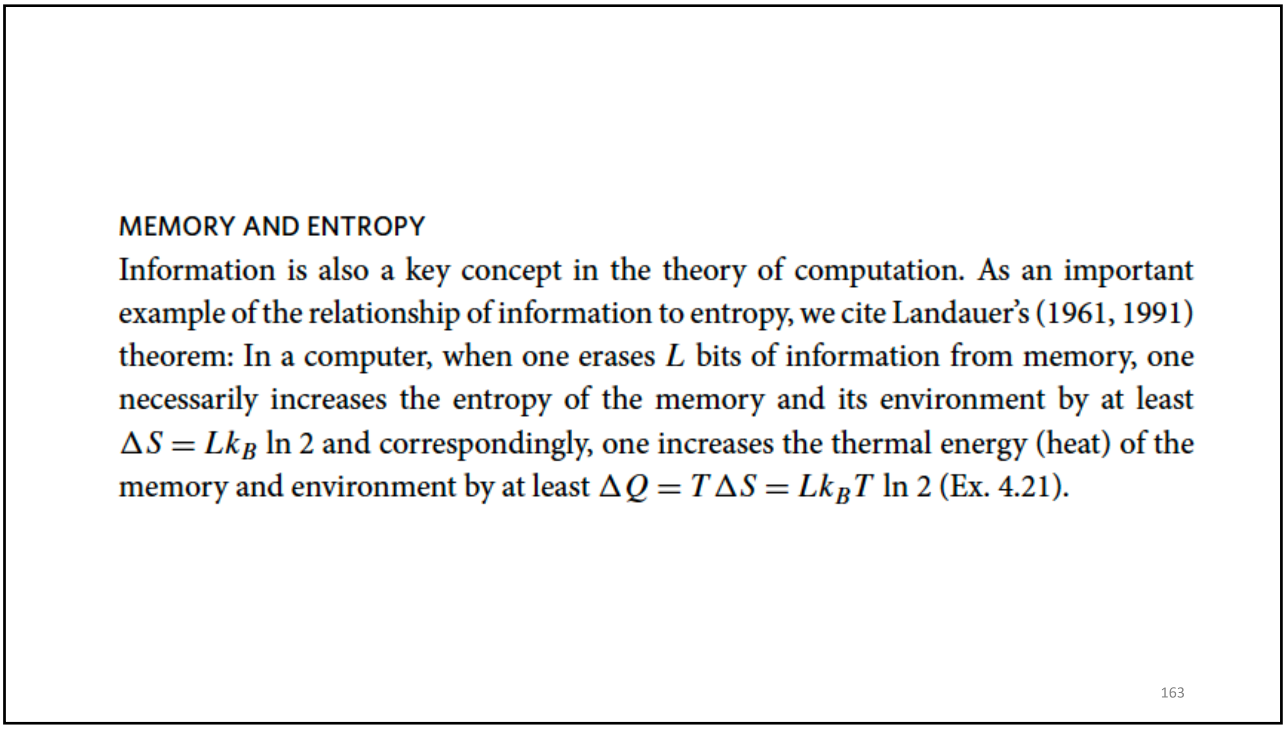


4.11.4 Some Properties of Information T2

Because of the similarity of the general formulas for information and entropy (both proportional to $\sum_n -p_n \ln p_n$), information has very similar properties to entropy. In particular (Ex. 4.18):

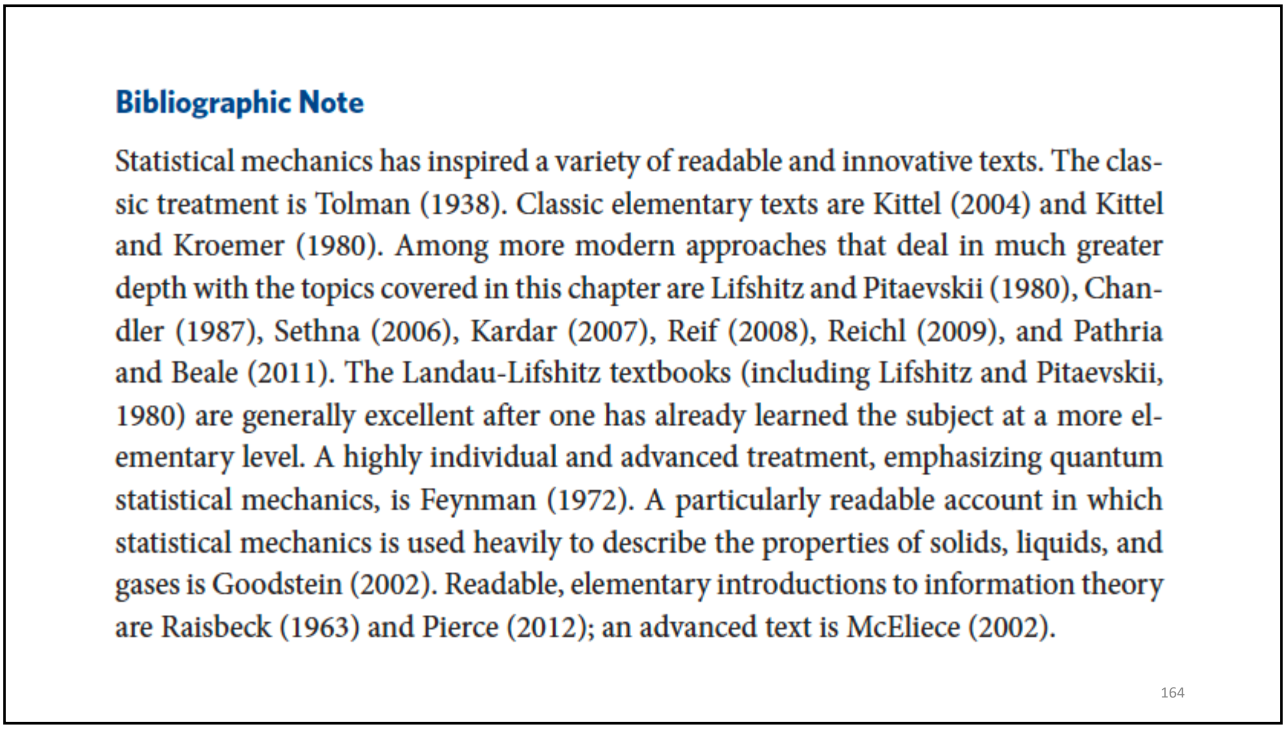
1. Information is additive (just as entropy is additive). The information in two successive, independent messages is the sum of the information in each message.
2. If the frequencies of occurrence of the symbols in a message are $p_n = 0$ for all symbols except one, which has $p_n = 1$, then the message contains zero information. This is analogous to the vanishing entropy when all states have zero probability except for one, which has unit probability.
3. For a message L symbols long, whose symbols are drawn from a pool of N distinct symbols, the information content is maximized if the probabilities of the symbols are all equal ($p_n = 1/N$), and the maximal value of the information is $I = L \log_2 N$. This is analogous to the microcanonical ensemble having maximal entropy.





MEMORY AND ENTROPY

Information is also a key concept in the theory of computation. As an important example of the relationship of information to entropy, we cite Landauer's (1961, 1991) theorem: In a computer, when one erases L bits of information from memory, one necessarily increases the entropy of the memory and its environment by at least $\Delta S = Lk_B \ln 2$ and correspondingly, one increases the thermal energy (heat) of the memory and environment by at least $\Delta Q = T\Delta S = Lk_BT \ln 2$ (Ex. 4.21).



Bibliographic Note

Statistical mechanics has inspired a variety of readable and innovative texts. The classic treatment is Tolman (1938). Classic elementary texts are Kittel (2004) and Kittel and Kroemer (1980). Among more modern approaches that deal in much greater depth with the topics covered in this chapter are Lifshitz and Pitaevskii (1980), Chandler (1987), Sethna (2006), Kardar (2007), Reif (2008), Reichl (2009), and Pathria and Beale (2011). The Landau-Lifshitz textbooks (including Lifshitz and Pitaevskii, 1980) are generally excellent after one has already learned the subject at a more elementary level. A highly individual and advanced treatment, emphasizing quantum statistical mechanics, is Feynman (1972). A particularly readable account in which statistical mechanics is used heavily to describe the properties of solids, liquids, and gases is Goodstein (2002). Readable, elementary introductions to information theory are Raisbeck (1963) and Pierce (2012); an advanced text is McEliece (2002).