

Do 'negative' temperatures exist?

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

1999 J. Phys. A: Math. Gen. 32 4279

(<http://iopscience.iop.org/0305-4470/32/23/307>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.105

The article was downloaded on 02/06/2010 at 07:33

Please note that [terms and conditions apply](#).

## Do ‘negative’ temperatures exist?

B H Lavenda

Università degli Studi, Camerino (MC) I-62032, Italy

E-mail: lavenda@camserv.unicam.it

Received 3 December 1998, in final form 29 March 1999

**Abstract.** A modification of the second law is required for a system with a bounded density of states and not the introduction of a ‘negative’ temperature scale. The ascending and descending branches of the entropy versus energy curve describe particle and hole states, having thermal equations of state that are given by the Fermi and logistic distributions, respectively. Conservation of energy requires isentropic states to be isothermal. The effect of adiabatically reversing the field is entirely mechanical because the only difference between the two states is their energies. The laws of large and small numbers, leading to the normal and Poisson approximations, characterize statistically the states of infinite and zero temperatures, respectively. Since the heat capacity also vanishes in the state of maximum disorder, the third law can be generalized in systems with a bounded density of states: the entropy tends to a constant as the temperature tends to either zero or infinity.

### 1. Introduction

‘Negative’ temperatures are supposed to occur in peculiar systems which exist in a state of incomplete equilibrium for a finite period of time [17, 19]. These systems are considered to violate the general proof that the temperature is positive [7]. In conventional thermodynamic systems the total energy is the sum of the internal energy,  $\bar{E}$ , and the kinetic energies of its internal constituents. And since the system is isolated, the total energy is constant. The entropy,  $S$ , is a function of the internal energy, which is the difference between the total energy and the kinetic energies of the internal constituents. Now, if the inverse temperature,

$$\beta = \left( \frac{\partial S}{\partial \bar{E}} \right)_Y \quad (1)$$

which is measured in energy units where Boltzmann’s constant is unity, and  $Y$  stands for a set of extensive variables held constant during the variation, could be negative then an increase in the entropy for a diminishing value of the internal energy would mean that the internal energy is converted into molecular kinetic energy. Since the end result would be spontaneous break-up of the system, systems with negative temperatures could not exist in a quiescent state of thermal equilibrium.

Suppose that the kinetic energy of molecular motion could be replaced by some sort of potential energy. Paramagnetic substances can be thought of as a collection of nuclear spins embedded in a crystalline lattice. Due to the separation in relaxation times, over which the nuclear spins interact and the spin–lattice interactions take place, one can consider the nuclear spins to be thermally isolated from the lattice for a finite period of time [17, 19]. At zero magnetic field, the spins point in all directions and the state is said to be completely

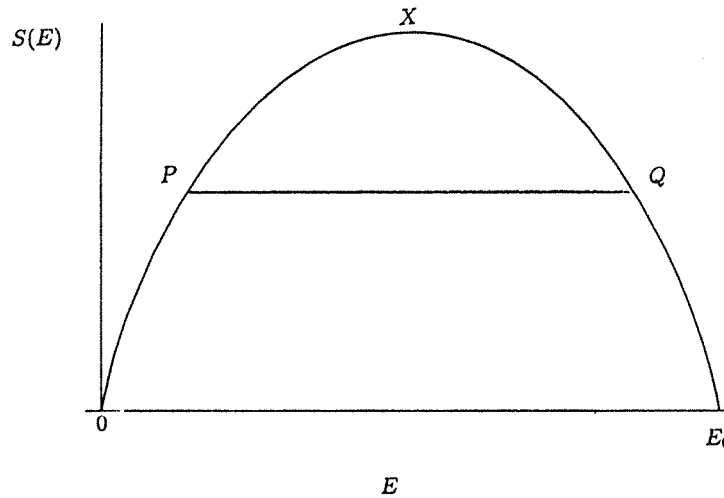
disordered. By applying a magnetic field, the spins tend to follow the field leading to a state of greater order; the spins which are furthest out of line with the field will possess the greatest energy. On reversing the field so quickly that the spins have no time to readjust and follow the field, the system will be brought into a nonequilibrium state which has a higher energy than the state prior to the reversal of the field. On the timescale of the spin–spin interactions we may expect that the spin system alone will achieve a state of metastable thermodynamic equilibrium with a uniform temperature. Removing the field adiabatically will have no effect upon the spin system over the shorter timescale, and over this timescale the spin system is said to manifest ‘negative’ temperatures. Only over the order of the longer timescale, where the spins can exchange energy with the lattice, will heat pass from the spin system at a ‘negative’ temperature and higher energy to the lattice with a positive temperature and lower energy in order that the combined spin–lattice system achieves a true state of thermodynamic equilibrium at a uniform, positive temperature.

Other systems that supposedly manifest ‘negative’ temperatures include laser oscillations in an inverted, amplifying medium [27], communication, information theory and language statistics [13], two-dimensional vortex fluids [14, 16], and a guiding-centre plasma [6]. Prior to the reversal of the field in paramagnetic materials, the higher the energy the less likely is it for the state to be populated. Upon reversal of the field, the decaying Boltzmann factor has become an exponentially increasing one, corresponding to what is formally a ‘negative’ temperature, so that the higher the energy the greater the probability of finding a state occupied by the system. It is for this reason that it is said that a state of ‘negative’ temperature is hotter than a state of positive temperature [20].

In the absence of the kinetic energy of the internal motions, the densities of states may not be unbounded so that negative temperatures do not necessarily imply the lack of convergence of thermodynamic quantities like the generating function. In the case where there are only a finite number of energy levels, themselves finite, the densities of states will, in fact, be bounded from above [7, p 213]. However, a bounded density of states poses problems in statistical mechanics because of the presence of a state of maximum entropy at a finite energy; consequently, the entropy is not a monotonically increasing function of the internal energy. The entropy versus energy curve will have the bell-shaped form shown in figure 1. States of ‘negative’ temperatures are associated with states lying along the descending branch of the entropy curve. Since the derivative (1) is no longer positive over the entire range of energies there is no longer an equivalence between the entropy maximum principle and the minimum energy principle [2].

The maximum  $X$  on the entropy curve in figure 1 corresponds to the state of maximum disorder, and has zero inverse temperature. The inverse temperature acts like an ‘order’ parameter: the state of complete disorder would require an enormous change in the internal energy to cause even a minute change in the entropy. As the system becomes more ordered, the order parameter,  $\beta$ , increases. If we consider a set of spins in a magnetic field, a sudden reversal of the field is sufficient to cause the system to jump from a state of positive temperature,  $P$ , to a state of ‘negative’ temperature,  $Q$ , at the same value of the entropy, as shown in figure 1 [3]. The state  $Q$  is said to be hotter than the state  $P$  because the spins which had the least energy in state  $P$  now have the greatest energy in  $Q$  and vice versa. Because they follow a Boltzmann distribution in state  $P$ , more states of lower energy are populated than of higher energy so that the population inversion in state  $Q$  really makes it quite hot [25]. One is presupposing that a state of higher energy means that it is hotter.

A single experiment is no longer sufficient to determine whether the entropy will decrease or increase with the internal energy. The lack of monotonicity of the entropy is disturbing from a thermodynamic point of view. Since the dividing line is the state of maximum disorder



**Figure 1.** Entropy as a function of the internal energy for Fermi–Dirac statistics. The bending over of the entropy curve is a hallmark of systems with a bound density of states. The entropy vanishes at 0 where  $T = 0$  and in the state of maximum energy  $E_0$  where again  $T = 0$ . The maximum state of disorder  $X$  corresponds to an infinite temperature. The ascending part of the curve from 0 to  $X$  corresponds to the usual Fermi statistics, while the descending curve has the same thermal equation of state except that  $-T$  appears in place  $T$ , but with  $T > 0$ . The states  $P$  and  $Q$  lie along an isothermal and isoentropic line.

at infinite temperature, how can the temperature increase beyond  $T = \infty$ ? Because the state of maximum disorder at  $T = \infty$  is thermodynamically indistinguishable with the state at  $T = -\infty$  [7, p 213]. Along the descending branch of the entropy curve, the temperature is increasing until it reaches the state of greatest energy at  $T = -0$ . Hence, it was advocated that two infinite ranges of temperatures are required [18].

The crux of the matter is whether magnetic energy can be considered as a form of heat energy. Since potential energy is not the heat that can be felt, it raises doubts as to whether it can be substituted for the uncontrollable energy buried in the kinetic energy of the molecules. Electric and magnetic fields can be added to the list of independent variables,  $Y$ ; they are, however, only 'pseudo-thermodynamic' quantities since no conservation law applies to them [24]. The inverse temperature (1) would then be defined at constant field leaving the internal energy to vary independently of the field. Indeed, the kinetic energy of molecular motion is not on par with magnetic energy; it does not appear in the set  $Y$ , and it is coupled to variations in the internal energy through energy conservation. Work processes involving pseudo-thermodynamic variables cannot be thought of the exchange of conserved quantities, so a decrease in the internal energy resulting in an increase in the entropy does not necessarily mean an increase in the magnetic field.

In the case of paramagnetic materials, variations in an applied magnetic field that are so quick as to leave the entropy of the magnetic system unchanged can cause variations in the temperature. An increase in the field produces an immediate increase in the temperature while decreases in the field produce corresponding decreases in the temperature [2, p 255]. Under adiabatic conditions, the change in the temperature  $dT$  due to a change in the applied field  $dH$  is given by

$$dT = \left( \frac{\partial T}{\partial H} \right)_S dH. \quad (2)$$

By continually decreasing the field, quasi-statically and adiabatically, extremely low temperatures can be reached. However, this necessarily implies that the entropy is an independent variable so that we are working in the energy representation. Recalling that the condition for transforming from the entropy to the energy representation is that (1) be positive, the above argumentation will not be valid for ‘negative’ temperatures since (1) will be negative over a finite range of energies. Any physical model which is used to investigate ‘negative’ temperatures must reflect the fact that the two definitions of heat capacity,  $T(\partial S/\partial T)_Y$  and  $(\partial \bar{E}/\partial T)_Y$ , will not coincide when (1) is negative.

The aim of this paper is to show that ‘negative’ temperatures do not exist. Rather, the second law must be changed to

$$-\beta = \left( \frac{\partial S}{\partial \bar{E}} \right)_Y < 0 \quad (3)$$

along the descending part of the entropy curve so that the temperature will still be positive. Specifically, we return to Ramsey’s model for ‘negative’ temperatures in section 2. In section 3, we compare the conventional Young inequality, that utilizes equations of state of the form (3), with its thermodynamic counterpart that employs hyperbolic equations of state (1). Since (3) applies along the descending branch of the energy curve, a thermodynamic ‘symmetry breaking’ will occur between the maximum entropy and minimum entropy representations. This will be discussed in section 4. In section 5, comparison will be made between the thermal distribution found for states lying along the descending branch of the entropy curve and positrons in relativistic hole theory. We present our conclusions in section 6.

## 2. A question of zero-point energy

In this section we return to the model used by Ramsey [20] to discuss ‘negative’ temperatures and show that different definitions of the conjugate variables of energy and inverse temperature result simply by choosing different zero-point energies.

In his discussion of ‘negative’ temperatures, Ramsey [20] considered a set of  $\eta$  equally spaced energy levels  $\varepsilon$ . Instead of following Ramsey by selecting the zero of energy midway between, we make allowance for different zero-point energy values. The single particle generating function is

$$Z_1(\beta; \eta) = \sum_{j=0}^{j=\eta} e^{-(j+\alpha)\beta\varepsilon} = e^{-(\eta-1+2\alpha)\beta\varepsilon/2} \frac{\sinh(\eta\beta\varepsilon/2)}{\sinh(\beta\varepsilon/2)} \quad (4)$$

where the parameter  $\alpha$  is related to a ‘zero-point’ energy. Ramsey studied a four level system. Actually, only two cases need be considered:  $\eta = 2$  and  $\eta = \infty$ , corresponding to Fermi–Dirac and Bose–Einstein statistics, respectively [8]. It is only in these two cases that the inverse temperature as a function of the energy can be expressed in closed form using, in the former case, the double-angle formula of hyperbolic sine. Since we want a system with a finite number of energy levels, we will deal here with the case  $\eta = 2$ . The logarithm of the generating function for a population of size  $r$  is

$$\ln Z_r(\beta) = r \left\{ \ln [2 \cosh(\frac{1}{2}\beta\varepsilon)] - \frac{1}{2}(1 + 2\alpha)\beta\varepsilon \right\}. \quad (5)$$

We now turn our attention to expressions for the average energy that are obtained by varying the value of the zero-point energy. For  $\alpha = 0$  we obtain

$$\frac{\partial}{\partial \beta} \ln Z_r = \frac{1}{2} E_0 (\tanh(\frac{1}{2}\beta\varepsilon) - 1) \quad (6)$$

where  $E_0 = r\varepsilon$  is the greatest attainable value of the energy. In order to achieve thermal equilibrium, we have to adopt the thermodynamic convention of defining the average energy as

$$\bar{E}_p = -\frac{\partial}{\partial\beta} \ln Z_r. \quad (7)$$

We then obtain the Fermi distribution,

$$\bar{E}_p = \frac{1}{2}E_0(1 - \tanh(\frac{1}{2}\beta\varepsilon)) = \frac{E_0}{e^{\beta\varepsilon} + 1} = \bar{n}_p\varepsilon \quad (8)$$

where  $\bar{n}_p$  is the average 'particle' occupation number. However, this is not the only possible choice of the parameter,  $\alpha$ .

For  $\alpha = -1$  there results

$$\frac{\partial}{\partial\beta} \ln Z_r = \frac{1}{2}E_0(\tanh(\frac{1}{2}\beta\varepsilon) + 1). \quad (9)$$

Now, in order to achieve a thermal equilibrium distribution we must define the average energy as:

$$\bar{E}_h = \frac{\partial}{\partial\beta} \ln Z_r. \quad (10)$$

In doing so, we find that the thermal distribution is the logistic distribution:

$$\bar{E}_h = \frac{1}{2}E_0(1 + \tanh(\frac{1}{2}\beta\varepsilon)) = \frac{E_0}{1 + e^{-\beta\varepsilon}} = \bar{n}_h\varepsilon \quad (11)$$

where  $\bar{n}_h$  is the average 'hole' occupation number. An analogy with semiconductor physics will help bring out the distinction between 'particles' and 'holes'.

In semiconductor physics, the Fermi distribution represents the thermal distribution of electrons, while the logistic distribution describes hole statistics [22]. For intrinsic semiconductors, the statistics of holes in the valence band matches that of electrons in the conduction band except for three notable differences. A decrease in the electron energy will cause an increase in the energy of the hole, particles and holes have opposite charges, and there is a maximum energy cut-off to the thermal hole distribution. Hole theory was also introduced by Dirac in his interpretation of negative energy states of the relativistic wave equation [5] (cf section 5)†.

Therefore, the phenomenon we are analysing is much more general than the reversal of a magnetic field in a spin system. It applies to all systems that obey Fermi statistics. The exclusion principle, which is the basis of the Fermi distribution, implicitly implies the presence of an energy gap separating particles from their antiparticles, electron from holes, and so on. The transitions between the two can be radiative, electronic, or magnetic. As a consequence of the energy gap, we should expect a threshold energy which is achieved by pumping in lasing materials where the rate of absorption is balanced by the rate of stimulated emission, the transition between thermal radiation and non-thermal radiation that is characteristic of luminescent materials [26], the energy separating the transition between the clustering of vortices of opposite sign and vortices of the same sign [14], and the energy of the state of maximum disorder in a paramagnetic material. The behaviour of the particles and their complements is mutually exclusive: an increase in energy of one implies a decrease in energy of the other, as will now be shown. We will continue to refer to paramagnetic materials, while realizing that the phenomenon is more widespread.

† What we will refer to as 'holes' can also be called antiparticles with positive energy, or antifermions [28]. The difference between antiparticles and particles, apart from their charge, is that their thermal distributions are given by (11) and (8), respectively.

If the energy  $\bar{E}_p$  of (8) corresponds to state  $P$  on the entropy curve in figure 1, then  $\bar{E}_h$  of (11) is the energy of state  $Q$  lying on the adiabat joining them. Their sum is the statement of energy conservation:

$$\bar{E}_p(\beta) + \bar{E}_h(\beta) = E_0. \quad (12)$$

Since the occupation numbers are the independent variables in quantum statistics [9], (12) necessarily coincides with the conservation of the number of particles and holes:

$$\bar{n}_p(\beta) + \bar{n}_h(\beta) = r. \quad (13)$$

It is just that the inverse temperature is defined differently in the two cases. Expression (13) is a statement of mutual exclusivity: the probability,  $\bar{n}_h/r$ , of an electron not occupying a state at energy  $\varepsilon$ , is one minus the probability that it is,  $(1 - \bar{n}_p/r)$ . Energy, or particle and hole, conservation implies that the temperature of states  $P$  and  $Q$  are the same. States with the same temperature have the same degree of disorder. So the two states  $P$  and  $Q$  must have the same entropy, which is what we originally used to connect the two states. The sole physical difference between the two states is that their energies are different; the difference being due to an applied external field.

According to the thermodynamic convention of defining the energy, (7), the Legendre conjugate of  $-\ln Z_r(\beta)$  is

$$S(\bar{E}) = \ln Z_r(\beta) - \beta \frac{\partial}{\partial \beta} \ln Z_r(\beta) = \ln Z_r(\beta) + \beta \bar{E}. \quad (14)$$

In the next section, we will appreciate that this definition ensures the existence of the canonical distribution in the case of an unbounded density of states (cf equation (25) below). Definition (10), in contrast, gives the dual to  $-\ln Z_r(\beta)$  as

$$S(\bar{E}) = \ln Z_r(\beta) - \beta \bar{E}. \quad (15)$$

Consequently, whereas (14) uses the conventional form of the second law, (1), (15) employs the modified form given by (3). In the next section, we show that this definition of the inverse temperature is valid over a range of energy values in systems possessing a bounded density of states (cf equation (29) below).

Solving (8) and (11) for their inverse functions result in

$$\beta = \pm \frac{1}{\varepsilon} \ln \left( \frac{E_0 - \bar{E}}{\bar{E}} \right) \quad (16)$$

respectively. If  $\beta$  is to be positive, then  $\bar{E} < E_0/2$  when the plus sign is chosen, while  $\bar{E} > E_0/2$  when the negative sign is selected. These correspond to the ascending and descending branches of the entropy curve, respectively. For no matter what sign is chosen in (16) their respective definitions, (1) or (3) in terms of the entropy derivative, both give

$$S(\bar{E}) = \text{const.} - (r - \bar{E}/\varepsilon) \ln(r - \bar{E}/\varepsilon) - (\bar{E}/\varepsilon) \ln(\bar{E}/\varepsilon) \quad (17)$$

upon integration. The constant of integration is determined by comparing (17) with the logarithm of the binomial coefficient in the case where the numbers  $r$  and  $\bar{E}/\varepsilon$  are large enough to warrant Stirling's approximation for their factorials. This determines the constant to be  $r \ln r$ . Expression (17) has the form of an entropy of 'mixing', having the shape an inverted parabola in figure 1 with a maximum energy  $E_0 = r\varepsilon$ .

Unlike the entropy and temperature, which have uniquely defined zeros, only differences in the internal energy are measurable. Thus, for the same entropy (17) and temperature (16) we

can have different expressions for the energy. Although these differ by an integral zero-point energy, the particle entropy:

$$\begin{aligned} S &= r \{ \ln[2 \cosh(\frac{1}{2}\beta\varepsilon)] - \frac{1}{2}\beta\varepsilon \tanh(\frac{1}{2}\beta\varepsilon) \} \\ &= r \ln(1 + e^{-\beta\varepsilon}) + \frac{1}{2}\beta E_0 [1 - \tanh(\frac{1}{2}\beta\varepsilon)] \\ &= \ln Z_r(\beta(\bar{E}_p)) + \beta(\bar{E}_p)\bar{E}_p \end{aligned}$$

and the hole entropy:

$$\begin{aligned} S &= r \{ \ln[2 \cosh(\frac{1}{2}\beta\varepsilon)] - \frac{1}{2}\beta\varepsilon \tanh(\frac{1}{2}\beta\varepsilon) \} \\ &= r \ln(1 + e^{\beta\varepsilon}) - \frac{1}{2}\beta E_0 [1 + \tanh(\frac{1}{2}\beta\varepsilon)] \\ &= \ln Z_r(\beta(\bar{E}_h)) - \beta(\bar{E}_h)\bar{E}_h \end{aligned}$$

are the same. The average energies  $\bar{E}_p$  and  $\bar{E}_h$  are given by (8) and (11), respectively.

The intermediate case where  $\alpha = -\frac{1}{2}$  shifts the entropy curve by an amount  $E_0/2$  to the left on the energy axis. The state of maximum disorder now corresponds to  $\bar{E} = 0$ . States lying on the ascending part of the entropy curve have negative energies, while those lying along the descending part have positive energies. In other words, the energy can be positive or negative whereas the temperature must always be positive. The derivative of the logarithm of the generating function,

$$\ln Z_r = r \ln[2 \cosh(\frac{1}{2}\beta\varepsilon)]$$

gives the inverse functions

$$\beta = \pm \frac{1}{\varepsilon} \ln \left( \frac{E_0 + 2\bar{E}}{E_0 - 2\bar{E}} \right) \tag{18}$$

depending on whether (10) or (7) is used, respectively. Since hole states correspond to energies  $\bar{E} > 0$ , while particle states to energies  $\bar{E} < 0$ , where  $|\bar{E}| \leq E_0/2$ , hole states lie above those of particle states. In both semiconductor physics and relativistic hole theory, in contrast, hole states lie below those of particles states, and in the latter theory hole states have negative energies. As we have seen, there is nothing less physical about negative energies than positive ones since it all depends on selecting the zero point energy. Only energy differences count, but they should be such that the temperature (18) always remains positive semi-definite.

Integration of either relation in (18), according to the definition of the second law as (3) or (1), coincides in the unique expression

$$S = r \ln r - \frac{1}{2}(r + 2\bar{E}/\varepsilon) \ln[\frac{1}{2}(r + 2\bar{E}/\varepsilon)] - \frac{1}{2}(r - 2\bar{E}/\varepsilon) \ln[\frac{1}{2}(r - 2\bar{E}/\varepsilon)] \tag{19}$$

for the entropy, where the constant of integration has been set equal to  $r \ln(r/2)$ . The entropy (19) is that of a one-dimensional symmetric random walk, where  $r$  represents the total number of jumps and  $2\bar{E}/\varepsilon$  is the excess number of positive jumps. The number of ways of getting  $2\bar{E}/\varepsilon$  excess positive jumps from a total of  $r$  jumps is  $\binom{r}{\frac{1}{2}(r+2\bar{E}/\varepsilon)}^\dagger$ . This model has been used as a combinatorial definition of the entropy reduction [10]; the reduction in entropy as measured from the state of maximum disorder with the greatest entropy. In the limit  $|\bar{E}| \ll E_0/2$ , the entropy (19) reduces to the characteristic quadratic form:

$$S(n) - r \ln 2 = -\frac{1}{2} \ln(\frac{1}{2}\pi r) - 2n^2/r \tag{20}$$

where we have set  $n = E/\varepsilon$ , and whose average value is  $\bar{n} = \bar{E}/\varepsilon$ . For equal *a priori* probabilities of positive and negative jumps, the probability distribution,

$$P(n) = \binom{r}{\frac{1}{2}(r+2n)} \left(\frac{1}{2}\right)^r = \exp(S(n) - S_0) = \sqrt{\frac{2}{\pi r}} e^{-2n^2/r} \tag{21}$$

$\dagger$  If  $\frac{1}{2}(r+2\bar{E}/\varepsilon)$  is not an integer, the binomial coefficient is interpreted as zero.



tends to the normal distribution in the limiting case of large  $r$ . In other words, the quadratic approximation of the entropy reduction, (20), determines the normal form of the probability distribution *exactly* when Stirling's formula,  $x! \sim \sqrt{2\pi x} x^x e^{-x}$  is used for the factorials. The negative of the logarithm of the *a priori* probability,  $(\frac{1}{2})^r$ , is the maximum entropy,  $S_0 = r \ln 2$ , and the entropy reduction  $S(n) - S_0$  determines the probability distribution according to the generalized form of Boltzmann's principle for extreme value distributions [10]. It is essential to bear in mind that an arbitrary constant does not enter into the expression relating the entropy to the probability distribution, which is not the 'thermodynamic' probability, which is an extremely large number, but, rather, a normalized probability distribution. The absolute value of the constant is related to a generalization of the third law, which will be discussed in section 4.

### 3. The thermodynamic Young inequality

A geometrical inequality known as Young's inequality deals with the sums of the integrals of inverse functions which correspond to thermodynamic conjugate variables, like energy and inverse temperature. This will allow us to obtain inequalities on the sum of the Legendre duals, the entropy and the logarithm of the generating function. Furthermore, Young's inequality and its thermodynamic generalization will permit us to discriminate between the two forms of the second law, (1) and (3), in terms of the type of exponential probability distributions that they give rise to.

A generalization of Young's theorem states [15] that if  $f(x)$  is a continuous and *increasing* function on the interval containing the points  $x = 0$  and  $x = a$  (where  $a > 0$  or  $a < 0$ ) such that  $f(0) = 0$  and  $f(x) \rightarrow \infty$  as  $x \rightarrow \infty$ , then its inverse  $f^{-1}$  exists and has the same properties as  $f$ . If  $b$  belongs to the domain of  $f^{-1}$  then the inequality holds

$$ab \leq \int_0^a f(x) dx + \int_0^b f^{-1}(y) dy$$

with equality iff  $b = f(a)$ . Young's inequality can be given a simple geometric proof which can be found in any text on convex analysis.

As it stands, Young's inequality, or its generalization, cannot be applied to thermodynamics. Equations of state relating conjugate thermodynamic variables are hyperbolic so that they fail to satisfy the condition  $f(0) = 0$ . Fortunately, this condition can be relaxed so that we may consider  $f$  on an interval containing the points  $x = c$  and  $x = q$ . Young's generalized inequality is [15, p 243]

$$qs - cf(c) \leq \int_c^q f(x) dx + \int_{f(c)}^s f^{-1}(y) dy \quad (22)$$

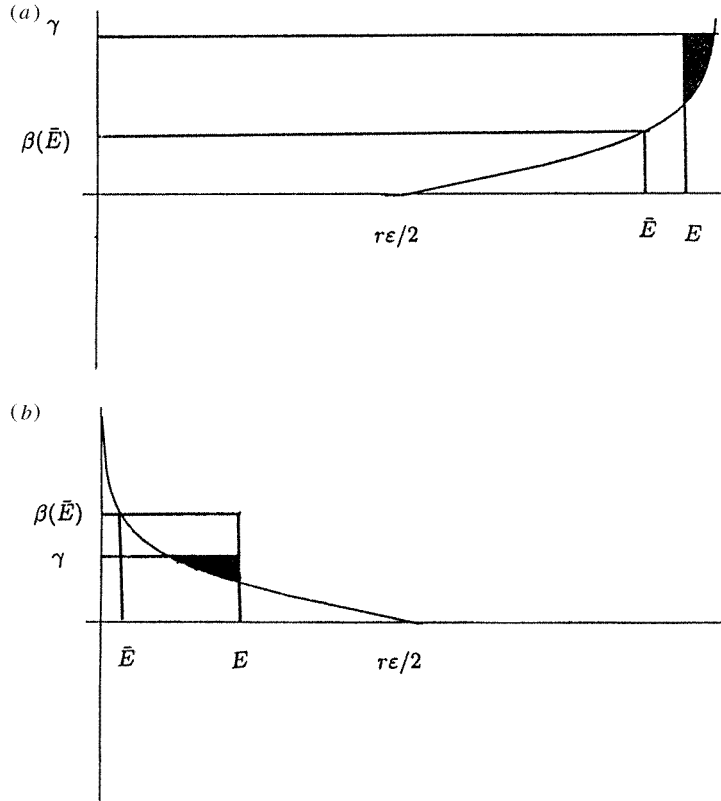
provided  $s$  belongs to the domain of  $f^{-1}$ , and the equality holds iff  $s = f(q)$ .

In conventional thermodynamics, there is the additional problem that the thermodynamic state function,  $f$ , is a strictly *decreasing* function of its conjugate variable. Only in the case of a bounded density of states can  $f$  be an increasing function over a limited range of values of its conjugate variable. In that case, Young's inequality is given in its usual form

$$\gamma E - \beta(\bar{E})\bar{E} \leq \int_{\bar{E}}^E \beta(x) dx + \int_{\beta(\bar{E})}^{\gamma} \beta^{-1}(y) dy \quad (23)$$

with equality when  $\gamma = \beta(E)$ . A graphic interpretation of inequality (23) is given in figure 2(a). In the case of an unbounded density of states  $f$  is continuous and decreasing so that the inequality in (23) has to be reversed, as shown graphically in figure 2(b).

The reverse of inequality (23) is applicable to all thermodynamic systems with an unbounded density of states,  $e^{S(E)}$ . The reason is that a rapidly increasing density of states



**Figure 2.** (a) A graphical interpretation of the generalized Young inequality for a bounded density of states. (b) A graphical interpretation of the generalized Young inequality, with the inequality reversed, for an unbounded density of states.

must be overpowered by an even more rapidly decreasing exponential factor,  $e^{-\gamma E}$  with  $\gamma > 0$ , in order to yield a finite integral,  $Z_r(\gamma) = \int_0^\infty e^{-\gamma E} e^{S(E)} dE$ . This can be demonstrated by rewriting inequality (23) as [12]

$$\begin{aligned} \gamma E - \beta(\bar{E})\bar{E} &\geq \int_E^{\bar{E}} \frac{\partial S}{\partial x} dx + \int_\gamma^{\beta(\bar{E})} \frac{\partial}{\partial y} \ln Z_r dy \\ &= S(E) - S(\bar{E}) + \ln Z_r(\beta(\bar{E})) - \ln Z_r(\gamma). \end{aligned} \tag{24}$$

Since  $\ln Z_r(\beta(\bar{E}))$  and  $S(\bar{E})$  are connected by the Legendre transform, (14), inequality (24) reduces to  $S(E) - \ln Z_r(\gamma) \leq \gamma E$ , or equivalently to

$$P_p(E; \gamma) = e^{S(E)} \frac{e^{-\gamma E}}{Z_r(\gamma)} \leq 1 \tag{25}$$

where the equality holds when  $\gamma = \beta(\bar{E})$ . This is the physical content contained in the reverse of inequality (23), and (25) is none other than the canonical distribution, which belongs to the exponential family of distributions, parameterized by  $\gamma$  which is an estimate of the inverse temperature [9, chapter 4]. The average energy is the most probable value of the energy, and this overwhelmingly maximizes the probability distribution (25).

As an illustration of the thermodynamic Young inequality, (24), consider the inverse of the particle energy (7),

$$\beta = \frac{2}{\varepsilon} \tanh^{-1} \left( 1 - \frac{2\bar{E}_p}{E_0} \right) \quad (26)$$

on the interval  $[0, E_0/2]$ . Expression (26) is clearly a *decreasing* function of its argument, and inequality (24) reads

$$\begin{aligned} \gamma E - \beta(\bar{E}_p)\bar{E}_p &\geq \frac{1}{\varepsilon} \int_{\bar{E}_p}^E \ln \left( \frac{E_0 - x}{x} \right) dx - \frac{1}{2} E_0 \int_{\gamma}^{\beta(\bar{E}_p)} \left( 1 - \tanh \left( \frac{1}{2} \varepsilon y \right) \right) dy \\ &= S(E) - S(\bar{E}_p) + \ln Z_r(\beta(\bar{E}_p)) - \ln Z_r(\gamma). \end{aligned}$$

On the strength of the Legendre transform (14), the inequality reduces to (25).

However, the density of states,  $\binom{r}{n}$ , is bounded and, consequently, we cannot consider all values of the energy on the positive half axis, but only those up to the maximum energy,  $E_0 = r\varepsilon$ . According to Newton's binomial theorem, the generating function is

$$Z_r(\gamma) = \sum_{n=0}^r e^{-\gamma \varepsilon n} \binom{r}{n} = (1 + e^{-\gamma \varepsilon})^r \quad (27)$$

so it would seem that it really makes no difference whether  $\gamma$  is positive or negative. Nevertheless, we have to limit the interval to  $[0, E_0/2]$  in order to apply the thermodynamic Young inequality (24). Over this interval, the density of states is an increasing function of the average energy, (26) is positive, and if  $\gamma$  is to be a consistent estimate of  $\beta$ , it too has to be positive.

As an illustration of the conventional form of Young's inequality, (23), consider the inverse of the hole energy (10),

$$\beta = \frac{2}{\varepsilon} \tanh^{-1} \left( \frac{2\bar{E}_h}{E_0} - 1 \right) \quad (28)$$

on the interval  $[E_0/2, E_0]$ . Since (28) is an *increasing* function the conventional Young inequality (23) reads

$$\begin{aligned} \gamma E - \beta(\bar{E}_h)\bar{E}_h &\leq \int_{\bar{E}_h}^E \ln \left( \frac{x}{E_0 - x} \right) dx - \frac{1}{2} E_0 \int_{\gamma}^{\beta(\bar{E}_h)} \left( 1 + \tanh \left( \frac{1}{2} \varepsilon y \right) \right) dy \\ &= -S(E) + S(\bar{E}_h) - \ln Z_r(\beta(\bar{E}_h)) + \ln Z_r(\gamma). \end{aligned}$$

The inequality is graphically shown in figure 2(a). The Legendre transform (15) reduces the inequality to

$$P_h(E; \gamma) = e^{S(E)} \frac{e^{\gamma E}}{Z_r(\gamma)} \leq 1. \quad (29)$$

What fixes the sign of the parameter in (29) is that on the half-interval  $[E_0/2, E_0]$ , the entropy, and consequently the density of states, is a decreasing function of the average energy, while the logarithm of the generating function is a monotonically increasing function of its argument. Consequently, any estimate of the true value of the inverse temperature (28) must itself be positive. The logarithm of the generating function,

$$Z_r(\gamma) = \sum_{n=0}^r e^{\gamma \varepsilon n} \binom{r}{n} = (1 + e^{\gamma \varepsilon})^r \quad (30)$$

differs from the logarithm of (27) by a positive, constant term  $\gamma E_0$ . In both cases, (25) and (29), the entropy can be determined to within a constant, and it is given by (17). Either definition

(1) or (3) of the inverse temperature can be used in (17). Whereas the entropy is continuous over the entire range of energy values, two thermal equations of state are required, one for particles and the other for holes.

The canonical distributions, (25) and (29), both correspond to binomial distributions. The distinction between the two lies in the expressions for the *a priori* probabilities,  $p$  and  $q$ , where  $p + q = 1$ , and not in the binomial coefficient. This is what leaves the entropy (17) invariant. Casting (25) as a binomial distribution,

$$P_p(E; \gamma) = e^{S(E)} \frac{e^{-\gamma E}}{Z_r(\gamma)} = \binom{r}{E/\varepsilon} p^{E/\varepsilon} q^{r-E/\varepsilon} \quad (31)$$

identifies  $q = (1 + e^{-\gamma\varepsilon})^{-1}$  and  $p = (1 + e^{\gamma\varepsilon})^{-1}$ . In contrast, for (29) we obtain

$$P_h(E; \gamma) = e^{S(E)} \frac{e^{\gamma E}}{Z_r(\gamma)} = \binom{r}{E/\varepsilon} p^{E/\varepsilon} q^{r-E/\varepsilon} \quad (32)$$

showing that the *a priori* probabilities have been interchanged:  $q = (1 + e^{\gamma\varepsilon})^{-1}$  and  $p = (1 + e^{-\gamma\varepsilon})^{-1}$ . Although the two binomial distributions have the same entropy, the fact that the *a priori* probabilities have been interchanged means that two thermal equations of state are required which are mirror images of one another upon temperature reversal, just like the pair of *a priori* probabilities.

In the case of a symmetrical random walk, where the energy (jump) can be positive or negative, the probability distribution is

$$P(E; \gamma) = e^{S(E)} \frac{e^{\pm\gamma E}}{Z_r(\gamma)} = \binom{r}{\frac{1}{2}(r+2E/\varepsilon)} p^{\frac{1}{2}(r+2E/\varepsilon)} q^{\frac{1}{2}(r-2E/\varepsilon)}$$

where the  $\pm$  refer to the positive and negative energy states on the falling and rising parts of the entropy curve, respectively. Although the *a priori* probabilities are the same as before, depending on the choice of the sign in the exponent, the entropy has changed and is now given by (19). According to Pascal's triangle, the maximum value coincides with the middle term, corresponding to the state  $\bar{E} = 0$ . It is further maximized with equal *a priori* probabilities,  $p = q = \frac{1}{2}$ . Maximum probability and maximum entropy coincide in the state of infinite temperature since this is the most disordered state. According to the central limit theorem, the distribution of particles + holes population will surely tend to the normal distribution as the population size  $r$  increases without limit. In the next section this will be contrasted with the case where the *a priori* probabilities tend to zero while the population,  $r$ , again increases without limit. This will be seen to be the case of infinitely small temperatures which fall under another limit theorem: the law of small numbers. We will therefore conclude that if the population is allowed to grow without limit, the extremes in the temperature lead to different different limit laws: the law of large numbers in the limit as  $T \rightarrow \infty$ , and the law of small numbers in the limit as  $T \rightarrow 0$ . Both limits will be equipped with their own statements of the third law.

#### 4. Thermodynamic symmetry-breaking

Since  $(\partial S/\partial \bar{E})_Y$  is negative over a limited range of values of the energy, the internal energy and entropy representations are no longer equivalent in the sense that the maximum entropy principle implies the minimum energy principle [2]. In fact, when  $(\partial S/\partial \bar{E})_Y$  is not positive one of the two expressions for the heat capacity  $T(\partial S/\partial T)_Y$  or  $(\partial \bar{E}/\partial T)_Y$  becomes negative, thus violating a fundamental thermodynamic stability criterion. A *thermodynamic symmetry-breaking* is said to have occurred [10, pp 83–93].

Higher energies do not necessarily imply higher temperatures and larger entropies. It is incorrect to identify the potential energy resulting from the application of a magnetic field as a form of heat energy [3]. Long ago, when it was in vogue to attempt a derivation of the second law from the mechanical principle of least action, it was recognized that two types of coordinates were necessary to specify a thermodynamic system: ‘controllable’ coordinates ‘which fix the geometrical, strain, electric and magnetic configuration of the system,’ and ‘unconstrainable’ coordinates which fix ‘the separate moving parts of the systems whose kinetic energy constitutes the temperature of the body’ [23]. No control can be exerted over latter coordinates individually.

We have already mentioned how an adiabatic reversal of the magnetic field supposedly enables one to jump from a state of positive temperature  $P$  into state a state of negative temperature  $Q$  along the adiabatic connecting both sates in figure 1 [3]. If the effect of the magnetic field were to increase the temperature, we would expect state  $Q$  to be more disordered than state  $P$ , simply because it is hotter. But, the disorder of the two states is exactly the same because the line connecting them is an adiabat. And as a result of energy conservation for any pair of states connected by the adiabat, (12), the temperatures of the two states will be the same. Otherwise, the total energy would not be independent of the temperature. What is different in the states  $P$  and  $Q$  is their values of the internal energy, so that the external field has a purely mechanical effect upon the system. Clearly then, the magnetic energy, or for that matter any potential energy, cannot be considered as a form of heat energy.

Introducing either of the thermal equations of state (8) or (11) into the fundamental relation (17) gives

$$S(\beta) = r \left\{ \frac{\ln(1 + e^{-\beta\varepsilon})}{1 + e^{-\beta\varepsilon}} + \frac{\ln(1 + e^{\beta\varepsilon})}{1 + e^{\beta\varepsilon}} \right\}. \quad (33)$$

Since the entropy is invariant under  $\beta$  reversal,  $\beta \rightarrow -\beta$ , the temperature of the two states  $P$  and  $Q$  will be the same, although they will have an energy difference of

$$\Delta \bar{E} = \bar{E}_h - \bar{E}_p = E_0 \tanh\left(\frac{1}{2}\beta\varepsilon\right).$$

In the limit as  $\beta \rightarrow \infty$ , the energy gap becomes  $\Delta \bar{E} \rightarrow E_0$ , while, in the opposite limit as  $\beta \rightarrow 0$ , the energy gap vanishes.

Taking the derivative of (33) with respect to the temperature gives

$$C_Y := T \left( \frac{\partial S}{\partial T} \right)_Y = r \left( \frac{1}{2}\beta\varepsilon \right)^2 \operatorname{sech}^2\left(\frac{1}{2}\beta\varepsilon\right) \quad (34)$$

as the expression for the heat capacity. In the limit  $\Delta \bar{E}/E_0 \ll 1$ , the heat capacity (34) is approximated by

$$C_Y \approx r \left( \frac{\Delta \bar{E}}{E_0} \right)^2 \left[ 1 - \left( \frac{\Delta \bar{E}}{E_0} \right)^2 \right].$$

This has the form of a Schottky heat capacity, whose anomalous peak is due to the presence of an intrinsic gap width. Both expressions (33) and (34) are symmetrical in  $\beta$ , and therefore cannot distinguish whether the temperature is positive or negative. If the two states are indistinguishable with respect to the sign of  $\beta$ , how can ‘negative’ temperatures be hotter than positive ones? In other words, if we place two systems with the same entropy,  $P$  and  $Q$  in figure 1, in thermal contact, heat will not flow from the ‘hotter’ state  $Q$  to the ‘colder’ state  $P$  because they are at the same temperature. According to the second law, if the two systems are placed in thermal contact which are at the same temperature, there will be no increase in entropy. If two identical systems have the same entropy and energy, they must necessarily be

at the same temperature. But, in open systems the energy of the two systems can be different while possessing the same values of the entropy and temperature.

The expression for the heat capacity (34) is valid for both increasing and decreasing parts of the entropy curve. The heat capacity displays a peak at around  $\beta\varepsilon \approx 1$  and tends to zero for  $\beta = 0$  and  $\beta = \infty$ . According to either of the thermal equation of states, (8) or (11),  $\bar{E}$  tends to half maximum,  $E_0/2$  as  $\beta \rightarrow 0$ . On the one hand, in the state  $\beta = 0$  the temperature is so high and the entropy is at its maximum so that it will be practically independent of the temperature. Consequently, the heat capacity vanishes. On the other hand, at  $\beta = \infty$ , both the entropy and the particle energy (8) vanish, while the hole energy (11) reaches its largest value,  $E_0$  with the entropy returning to zero<sup>†</sup>. Therefore, the heat capacity vanishes both in the states of lowest and highest energies.

This constitutes the most important generalization of the third law for systems with a bounded density of states: *the heat capacity tends to zero at infinite temperature as well as at zero temperature*. In other words, there is a saturation effect in which the entropy tends to its maximum value,  $r \ln 2$ , independent of the temperature. The evolution of the entropy as a function of temperature is shown in figure 3(a). The fact that the entropy tends to a constant both at zero and infinite temperatures is responsible for the anomalous peak in the heat capacity which is shown in figure 3(b). Although the zero temperature limit is the conventional statement of the third law, the infinite temperature limit can be used as a generalized statement of the third law for a constrained system, like the one under investigation, that has a bounded density of states [11]. In other words, no arbitrary constant of integration enters the entropy expression: absolute and not relative differences in the entropy can be measured from the state of maximum, as well as the state of minimum, disorder. Consequently, no arbitrary constant of integration will appear in the generalized Boltzmann principle (21): the entropy reduction determines the entire probability distribution, and not just a part of it as in Boltzmann's original formulation for systems with an unbounded density of states.

The distinction between the two states of absolute zero, at lowest,  $\bar{E} = 0$ , and highest,  $\bar{E} = E_0$ , energies, can be made in by introducing the fugacity,  $z := e^{\mu/T}$ , where  $\mu$  is the chemical potential. The average occupation number for particle states is

$$\bar{n}_p = \frac{r}{e^{(\varepsilon-\mu)/T} + 1}. \quad (35)$$

It tends to  $r$  as  $T \rightarrow 0$  for all those values of the energy,  $\varepsilon < \mu$ , while for energies  $\varepsilon > \mu$  the occupation number tends to zero. The chemical potential at absolute zero is the limiting energy of the particles, and it is necessarily positive. All particle states with energies up to  $\mu$  are completely filled while all single-particle states with energies  $\varepsilon > \mu$  are empty.

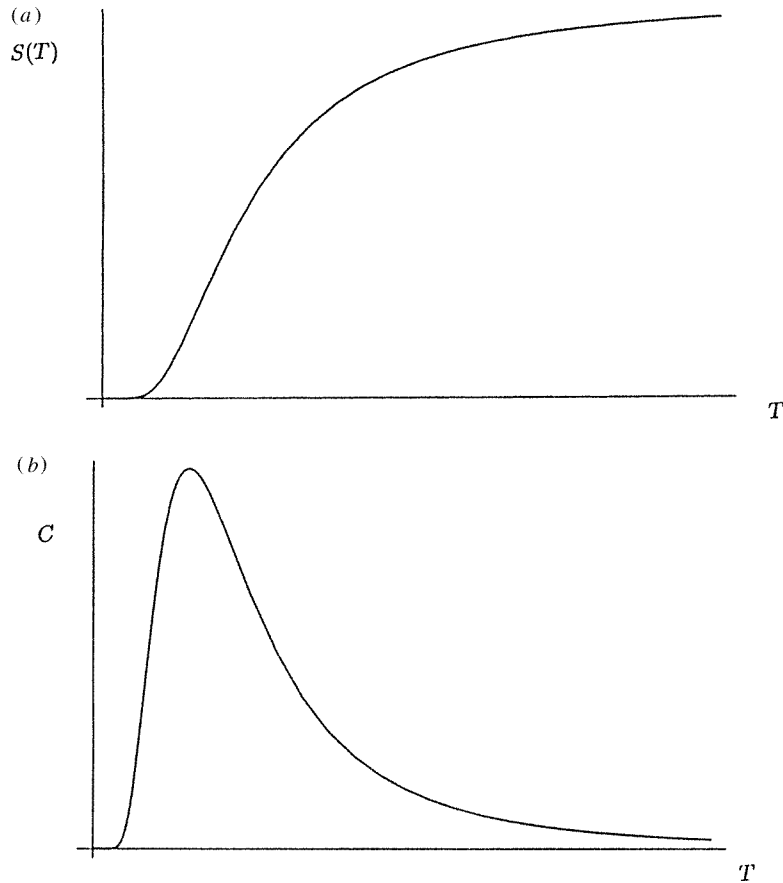
Alternatively, the average occupation number for holes is

$$\bar{n}_h = \frac{r}{1 + e^{-(\varepsilon-\mu)/T}} \quad (36)$$

in the presence of a finite value of the chemical potential. As  $T \rightarrow 0$ , the logistic distribution (36) tends to zero for all  $\varepsilon < \mu$ , and to  $r$  for  $\varepsilon > \mu$ . Consequently, we get the inverse step function, indicating that all single-hole states up to the Fermi energy,  $\mu$ , are completely empty, while all single-hole states with energies greater than the Fermi energy are filled. In order to bring the system into its highest quantum state at absolute zero, it is necessary to fill all hole states with particles below the Fermi energy.

The Fermi sphere now comprises the 'cavity,' and the holes occupy an annular region from the Fermi energy up to the maximum,  $E_0$ , that is characteristic of the metal. This is the system's highest quantum state, which is reached at  $T = 0$ , where there is a complete

<sup>†</sup> This statement has to be qualified when a finite chemical potential is introduced, as we do in (35) and (36) below.



**Figure 3.** (a) The entropy as a function of temperature. (b) The heat capacity curve with an anomalous peak.

segregation between particles and holes. The entropy vanishes at  $T = 0$ , for  $\bar{E}_p = 0$  and  $\bar{E}_h = E_0$ , in the case  $\varepsilon > \mu$ , and for  $\bar{E}_p = E_0$  and  $\bar{E}_h = 0$  in the opposite case where  $\varepsilon < \mu$ . This is analogous to the clustering of parallel vortices of the same sign in unsteady two-dimensional flow. Higher energy states can be achieved by segregating the vortices into clusters of like kind. Onsager [14] reasoned that since these states are of greater order, and consequently lower entropy, they must manifest a 'negative' temperature because they require greater energy. However, the segregation of like particles also occurs at minimum energy because it takes energy to create a particle excitation outside of the Fermi sphere leaving a hole behind. Consequently, the case of complete segregation of particles and holes occurs at the greatest and smallest energies.

In the limit as  $T \rightarrow 0$ , the *a priori* probability  $p \rightarrow 0$  for particle energies  $\varepsilon > \mu$ . Furthermore, if  $r$  is allowed to increase without limit we witness the binomial distribution (31) transforming into the Poisson distribution,

$$P_p(n) = \frac{\bar{n}^n}{n!} e^{-\bar{n}}$$

for the probability that there are  $n = E/\varepsilon$  particles, whose average value is  $\bar{n} = rp$ , in energy states  $\varepsilon > \mu$ . A Poisson distribution for holes is also obtained in the same limit

for the probability of occupying energy states  $\varepsilon < \mu$ . Only for large values of  $\bar{n}$  can the Poisson distribution be approximated by the normal distribution, so that at extremely low temperatures, where  $p$  tends to zero, thermal excitations above the Fermi energy by particles, or hole excitations below the Fermi energy, cannot be approximated by a normal distribution. Therefore, thermal excitations at low temperatures follow the classical law of rare events, in contrast to fluctuations about the state of maximum entropy which follow the normal error law (21).

The total particle energy is obtained by integrating the particle distribution,  $\bar{n}_p = r/(e^{\varepsilon/T} + 1)$ , over all energies  $\varepsilon$ , where, for simplicity, we again suppress the chemical potential. We then obtain

$$E_p = r \int_0^{\infty} \frac{d\varepsilon}{e^{\varepsilon/T} + 1} = -rT \ln(1 + e^{-\varepsilon/T})|_0^{\infty} = rT \ln 2.$$

The energy  $T \ln 2$  represents the minimum energy necessary to transmit one bit of information [1]†. For a temperature of 300 K, this corresponds to a frequency of  $4.3 \times 10^{12}$  Hz lying at the lower end of the far-infrared. If we want to use frequencies larger than this we must consider the hole energies.

Integrating the hole distribution,  $\bar{n}_h = r/(1 + e^{-\varepsilon/T})$ , over all energies results in a divergent integral. Consequently, we must introduce a maximum cut-off,  $\varepsilon_{\max}$ , just as in the Debye model of a crystalline solid at low temperatures. Analogously, in semiconductors,  $\varepsilon_{\max}$  would correspond to the valence band edge. The physical origin of the cut-off is that only in systems with a finite number of degrees of freedom can the entropy decrease as the internal energy increases over a finite range of values. The total energy is

$$E_h = r \int_0^{\varepsilon_{\max}} \frac{d\varepsilon}{e^{-\varepsilon/T} + 1} = rT [\ln(1 + e^{\varepsilon_{\max}/T}) - \ln 2].$$

For energies  $\varepsilon_{\max} \gg T$ , this reduces to  $E = r\varepsilon_{\max}$ , independent of the temperature. The hole energy is therefore much larger than the particle energy. The characteristic largest frequency for particles is  $T \ln 2/\hbar$ , whereas for holes, it is  $\varepsilon_{\max}/\hbar$ , with  $\varepsilon_{\max} \gg T \ln 2$ . The quantity  $T \ln 2$  constitutes the lower limit for detecting a quantum of energy above the limit of thermal noise [1]. This corresponds to the largest entropy per particle,  $\ln 2$ , in the state of maximum disorder.

## 5. Comparison with relativistic hole theory

The relativistic wave equation admits solutions for which the energy (and not the temperature) is negative. These negative energy states have caused a great deal of concern in quantum mechanics because an applied field can cause transitions from a state of positive energy to a state of negative energy. Once the electron reaches a negative energy state it can continue to lower its energy by emitting photons without limit. In order to avoid the conclusion that positive energy electrons can jump into negative energy states, Dirac assumed that all negative energy states, comprising energies between  $-mc^2$  to  $-\infty$ , are filled. Moreover, they do not contribute to the charge of the system because the charge is uniformly distributed [5]. High energy, rapidly varying electromagnetic fields enable an electron to jump into a positive energy state leaving a hole, or a negative energy positive electron, behind. The energy gap separating positive and negative energy states is  $2mc^2$ , and this rest energy can be given up in the form of light when an electron-positron pair annihilate each other which can be pictured as an electron jumping into a hole representing the positive electron.

† This is deduced from the equality  $e^{-\varepsilon/T} = 1 - e^{-\varepsilon/T} = \frac{1}{2}$  of the probabilities of finding an energy level  $\varepsilon$  occupied and vacant.



Although this is the conventional formulation of relativistic hole theory, it is not the fact that the energy is negative but, rather, that there are *two* values of the energy which are admissible for each 'particle'. Even in the nonrelativistic theory of intrinsic semiconductors there are two energy states, one in the valence and the other in the conduction band, which are connected by an electronic transition. When an electron has been raised to the conduction band, a hole has been left behind in the valence band having all the properties of an electron except that it has positive charge. The energy gap of  $2mc^2$  corresponds to the intrinsic energy gap separating the two bands.

The problem with the formulation of relativistic hole theory is the association of the negative energy hole with the positive energy positron. The occupation number for the positive energy positrons is converted into the occupation number of negative energy holes by changing the sign of the energy. We will argue that it is only the latter distribution that is thermodynamically admissible.

The thermal equilibrium distributions for positive energy electrons and positive energy positrons are [4]

$$\bar{n}(e^-) = z \frac{\partial}{\partial z} \ln(1 + ze^{-\varepsilon_p/T}) = \frac{r}{z^{-1}e^{\varepsilon_p/T} + 1} \quad (37)$$

and

$$\bar{n}(e^+) = -z \frac{\partial}{\partial z} \ln(1 + z^{-1}e^{-(2mc^2 + \varepsilon_h)/T}) = \frac{r}{ze^{(2mc^2 + \varepsilon_h)/T} + 1} \quad (38)$$

respectively, where  $\varepsilon_p$  and  $2mc^2 + \varepsilon_h$  are the energies of the states of an electron and positron. The generating functions in (37) and (38) do not differ by an integral zero point, as do (6) and (9). Expression (38) represents the occupation number of positive energy positrons, which by reversing the sign of the energy is converted into the logistic distribution (36) for holes. As it stands, (38) would have the mean occupation number *decreasing* with the chemical potential, rather than increasing with it. It would thus violate a fundamental thermodynamic stability criterion.

The same criticism could be lodged against the logistic distribution (36). Yet, it must be remembered that hole states lie along the descending part of the entropy curve, and for quantum statistics it is the occupation number that is the independent variable and not the energy, which is related linearly to it,  $\bar{E} = \bar{n}\varepsilon$ . For quantum statistics, the modified version of the second law, (3), corresponds to

$$\frac{dS}{d\bar{n}} = \left( \frac{\partial S}{\partial \bar{n}} \right)_{\bar{E}} + \left( \frac{\partial S}{\partial \bar{E}} \right)_{\bar{n}} \frac{d\bar{E}}{d\bar{n}} = \frac{\mu - \varepsilon}{T}. \quad (39)$$

Clearly, it is the distribution of holes (36) which satisfies (39) and not the distribution of positive energy positrons, (38). Put slightly differently, it is the number of electrons and holes at any given energy which is conserved, since their distributions are complementary to each other. Without number conservation we would be at a loss to interpret the chemical potential,  $\mu$ .

According to the conventional formalism, the basic interaction in electrodynamics is such that the *difference* between the number of positive energy electrons, (37), and the number of positive energy positrons, (38), be conserved [4]:

$$\bar{n}(e^-) - \bar{n}(e^+) = \text{const.} \quad (40)$$

It is this condition that supposedly determines the fugacity [4]. In the hole theoretic interpretation, one equates  $\bar{n}(e^-) = \bar{n}_p(e^-_{\varepsilon > 0})$ , while [21]

$$\bar{n}_h(e^-_{\varepsilon < 0}) = -\bar{n}(e^+) + \text{constant background} \quad (41)$$

which are given by (35) and (36), respectively. The 'background' is the state in which all negative kinetic energy levels are filled. The second definition is tantamount to reversing the sign of the positive energy electrons in (38) so that it would coincide with the logistic distribution (36). Since it is the difference between the number of electrons and positrons that is conserved, (40), a negative sign is needed in (41) to get the conservation of the sum of particles and holes (13). The condition on the chemical potential that was implied by (40) has disappeared because the sum of the particle and hole distributions is an identity. Consequently, the two conservation conditions, (13) and (40), are not equivalent.

Moreover, the lack of symmetry between (37) and (38) enables the average number of positrons to vanish in the limit as  $T \rightarrow 0$ , while the vanishing or nonvanishing of (37) depends upon the sign of the difference  $\varepsilon_p - \mu$  in the same limit. Consequently, when we take the  $T \rightarrow 0$  limit in (41),  $\bar{n}(e^+)$  vanishes, whereas the occupation number of holes vanishes for energies less than  $\mu$ , while it tends to  $r$  for energies greater than  $\mu$ . Hence, the 'constant background' is less than constant. This leads us to the conclusion that the logistic distribution for positive energy holes (36) is the physically meaningful thermal distribution and not the thermal distribution of positive energy positrons, (38).

## 6. Conclusions

The aim of this paper has been to show that there is no need, or reason, to introduce a negative temperature scale. 'Negative' temperatures can be replaced by positive ones simply by modifying the second law (1) to read (3) for states lying on the descending branch of the entropy curve in systems with a bounded density of states.

We chose Ramsey's model [20] of 'negative' temperatures, but made allowance for different zero-point energies. Different choices of the zero-point energy gave rise to different definitions of the expected energy in terms of the derivative of the logarithm of the generating function. Since the logarithm of the generating function and the entropy are duals in the sense of the Legendre transform, a change in definition of the expected value of the energy, in terms of the derivative of the logarithm of the generating function, causes a corresponding change in the definition of the inverse temperature in terms the derivative of the entropy. It is the modification of the second law along the descending part of the entropy curve, (3), which is required and not two temperature scales both having infinite ranges [18].

Systems that differ by an integral zero-point energy have the same entropy but have different thermal equations of state. The resulting equations of state, (8) and (11), or, equivalently, the expected number of elements and their complements as a function of temperature, show a behaviour akin to particles and holes in semiconductors, or particles and antiparticles in relativistic quantum theory. The only difference is that hole states lie above particle states energetically, while hole states in semiconductors and antiparticle states in the Dirac sea lie below those of particles. Therefore, for communication purposes, hole states are to be preferred. The thermal distribution for particles for states lying on the ascending branch of the entropy curve is the Fermi distribution, (8), while the thermal distribution for holes on the descending branch obey the logistic distribution, (11). The sum of the two distributions is constant guaranteeing both the conservation of energy, (12), and the total number of particles and holes, (13).

Since the entropy and the logarithm of the generating function are Legendre duals, their derivatives are inverses of one another. One would naturally think of their satisfying Young's inequality were it not for the condition that the equation of state for the inverse temperature is hyperbolic and hence does not vanish at the origin as required by the usual form of Young's inequality. Therefore, it was necessary to generalize Young's inequality to include the case

where the thermodynamic state function is a strictly *decreasing* function of its argument. This is the usual thermodynamic situation which is covered by the conventional form of the second law (1). Exponentiating both sides of the thermodynamic Young inequality leads to a probability distribution (25), belonging to the exponential family of distributions. The prior distribution, or the density of states, is unbounded from above and thus requires an exponentially decreasing function of the energy to cut off the power growth of the density of states, thereby ensuring the existence of its integral over all energies. The coefficient of the energy in the exponential is the inverse temperature, which necessarily must be positive.

Yet, in the case under discussion the density of states is bounded. Consequently, the unconventional thermodynamic Young inequality can only apply to the ascending branch of the entropy curve. Along the descending part of the curve the conventional form of Young's inequality applies since the inverse temperature is a strictly *increasing* function of the energy. The density of states now decreases as an inverse power of the energy, and although Young's inequality again yields a distribution belonging to the exponential family, (29), the coefficient in the exponentially increasing factor is positive since the generating function is now a monotonically increasing function of the inverse temperature. Hence, we concluded that a modification of the second law is required along the descending branch of the entropy curve rather than a new temperature scale. Moreover, we appreciated that whereas two thermal equations of state are required, the entropy (19) is unique and insensitive to whether the inverse temperature is defined according to its usual definition (1) or its negative (3).

The two distributions belonging to the exponential family can be cast as two binomial distributions. The binomial coefficient, whose logarithm is the entropy, is the same in both cases. But what changes are the expressions for the *a priori* probabilities: they reflect the same mirror symmetry under temperature reversal that the two thermal equations of state, (8) and (11), manifest. This provides additional support for the need of a modification of the second law for hole states. Moreover, we found that the state of maximum disorder has equal *a priori* probabilities leading to the coincidence of maximum probability with maximum entropy. As the population,  $r$ , is allowed to increase without limit, the probability distribution tends to the normal distribution, (21), because the *a priori* probabilities,  $p$  and  $q$ , are equal. The error of the normal approximation will be small if  $rpq \gg 1$ . Alternatively, if the total population is large while the probability of success,  $p$ , is small such that their product  $rp = \bar{n}$  is small, but finite, another limit law applies: the law of small numbers leading to the Poisson approximation. Whereas the normal approximation applies to the state of infinite temperature and maximum disorder, the Poisson approximation applies to the states of zero temperature at lowest and highest energies.

States along the descending branch of the entropy curve are characterized by a thermodynamic symmetry breaking in which the equivalence between the maximum entropy and minimum energy representations is destroyed. In the region where (3) holds, it is evident that both  $T(\partial S/\partial T)_Y$  and  $(\partial \bar{E}/\partial T)_Y$  cannot both yield positive definite expressions for the heat capacity. And since this is a necessary condition for thermodynamic stability, only one of the two representations will survive. The fact that the thermal distribution of holes is governed by the logistic distribution (36) implies that  $(\partial \bar{E}/\partial T)_Y$  will be negative along the descending branch of the entropy curve. Furthermore, since the heat capacity vanishes at the extremes in the temperature, there will be a statement of the third law both a zero and infinite temperatures.

The adiabat which intersects the bell-shaped entropy curve connects two states whose entropies are not only equal but also their heat capacities are equal. When either (1) or (3) is used to eliminate the energy in the expression for the entropy, (19), the same expression for the entropy as a function of temperature results, (33). Both the entropy (33), and heat capacity (34), are even functions of the inverse temperature and cannot distinguish between plus and minus

temperatures. The conservation of energy (12), or, equivalently, the conservation of particles and holes, (13), at any temperature requires the temperatures of the two states connected by the adiabat to be one and the same. Furthermore, both states possess the same degree of disorder because they have the same entropy. But what are different are the energies of the two states and the different definitions used to define their temperature. Since the field affects only the internal energy, we are treating a purely mechanical effect, and this can happen only in open thermodynamic systems.

### Acknowledgments

This work was sponsored, in part, by contributions from MURST 60% and ex-40%.

### References

- [1] Brillouin L 1962 *Science and Information Theory* 2nd edn (New York: Academic) pp 184ff
- [2] Callen H B 1985 *Thermodynamics* 2nd edn (New York: Wiley) p 31
- [3] Dyson F 1954 *Sci. Am.* September 58–63
- [4] Fowler R H 1938 *Statistical Mechanics* (Cambridge: Cambridge University Press) p 653
- [5] Heitler W 1954 *The Quantum Theory of Radiation* (Oxford: Clarendon) pp 110–14
- [6] Joyce G and Montgomery D 1973 *J. Plasma Phys.* **10** 107–21  
Montgomery D and Joyce G 1974 *Phys. Fluids* **17** 1139–45
- [7] Landau L D and Lifshitz E M 1969 *Statistical Physics* 2nd edn (Oxford: Pergamon) pp 36–7
- [8] Lavenda B H and Dunning-Davies J 1989 *J. Math. Phys.* **30** 1117–21
- [9] Lavenda B H 1991 *Statistical Physics: A Probabilistic Approach* (New York: Wiley-Interscience) ch 6
- [10] Lavenda B H 1995 *Thermodynamics of Extremes* (Chichester: Albion) pp 114–15
- [11] Lavenda B H and Dunning-Davies J 1995 *Nuovo Cimento B* **110** 265–76
- [12] Lavenda B H 1999 Inequalities and irreversibility *Nuovo Cimento B* at press
- [13] Mandelbrot B B 1970 *Critical Review of Thermodynamics* ed E B Stuart *et al* (Baltimore, MD: Mono Book) pp 230–2
- [14] Onsager L 1949 *Suppl. Nuovo Cimento* **6** 279
- [15] Pečarić J E, Proschan F and Tong Y L 1992 *Convex Functions, Partial Orderings and Statistical Applications* (San Diego, CA: Academic) pp 240–2
- [16] Pointin Y B and Lundgren T S 1976 *Phys. Fluids* **19** 1459–70  
Lundgren T S and Pointin Y B 1977 *Phys. Fluids* **20** 356–63
- [17] Pound R V 1951 *Phys. Rev.* **81** 156  
Ramsey N F and Pound R V 1951 *Phys. Rev.* **81** 278
- [18] Proctor W G 1977 *Sci. Am.* 78–85
- [19] Purcell E M and Pound R V 1951 *Phys. Rev.* **81** 279
- [20] Ramsey N V 1956 *Phys. Rev.* **103** 20
- [21] Sakurai J J 1967 *Advanced Quantum Mechanics* (Redwood City, CA: Addison-Wesley) pp 131–44
- [22] Shockley W 1950 *Electrons and Holes in Semiconductors* (Princeton, NJ: Van Nostrand) ch 15
- [23] Thomson J J 1888 *Applications of Dynamics to Physics and Chemistry* (London: Macmillan) p 94
- [24] Tisza L 1966 *Generalized Thermodynamics* (Cambridge, MA: MIT) p 123
- [25] Wilks J 1961 *The Third Law of Thermodynamics* (London: Oxford University Press) section 10.2
- [26] Würfel P 1982 *J. Phys. C: Solid State Phys.* **15** 3967–85
- [27] Yariv A 1989 *Quantum Electronics* 3rd edn (New York: Wiley) p 577
- [28] Ziman J M 1969 *Elements of Advanced Quantum Theory* (Cambridge: Cambridge University Press) pp 48–52