Reprinted from
Modern Developments in Thermodynamics
Edited by B. Gal-Or
Israel Universities Press, Jerusalem
John Wiley & Sons, New York · Toronto

Thermodynamics and Cosmology

J.V. NARLIKAR*
Institute of Theoretical Astronomy
University of Cambridge
Cambridge, England

ABSTRACT

Arguments are advanced to show that the time-asymmetry present in thermodynamics is likely to be strongly related to cosmology. It is also shown that not all models of the expanding universe provide a satisfactory link between thermodynamics and cosmology. Of the models usually discussed by the cosmologist only the steady state model has the required properties.

INTRODUCTION

It is usual to regard thermodynamics as a branch of local physics. Many thermodynamical discussions begin with such statements as "suppose we isolate the system from its surroundings . . ." or "enclose the system in a light-tight box . . ." etc. While statements like these simplify the mathematical discussion and are often good approximations to reality, they tend to obscure the important part played by the cosmological boundary conditions. The purpose of this article is to emphasize this part. In the present section we describe ideas which are relatively better known. In the following two sections we shall turn to the less conventional views on this subject.

The universe, as we know it today, is made up of galaxies, radio sources, quasi-stellar objects, etc. These objects differ considerably from one another in their internal structure and physical behaviour. Yet they have one feature in common. They radiate energy in some part or the other of the electromagnetic spectrum, ranging from the long radio waves through microwave, infra-red, optical, ultra-violet, X-ray to the very high energy γ-rays. To what extent is this radiation relevant to us on the Earth?

This question was raised as long ago as 1826 by Olbers [1], and even earlier by Halley [2], and has come to be known as the Olbers paradox. According to a simple calculation made by Olbers, the night sky should not be dark—but immensely

* Now at Tata Institute of Fundamental Research, Bombay 5, India.

bright. The simplest version of this argument is as follows. Suppose we have an infinitely old and infinite in extent universe with Euclidean geometry, and with a uniform distribution of radiating objects. Suppose there are n such objects per unit volume and that each object has the same luminosity L. The flux of radiation received from a typical source located at distance r from the Earth has the magnitude

$$\frac{L}{4\pi r^2}. \tag{1}$$

There are

$$4\pi r^2 dr \cdot n \tag{2}$$

such objects in a shell of thickness dr. The total contribution from all such objects is the product of (1) and (2), i.e.,

$$nL dr.$$

Clearly the total contribution from the entire universe would then be infinite.

However, we can obtain a finite answer by assuming a finite size for each object. Suppose each object has a cross-section A normal to the line of sight from the Earth. So an object at a distance r would subtend a solid angle

$$\Omega = \frac{A}{r^2}. \tag{4}$$

This object will block radiation from all objects beyond it and lying in the same solid angle Ω. Total blockage of radiation will result when the entire solid angle 4π has been covered. This will happen when we go out to a distance $\sim R$, where

$$\int_0^R 4\pi r^2 n \cdot \frac{A}{r^2} dr = 4\pi, \quad \text{i.e.,} \quad R = \frac{1}{An}. \tag{5}$$

The contribution from all sources within the distance R is given by

$$\int_0^R nL dr = \frac{L}{A}. \tag{6}$$

This is of the order of the surface brightness of each object.

Even this finite result is too high, however! If we take a typical radiator to be like the Sun, we would have a sky brightness temperature of 6000°K. Clearly something must be wrong somewhere. Various ways were suggested to get round this paradox, but none so satisfactory as that provided by modern cosmology. For instance, absorption of distant radiation en-route does not help; it merely heats up the absorbing matter and reproduces the same problem as before. A somewhat artificial resolution can be obtained by chopping up the universe in space or time or both. A detailed discussion of these ideas is given by Bondi [3].

How does modern cosmology resolve this paradox? In the late twenties, Hubble [4] made the important discovery that radiation from distant galaxies is systematically redshifted. A simple minded explanation in terms of Doppler effect implies that these galaxies are receding from us. Hubble also found that the speed of recession or, more correctly, the redshift z is proportional to the distance r of the galaxy:

$$z = \frac{H}{c} \cdot r. \tag{7}$$

(7) is called Hubble's law. H is the constant of proportionality, known as Hubble's constant. c is the velocity of light. The value of H is usually taken around $3 \cdot 10^{-18} \mathrm{sec}^{-1}$, although recent observations [5] may very well require only about half this value.

The redshift has the important effect of modifying some of the formulae used in Olbers's calculation. For instance, (1) is modified to

$$\frac{L}{4\pi r^2 (1 + z)^2}. \tag{8}$$

The factor $(1 + z)^{-2}$ produces convergence when we calculate the total flux at the Earth. Also, the calculated finite value of sky brightness is very low and entirely consistent with observations.

Thus a result of importance to local thermodynamics has been obtained by appealing to the large scale structure of the universe! Olbers's paradox may be worded differently in thermodynamic language. A static Euclidean universe with a uniform distribution of radiating sources, will eventually reach thermodynamic equilibrium when the radiation temperature everywhere is the same as that of each source. Why is the universe not in that state?

The present state of the universe is far from thermodynamic equilibrium. In formulating the steady state model of the universe Bondi and Gold [6] made use of the connection between thermodynamics and cosmology in the following way. The starting point of the steady state theory is the perfect cosmological principle (P.C.P.) which states that the universe in the large is homogeneous in space and unchanging in time. If we have such a universe in static state (with all galaxies at rest relative to one another) we would be back with Olbers's paradox. If the universe were contracting, the blueshifted radiation would make matters even worse. The only possibility left is that the universe must expand. Thus using the P.C.P. and the observed local thermodynamic dis-equilibrium, Bondi and Gold were able to *deduce* that the universe expands.

These ideas serve to emphasize the important role the universe appears to play in thermodynamics. Briefly, we might say that the expanding universe acts like a sink for the radiation emitted by the various sources—from the atom to the vast radio galaxies. We have, however, yet to examine how the radiation itself behaves in the expanding universe. Does it necessarily go from the source to the sink?

THE NATURE OF RADIATION: CLASSICAL CONSIDERATIONS

The radiation discussed in thermodynamics refers to some part or another of the electromagnetic spectrum. Recently there has been some indication of the existence of gravitational radiation [7]. This, if it exists, will add an important consideration to astrophysical processes, although its effects in the laboratory will be small. In the following discussion we shall assume radiation to be electromagnetic in nature.

Classical electromagnetic theory is usually discussed in terms of Maxwell's equations. These make use of two interdependent entities—the electromagnetic field and the electric charges. In free space with electric charge density ρ and current density $\mathbf{j}$, the electric field $\mathbf{E}$ and magnetic field $\mathbf{H}$ are given by

$$\left.\begin{aligned} &\nabla \cdot \mathbf{E} = 4\pi\rho, \qquad \nabla \times \mathbf{H} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} = \frac{4\pi}{c}\mathbf{j} \\ &\nabla \cdot \mathbf{H} = 0, \qquad \nabla \times \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{H}}{\partial t} = 0. \end{aligned}\right\} \tag{9}$$

These are Maxwell's equations in Gaussian units. They are supplemented by the Lorentz-force equation which states that the force on an electric charge e moving with velocity $\mathbf{v}$ in the field $(\mathbf{E}, \mathbf{H})$ is given by

$$e\left[\mathbf{E} + \frac{v \times \mathbf{H}}{c}\right]. \tag{10}$$

It is convenient to write $\mathbf{E}$ and $\mathbf{H}$ in terms of potentials $\mathbf{A}$ and ϕ:

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} - \nabla\phi, \qquad \mathbf{H} = \nabla \times \mathbf{A}, \qquad \nabla \cdot \mathbf{A} + \frac{1}{c}\frac{\partial \phi}{\partial t} = 0. \tag{11}$$

It is easy to see that (9) then reduces to

$$\Box\,\phi = 4\pi\rho, \qquad \Box\,\mathbf{A} = \frac{4\pi}{c}\mathbf{j} \qquad \left(\Box \equiv \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right). \tag{12}$$

These results look more elegant in the four dimensional language of special relativity. We shall, however, continue to use the three-dimensional format.

The radiation arises from the solutions of (12). Given ρ and $\mathbf{j}$ we can calculate $\mathbf{A}$ and ϕ and hence $\mathbf{E}$ and $\mathbf{H}$. It is usual to write the solution in the form:

$$\left.\begin{aligned} \phi_{\text{ret}}(t, \mathbf{r}) &= \int \frac{\rho\left(t - \dfrac{|\mathbf{r} - \mathbf{r}_1|}{c}, \mathbf{r}_1\right)}{|\mathbf{r} - \mathbf{r}_1|} d^3\mathbf{r}_1, \\ \mathbf{A}_{\text{ret}}(t, \mathbf{r}) &= \int \frac{\mathbf{j}\left(t - \dfrac{|\mathbf{r} - \mathbf{r}_1|}{c}, \mathbf{r}_1\right)}{c|\mathbf{r} - \mathbf{r}_1|} d^3\mathbf{r}_1. \end{aligned}\right. \tag{13}$$

The suffix 'ret' indicates that we are using the retarded solutions. Thus ϕ_{ret} is due to disturbances arising earlier from the source. At large distance from the source, $\mathbf{E}$ and $\mathbf{H}$ fall off as r^{-1} and hence the flux of radiation falls off as r^{-2}. This was the result used in (1).

It would appear that (13) "explains" how a source loses energy via radiation. To draw such a conclusion would be justified if (13) were the only solution of (12) or if all solutions of (12) correspond to loss of energy by the source. Neither is the case. The time-symmetry of Maxwell's equations tells us that we could obtain other solutions simply by reversing the sign of t. Such solutions are called advanced solutions:

$$\left.\begin{aligned} \phi_{\text{adv}}(t, \mathbf{r}) &= \int \frac{\rho\left(t + \dfrac{|\mathbf{r} - \mathbf{r}_1|}{c}, \mathbf{r}_1\right)}{|\mathbf{r} - \mathbf{r}_1|} d^3\mathbf{r}_1, \\ \mathbf{A}_{\text{adv}}(t, \mathbf{r}) &= \int \frac{\mathbf{j}\left(t + \dfrac{|\mathbf{r} - \mathbf{r}_1|}{c}, \mathbf{r}_1\right)}{c|\mathbf{r} - \mathbf{r}_1|} d^3\mathbf{r}_1. \end{aligned}\right. \tag{14}$$

Here ϕ and $\mathbf{A}$ at time t are due to source disturbances at *later* times. In this solution also $\mathbf{E}$ and $\mathbf{H}$ drop off as r^{-1} for large r. But this solution represents incoming rather than outgoing radiation. A more general solution can be obtained by combining (13) and (14):

$$\left.\begin{aligned} \phi &= \alpha\phi_{\text{ret}} + (1 - \alpha)\,\phi_{\text{adv}} \\ \mathbf{A} &= \alpha\mathbf{A}_{\text{ret}} + (1 - \alpha)\,\mathbf{A}_{\text{adv}} \end{aligned}\right\} \quad 0 \leqq \alpha \leqq 1, \tag{15}$$

in which there is incoming and outgoing radiation.

So it appears that we need $\alpha = 1$ to have purely outgoing radiation. This case corresponds to retarded solutions. Is there any reason why we should have retarded solutions only in nature? The argument that we must have retarded solutions in order to have causality is no argument at all. It simply transfers the question to "why causality?" Clearly, what is needed is some overriding reason for $\alpha = 1$, which may be beyond pure electrodynamics. Can such a reason be found in thermodynamics or cosmology? Both these subjects deal with time-asymmetric phenomena, and what we are looking for here, viz., $\alpha = 1$, is also a time-asymmetric situation. In the Maxwellian picture described above there is no such connection. The main reason is that the theory admits too many solutions! We can always choose the solution we want by making use of the free-field equations:

$$\Box \phi = 0, \qquad \Box \mathbf{A} = \mathbf{0}. \tag{16}$$

What we need is a theory where so much choice or room for manoeuvre is not available. In the rest of the section we shall describe such a theory.

The theory in its original form appeared as Coulomb's inverse-square law of attraction or repulsion between charges e, e'

$$\frac{ee'}{r^2}. \tag{17}$$

Here the charges interact directly without the medium of a field. This law of instantaneous action at a distance was later modified [8, 9, 10] and expressed in a language suitable for special relativity. The formal structure of such a theory has been discussed elsewhere [11]. We shall give the outline of its physical contents. The theory considers only electric charges as basic entities. Such fields as there are are traced to their sources—the moving charges. Thus there are no fields which satisfy (16). This greatly reduces the number of degrees of freedom of the fields. The electric charges influence each other with action propagating with the speed of light. Since action and reaction are equal and opposite, a retarded action implies advanced reaction and vice versa. Thus advanced and retarded fields occur with equal status. In terms of (15) we only have the case $\alpha = \frac{1}{2}$. Henceforth we shall write the field arising from a charge a as

$$\tfrac{1}{2}\{F^{(a)}_{\text{ret}} + F^{(a)}_{\text{adv}}\}. \tag{18}$$

Here F stands for a typical electric or magnetic field component.

At first sight the restriction $\alpha = \frac{1}{2}$ seems to be a great drawback, so much so that when the theory was formulated this way it appeared doomed from the start. However, the restriction was turned into an advantage by Wheeler and Feynman [12] in 1945. Their ingenious argument is summarized below.

Consider an electric charge a at $r = 0$. Its radiative electric field at a distance r will be proportional to r^{-1}. A charge b at this distance will therefore move with an acceleration $\propto r^{-1}$. Its response field back at a will fall off by a further factor r^{-1}. Also, if the outgoing disturbance from a is retarded, the reaction from b will be advanced and will arrive at a at the moment the former left a! Hence a charge b at a distance r produces an instantaneous reaction at a proportional to r^{-2} *no matter how large r is*. Since the number of such particles b in a shell of thickness dr is proportional to r^2, we seem to be in an Olbers-type situation. However, when we take the phases of the disturbances into account, and the absorption produced by the intergalactic medium, we arrive at a finite value for the response from the universe to the motion of a. This value, as calculated by Wheeler and Feynman, is simply

$$\tfrac{1}{2}\{F^{(a)}_{\text{ret}} - F^{(a)}_{\text{adv}}\}. \tag{19}$$

The sum of (18) and (19) produces the required retarded field. Also, (19) is the radiative reaction which causes a damping of the motion of the charge.

The awkward advanced effects are thus damped out by the universe and only the observed retarded effects are present. The theory has the advantage over Maxwell's in that the value $\alpha = 1$ does not have to be assumed, but can be related to the nature of the universe.

This argument, though satisfactory, is not complete. This is because the model of the universe assumed by Wheeler and Feynman was static and hence time-symmetric. This means a change of t to $-t$ should leave the argument invariant. This has the effect of producing a radiative reaction opposite to (19):

$$\tfrac{1}{2}\{F^{(a)}_{\text{adv}} - F^{(a)}_{\text{ret}}\}, \tag{20}$$

thus leading to the case $\alpha = 0$. How do we avoid this possibility?

Wheeler and Feynman achieved this by introducing asymmetrical initial conditions, of the type that lead to the thermodynamic arrow of time. This is done as follows. The solution (19) assumes that the charge b is at rest *before* the retarded wave from a hits it. This seems reasonable enough. But its time-reversed part requires b to be at rest *after* the advanced wave from a hits it. This must not only be true of b but of all other particles in the universe. Although an initial condition, in which all charges $b \neq a$ moved in such a way before the disturbance from a hit them that after the hit they were brought to rest, is possible, it is extremely unlikely from the viewpoint of statistical mechanics. Hence the case $\alpha = 0$ should be rejected.

What part does the universe play in the Wheeler–Feynman theory? It acts as a perfect absorber. That is, it absorbs all electromagnetic disturbances emanating from a typical charge in it. For retarded solutions to be consistent the universe must act as a perfect absorber in the future light cone. For advanced solutions to be consistent it must be a perfect absorber in the past. A static Euclidean universe is a perfect absorber in the past as well as in the future. This results in the ambiguous situation described above and requires recourse to the thermodynamic arrow of time.

This ambiguity would not be present if the universe were a perfect absorber either in the past or the future but not in both. Are there any models of the universe which have this property? This question was answered in the affirmative by Hogarth [13]. He pointed out that unlike the static universe, an expanding universe was not time-symmetric. Thus the future and past absorbers in such a universe may not be identical in their properties.

Hogarth and others [14, 15] have examined the Wheeler–Feynman theory in various models of the expanding universe. The main results of this work can be described in terms of three types of cosmological models.

(*a*) In ever expanding Friedmann models (20) and not (19) is consistent. In such universes only advanced waves exist.

(*b*) In Friedmann models with expansion followed by contraction, both (19) and (20) are consistent as in the static case. The considerations of the next section, however, rule out this case also.

(*c*) In ever expanding models with continuous creation of matter, e.g. the steady state model, (19) but not (20) is valid. This type of model therefore agrees with the observed retarded solutions.

It is not difficult to see why different models respond differently in this way. In the case of retarded solutions, we are concerned with the absorption of waves going into the future, i.e., waves which are redshifted. Although these waves are easy to absorb, the matter density in the future tends to zero in the (*a*) type models and there is only incomplete absorption. In (*b*) and (*c*) type models there is enough matter to absorb the future going waves. In type (*b*) the future going waves are subsequently blueshifted. In the case of advanced solutions we are concerned with absorption in the past. This means absorption of blueshifted waves. In (*a*) and (*b*) but not in (*c*) there is enough matter in the past to absorb these waves.

THE NATURE OF RADIATION: QUANTUM CONSIDERATIONS

So far we have considered radiation in its classical aspects. Its more interesting properties are, however, provided by quantum theory, such as spectral lines, the Planck radiation law, etc. These have been thoroughly investigated in the Maxwell theory long ago. How are these quantum phenomena explained in the Wheeler–Feynman theory? We shall describe below some of the recent developments [16, 17].

We begin with the example of a spectral line, arising, say, when an atomic electron jumps spontaneously from an upper energy level E_2 to a lower one E_1. In Maxwell's theory this is explained by field quantization. In quantum field theory even the state of "zero" field, i.e., vacuum, is a non-trivial one. Owing to uncertainty principle, there are fluctuations even in vacuum. These are responsible for the downward jump of the electron.

In the action at a distance theory such a point of view is unacceptable. Here fields do not exist on their own and so cannot be quantized as independent entities.

Such fluctuations as exist in these fields must be traced to the motions of their sources which *can* be quantized. So how do we explain an atomic transition when there are no inducing fields present?

The paradoxical situation is resolved if we take into account the role of the universe as an absorber. Suppose, to begin with, the electron does jump down from E_2 to E_1. In so doing it accelerates and radiates. The universe responds to this radiation and produces a reaction like (19). It is this reaction which causes the electron to jump down in the first place! So we do not have a primary cause but rather a chicken and egg situation. Classically we might have argued that the electron could have sat at E_2 permanently, and there would have been no radiation. However, in quantum mechanics, the electron does not follow a unique and definite path (fixed by, say, the principle of stationary action). Rather, it follows any of a large number of geometrically possible paths, each one with a prescribed probability amplitude. Some paths would lead the electron from E_2 to E_1.

So there exists a finite probability for downward transition. Detailed calculation shows this to be the probability actually observed.

The question arises, why not have spontaneous upward transitions, say, from E_1 to E_2? Although there are geometrical paths from E_1 to E_2 which the electron could follow, these would involve gain of energy and this must come from the universe. However, if we live in type (*c*) models, the universe acts as absorber, not emitter. Hence such situations turn out to have zero probability. In universe of type (*a*), however, the opposite will take place!

Detailed examination [16] of the process reveals that the universe in the future should be in ground state. When the retarded signal from a typical atomic source is emitted, it is absorbed by an absorber particle making an upward transition. The particle subsequently goes back to ground state by losing its energy through collisions in the intergalactic medium. This state of the universe is often called the cold environment.

Such a cold environment is not possible in the models of type (*b*) of the previous section. Hence, although such models are able to absorb in the classical sense, they are unable to do so in the quantum sense.

It is interesting to see how the probability of spontaneous transition downward serves to define the unit of radiation, or the photon. Suppose we have an atomic electron in an incident field of other charges. The probability for downward transition $E_2 \to E_1$ induced by this field is the same as that for the upward transition $E_1 \to E_2$. The total downward transition probability is, however, greater because of spontaneous transition.

The ratio of the two probabilities may be written as

$$\frac{P(E_2 \to E_1)}{P(E_1 \to E_2)} = \frac{n+1}{n}, \qquad n > 0. \tag{21}$$

We then say that the incident field has n units of radiation of frequency $\nu = (E_2 - E_1)/h$. In the language of quantum field theory these units are nothing but photons.

Local thermodynamics tells us that in an enclosure with thermodynamic equilibrium at temperature T the number of atoms with electrons in state E_1 is related to the number of atoms with electrons in state E_2 by

$$\exp[(E_2 - E_1)/kT] = \exp[h\nu/kT], \tag{22}$$

where k is the Boltzmann constant. From (21) and (22) we get

$$\frac{n+1}{n} = \exp\left[\frac{h\nu}{kT}\right]. \tag{23}$$

The interesting point about (23) is that the left hand side owes its origin to time asymmetry in electrodynamics and cosmology whereas the right hand side is thermodynamic in origin.

In the above simple example we see the beginning of a theory which links the arrows of time in cosmology and thermodynamics. The link so far is not as strong as in the case of electrodynamics and cosmology. In the latter case we can assert that electromagnetic radiation exists because there is a suitably absorbing expanding universe. In the former case we have been able to show that the time asymmetries in thermodynamics are consistent with those in cosmology. By examination of detailed cases like the Planck radiation law above, we may be able to strengthen this link.

REFERENCES

[1] OLBERS, *Bode's Jahrbuch*, 110 (1826).
[2] HALLEY, E., *Phil. Trans. Roy. Soc. (London)* **31**, (1720).
[3] BONDI, H., *Cosmology*, Cambridge University Press, 1961.
[4] HUBBLE, E.P., *Proc. Natl. Acad. Sci.*, **15**, 168 (1929).
[5] SANDAGE, A. and TAMMAN, G. (private communication).
[6] BONDI, H. and GOLD, T., *Mon. Nat. Roy. Astr. Soc.* **108**, 252 (1948).
[7] WEBER, J., *Phys. Rev. Letters* **22**, 1320 (1969).
[8] WHEELER, J.A. and FEYNMAN, R.P., *Rev. Mod. Phys.* **21**, 425 (1949).
[9] SCHWARZSCHILD, K., *Gottinger Nachrichten* **128**, 132 (1903).
[10] TETRODE, H., *Z. f. Physik* **10**, 317 (1922).
[11] FOKKER, A.D., *Z. f. Physik* **58**, 386 (1929); *Physica* **9**, 33 (1929); **12**, 145 (1932).
[12] WHEELER, J.A. and FEYNMAN, R.P., *Rev. Mod. Phys.* **17**, 157 (1945).
[13] HOGARTH, J.E., *Proc. Roy. Soc.* **A 267**, 365 (1962).
[14] HOYLE, F. and NARLIKAR, J.V., *Proc. Roy. Soc.* **A 277**, 1 (1963).
[15] ROE, P.E., *Mon. Not. Roy. Astr. Soc.* **144**, 219 (1969).
[16] HOYLE, F. and NARLIKAR, J.V., *Ann. Phys.* **54**, 207 (1969).
[17] HOYLE, F. and NARLIKAR, J.V., *Ann. Phys.* **62**, 44 (1971).