13 August 1982, Volume 217, Number 4560 SCIENCE

Entropy in an Expanding Universe

Steven Frautschi

Summary. The question of how the observed evolution of organized structures from initial chaos in the expanding universe can be reconciled with the laws of statistical mechanics is studied, with emphasis on effects of the expansion and gravity. Some major sources of entropy increase are listed. An expanding "causal" region is defined in which the entropy, though increasing, tends to fall further and further behind its maximum possible value, thus allowing for the development of order. The related questions of whether entropy will continue increasing without limit in the future, and whether such increase in the form of Hawking radiation or radiation from positronium might enable life to maintain itself permanently, are considered. Attempts to find a scheme for preserving life based on solid structures fail because events such as quantum tunneling recurrently disorganize matter on a very long but fixed time scale, whereas all energy sources slow down progressively in an expanding universe. However, there remains hope that other modes of life capable of maintaining themselves permanently can be found.

If gravity is neglected, matter in a closed box approaches equilibrium—the state of maximum entropy. In the past century some people applied this description to the universe, arriving at a gloomy picture called the "heat death of the universe" in which the state of maximum entropy would eventually be reached. A look at our present picture of the history of the universe reveals a remarkably different and more interesting situation. In the beginning there is a hot gas, nearly homogeneous and in thermal equilibrium. As it expands it breaks into clumps of matter—galaxies, stars, planets, rocks, dust, and gas—with a wide range of temperatures. Some of these objects develop highly organized structures and, on at least one planet, self-replicating structures called "life" develop. Finally, a form of life emerges with the capability to ask questions about these systems.

The questions we will consider in this article are:

1) How can the observed evolution of organized structures from chaos be reconciled with the second law of thermodynamics?

2) Quantitatively, what are the main sources of entropy increase in the universe?

3) Will the heat death eventually occur, and if so in what form?

4) If the heat death does not occur, is sufficient free energy available to maintain life forever?

None of these questions could have been answered on the basis of physics known in the 19th century. Indeed, a good deal of the picture could not be filled in until J. D. Bekenstein and S. W. Hawking deduced the entropy of black holes, and their radiation properties, in the early 1970's.

As our topic is extremely speculative, it has been treated in only a few research works (*1–4*). Two very interesting general references are a book by Davies (*5*) and lectures by Dyson (*6*).

Assumptions and Basic Formulas

We will adopt the following assumptions and specializations:

1) The present standard laws of physics remain valid into the indefinite past and future.

2) The universe remains approximately homogeneous and isotropic and is thus describable by a Friedmann model (this assumption is not needed to establish many of our conclusions, and it may fail at late times, but we make it to establish a framework for quantitative estimates).

3) The universe expands without limit [that is, in questions dealing with the future we consider the open ($k = -1$) and critical ($k = 0$) Friedmann models but not the closed ($k = +1$) model, which presents a very different set of issues].

We will make repeated use of the formula for entropy in two limiting cases where it is known exactly.

Case 1. For a gas of N free particles with temperature T in a volume V

$$S = k\ln\,(\text{number of } N\text{-particle states}) \simeq k\ln\,(\text{number of 1-particle states})^N \qquad (1)$$

Evaluating the one-particle phase space, one finds (*7*) for particles of mass m

$$S = kN\ln\left[\frac{V}{N}\left(\frac{2\pi mkT}{h^2}\right)^{3/2} e^{5/2}\right] \qquad (2)$$

where k is the Boltzmann constant and h is Planck's constant, and a similar formula for massless particles. I use the simple approximation

$$S \simeq kN \qquad (3)$$

which is normally accurate within two orders of magnitude because, as noted by Fermi, all large logs are $\lesssim 100$ even in cosmology.

Case 2. For a black hole of mass $M_{\rm BH}$ (*8, 9*)

$$S = \frac{4\pi k\,G\,M_{\rm BH}^2}{\hbar c} \qquad (4)$$

The author is a professor in the Department of Physics, California Institute of Technology, Pasadena 91125.

0036-8075/82/0813-0593$01.00/0

where G is the gravitational constant and $\hbar = h/2\pi$. Numerically, the black hole entropy is of order k for the Planck mass M_{Planck} defined by

$$\frac{GM^2_{\text{Planck}}}{\hbar c} \simeq 1 \qquad (5)$$

(The Planck mass, of magnitude $M_{\text{Planck}} \simeq 10^{19}$ billion electron volts $\simeq 10^{-5}$ gram, is the smallest black hole mass allowed by quantum mechanics.) The entropy of a black hole with the mass of a star ($\simeq 10^{57}$ GeV) or larger is thus enormous.

One way to think of these huge entropies (8, 9) is as a measure of the total number of quantum levels inside the black hole, which (being undetectable) are all equally probable. From another point of view (9), the black hole eventually decays through Hawking radiation (10, 11) into quanta characterized by the Hawking temperature

$$kT_{\text{BH}} \simeq \hbar c/R_{\text{BH}} \qquad (6)$$

After decay the entropy of the radiated quanta can be estimated by means of case 1:

$$S(\text{Hawking radiation}) \simeq kN(\text{quanta}) \simeq \frac{kM_{\text{BH}}c^2}{kT_{\text{BH}}} \simeq \frac{kM_{\text{BH}}cR_{\text{BH}}}{\hbar} \qquad (7)$$

Since the black hole radius is given by

$$R_{\text{BH}} = 2M_{\text{BH}}/c^2 \qquad (8)$$

(from $GM^2_{\text{BH}}/R_{\text{BH}} \doteq M_{\text{BH}}c^2$), we have crudely

$$S(\text{Hawking radiation}) \simeq \frac{kGM^2_{\text{BH}}}{\hbar c} \qquad (9)$$

Thus the entropy in the black hole eventually appears in the directly detectable form of a huge number of low-energy quanta. The mass dependence has the surprising quadratic form $S \sim M^2_{\text{BH}}$ because the emitted quanta have wavelength $\lambda \simeq R_{\text{BH}} \sim M_{\text{BH}}$ and energy per quanta $\sim M^{-1}_{\text{BH}}$.

Armed with these formulas for entropy, we proceed to tackle our list of physical questions.

How Can the Evolution of Structure Be Reconciled with the Second Law?

The universe differs from a closed nongravitating box in three key respects: expansion, the long-range nature of gravity, and the interplay of relaxation times with the expansion rate.

Expansion. In general, entropy need not be conserved during an expansion even if the system remains in equilibrium (12). Consider, for example, nucleosynthesis in the big bang. We adopt an oversimplified picture of N free nucleons coupled to photons in a box (leptons are ignored). If the nucleons bind into alpha particles, the number of free particles is at first sight reduced by a factor of 4, tending to reduce the entropy. However, the 7-MeV energy release per nucleon is available to make new photons, tending to raise the entropy. The quantitative comparison, using Eq. 3, is

$$S(\text{n,p gas}) \simeq kN \qquad (10)$$

$$S(\alpha,\gamma \text{ gas}) \simeq k(N_\alpha + N_\gamma) \simeq k\left[\frac{N}{4} + \frac{N\,(7\text{ MeV})}{3kT}\right] \qquad (11)$$

(Here n, p, α, and γ denote neutrons, protons, alpha particles, and photons, respectively.) At $kT > 7$ MeV, the entropy of dissociated nucleons is higher and few α's are formed. At a critical T determined by

$$1 = \frac{1}{4} + \frac{7\text{ MeV}}{3\,kT} \qquad (12)$$

(that is, $kT \simeq 28/9$ MeV in our oversimplified model) the photon number has risen sufficiently to make α formation worthwhile from an entropy standpoint. Thus, as the universe cools by expansion, the favored state changes from free neutrons and protons to condensed nuclei (α, and later at a lower T, Fe) and the change results in an entropy increase.

The question now arises: granted that cooler, less dense states may have different entropies, what guarantees that their entropies will be larger rather than smaller? A partial answer, recently advocated by Bekenstein (13), is the following. If $S \sim kN$, the main way to increase S is to make more quanta. Since

$$N = E(\text{total})/E(\text{per quanta}) \qquad (13)$$

the way to make more quanta is to split particles into lower energy particles—preferably massless, since otherwise the process terminates. But the minimum energy per massless particle is

$$E_{\min} = cp_{\min} \simeq c\hbar/R \qquad (14)$$

where R is the radius of the volume under discussion. Thus

$$\frac{\max S}{E} \simeq \frac{kN_{\max}}{E} \simeq \frac{k}{E_{\min}} \simeq \frac{2\pi kR}{c\hbar} \qquad (15)$$

which increases as the system expands.

Long-range nature of gravity. Standard statistical mechanics is based implicitly on an assumption of short-range forces among particles. It is not, strictly speaking, valid in the presence of the long-range unshielded gravitational force (although it is valid to extremely good approximation over times short compared to the gravitational relaxation time). For example, in the standard statistical mechanics of a gas in equilibrium in a box, the intensive quantities such as temperature and pressure are uniform. But a sufficiently large box of galactic gas, initially in equilibrium, eventually clumps into stars under the action of gravity. The intensive quantities are now distributed nonuniformly. One might object that if the box is big enough to contain many stars, at least a coarse-grained uniformity survives. But even this is only temporary since the long-range gravitational force eventually introduces more clumping on all distance scales.

We can express this tendency quantitatively in the limiting case of black holes, where the entropy is exactly known. Using Eq. 4, we see that the entropy of one black hole of mass NM exceeds the entropy of N black holes each of mass M

$$\frac{S(NM)}{NS(M)} = N \qquad (16)$$

Thus entropy favors one large black hole over many small ones no matter how big the box; the bigger the box, the more extreme the nonuniformity.

Returning to the initial homogeneous hot gas in the early big bang, we now see that it would have been unstable even in the absence of expansion because, although its "thermal" and "chemical" entropy were maximized, its "gravitational" entropy was very small (5, 14). This goes far toward explaining the seeming paradox of how an initially homogeneous gas has been able to undergo such extensive development of structure. But it gives rise to another question: why was the gravitational entropy so small at early times?

At present we have no idea why the early universe was so homogeneous over distance scales which were not then within each other's light horizons. But suppose we accept this as given and introduce the concept of a "causal region," all parts of which can influence one another causally during (say) a doubling time of the expansion. We use doubling time because the temperature and density of a Friedmann model remain roughly fixed over this time scale.

By considering the entropy of a causal region we can gain a fresh perspective on the course of events. At any given time t, the maximum entropy obtainable from black hole formation in a causal region is

$$\max S_{\rm BH}(t) \simeq \frac{kGE_{\rm c}^2(t)}{\hbar c^5} \quad (17)$$

$E_{\rm c}(t)$ and $R_{\rm c}(t)$ are the energy and radius of the causal volume. Since $R_{\rm c}(t)$ grows as t and the energy density $\rho(t)$ falls as t^{-2} in the early universe, $E_{\rm c}(t)$ grows as

$$E_{\rm c}(t) \simeq \rho(t)R_{\rm c}^3(t) \sim t \quad (18)$$

Thus the maximum gravitational entropy in a causal region grows rapidly—more rapidly than the energy contained in the region. This relentless growth of max $S_{\rm BH}/E_{\rm c}$ is a particularly good example of the Bekenstein relation max $S/E \sim R$.

Turning this around, the maximum gravitational entropy in a causal region shrinks rapidly as we go backward in time. Taking $R_{\rm c}(t) \simeq ct$ and putting in numbers, we find

$$\max S_{\rm BH} \simeq k\left(\frac{t}{10^{-43}\ \rm sec}\right)^2 \quad (19)$$

The reference time 10^{-43} second is the time when the causal region contained just 10^{19} GeV of energy, corresponding to one Planck mass. If, instead of a black hole, the causal region contained free particles in thermal equilibrium, then the typical energy per particle was

$$kT \simeq 10^{19}\ {\rm GeV}\left(\frac{10^{-43}\ \rm sec}{t}\right)^{1/2} \quad (20)$$

and the particle entropy within the causal region was

$$S_{\rm particle} \simeq kN_{\rm particle} \simeq k\left(\frac{kT}{\hbar c}\right)^3 R_{\rm c}^3$$

$$\sim t^{-3/2}t^3 \simeq k\left(\frac{t}{10^{-43}\ \rm sec}\right)^{3/2} \quad (21)$$

A comparison of Eqs. 21 and 19 reveals two important points. First, the particle entropy grows more slowly than the maximum entropy. Second, if we are so bold as to extrapolate back to 10^{-43} second, when kT was of the order of the Planck energy, we find that the particle entropy $\simeq$ max $S_{\rm BH}$. At that moment a system of particles in thermal equilibrium was only marginally unstable against gravitational collapse; the entropy (while absolutely small) was of the same order as its possible maximum. Thus we have understood why the initial gravitational entropy within a causal volume was small. The remaining question is: why did the gravitational entropy fail to grow as fast as max $S_{\rm BH}$?

Relaxation times. The third major difference between the universe and a closed nongravitating box is that the universe falls out of equilibrium unless its relaxation time is less than the doubling time over which it expands appreciably. This condition becomes progressively harder to fulfill as the system thins out.

One well-known example is that the nucleon gas does not have time to nucleosynthesize all the way to the nuclear energy minimum at iron as it cools. The process is cut short by the decay of the free neutrons at 10^3 seconds, by the gap at total nucleon number $A = 5$, by thinning densities, and by the Coulomb barrier in proton-proton reactions, leaving a mixture of primarily p's and α's. Nucleosynthesis resumes only much later when matter reconcentrates in stars, and even then it fails to achieve rapid or full completion.

The failure of black holes to form or, if preexistent, to grow at maximal rates is another example. As long as the surrounding gas is ultrarelativistic, development and growth of local density fluctuations that might lead to black holes is inhibited by the high pressure, which tends to blow them apart. Preexistent black holes decay through Hawking radiation and disappear within a time

$$\tau_{\rm Hawking} \simeq 10^{-43}\ {\rm sec}\left(\frac{M_{\rm BH}}{M_{\rm Planck}}\right)^3 \quad (22)$$

which is very slow for star-sized black holes, but is immediate for black holes initially present on the scale of the Planck mass. So right from the beginning at 10^{-43} second, and certainly later at times when the physics is better understood, gravitational entropy in a causal region fails to keep pace with its maximum potential value.

We have thus come to a conclusion which stands the closed 19th-century model on its head. Far from approaching equilibrium, the expanding universe as viewed in a succession of causal regions falls further and further behind achieving equilibrium. This gives ample scope for interesting nonequilibrium structures to develop out of initial chaos (*15*), as has occurred in nature.

Numerical Estimates of Entropy Increase in a Model Universe

If a homogeneous, isotropic space filled with pure blackbody radiation or pure pressureless nonrelativistic gas expands, a comoving volume expanding with the space contains a constant number of quanta with constant entropy (*16*). Thus comoving volume is convenient for measuring the actual growth of entropy, whereas the causal volume was useful for considering the maximum possible rate of entropy growth.

To get a feeling for the numbers involved, let us consider some major sources of entropy increase in a comoving volume. We adopt a simple Friedmann model in which the universe is initially filled with radiation and devoid of black holes. We start a second or so after the big bang, when experimentally well-established laws of physics already apply and the radiation is salted with nucleons in the present ratio of about

$$n_\gamma/n_{\rm N} \simeq 10^9 \quad (23)$$

where $n_{\rm N}$ is the number of nucleons. We further assume that the eventual departures from homogeneity are limited to scales no larger than, say, superclusters of galaxies, an assumption which limits the size of the black holes that may form.

In the radiation-dominated universe, the scale of a comoving volume grows as

$$V_{\rm comoving} \simeq R^3_{\rm comoving} \sim t^{3/2} \quad (24)$$

where R is radius, while temperature and entropy follow Eqs. 20 and 21. Thus $S/V_{\rm comoving}$ is essentially constant during the radiation era (with modest increments from nucleosynthesis and various other events). The entropy is falling behind max $S_{\rm BH}$, however, throughout the radiation era (up to 10^{11} seconds) at the rate implied by Eqs. 19 and 21.

$$S/\max S_{\rm BH} \sim (10^{-43}\ {\rm sec}/t)^{1/2} \quad (25)$$

The situation changes when photons decouple at about 10^{11} seconds, allowing stars and galaxies to form. The clumping into gravitational potential wells and the resumption of nucleosynthesis within stars release energy that can be degraded into large numbers of low-energy quanta. The resulting entropy gains for several significant processes are listed in Table 1 and discussed below.

Entropy increase in stars (5). Nucleosynthesis near the center of a star releases about 7 MeV per nucleon. Part of the energy goes into neutrinos (ν's), which escape immediately, resulting in a modest entropy increase (several ν's per nucleon). The rest of the energy goes into γ's and positrons (e^+'s), which annihilate into γ's. These cannot escape immediately, so their energy is thermalized. The energy gradually flows outward through zones of decreasing temperature, with entropy steadily increasing as the photons degrade in energy. Finally the energy reaches the surface, where the temperature is of or-

Table 1. Major entropy increases in a comoving volume at times greater than 1 second for a model in which gravitational binding does not extend beyond superclusters of galaxies and black holes are initially absent. The increase listed for positronium formation and decay is the minimum estimate of Page and McKee (4) and applies to a $k = 0$ Friedmann model only.

Event	Duration (years)	Increase over entropy of blackbody radiation
Nucleosynthesis in stars	10^{11}	10^{-2}
Formation of stellar black hole	10^{11}	$\leq 10^{10}$
Formation of galactic black hole	$\leq 10^{20}$	$\leq 10^{21}$
Collapse of supercluster of galaxies into black hole	$\leq 10^{20}$	$\leq 10^{24}$
Nucleon decay	$\geq 10^{31}$	10^{3}
Flow of cold matter by quantum tunneling (if nucleons stable)	10^{65}	10^{6}
Black hole decay	$\leq 10^{106}$	$\leq 10^{24}$
Positronium formation and decay (if nucleons unstable)	$\geq 10^{116}$	$\geq 10^{13}$
Quantum tunneling of nuclei to iron (if nucleons stable)	10^{1500}	10^{15}
Quantum tunneling of matter into black holes (if nucleons stable)	$10^{10^{26}}$	10^{18}

der 5000 K or about 1 eV per photon for a typical star such as our sun. Thus 5×10^6 photons are radiated per original nucleon, for an entropy gain of 5×10^6 per nucleon.

Entropy increase on the earth. The entropy increase on the earth can be estimated in a similar way. The main energy source is solar radiation. Photons arriving from the sun have energies corresponding to T (solar surface) $\simeq$ 5000 K, whereas photons radiated by the earth have energies corresponding to T (earth surface) $\simeq$ 300 K. Since arriving and departing radiation is in approximate energy balance, about 5000/300 $\simeq$ 17 photons leave the earth per arriving photon for an entropy gain of 17 per arriving solar photon. Note that this increase, so crucial for life on the earth, is very minor on a cosmic scale.

Black hole formation. Much more substantial entropy production, overshadowing for the first time the 10^9 photons per nucleon in the blackbody radiation, occurs in black hole formation. From Eqs. 4 and 5, we found

$$S_{BH} \simeq k\,(M_{BH}/10^{19}m_p)^2 \qquad (26)$$

where m_p is the proton mass. A typical star contains about 10^{57} nucleons. In its youthful gaseous form it thus has an entropy

$$S_{star} \simeq 10^{57}k \qquad (27)$$

If it later evolves into a black hole its entropy becomes, according to Eq. 26,

$$S(\text{BH of } 10^{57}\, m_p) \simeq 10^{76}k \qquad (28)$$

an increase of 10^{19} per nucleon. On a time scale of 10^{11} years, a substantial fraction of all stars (and thus all nucleons) may generate entropy increases in this manner.

On a longer time scale, the entire galaxy is thought to evolve into a large central black hole. While the dominant mechanism and time scale are not well understood, gravitational Rutherford scattering of stars is certain to knock many stars out of the galaxy, and others down into its center, within about 10^{20} years (3, 17). The result is evaporation of a large fraction [Dyson (6) estimates 90 to 99 percent] from the outer regions, and collapse of the rest into a large central black hole with mass up to 10^{11} solar masses ($M_\odot$) $\simeq 10^{68}\, m_p$. The entropy of the large galactic black hole thus ranges up to

$$S(\text{BH of } 10^{68}\, m_p) \simeq 10^{98}k \qquad (29)$$

an increase of 10^{30} per nucleon on a time scale of 10^{20} years. [This estimate is extremely uncertain, because the fraction evaporated may be much greater and because other mechanisms may act on a shorter time scale. For example, there are suggestions that our galaxy has already developed a $10^6\, M_\odot$ central black hole (18), and some quasars may already have central black holes with masses of 10^7 to $10^9\, M_\odot$ (19).]

Since we have assumed a maximum scale of gravitational binding—for instance, superclusters of galaxies—black hole formation eventually comes to an end in our model, with masses up to $10^{14}\, M_\odot \simeq 10^{71}\, m_p$ and black hole entropies up to $10^{104}k$.

Black hole decay. When the temperature of blackbody or other ambient radiation falls below the Hawking temperature (6), the black hole becomes a net radiator. According to Eq. 22, the time scale for black holes to radiate away all their energy ranges from 10^{64} years for black holes of one solar mass to 10^{106} years for black holes of $10^{14}\, M_\odot$. At these late times the emitted quanta seldom react further, so the entropy is essentially given by the number of emitted quanta as stated in Eq. 9. On our very rough scale, the resulting entropy increase is of the same order as the parent black hole entropy.

Other major sources of entropy increase. The stars, planets, rocks, dust, and gas that escape the gravitational binding of galaxies and clusters of galaxies by evaporation avoid incorporation into black holes and remain available for further evolution. One possible development is decay of the nucleons at $t \geq 10^{31}$ years. In the alternative case that nucleons are stable, Dyson (6) lists a number of events which would increase the entropy at very late times: (i) liquid flow of cold matter by quantum tunneling at 10^{65} years, (ii) fusion or fission of all nuclei to iron by quantum tunneling at 10^{1500} years, and (iii) quantum tunneling of all bodies larger than $10^{19}\, m_p$ (the size of a dust grain) into black holes in $10^{10^{26}}$ years.

In processes (i) and (ii), or in nucleon decay, the energy release occurs in a large number of independent local events distributed randomly over the interior of the body. These heat the body; the heat is carried to the surface and radiated away. As in the case of a normal star, the entropy gain depends on the surface temperature.

At sufficiently late times, outside energy inputs (including the flux of cosmic blackbody photons) become negligible. A body will be in thermal equilibrium when the power generated within it balances radiation from the surface. For a spherical blackbody with radius R, containing N nucleons that undergo energy release ϵ per nucleon on a time scale τ, the surface temperature T is fixed by

$$\frac{N\epsilon}{\tau} = 4\pi R^2 \sigma T^4 \qquad (30)$$

where σ is the Stefan-Boltzmann constant.

Nucleon decay. In this case $\epsilon \simeq m_p c^2$ (a fraction of the energy is lost to neutrinos, which escape without thermalizing). If we take $\tau = 10^{31}$ years, we find surprisingly high temperatures for degenerate matter: $T \simeq 50$ K for neutron stars, 2 K for a solar mass white dwarf, 0.1 K for the earth, and so forth (4, 20). The resulting entropy increase per nucleon, of order $k(m_p c^2/5kT)$, comes out to $5 \times 10^{10}k$ for neutron stars, $10^{12}k$ for a solar mass white dwarf, and so on: a substantial increase, but not competitive with formation of large black holes.

Liquid flow of cold matter. In this case the energy release from chemical rearrangement, settling of heavy elements toward the core, and so on, is specific to each type of body. It can never be more than a small fraction of Mc^2, and for objects of planetary mass or smaller one can take 10 eV per nucleon as characteristic (*21*). Inserting $\tau \simeq 10^{65}$ years into Eq. 30, we obtain surface temperatures about 10^{-11} cooler than those estimated for nucleon decay and entropy increases per nucleon about 1000 times greater than in the case of nucleon decay.

Quantum tunneling of nuclei to iron. In this case $\epsilon \simeq 1$ MeV and $\tau \simeq 10^{1500}$ years. The equilibrium temperature comes out below the lowest phonon energy for any cold body small enough to be stable against gravitational collapse, so it is the lowest phonon energy rather than T that sets the scale of radiation from the surface. The lowest phonon has wavelength $\simeq R$ and energy $\simeq hv/R$, where v is the velocity of sound (about 10^5 cm/sec) and R is the radius of the body. Thus the entropy gain per nucleon is

$$S \simeq \frac{k\epsilon}{hv/R} \simeq 10^{21}k\,\frac{R}{\text{km}} \qquad (31)$$

This is about $10^{26}k$ for the largest cold bodies such as Jupiter and $10^{24}k$ for black dwarfs. This is a huge gain, though still somewhat smaller than the gain attainable in a galactic black hole.

Quantum tunneling of matter into black holes. The tunneling time is shortest for the smallest black holes, of mass $10^{19}\ m_p$. Once formed, these decay instantly (10^{-43} second) into 10^{19}-GeV particles. These relativistic particles shoot their way out of small objects, but will be stopped by large objects, resulting in local heating. The analysis for large bodies proceeds as in the previous case with $\epsilon \simeq m_p c^2$. The resulting entropy gain is $10^{27}k$ for black dwarfs.

Positronium formation and decay. Free positrons are created in small quantities by black hole decay and in large quantities if the nucleons in gas and dust decay. In an open ($k = -1$) Friedmann model many of them never recombine, but in a critical ($k = 0$) Friedmann model they do eventually recombine into positronium (*3*). Page and McKee (*4*) pointed out that in the $k = 0$ model recombination occurs gradually, forming increasingly loosely bound positronium, which emits increasingly large numbers of photons in the course of cascading to the ground level. If nucleons are unstable, Page and McKee estimate that entropy increases of at least $10^{22}k$ per parent nucleon occur.

Summary of entropy increases and S/max S. Surveying the results in Table 1, we note that the entropy in a comoving volume approaches an asymptotic limit in our model, though only on an enormous time scale. The period of most rapid entropy increase begins with the formation of stars and ends with the formation of black holes on the largest scales that are bound.

Table 2. Time dependence of the ratio of the entropy of a free gas to max S in a causal region for a critical ($k = 0$) universe. The late universe is taken to remain matter-dominated because even if nucleons decay, enough electron pairs remain to constitute a significant part of the total energy density at almost all late times, according to Page and McKee (*4*).

Epoch	R_{causal}	R_{comoving}	$S_{\text{gas}} = \dfrac{V_{\text{causal}}}{V_{\text{comoving}}}$	$\dfrac{S_{\text{gas}}}{\max S}$
Radiation-dominated ($t < 10^3$ years)	$\sim t$	$\sim t^{1/2}$	$\sim t^{3/2}$	$\sim t^{-1/2}$
Matter-dominated ($t < 10^3$ years)	$\sim t$	$\sim t^{2/3}$	$\sim t$	$\sim t^{-1}$

It is also interesting to consider the ratio S/max S in a causal region as a function of time. In a critical ($k = 0$) universe, max S_{BH} grows as t^2 at all times. The entropy of a free gas grows as $t^{3/2}$ or t, depending on whether the gas is ultrarelativistic (radiation dominant) or nonrelativistic (matter dominant); the behaviors are summarized in Table 2. Adding in the various entropy-generating processes in Table 1, especially black hole formation, we arrive at the approximate curve for S/max S displayed in Fig. 1.

In an open ($k = 1$) universe, the time development is initially the same as in a critical universe. But later the radius of a comoving volume starts expanding at a rate $\sim t$. Once this happens, causal regions grow essentially no faster than comoving volumes on the largest scales, so S_{gas} and the number of particles in a causal region on the scale of the light horizon stop growing. In this relatively empty type of universe, black holes no longer maximize the entropy. Matter dominance is maintained even if nucleons decay, because free electron pairs annihilate only partially in an open universe (*3*). Given these circumstances, the Bekenstein limit max $S \simeq 2\pi kRE/c\hbar$ of Eq. 15 apparently cannot be realized. As pointed out by Page (*13*), a more realistic bound is that which would be obtained if all the matter could suddenly be transmuted into radiation: max $S \simeq k(RE/c\hbar)^{3/4}$. The actual matter-dominant S and E, on the other hand, become constant. Thus S/max S falls asymptotically as $R^{-3/4} \sim t^{-3/4}$, yielding a curve qualitatively similar to, though less steep than, Fig. 1.

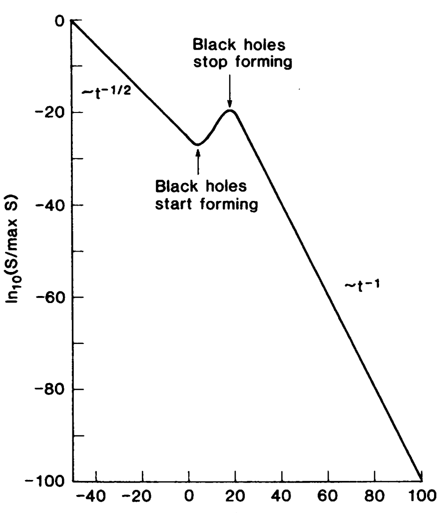

Fig. 1. Time development of S/max S in a causal region of a critical ($k = 0$) universe, assuming the model described in the text and no large entropy generation during the first second. The development of S/max S in an open ($k = -1$) universe is qualitatively similar, with $t^{-3/4}$ rather than t^{-1} behavior at large t.

It is apparent from Fig. 1 that the entropy in a causal region falls steadily further behind max S during most of cosmic history. S/max S does increase temporarily during the period of stellar and galactic black hole formation. Life as we know it develops during the same period, utilizing the much smaller but conveniently arranged entropy generation on a planet or planets situated near nucleosynthesizing stars.

Heat Death

As we have seen, modern cosmology does not terminate in the classical heat death of the 19th century. The classical heat death was characterized by statistical equilibrium of matter at constant temperature and entropy. An expanding universe never achieves equilibrium and never reaches a constant temperature.

Nevertheless, the conclusion reached in the previous section was equally gloomy. With our assumptions the expanding universe does "die" in the sense that the entropy in a comoving volume asymptotically approaches a constant limit.

Can this gloomy conclusion be avoided? What comes to mind immediately is our assumption that superclusters of galaxies are the largest gravitationally bound objects. This limited the size scale obtained for black hole formation. If the sequence of gravitationally bound objects were to continue indefinitely to all size scales, it seems possible that formation of progressively larger black holes might proceed indefinitely, causing the entropy per comoving volume to grow without limit, though never reaching its maximum. In such a universe there would be no heat death.

Detailed analysis of this possibility must be left for the future. The black holes must amalgamate on a fairly precise time schedule, with sufficient speed and regularity to prevent their disappearance by Hawking radiation at any stage. To bind on all scales without closing appears to require a nearly critical universe with kinetic and gravitational potential energy nearly balanced on all scales. But it is far from clear that the universe would remain forever of the Friedmann type (homogeneous and isotropic on sufficiently large scales) if binding existed on all scales.

Page and McKee (*4*), in addition to mentioning the possibility of limitless entropy generation by black hole amalgamation, pointed out that positronium formation on steadily increasing scales may generate a limitless supply of entropy. Whether the number of photons generated by cascade decay per positron is 10^{22}, the minimum used in Table 1, or limitless depends on details of the reheating and redissociation of positronium by cascade photons and on the energy balance between positronium and its radiation (*4*). Actually, continued black hole formation and positronium formation are related insofar as both involve binding of ever-larger structures in a long-range potential, both require a nearly critical universe to keep potential energy competitive with kinetic energy, and both involve long-range multibody interactions which are hard to estimate.

If Heat Death Does Not Occur, Is Free Energy Available to Maintain Life?

At root, discussions of the heat death are driven by our curiosity or anxiety concerning life, its eventual fate, and its significance. Life is notoriously resistant to definition, let alone mathematical analysis. But it does appear generally to require growth and development (that is, changes in entropy) as well as form and order (submaximal entropy). Thus arises the connection between the heat death, even in our modified version where the entropy per comoving volume approaches a low but changeless limit, and the end of life.

Following Dyson (*6*), we take as a necessary condition for the immortality of life

$$\int \frac{dS}{dt}\, dt = \infty \qquad (32)$$

Naturally, the entropy increase must involve real growth in a comoving volume, not just the counting of noninteracting quanta in larger and larger regions of space. As we have seen, the only known mechanisms for unlimited growth of entropy in a comoving volume are continued coalescence of black holes, or positrons and electrons, on all scales.

The entropy requirement (Eq. 32) is a necessary but not a sufficient condition for life. Dyson (*6*) entrusts the rest to a sufficiently resourceful intelligence. The challenges this intelligence must eventually meet include the required search for fresh free energy sources, the concomitant need for space travel, and the adjustment to progressively lower temperatures and lower rates of activity necessitated by the finiteness of all known free energy sources. There is also a less obvious problem: if kT fell below the lowest phonon excitation energy ΔE, all process rates, waste heat radiation, and information accumulation would die as $\exp(-\Delta E/kt)$ whereas at $\Delta E < kT$ they fall only as a power of T. Thus any life based on solid structures must eventually adjust to longer and longer length scales to keep $\Delta E < kT$.

There are several reasons why positronium formation would be a less useful perpetual free energy source for life than black hole amalgamation:

1) Black hole amalgamation, if it occurs, is capable of supplying greater entropy increases (*4*).

2) Black holes provide a concentrated source of free energy, whereas the photons radiated by positronium, being spread evenly over space, are harder to collect.

3) To meet the need for expanding length scales described above, life based on solid matter must continually search out and bring together more matter. This requires energy. Hawking radiation from a black hole can yield enough energy to meet this requirement, whereas positronium radiation cannot.

We therefore turn to black holes as the free energy source, and envision how life might attempt to maintain itself indefinitely, and even play a major role in shaping the universe. A sufficiently resourceful intelligence inhabiting a critical universe learns how to move black holes (*22*), bringing them together from increasingly widely separated locations and merging them to increase the entropy. The region from which the black holes are collected, which I will call the empire (E), has radius

$$R_E \sim t^p \qquad (33)$$

and energy

$$E_E \sim R_E^3\ \rho \sim t^{3p-2} \qquad (34)$$

(we take $\rho \sim t^{-2}$, as appropriate in a matter-dominated critical universe). Thus, if black holes have a large fraction of the energy, an efficient collection and towing system can concentrate the energy in a black hole with mass

$$M_{BH} \sim t^{3p-2} \qquad (35)$$

at time t.

In each doubling time, the volume of the empire increases by 2^{3p} so several new black holes come within its boundaries. These must be moved over distances $R_E \sim t^p$ within a time of order t, so the required towing velocity is of order

$$v \simeq R_E/t \sim t^{p-1} \qquad (36)$$

The energy cost of towing, of order $M_E v^2$ per doubling time, can be kept well below the energy release by Hawking radiation (which can be as high as $M_E c^2$); provided $p < 1$, the towing time is not much less than t, and the lifetime for evaporation by Hawking radiation is not much greater than t (*23*). [In the detailed scheme presented below, only $\sim t^{-1/3}$ of the Hawking radiation is captured by life, but $v^2 \sim t^{-4/9}$ ($p = 7/9$), so sufficient energy is available for towing.]

As already noted, the maintenance of life involves a compromise: the entropy must increase, but not so rapidly as to reach maximum. In our scenario one can be precise about the optimum rate of black hole formation. In view of Eqs. 22 and 35, the Hawking decay of the large black holes created in the empire proceeds with a lifetime

$$\tau \sim M_{BH}^3 \sim t^{9p-6} \qquad (37)$$

If $\tau << t$, the black holes evaporate before they can be merged, and the empire dies for lack of a concentrated free energy source. If $\tau >> t$, the black holes radiate very little of their energy during a doubling time, and the empire is starved for usable energy. For sustaining life, the optimum radiant lifetime (*24*) is $\tau \sim t$; that is,

$$p = 7/9 \qquad (38)$$

In this case

$$M_{\rm BH} \sim t^{1/3} \qquad (39)$$

and the entropy of the empire scales as

$$S_{\rm E} \sim t^{2/3} \qquad (40)$$

To collect the energy radiated by a black hole with (from Eqs. 6, 8, and 39)

$$T_{\rm BH} \sim R_{\rm BH}^{-1} \sim M_{\rm BH}^{-1} \sim t^{-1/3} \qquad (41)$$

intelligent life might inhabit a shell of radius $R_s \sim t^{1/3}$ surrounding the black hole. Waste heat would be radiated to outer space, which at blackbody temperatures $T_{\rm BB} \sim t^{-2/3}$ would always be colder. Mechanical stability requires a minimum thickness for the shell. A complete spherical shell would need an amount of material proportional to $R_s^2 \sim t^{2/3}$, but the total mass in the empire grows only as $t^{1/3}$. To hold the material requirements down to $M_s \sim t^{1/3}$, a Fuller dome construction utilizing fixed-thickness rods with length scaled as $l \sim t^{1/3}$ might be employed.

Equation 40 tells us that

$$dS_{\rm E}/dt \sim t^{-1/3} \qquad (42)$$

The material in the shell would cover a fraction $\sim R^{-1} \sim t^{-1/3}$ of the full solid angle, so it could absorb a fraction $\sim t^{-1/3}$ of the energy radiated by the black hole and generate entropy at the rate

$$dS_s/dt \sim t^{-2/3} \qquad (43)$$

According to Dyson (6), life would have problems of heat disposal which would require it to "hibernate" a fraction $[1 - g(t)]$ of the time. In Dyson's formulation $g(t)$ scales as the temperature of the life zone, $T_{\rm life}(t)$. This would prevent life from generating entropy continuously at a rate as high as $t^{-2/3}$. Nevertheless, as one sees by taking, for example,

$$g(t) \sim T_{\rm life}(t) \sim T_{\rm BH}(t) \sim t^{-1/3} \qquad (44)$$

life could produce entropy at a rate scaling as $t^{-2/3}$ during its active phases and t^{-1} overall, which would still allow its integrated entropy generation to go to infinity in our model. Thus the model seems to reach the goal of life without end, with the striking feature that life permanently modifies the overall environment to sustain itself, amalgamating black holes and raising the entropy above natural levels by a growing and eventually infinite factor.

Can this scenario survive closer scrutiny? Without getting into biological, chemical, or engineering details, we can find a fatal flaw in the system on basic physical grounds. In discussing entropy increases in matter at very late times, we identified several mechanisms such as (i) possible nucleon decay at $t \geq 10^{31}$ years, (ii) liquid flow of cold matter by quantum tunneling at $\simeq 10^{65}$ years, (iii) nuclear fusion by tunneling at 10^{1500} years, and (iv) quantum tunneling to black holes in $10^{10^{26}}$ years. Even if the nucleon is stable, the other processes are sure to occur eventually. They recurrently disorganize matter, necessitating repair work to maintain life, on a fixed time scale. Thus the power requirement for repair of the empire goes as $M_s(t) \sim t^{1/3}$. This is fatal because the power available from the strongest enduring source, black hole radiation, scales down with time as $dE/dt \sim t^{-2/3}$.

Although we have failed to find a viable scheme for preserving life based on solid structures, other forms of organization may be possible, as emphasized by Dyson. It stands as a challenge for the future to find dematerialized modes of organization (based on dust clouds or an e^+e^- plasma?) capable of self-replication. If radiant energy production continues without limit, there remains hope that life capable of using it forever can be created.

References and Notes

1. P. C. W. Davies, *Mon. Not. R. Astron. Soc.* **161** 1 (1973).
2. J. N. Islam, *Q.J.R. Astron. Soc.* **18**, 3 (1977); *Sky Telesc.* **57**, 13 (1979).
3. J. D. Barrow and F. J. Tipler, *Nature (London)* **276**, 453 (1978).
4. D. N. Page and M. R. McKee, *Phys. Rev. D* **24**, 1458 (1981); *Nature (London)* **291**, 44 (1981).
5. P. C. W. Davies, *The Physics of Time Asymmetry* (Univ. of California Press, Berkeley, 1974), chaps. 4 and 7.
6. F. J. Dyson, *Rev. Mod. Phys.* **51**, 447 (1979).
7. K. Huang, *Statistical Mechanics* (Wiley, New York, 1963), p. 154.
8. J. D. Bekenstein, *Phys. Rev. D* **7**, 2333 (1973).
9. S. W. Hawking, *ibid.* **13**, 191 (1976).
10. ______, *Nature (London)* **248**, 30 (1974).
11. ______, *Commun. Math. Phys.* **43**, 199 (1975).
12. R. C. Tolman, *Relativity, Thermodynamics, and Cosmology* (Oxford Univ. Press, New York, 1934), section 175.
13. J. D. Bekenstein, *Phys. Rev. D* **23**, 287 (1981). Various aspects of Bekenstein's detailed formulation have been criticized by W. G. Unruh and R. M. Wald [*ibid.* **25**, 942 (1982)]; D. N. Page ["Comment on a universal upper bound on the entropy-to-energy ratio for bounded systems," Pennsylvania State University preprint (1981)]; and J. Ambjørn and S. Wolfram ["Properties of the vacuum: 1. Mechanical and thermodynamic," California Institute of Technology preprint 68-855 (1981)].
14. R. Penrose, in *General Relativity: An Einstein Centenary*, S. W. Hawking and W. Israel, Eds. (Cambridge Univ. Press, Cambridge, 1979), chap. 12.
15. D. Layzer, *Sci. Am.* **233** (No. 6), 56 (1975).
16. R. C. Tolman, *Relativity, Thermodynamics, and Cosmology* (Oxford Univ. Press, New York, 1934), sections 168 and 169.
17. A. P. Lightman and S. L. Shapiro, *Rev. Mod. Phys.* **50**, 437 (1978).
18. T. R. Geballe, *Sci. Am.* **241** (No. 1), 60 (1979); J. H. Lacy, C. H. Townes, T. R. Geballe, D. J. Hollenbach, *Astrophys. J.* **241**, 132 (1980).
19. D. Lynden-Bell, *Nature (London)* **223**, 690 (1969).
20. G. Feinberg, *Phys. Rev. D* **23**, 3075 (1981); D. A. Dicus, J. R. Letaw, D. C. Teplitz, V. L. Teplitz, *Astrophys. J.* **252**, 1 (1982). It is interesting to trace out the decay of a cold neutron star at times comparable to the nucleon lifetime and later. As it loses mass, it slowly moves along the cold matter equilibrium curve down to the minimum neutron star mass, where an explosive expansion occurs. Subsequently, the star settles down onto the white dwarf branch of the cold matter equilibrium curve, expands along it slowly until it reaches the size of Jupiter, and then contracts to earth-sized, asteroid-sized, and rock-sized objects before disappearing. Meanwhile the surface temperature is falling; for example, in the post-Jupiter stage of essentially constant density, $N \sim \exp(-t/\tau)$, $R \sim \exp(-t/3\tau)$, and $T \sim \exp(-t/12\tau)$ according to Eq. 30.
21. D. J. Stevenson, private communication.
22. The first suggestion that black holes be towed and used as energy sources, albeit on a more modest scale, was made by L. Wood, T. Weaver, and J. Nuckolls [*Ann. N.Y. Acad. Sci.* **251**, 623 (1975)].
23. In the case of positronium, the collectible mass in a region growing as $R \sim t^p$ is again $M \sim R^3\rho \sim t^{3p-2}$, and the towing energy requirement per doubling time is again $Mv^2 \sim t^{5p-4}$. However, the radiant energy collection per doubling time by life based on a thin sheet of solid mass with area $A \sim M \sim t^{3p-2}$ is only ρ (radiation) $At \sim t^{3p-3}$. For $p > 1/2$, the towing energy requirement exceeds the energy collection. Since the collection region must grow faster than a comoving volume to supply the necessary perpetual growth in mass, $p > 2/3$ is required and energy collection is inadequate.
24. This optimum applies independent of whether the black holes are brought together by intelligent intervention or coalesce spontaneously. In the latter case, life would still have to perform some towing of matter to engineer the growth of the inhabited shell, and in general the movement of the black holes would need to be accelerated or braked to optimize their rate of coalescence.

25. Supported in part by U.S. Department of Energy contract DE-AC-03-81-ER40050. Much of the foregoing work was done at the Aspen Center for Physics, where the interdisciplinary atmosphere played a vital role in stimulating this study. I particularly thank D. Gross, P. Ramond, I. Redmount, K. Thorne, S. Wolfram, and A. Zee for informative and encouraging discussions.

Science

Entropy in an Expanding Universe

Steven Frautschi

Science, 217 (4560), • DOI: 10.1126/science.217.4560.593

Science (ISSN 1095-9203) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.