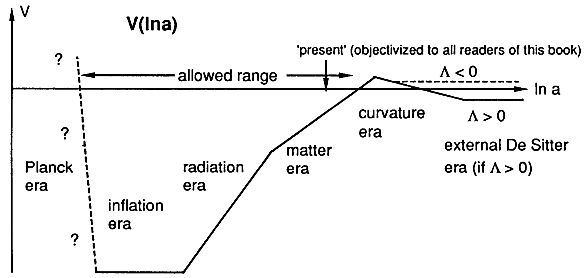

Fig. 5.6. Schematic behavior of the 'potential energy' for the dynamics of $\ln a$ in the case of positive spatial curvature. Since only regions with positive kinetic energy $E - V = -V > 0$ are allowed, turning points of the cosmic expansion would arise at values of $V = 0$. An upper turning point would lead to recontraction, while a lower turning point describes a 'bouncing' universe (without big bang or big crunch singularities)

tive positive cosmological constant of unknown origin ('dark energy') already contributes two thirds of the potential $V(a)$.

5.3.1 Instability of Homogeneity

While the Friedmann model is an exact solution of the Einstein equations, and apparently a reasonable approximation to the very large scale behavior of the real Universe, it is not stable against density fluctuations (as discussed in the introduction to this chapter and in Sect. 5.1). This local instability cannot be compensated by a *global* force, such as a cosmological constant. It is in fact successfully used to explain the formation of stars, galaxies, galaxy clusters, possibly larger structures, and eventually black holes in the present Universe. Thereby, the *assumed* initial symmetries of the Friedmann universe must be dynamically broken. In classical physics, density fluctuations would be microscopically determined (Sect. 3.4). In quantum theory they may also result from an indeterministic (genuine or apparent) collapse of the wave function, induced by decoherence (see Calzetta and Hu 1995, Kiefer, Polarski and Starobinsky 1998, and Sect. 6.1). A similar quantum effect is known to limit the retardation of symmetry-breaking phase transitions (their hysteresis). The onset of these primordial structures of the Universe is now believed to be observed in the cosmic background radiation.

The arrow of time characterizing these irreversible processes is thus again based on an improbable (but 'simple') cosmic initial condition: homogeneity. When Boltzmann (1896) discussed the origin of the Second Law in the context of an infinite and eternal universe, he had to conclude that we, here and now, are living in the aftermath of a gigantic cosmic fluctuation. Its maximum (that is, a state of very low entropy) must have occurred in the distant past in order to explain the existence of fossils and other documents in terms of causal history and evolution (see Sect. 3.5).

How improbable is the novel initial condition of homogeneity that Boltzmann did not even recognize as an essential assumption? We may calculate its probability by means of Einstein's relation (3.56) if we know the entropy of the most probable state. The entropy of a non-degenerate homogeneous physical state in local equilibrium is proportional to the number of particles, N. All other parameters enter this expression only logarithmically – as exemplified for the ideal gas in (3.14). In the present Universe, the number of photons contained in the 2.7 K background radiation exceeds that of massive particles by a factor 10^8. The entropy of a finite 'standard universe' of 10^{80} baryons (now often regarded as no more than a 'bubble' in a much larger or infinite universe) would therefore possess an entropy of order 10^{88} plus a small but important contribution resulting from gravitating objects. Most of this entropy must therefore have been produced in the early Universe by the creation of photons and other particles, which are strongly entangled in a chaotic way.

However, the present entropy is far from its maximum that would be achieved by the production of black holes. In Planck units, the horizon area of a neutral and spherical black hole of mass M is given by $A = 4\pi(2M)^2$. Its entropy according to (5.15) thus grows with the *square* of its mass,

$$S_{\text{bh}} = 4\pi M^2 \; . \tag{5.23}$$

Merging black holes will therefore produce an enormous amount of entropy. If the standard universe of 10^{80} baryons consisted of 10^{23} solar mass black holes (since $M_{\text{sun}} \approx 10^{57} m_{\text{baryon}}$), it would already possess a total entropy of order 10^{100}, that is, 10^{12} times its present value. If most of the matter eventually formed a single black hole, this value would increase by another factor of 10^{23}. The probability for the present, almost homogeneous universe is therefore a mere

$$p_{\text{hom}} \approx \frac{\exp(10^{88})}{\exp(10^{123})} = \exp(10^{88} - 10^{123}) \approx \exp(-10^{123}) \tag{5.24}$$

(Penrose 1981), indistinguishable in this approximation from the much smaller probability at the big bang. Gravitational contraction thus offers an enormous further entropy capacity to assist the formation of structure and complexity.

This improbable initial condition of homogeneity as an origin of thermodynamical time asymmetry is different from attempts (see Gold 1962) to derive this arrow from a homogeneous expansion of the Universe in a causal manner (see Price 1996 and Schulman 1997 for critical discussions). While it is true that non-adiabatic expansion of an equilibrium system may lead to a retarded non-equilibrium, this would equally apply to non-adiabatic *contraction* in our causal world. The growing space (and thus phase space, representing increasing entropy capacity) cannot form the master arrow of time, since it is insufficient to explain causality (the absence of any advanced correlations). Non-adiabatic *compression* of a vessel would lead to retarded pressure waves emitted from the walls, but not to a reversal of the thermodynamical arrow. The entropy capacity of gravitational contraction is far more important than homogeneous expansion, but probably not very relevant for the very early stages of the Universe.

There are other examples of *using* causality in thermodynamical arguments rather than *deriving* it in this cosmic scenario. For example, Gal-Or (1974) discussed retarded equilibration due to the slow nuclear reactions in stars. Even though nuclear fusion controls the time scale and energy production during most stages of stellar contraction, it presumes a strong initial non-equilibrium.

5.3.2 Inflation and Causal Regions

The finite age of an expanding universe that starts from an initial singularity (a big bang) leads to the consequence that the backward light cones of two events may not overlap. These events would then not be causally connected. A sphere formed by the light front originating in a point-like event at the big bang, where $a(0) = 0$, is therefore called a *causality horizon*. Its radius $s(t)$ at Friedmann time t is given by

$$s(t) = \int_0^t \frac{a(t)}{a(t')} \mathrm{d}t' \, . \qquad (5.25)$$

In a matter- or radiation-dominated universe, this integral would converge for $t' \to 0$, and thus define a finite horizon size. Only *parts* of the Universe may then be causally connected – excluding even readily observable distant pairs of objects that strongly indicate a simultaneous origin.

In particular, the homogeneity of the universe on the large scale would thereby remain causally unexplained. This *horizon problem* was the major motivation for postulating a phase transition of the vacuum or another mechanism of quantum fields that would lead to a transient cosmological constant, and thus to an early de Sitter era. In an exponentially expanding universe, the big bang singularity could in principle be shifted arbitrarily far into the past – depending on the duration of this era. However, in an extremely short time span (of the order of 10^{-33} s), the universe, and with it all causality horizons, would have been *inflated* by a huge factor that was sufficient for the sources of the whole now observable cosmic background radiation to be causally connected (Linde 1979). On the other hand, since causality horizons started with zero radius, this would explain the initial absence of nonlocal correlations and entanglement, provided they were assumed to *require* a causal origin.

Measurements of the cosmic background radiation indicate that an inflation era did in fact occur. Since the corresponding repulsive force counteracts gravity, it has also been conjectured to have driven the universe into a state of homogeneity in a causal manner. This *cosmological no-hair conjecture* is supported by a theorem of Hawking and Moss (1982). However, this theorem remains insufficient for the required purpose, since the global effect of a cosmological constant cannot generally force *local* gravitating systems, in particular black holes, to expand into a state of homogeneity. Proofs of the cosmic no-hair theorem had therefore to exclude positive spatial curvature. (Expanding white holes would require acausally incoming advanced radiation, as explained in Sect. 5.1.)

Since a cosmological constant that was simulated by a phase transition of the vacuum would depend on the local density, it may at least overcompensate the effect of gravity until strong inhomogeneities begin to form. This may *partly* explain the homogeneity of the observed part of our universe. It can be described by saying that the Weyl tensor 'cooled down' as a consequence of this spatial expansion – similar to the later red-shifting of the primordial electromagnetic radiation. While these direct implications of the expansion of the universe define reversible phenomena, equilibration during the radiation era or during the phase transition would be irreversible in the statistico-thermodynamical sense (based on microscopic causality).

This explanation of homogeneity is incomplete as it has to presume the absence of *strong* initial inhomogeneities (abundant initial black holes, in particular). In order to work in a deterministic theory, it would furthermore require the state that precedes inflation to be even less probable than the homogeneous state after inflation.

Similar inflationary scenarios have been discussed in various hypothetical models of quantum cosmology (see Caroll and Chen 2004, and Chap. 6).

5.3.3 Big Crunch and a Reversal of the Arrow

These questions may also be discussed by means of a conceivable recontracting universe. A consistent analysis of the arrow of time for this case is helpful regardless of what will happen to our own Universe. Would the thermodynamical arrow have to reverse direction when this universe starts recontracting towards the big crunch after having reached maximum extension? The answer would have to be 'yes' if the cosmic expansion represents the master arrow, but it is often claimed to be 'no' on the basis of causal arguments if they are continued into this region. For example, some authors argued that the background radiation would reversibly heat up during contraction (blue-shifting), while the temperature gradient between interstellar space and the fixed stars would first have to be inverted in order to reverse stellar evolution long after the universe had reached its maximum extension. However, this argument presupposes the overall validity of the 'retarded causality' in question, that is, the absence of future-relevant correlations in the contraction phase. It would be justified if the relevant initial condition held at only one 'end' of this otherwise symmetric cosmic history. The absence or negligibility of any anti-causal events in our present epoch seems to indicate either that our Universe is thermodynamically asymmetric in time, or that it is still 'improbably young' in comparison to its total duration.

Paul Davies (1984) argued in a similar causal manner that there can be no reversed inflation leading to a homogeneous big crunch, since correlations which would be required for an inverse phase transition have to be excluded for being extremely improbable. Instead of a homogeneous big crunch one would either obtain locally re-expanding 'de Sitter bubbles' forming an inhomogeneous 'bounce', or inhomogeneous singularities at variance with a reversed Weyl tensor condition, or both. This probability argument fails, however, if the required correlations are *caused in the backward direction of time* by a final condition that was thermodynamically a mirror image in time of the initial one (see also Sect. 6.2.3). Similarly, if the big bang was replaced by a non-singular *homogeneous bounce* by means of some kind of 'Planck potential' (Fig. 5.6), entropy must have decreased prior to the bounce. In particular, decoherence would have to be replaced by recoherence in all contraction eras. In this case, an observer complying with the Second Law would always experience an expanding universe; the sign of the dynamical time parameter used in this description is merely formal (see Sect. 5.4).

On the other hand, a low entropy big bang *and* an equivalent big crunch may lead to severe consistency problems, since the general boundary value problem (Sect. 2.1) allows only one complete (initial or final) condition. Although the requirement of low entropy is not a complete boundary condition, statistically independent two-time conditions would lead to the square of the already very small probability of (5.24), that is,

$$p_{\text{two-time}} = p_{\text{hom}}^2 \approx \left[\exp(-10^{123})\right]^2 \approx \exp(-10^{123.301}) \ . \tag{5.26}$$

The RHS appears as a small correction to (5.24) only because of this double-exponential form, although an element of phase space corresponding to (5.26) could now easily be much smaller than a Planck cell (see Zeh 2005b). A two-time boundary condition of homogeneity may thus be inconsistent with 'ergodic' quantum cosmology (that would have to include the repeated formation and decay of black holes, which contribute most of phase space).

The consistency of general two-time boundary conditions has been investigated for simple deterministic systems (see Cocke 1967 and Schulman 1997). Davies and Twamley (1993) discussed the more realistic situation of classical electromagnetic radiation in an expanding and recollapsing universe. According to their estimates, our Universe will remain essentially transparent all the way between the two opposite radiation eras (in spite of the reversible red- and blue-shifting over many orders of magnitude in between) – in contrast to ergodic assumptions used in (5.26). Following a suggestion by Gell-Mann and Hartle, they concluded that light emitted *causally* by all stars before the 'turning of the tide' propagates freely until it reaches the time-reversed radiation era – thus giving rise to an asymmetric history of this universe.

David Craig (1996) argued on this basis, but by *assuming* a thermodynamically time-symmetric universe, that the night sky at optical frequencies should contain an almost homogeneous component that represents the advanced radiation from stars existing during the contraction era. It should be observable as a non-Planckian high frequency tail in the isotropic background radiation with a total intensity at least equalling that of the light now observed from all stars and galaxies in our past – but probably much higher because of the advanced light corresponding to that which will have to be produced until the turning point is reached. However, since classical radiation would preserve all information about its origin, it is inconsistent with a time-reversed absorber (the opposite radiation era), that allows only its thermal radiation in *its* causal future (Sect. 2.2). Craig also concluded that the intensity of the thermal part of the background radiation would be doubled because of the two radiation eras, but this does not seem to be required, since the 'two' *thermal* components may be identical. (Retarded and advanced fields do not add – see Sect. 2.1 – but they must be consistent with one another.) Only in the non-thermal frequency range can retarded and advanced radiation be conceptually distinguished and thus carry information about their origin.

These conclusions have to be modified in an essential way when the quantum aspect of electromagnetic radiation is taken into account. The information content of radiation consisting of photons is limited, as first emphasized by Brillouin (1962). This consequence had also turned out to be important for Borel's argument of Sect. 3.1.2 – see footnote 4 of Chap. 3. Each photon, even if emitted into intergalactic space as a spherical wave, disappears from the whole quasi-classical universe as soon as it is absorbed *somewhere*. A reversal of this process would again require recoherence, that is, the superposition of many Everett branches. This argument requires consistent quantum cosmology (Chap. 6), where initial or final conditions can only affect the total, unitarily evolving Everett wave function. If the Schrödinger dynamics was instead modified by means of a collapse of the wave function (as implicitly assumed also for Gell-Mann and Hartle's 'histories'[2]), the corresponding new

[2] Gell-Mann and Hartle (1994) discussed quantum mechanical 'histories', which are defined in terms of time-ordered series of projections in Hilbert space. These *individual* histories are thus equivalent to successions of stochastic collapse events (global quantum jumps) – even though a collapse is not explicitly used. The authors nonetheless discussed the possibility of a thermodynamically time reversal-symmetric cosmic history by presuming a final condition that is similar to the initial one. This proposal is based on the equivalence of the upper and lower diagrams of Fig. 4.4, but neglects the asymmetric structure (4.56) of a collapse, which would have to include all retarded entanglement with 'information gaining systems'. Therefore, it leads to insurmountable problems as soon as one attempts to justify the probabilistic interpretation ('consistent histories') by an in practice irreversible decoherence process (see Fig. 4.5). Time reversal symmetry could be restored in the contraction era only by means of a complete process of recoherence. This would not only have to include those Everett components that have been disregarded by the Hilbert space projections which lead to individual measurement outcomes, and in this way define quasi-classical 'histories' as a *partial* quantum reality. It should also require components that have to be regarded as being retro-caused in the future.

dynamical law would have to be reversed, too, in order to save a thermodynamically time-symmetric (but now indeterministic) universe.

This problem of consistent cosmic two-time boundary conditions will assume a conceptually quite novel form in the context of quantum gravity, where any fundamental concept of time disappears from the description of a closed universe (Sect. 6.2).

5.4 Geometrodynamics and Intrinsic Time

In general relativity, the 'block universe picture' is traditionally preferred to a dynamical description, as its unified spacetime concept is then manifest. So it took almost half a century before its dynamical content was sufficiently understood, in particular by means of its Hamiltonian form, invented by Arnowitt, Deser and Misner (1962). This approach, which is essential for a quantization of the theory, has not always been welcomed, as it seems to destroy the beautiful relativistic spacetime concept by reintroducing a 3+1 (space and time) representation. However, only in this *form* can the dynamical content of general relativity be fully appreciated (see Chap. 21 of Misner, Thorne and Wheeler 1973). A similarly symmetry-violating form in spite of Lorentz invariance is known for the electromagnetic field when described in the Coulomb gauge by the vector potential $\boldsymbol{A}$ as the dynamical field configuration on a space-like hypersurface of Minkowski spacetime.

This dynamical reformulation requires the separation of unphysical gauge degrees of freedom (which in general relativity simply represent the choice of coordinates), and the skillful handling of boundary terms. The result of this technically demanding procedure turns out to have a simple interpretation. It describes the *dynamics of the spatial geometry* ('three-geometry') ${}^{(3)}G(t)$, that is, a propagation of the intrinsic curvature on space-like hypersurfaces with respect to a time coordinate t that labels a foliation of the spacetime arising dynamically in this way. This foliation has to be *chosen* simultaneously with the construction of the solution. The extrinsic curvature, which describes the embedding of the three-geometries into spacetime, is represented by the corresponding canonical momenta. The configuration space of three-geometries ${}^{(3)}G$ has been dubbed *superspace* by Wheeler, since the form of its kinetic energy defines a metric. Trajectories in this superspace define four-dimensional spacetime geometries ${}^{(4)}G$.

This 3+1 description may appear ugly not only as it hides Einstein's beautiful spacetime concept, but also since the foliation of a given ${}^{(4)}G$ by means of space-like hypersurfaces, on which ${}^{(3)}G(t)$ is defined, is quite arbitrary. Many trajectories ${}^{(3)}G(t)$ therefore represent the same spacetime ${}^{(4)}G$, which is absolutely defined. It is only in special situations – such as for the FRW metric (5.20) – that there may be a 'preferred choice' of coordinates, which then reflect their exceptional symmetry. The time coordinate t, characterizing a foliation, is just one of the four arbitrary (physically meaningless) spacetime