The generic nature of the global and non-entropic arrow of time and the dual role of the energy–momentum tensor

2004 J. Phys. A: Math. Gen. 37 4445

(http://iopscience.iop.org/0305-4470/37/15/012)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **37** (2004) 4445–4463 PII: S0305-4470(04)70170-5

The generic nature of the global and non-entropic arrow of time and the dual role of the energy–momentum tensor

Mario Castagnino[1] **and Olimpia Lombardi**[2]

[1] CONICET—Instituto de Astronomía y Física del Espacio, Casilla de Correos 67, Sucursal 28, 1428 Buenos Aires, Argentina
[2] CONICET—Universidad Autónoma de Madrid, Ctra. Colmenar Km 15, 28049 Madrid, Spain

Received 6 October 2003
Published 29 March 2004
Online at stacks.iop.org/JPhysA/37/4445 (DOI: 10.1088/0305-4470/37/15/012)

Abstract

In this paper we adopt a generic, global and non-entropic approach to the problem of the arrow of time, according to which the arrow of time is a generic, intrinsic and geometrical property of spacetime. We demonstrate that the arrow of time so defined is generic in the sense that any spacetime with physically reasonable properties (e.g. time-orientability and global time) will be endowed with an arrow of time. The only exceptions are very special cases belonging to a subset of zero measure of the set of all possible spacetimes. We also show the dual role played by the energy–momentum tensor in the context of our approach. On one hand, the energy–momentum tensor is the intermediate step that permits us to turn the geometrical time-asymmetry of the universe into a local arrow of time manifested as a time-asymmetric energy flow. On the other hand, the energy–momentum tensor supplies the basis for deducing the time-asymmetry of quantum field theory, posed as an axiom in this theory.

PACS number: 98.80.Dr

1. Introduction

Since the nineteenth century, the problem of the arrow of time has been one of the most controversial questions in the foundations of physics and one of the main concerns of many physicists. In the last few years we have been very interested in this problem and have devoted several papers to the cosmological origin of the arrow of time [1]. This research has reached its culmination in papers [2] and [3], where we have presented a comprehensive formulation of our view on the subject based on a generic, global and non-entropic approach, according to which the arrow of time is a generic, intrinsic and geometrical property of spacetime.

The main aim of this paper is twofold. First, we will demonstrate the generic nature of the global and non-entropic arrow of time. Here we will provide a demonstration more general than that presented in [2] and [3], where the generic nature of the arrow of time was proved in the particular case of Friedmann–Lemâitre–Robertson–Walker (FLRW) models. Second, we will show the dual role played by the energy–momentum tensor in the context of our approach. On one hand, the energy–momentum tensor is the intermediate step that allows us to turn the geometrical time-asymmetry of the universe into a local arrow of time manifested as a time-asymmetric energy flow. On the other hand, the energy–momentum tensor supplies the basis for deducing the time-asymmetry of quantum field theory posed as an axiom in this theory.

We are completely aware of the fact that this global and non-entropic approach is very different from the usual opinions on the subject. Most physicists will find it strange in light of traditional assumptions according to which the arrow of time is inescapably linked to entropy and complexity notions of thermodynamics. In this paper we will try to show that a different perspective is possible, and that it is reasonable to pay attention to a new proposal in seeking an answer to a problem as controversial as the arrow of time. In order to reach this goal we devote the first section of the paper to the justification of our global and non-entropic approach. On this basis, the paper is organized as follows. In section 2 the problems of irreversibility and the arrow of time are precisely stated, and the global and non-entropic character of our approach is justified in contrast with traditional approaches. Section 3 is devoted to the introduction of the definition of the global and non-entropic arrow of time and the presentation of the theorem that proves the generic nature of such an arrow. In section 4 the arguments developed in the previous section are applied to the case of FLRW models. In section 5 we show the role played by the energy–momentum tensor in translating the global geometrical arrow into the local level. Finally, in section 6 we show how the energy–momentum tensor can be used to justify the time-asymmetry postulate of quantum field theory on global grounds.

2. The justification of the local and non-entropic approach

2.1. Irreversibility versus arrow of time

Traditionally, the problems of irreversibility and the arrow of time have been wrongly identified, as if irreversibility was the clue for understanding the origin and the nature of the arrow of time. On the other hand, the concepts of irreversibility and time-reversal invariance are invoked in the discussions about the arrow of time, but usually with no elucidation of their precise meanings. For this reason, we will begin by providing some necessary definitions.

Definition 1. *A dynamical equation is* time-reversal invariant *if it is invariant under the transformation $t \to -t$; as a result, for each solution $f(t)$, $f(-t)$ is also a solution.*

Definition 2. *A solution of a dynamical equation is* reversible *if it corresponds to a closed curve in phase space*[3].

It is quite clear that both concepts are different to the extent that they apply to different mathematical entities: time-reversal invariance is a property of dynamical equations (laws) and, *a fortiori*, of the set of their solutions (evolutions); reversibility is a property of a single

[3] This definition can be easily generalized to statistical and quantum cases: an evolution is irreversible if it has a final (or initial) equilibrium state, namely, a 'point of no return'.

solution of a dynamical equation. Furthermore, they are not even correlated: for instance, time-reversal invariant equations can have irreversible solutions[4].

When the fact that these two concepts apply to different mathematical entities is recognized, *the problem of irreversibility* can be clearly stated: *how to explain irreversible evolutions in terms of time-reversal invariant laws*. The difference between both concepts also shows that the problem of irreversibility has a simple conceptual answer: nothing prevents a time-reversal invariant equation from having irreversible solutions. Of course, a great deal of theoretical work is needed for obtaining irreversible evolutions from an underlying time-reversal invariant dynamics[5]. For instance, much effort has been directed towards explaining the irreversible behaviour of quantum systems in terms of the time-reversal invariant quantum theory (see, e.g., [4]). Here we only mean that, in order to face the problem of irreversibility, the question of the arrow of time does not need to be addressed. In fact, the distinction between the two directions of time is *usually presupposed* when the irreversible evolutions are conceived as processes going from non-equilibrium to equilibrium or from preparation to measurement *towards the future*.

By contrast, the problem of the arrow of time arises when we seek a physical correlate of the intuitive asymmetry between past and future. The main difficulty to be encountered in solving this problem is the fact that there is nothing in physical evolution laws that distinguishes, in a non-arbitrary way, between past and future as we conceive them in everyday life. It might be objected that physics implicitly assumes this distinction with the use of temporally asymmetric expressions, such as ‘future light cone’, ‘initial conditions’, ‘increasing time’ and so forth. However, this is not the case: in physics the distinction between past and future is *conventional*, that is, only based on assigning different names to formally identical objects. The problem of the arrow of time, in contrast, requires us to find a *substantial* difference [5, 6] between the two temporal directions (for a full discussion of this subject, see [2]). This means that we must address the problem from a perspective purged of our temporal intuitions: as Price [7] claims, it is necessary to stand at a point outside time, and thence to regard reality in atemporal terms.

But then, what does ‘the arrow of time’ mean when we accept this constraint? Of course, the traditional expression coined by Eddington has only a metaphorical sense: its meaning must be understood by analogy. We recognize the difference between the head and the tail of an arrow on the basis of its geometrical properties; therefore, we can substantially distinguish between both directions, head-to-tail and tail-to-head, *independently of our particular perspective*. Analogously, we will conceive *the problem of the arrow of time* in terms of *the possibility of establishing a substantial distinction between the two directions of time on the basis of exclusively physical arguments*.

2.2. A generic, global and non-entropic approach

(i) *Why global?* The traditional local approach owes its origin to the attempts to reduce thermodynamics to statistical mechanics: in this context, the usual answer to the problem of

[4] For instance, the equation of motion of a pendulum with Hamiltonian

$$H = \frac{1}{2} p_\theta^2 + \frac{K^2}{2} \cos\theta$$

is *time-reversal invariant*, namely, it is invariant under the transformation $\theta \rightarrow \theta$, $p_\theta \rightarrow -p_\theta$; however, whereas the trajectories within, above and below the separatrices are reversible since they are closed curves, the separatrices themselves are *irreversible* since the corresponding motions tend asymptotically to the point where the separatrices intersect themselves and the $p_\theta = 0$ axis.

[5] Usually it is necessary to use coarse-grainings, traces, restricted algebras, etc.

the arrow of time consists in defining the future as the direction of time in which entropy increases. However, already in 1912 Ehrenfest and Ehrenfest [8] had noted that, if the entropy of a closed system increases towards the future, such an increase is matched by a similar one in the past evolution of the system. The argument can also be posed in different terms. Let us assume that we have solved the irreversibility problem; so we have the description of all the irreversible evolutions, say, decaying processes, of the universe. However, since we have not yet established a substantial difference between both directions of time, there is no way to decide towards which temporal direction *each* decay proceeds. Of course, we would obtain the arrow of time if we could coordinate the processes in such a way that all of them parallely decay towards the same temporal direction. But this is precisely what local physics cannot offer: only by means of global considerations can all the decaying processes be coordinated. This means that the global arrow of time plays the role of the background scenario where we can meaningfully speak of the temporal direction of irreversible processes, and this scenario cannot be built up by means of local theories that only describe phenomena confined in small regions of spacetime.

(ii) *Why non-entropic?* In the late nineteenth century, Boltzmann offered the first global approach to the problem of the arrow of time [9]. Since this seminal work, many authors have related the temporal direction past-to-future to the gradient of the entropy function of the universe (see, e.g., [10, 11]). This global entropic approach rests on two assumptions: it is possible to define entropy for a complete cross-section of the universe, and there is an only time for the universe as a whole. However, both assumptions involve difficulties. In the first place, the definition of entropy in cosmology is still a very controversial issue: there is no consensus regarding how to define a global entropy for the universe. In fact, it is usual to work only with the entropy associated with matter and radiation because there is no clear idea yet about how to define the entropy due to the gravitational field. In the second place, when general relativity comes into play, time cannot be conceived as a background parameter which, as in pre-relativistic physics, is used to mark the evolution of the system. But there is a further conceptual argument for abandoning the global entropic approach. As is well known, a given value of entropy is compatible with many configurations of a system: entropy is a phenomenological property. But, if the arrow of time reflects a substantial difference between both directions of time, it is reasonable to consider it as an intrinsic property of time, or better, of spacetime, and not as a secondary feature depending on a phenomenological property. For these reasons we follow Earman's '*Time Direction Heresy*' [12], according to which the arrow of time is an intrinsic, geometrical property of spacetime embodied in $g_{\mu\nu}(x)$, which does not need to be reduced to non-temporal features. In other words, the geometrical approach to the problem of the arrow of time has conceptual priority over the entropic approach, since the geometrical properties of the universe are more basic than its thermodynamic properties.

(iii) *Why generic?* The global entropic approach, which defines the future as the direction of time in which the entropy of the universe increases, has to posit a low-entropy initial state of the universe. Then, the problem is pushed back to the question of why that initial state has low entropy. But a low-entropy initial state is extraordinarily improbable in the collection of all possible initial states. According to Davies [11], the expansion of the early universe during the first few minutes caused it to depart from equilibrium; but this fact requires us to suppose that the gravitational field at the time of the big bang was in a much-less-than-maximum entropy state. Penrose [13], in turn, suggests that the early universe begins with a high degree of spatial homogeneity and isotropy; but this uniform distribution is actually a low-entropy state when gravity is taken into account, since the purely attractive force of gravity is such that a uniform distribution is highly improbable. In either case, the global entropic approach is committed to explain an extraordinarily improbable initial condition such as the low-entropy or the near-flatness of the spacetime. In contrast, in our global and non-entropic approach there are no improbable conditions which require an explanatory account. We will prove that, if the spacetime has some minimal topological properties such as time-orientability and global time (which, anyway, are also necessary conditions for the existence of the arrow of time in the entropic approach), the arrow of time is a *generic* feature of the universe. This means that, in the collection of all possible spacetimes, the universes endowed with the global and non-entropic arrow of time are overwhelmingly probable: the non-existence of the arrow of time is what requires an extraordinarily fine tuning of all the variables of the universe.

3. The definition of the global and non-entropic arrow of time

3.1. Conditions: time-orientability and global time

As is well known, many different topologies are consistent with Einstein's field equations. In particular, there arises the possibility of spacetime being curved along the spatial dimension in such a way that, e.g., the spacelike sections of the universe become the three-dimensional analogue of a Möbius band; in this case it is said that the spacetime is non-time-orientable.

Definition 3. *A spacetime is* time-orientable *if there exists a continuous non-vanishing vector field $\gamma^{\mu}(x)$ on it which is non-spacelike everywhere.*

By means of this field, the set of all semicones of the manifold can be split into two equivalence classes conventionally called C_+ and C_-: the semicones of C_+ contain the vectors of the field and the semicones of C_- do not contain them. On the other hand, in a non-time-orientable spacetime it is possible to transform a future pointing timelike vector into a past pointing timelike vector by means of a continuous transport that always keeps non-vanishing timelike vectors timelike; therefore, the distinction between future semicones and past semicones is not definable in a univocal way on a global level. This means that the time-orientability of spacetime is a precondition for defining a global arrow of time since, if spacetime is not time-orientable, it is not possible to distinguish between the two temporal directions for the universe as a whole[6].

But time-orientability still does not guarantee that we can talk of *the* time of the universe, that is, a time that synchronizes the clocks fixed to all the particles of the universe. In fact, a spacetime may be such that it is not possible to partition the set of all events into equivalent classes such that: (i) each is on a spacelike hypersurface and (ii) the hypersurfaces can be ordered in time. There is a hierarchy of conditions which, applied to a time-orientable spacetime, avoid 'anomalous' temporal features (see [16]). One of the strongest conditions is the existence of a global time function, which increases along every future directed non-spacelike curve.

[6] Nevertheless, not all accept this conclusion. For instance, Mattews [14] claims that a spacetime may have a regional but not a global arrow of time if the arrow is defined by means of local considerations. But this opinion implies ignoring the principles of *universality* (according to which the laws of physics are valid at all points of the spacetime) and of *uniqueness* (which states that there is only one universe and completely disconnected spacetimes are not allowed), unquestioningly accepted in contemporary cosmology (for a full argument, see [2]). Although there are quantum cosmologies exhibiting disconnected spacetimes, such models only play an explanatory role since they are unvarifiable in principle. Anyway, even if disconnected spacetimes were allowed, each connected region could be considered a universe in itself. This fact is relevant since we are interested in explaining the arrow of time of our own connected universe (for a different opinion, see [15] and our criticism in [2]).

Definition 4. *A* global time function *on the Riemannian manifold M is a function* $t : M \rightarrow \mathbb{R}$ *whose gradient is timelike everywhere.*

The existence of such a function guarantees that the spacetime can be globally split into hypersurfaces of simultaneity which define a *foliation* of the spacetime (see [17]). Of course, the existence of a global time imposes a significant topological and metric limitation on the spacetime. But with no global time, there is not a single time of the universe and, therefore, it is nonsensical to speak of the two directions of time for the universe as a whole. Therefore, the possibility of defining a global time is a precondition for meaningfully speaking of a global arrow of time.

This fact provides an additional argument against the global entropic approach, which takes for granted the possibility of defining the entropy function of the universe. This amounts to the assumption that the spacetime can be partitioned in spacelike hypersurfaces on which the entropy of the universe can be defined (that is, the spacetime has a global time function). When the possibility of spacetimes with no global time is recognized, it is difficult to deny the conceptual priority of the geometrical structure of spacetime over its entropic features in the context of our problem.

3.2. Time-asymmetry

It is quite clear that time-orientability and existence of global time are merely necessary conditions for defining the global arrow of time: they do not provide a physical, substantial criterion for distinguishing between the two directions of time. As we will see, such a distinction requires the time-asymmetry of the universe.

It is usually accepted that the obstacle to the definition of the arrow of time relies on the fact that the fundamental laws of physics are time-reversal invariant[7]. Nevertheless, this common viewpoint can be objected to when the concept of time-symmetry is clearly elucidated and compared with the concept of time-reversal invariance: whereas time-reversal invariance is a property of dynamical equations (laws), time-symmetry is a property of a single solution (evolution) of a dynamical equation.

Definition 5. *A solution* $f(t)$ *of a dynamical equation is time-symmetric if there is a time* t_S *such that* $f(t + t_S) = f(t - t_S)$.

Of course, the time-reversal invariance of an equation does not imply the time-symmetry of its solutions: a time-reversal invariant law may be such that all or most of the possible evolutions relative to it are individually time-asymmetric. Price [7] illustrates this point with the familiar analogy of a factory which produces equal numbers of left-handed and right-handed corkscrews: the production as a whole is completely unbiased, but each individual corkscrew is asymmetric.

In the context of general relativity, the concept of time-symmetry cannot be defined in such simple terms, but it can be formulated by means of the concept of time isotropy:

Definition 6. *A time-orientable spacetime* (M, g) *(where M is a four-dimensional Riemannian manifold and g is a Lorentzian metric for M) is* time isotropic *if there is a diffeomorphism d of M onto itself which reverses the temporal orientations but preserves the metric g.*

On the basis of this definition, the concept of time-symmetry for a spacetime which admits a global time can be characterized as follows.

[7] The exception is the law that rules weak interactions. We will return to this issue in the last section.

Definition 7. *A time-orientable spacetime which admits a global time is* time-symmetric *with respect to some spacelike hypersurface* Σ *corresponding to a constant value of the global time if it is time isotropic and the diffeomorphism* d *leaves the hypersurface* Σ *fixed.*

Intuitively, this means that, from the hypersurface Σ, the spacetime *looks the same* in both temporal directions. Therefore, if a time-orientable spacetime is time-asymmetric, we will not find a spacelike hypersurface Σ which splits the spacetime into two 'halves', one being the temporal mirror image of the other with respect to their intrinsic geometrical properties.

But, how does this time-asymmetry allow us to choose a time orientation of the spacetime? As we have seen, in a time-orientable spacetime a continuous non-vanishing non-spacelike vector field $\gamma^{\mu}(x)$ can be defined all over the manifold. Up to now, the universe is *time-orientable* but not yet *time oriented*, because the distinction between $\gamma^{\mu}(x)$ and $-\gamma^{\mu}(x)$ is just conventional. Now time-asymmetry comes into play: in a time-asymmetric spacetime, *any* hypersurface Σ splits the manifold into two sections that are different from each other: the universe on one side of Σ is *substantially* different from the universe on the other side. We can choose any point $x_0 \in \Sigma$ and conventionally consider that the non-spacelike vector $\gamma^{\mu}(x_0)$ points towards one side of Σ and that $-\gamma^{\mu}(x_0)$ points towards the other or vice versa: in any case we have established a substantial difference between $\gamma^{\mu}(x_0)$ and $-\gamma^{\mu}(x_0)$ since these vectors point to substantially different parts of the universe. We can conventionally call the direction of $\gamma^{\mu}(x_0)$ 'future' and the direction of $-\gamma^{\mu}(x_0)$ 'past' or vice versa, but in any case past is substantially different from future. Now we can extend this difference to the whole continuous fields $\gamma^{\mu}(x)$ and $-\gamma^{\mu}(x)$ obtained by making the global continuations of $\gamma^{\mu}(x_0)$ and $-\gamma^{\mu}(x_0)$ permitted by the definition of time-orientability: in this way, the time orientation of the spacetime has been established. Since the field $\gamma^{\mu}(x)$ is defined all over the manifold, it can be used locally at each point x to define the future and the past semicones: for instance, if we have called the direction of $\gamma^{\mu}(x)$ 'future', $C_{+}(x)$ contains $\gamma^{\mu}(x)$ and $C_{-}(x)$ contains $-\gamma^{\mu}(x)$.

3.3. The generic nature of the arrow of time

As Savitt [18] correctly points out, there are two different questions involved in the problem of the arrow of time:

- The 'how possible' question: how is it possible to formulate a time-asymmetric model by means of time-reversal invariant laws?
- The 'how probable' question: what is the reason to suppose that time-asymmetry is probable?

We have already shown that the time-reversal invariance of laws is not an obstacle to the construction of time-asymmetric models where the substantial distinction between the two directions of time can be established. In this subsection we will address the second question by proving the *main result* of our argumentation: in time-orientable spacetimes endowed with a global time, *time-asymmetry is generic* to the extent that time-symmetric solutions of the universe equations have measure zero in the corresponding phase space.

We will begin with a simple case. Let us consider a time variable function $x(t)$ defined by a time-reversal invariant second-order equation $\ddot{x}(t) = F[x(t)]$. By definition of time-reversal invariance, if $x_1 = x(t)$ is a solution of the equation, then $x_2 = x(-t)$ is also a solution. In the general case, the initial conditions at $t = 0$ of both solutions are $x_1(0) = x_2(0)$ and $\dot{x}_1(0) = -\dot{x}_2(0)$. But in the particular case in which $\dot{x}_1(0) = -\dot{x}_2(0) = 0$, both solutions $x_1(t)$ and $x_2(t)$ coincide, leading to a single solution $x_1(t) = x_2(-t) = x(t)$ which is time-symmetric with respect to the axis $t = 0$. This means that whereas the space of the initial conditions corresponding to all the possible solutions is a two-dimensional space defined by $(x(0), \dot{x}(0))$, the space of the initial conditions corresponding to time-symmetric solutions is a one-dimensional space defined by $(x(0), \dot{x}(0) = 0)$. If the dynamical equation describes the time evolution of a typical system of classical mechanics, the time-symmetry conditions become $(q(0), p(0) = 0)$. Therefore, we can conclude either that the subset of time-symmetric solutions has measure zero in the set of all possible solutions (this claim cannot be generalized to spaces with no measure) or that the space of time-symmetric solutions is a proper subspace of the space of all possible solutions (this conclusion is also valid in infinite-dimensional spaces). Nevertheless, in both cases the central idea is the same: *time-asymmetry is generic*, and time-symmetry is (extraordinarily) specific.

The argument developed for a simple case can now be extended to general relativity and spacetimes governed by the time-reversal invariant Einstein field equations. Since time-orientability and existence of a global time are necessary conditions for defining the arrow of time, we will focus our attention on spacetimes describable by the ADM formalism, where the foliability of the spacetime is assumed[8]. In this context, a 4-geometry of a spacetime is said to be time-symmetric when there exists a spacelike hypersurface Σ such that the extrinsic curvature vanishes at all its points: $K_{ij} = 0$ $(i, j = 1, 2, 3, g = \det g_{ij})$ [19]. On the other hand, the initial conditions are given by the following Cauchy data: g_{ij} (that is, the geometry of Σ) and its time derivative $\dot{g}_{ij} = \frac{\partial g_{ij}}{\partial t}$ (where t is the global time variable of the ADM metric). But these data amount to having g_{ij} and its conjugated momentum π^{ij}, which can be computed as (see [19] equation (21.91)):

$$\pi^{ij} = g^{1/2}(g^{ij} g_{hk} K^{hk} - K^{ij}). \tag{1}$$

According to this equation, if the time-symmetry condition $K_{ij} = 0$ holds, then $\pi^{ij} = 0$. This means that the Cauchy data corresponding to spacetimes that are time-symmetric with respect to Σ are given by g_{ij} arbitrary and $\pi^{ij} = 0$ (in complete agreement with the simple case presented above where $q(0)$ is arbitrary and $p(0) = 0$). Let us denote by $\mathcal{E}$ the space of all the possible Cauchy data, $\{g_{ij}(x^1, x^2, x^3), \pi^{ij}(x^1, x^2, x^3)\}$, and by $\mathcal{E}_S$ the space of the Cauchy data that lead to time-symmetric geometries, $\{g_{ij}(x^1, x^2, x^3), \pi^{ij}(x^1, x^2, x^3) = 0\}$.[9] Analogous to the previous case we can conclude that $\mathcal{E}_S$ is a proper subspace of $\mathcal{E}$ and, therefore, *time-asymmetric spacetimes are generic* while time-symmetric spacetimes are extraordinarily specific.

As we can see, this global geometrical approach does not have to face the problem of explaining overwhelmingly improbable initial conditions in order to account for the arrow of time. In contrast to the global entropic approach, from this perspective the arrow of time is a completely generic feature which does not require an exceptional fine tuning of all the variables of the universe.

4. Application to FLRW models

In the previous section we have considered the general conditions necessary for the existence of a global and non-entropic arrow of time. Nevertheless, the argumentation would not be complete if we did not apply those conditions to the particular case of our universe.

[8] This demonstration could be extended to more general spacetimes by using the framework of [16] (p 251), but this is not relevant to our case since we are interested in the problem of the arrow of time.

[9] Although we have discussed the case of empty universes, the demonstration can be easily generalized to universes with a scalar field or an electromagnetic field [19], and to many other solutions of the Cauchy problem [20, 21].

As is well known, on the basis of the cosmological principle and the assumption of expansion, the (averaged) metric of the spacetime can be represented in the simple Robertson–Walker form:

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\,\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2\right). \tag{2}$$

With the Robertson-Walker metric, Einstein's field equations can be solved: the corresponding solutions describe the isotropic and homogeneous Friedmann–Lemâitre–Robertson–Walker models, which are the standard models of present-day cosmology.

In the following subsections we will show that FLRW models satisfy the necessary conditions for the existence of a global and non-entropic arrow of time, and we will define such an arrow in this particular case.

4.1. Time-orientability

FLRW models provide a simple answer to the question of the first condition for the existence of the arrow of time since they are time-orientable. Moreover, astronomical observations provide empirical evidence that makes implausible the non-time-orientability of our spacetime. In particular, there is no astronomical observation of temporally inverted behaviour in some (eventually very distant) region of the universe.

According to contemporary cosmology, at the decoupling time (400 000 years after the big bang) the universe was essentially composed of light elements such as H and He (with traces of isotopes of these elements and of Li and Be) and was virtually free of ions. This matter condensed in stars, where heavy elements were formed. The explosion of these stars as supernovae scattered the heavy elements producing clouds which, in turn, condensed giving rise to a second generation of stars. Analogously, a third generation appeared, and so forth. We have indirect evidence of the first generation (known as population III), appearing 200 million years after the big bang, by the observed reionization of the universe which can be explained only by the presence of stars during that period [22]. Then, two new generations (populations II and I) followed. This is the history as manifest in all the universe.

- In this history, the evolutions of the different generations of supernovae always follow the same pattern (let us say, from birth to death), and there is no trace of a time-inverted pattern in the visible universe. This is a relevant fact in the context of our problem since supernovae are the markers or standard candles used to measure the more distant galaxies ($z = 1$ and even more) [23].
- In turn, the peak in the quasar formation rate took place when the universe was 3000 million years old ($z \sim 2$), and this rate has always decreased since then (as shown by astronomical observations in all spatial directions). This fact can be considered as a good indication that the universe is time-orientable, at least up to the corresponding distances.
- Furthermore, the process of decoupling between matter and radiation occurred 400 000 years after the big bang ($z = 1000$, the limit of the visible universe in all spatial directions) and never happened again. This fact also counts as evidence for the time-orientability of the universe (from 400 000 years after the big bang up to the present).

4.2. Global time and cosmic time

As we have seen, the existence of a global time is a precondition for the arrow of time. Nevertheless, cosmological models may have an even stronger temporal condition. In fact, a global time function does not yet permit us to define the notion of simultaneity in a univocal manner and with physical meaning. In order to avoid ambiguities in the notion of simultaneity, we must choose a particular foliation. The foliation τ such that there exists a continuous set of worldline curves which are orthogonal to all the hypersurfaces $\tau = \text{const}$ is the proper choice, because orthogonality recovers the notion of simultaneity of special relativity for small regions (tangent hyperplanes) of the hypersurfaces $\tau = \text{const}$ (for the necessary conditions, see [19]). However, even if this condition selects a particular foliation, it permits that the proper time interval between two hypersurfaces of simultaneity be relative to the particular worldline considered for computing it. If we want to avoid this situation, we must impose an additional constraint: the proper time interval between two hypersurfaces $\tau = \tau_1$ and $\tau = \tau_2$ must be the same when measured on any orthogonal worldline curve of the continuous set mentioned above. In this case, the metric results

$$\mathrm{d}s^2 = \mathrm{d}t^2 - h_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j. \tag{3}$$

Definition 8. *When the metric of the spacetime can be expressed as* $\mathrm{d}s^2 = \mathrm{d}t^2 - h_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j$, t *is the* cosmic time *and* $h_{ij} = h_{ij}(t, x^1, x^2, x^3)$ *is the three-dimensional metric of each hypersurface of simultaneity.*

Only the existence of a cosmic time guarantees that all the processes of the universe can be coordinated by a single time, recovering the temporal structure of pre-relativistic physics where the history of the universe as a whole can be conceived as the temporal sequence of its instantaneous states.

These considerations can be easily applied to the case of FLRW models: since FLRW models are spatially homogeneous and isotropic on the large scale, it is possible to find a family of spacelike hypersurfaces which can be labelled by the proper time of the worldlines that orthogonally thread through them: these labels define the cosmic time, which is represented by the variable t, and the scale factor a is a function only of t. Therefore, to the extent that FLRW models have a cosmic time, *a fortiori*, they satisfy the weaker condition of admitting a global time function.

Let us note that the existence of a cosmic time imposes too strong a condition on realistic cosmological models. For instance, a spacetime with cosmic time does not admit, for instance, a local Schwarzschild metric; this fact, in turn, precludes the occurrence of well-known physical phenomena, such as the existence of solar systems or the gravitational lensing effect. This means that FLRW models are valid only on a large scale and in an averaged sense, but must be modified when local cosmological phenomena have to be explained. Nevertheless, this kind of modification does not count against the existence of a global and non-entropic arrow of time, since such an arrow requires that the spacetime admits a global time function and this condition holds even when local cosmological phenomena are taken into account.

It is interesting to emphasize how these considerations provide a further argument against the global entropic approach, which relies on the possibility of defining the entropy gradient of the universe, that is, the gradient of the entropy function $S(t)$. In the standard formulations, the variable t in the entropy function $S(t)$ denotes the time that plays the role of a dynamical parameter in the evolution of the universe. But, as we have seen, only the existence of a cosmic time guarantees that the temporal structure of pre-relativistic physics can be recovered. Therefore, if the universe has no cosmic time t, the function $S(t)$ cannot be defined. This means that, in order to define the arrow of time, the global entropic approach imposes a condition so strong that would preclude well-known cosmological phenomena.

4.3. Time-asymmetry

When, as in the case of FLRW models, the spacetime admits a cosmic time, the issues of time-reversal invariance and time-asymmetry can be presented in more familiar terms than those used in the previous sections. In fact, when the spacetime has a cosmic time, Einstein's field equations are time-reversal invariant in the sense that if $h_{ij}\,(t, x^1, x^2, x^3)$ of equation (3) is a solution, $h_{ij}\,(-t, x^1, x^2, x^3)$ is also a solution. But the time-reversal invariance of these equations does not prevent us from describing a time-asymmetric universe whose spacetime is asymmetric regarding its geometrical properties along the cosmic time. In this case, the concept of time-symmetry can be defined as follows.

Definition 9. *A time-orientable spacetime which admits a cosmic time t is* time-symmetric *with respect to some spacelike hypersurface $t = t_S$, where t_S is a constant, if it is time isotropic and the diffeomorphism d leaves the hypersurface $t = t_S$ fixed.*

On the basis of these considerations, it is easy to see that in FLRW models the time-symmetry of spacetime may manifest itself in two different ways depending on whether the universe has singular points at one or both temporal extremities[10]. Big bang–big chill universes are manifestly time-asymmetric: since the scale factor $a(t)$ increases with the cosmic time t, there is no hypersurface $t = t_S$ from which the spacetime looks the same in both temporal directions. In big bang–big crunch universes, in contrast, $a(t)$ has a maximum value: therefore, the spacetime might be time-symmetric about the time of maximum expansion. These are the cases of some FLRW models with dust and radiation. In more general cases (e.g. inflationary models) different fields can represent the matter–energy of the universe. Many interesting results have been obtained, for instance, by modelling matter–energy as a set of scalar fields $\phi_k(t)$: homogeneity is retained and the time-reversal invariance of the field equations is given by the fact that, if $[a(t), \phi_k(t)]$ is a solution, $[a(-t), \phi_k(-t)]$ is also a solution. In these cases, if there is a time t_{ME} of maximum expansion, the scale factor $a(t)$ may be such that $a(t_{\mathrm{ME}} + t) \neq a(t_{\mathrm{ME}} - t)$ (see, for instance, the models in [24]). This means that a big bang–big crunch universe is generically a time-asymmetric object with respect to the metric of the spacetime, and this asymmetry, essentially grounded on geometrical considerations, allows us to distinguish between the two directions of the cosmic time, independently of entropic considerations.

With regard to the generality of time-asymmetry in the particular case of FLRW models, we have proved in [2] the vanishing probability of perfect time-asymmetry: that demonstration was generalized in subsection 3.3 for time-orientable spacetimes admitting a global time but which may lack a cosmic time.

5. From geometry to energy flow: the first role of the energy–momentum tensor

5.1. The arrow of time as energy flow

Up to now, the arrow of time was defined in terms of the substantial difference between the vector fields $\gamma^\mu(x)$ and $-\gamma^\mu(x)$, grounded on the time-asymmetry of the spacetime. But $\gamma^\mu(x)$ was characterized merely as the vector field that must exist for the time-orientability of spacetime. The question now is whether the arrow of time can be defined in a physical way, that is, by means of a mathematical object that can be interpreted not only geometrically but also in terms of the more familiar magnitudes of physics.

As is well known, the energy–momentum (or gravitational stress) tensor $T_{\mu\nu}$ represents the density and the flow of energy and momentum at each point of the spacetime. Then, it would be desirable to define the vector field $\gamma^\mu(x)$ in terms of $T_{\mu\nu}$ in order to endow it with a

[10] This depends on the values of the factor k and of the cosmological constant Λ.

physical meaning[11]. Although this task cannot be accomplished in a completely general case, it is possible to define the arrow of time in terms of $T_{\mu\nu}$ when the energy–momentum tensor satisfies the *dominant energy condition* everywhere (see [16, 25]).

Definition 10. *The energy–momentum tensor satisfies the* dominant energy condition *if, in any orthonormal basis, the energy component dominates the other components of $T_{\alpha\beta}$:*

$$T^{00} \geqslant |T^{\alpha\beta}| \qquad \text{for each } \alpha, \beta$$

where α, β indicate the indices of the orthonormal bases.

This means that, to any observer, the local matter–energy density appears non-negative and the energy flow is non-spacelike. The dominant energy condition does not impose a very strong constraint, since it holds for almost all known forms of ordinary matter[12].

Let us consider a continuous orthonormal basis field $\left\{V^{\mu}_{(\alpha)}(x)\right\}$ (a tetrad or *vierbein*) consisting of four unitary vectors $\left(V^{\mu}_{(0)}(x), V^{\mu}_{(1)}(x), V^{\mu}_{(2)}(x), V^{\mu}_{(3)}(x)\right)$. In this basis, $g_{\mu\nu}V^{\mu}_{(\alpha)}V^{\nu}_{(\beta)} = \eta_{\alpha\beta}$ are the coordinates of the local Minkowski metric tensor, and $T_{\mu\nu}V^{\mu}_{(\alpha)}V^{\nu}_{(\beta)} = T_{\alpha\beta}$ are the coordinates of the energy–momentum tensor. Then, $T^{0\alpha}V^{\mu}_{(\alpha)}$ can be conceived as a vector representing the energy flow, whose coordinates in that basis are the $T^{0\alpha}$. Now, if $T^{00} \geqslant |T^{\alpha\beta}|$, then $T^{00} \geqslant |T^{0\alpha}|$. In turn, $T^{00} \geqslant |T^{0\alpha}|$ implies that $T^{0\alpha}V^{\mu}_{(\alpha)}$ is non-spacelike. On the other hand, if the manifold and the orthogonal basis field are continuous, $g_{\mu\nu}$ is continuously defined over it and, provided that the derivatives of $g_{\mu\nu}$ are also continuous, $T^{\mu\nu}$ ($T^{\alpha\beta}$) and, then, $T^{0\mu}$ ($T^{0\beta}$) are also continuously defined all over the manifold. Therefore, it seems that, in the regions where the density of matter–energy is non-zero, we have found a physical correlate of the continuous non-vanishing non-spacelike vector field $\gamma^{\mu}(x) = T^{0\alpha}(x)V^{\mu}_{(\alpha)}(x)$. The drawback of this conclusion is that $T^{0\alpha}V^{\mu}_{(\alpha)}$ is not strictly a vector, since it is not transformed as a vector by the Lorentz transformations. Strictly speaking, at each point x of the spacetime $T^{0\alpha}(x)V^{\mu}_{(\alpha)}(x)$ is a tetra-magnitude which represents the energy flow only in the basis $\left\{V^{\mu}_{(\alpha)}(x)\right\}$; thus, it seems that it cannot directly play the role of $\gamma^{\mu}(x)$ as initially desired.

Nevertheless, the fact that the energy flow cannot be represented by a vector is not an obstacle to the definition of the arrow of time in terms of such a flow. The dominant energy condition poses a *covariant* condition: if the energy flow is non-spacelike in a reference frame, it is non-spacelike in all reference frames. This means that irrespective of which orthonormal basis $\left\{V^{\mu}_{(\alpha)}(x)\right\}$ is chosen, the energy flow in that basis, represented by $T^{0\alpha}(x)V^{\mu}_{(\alpha)}(x)$, can be used to define the arrow of time. The usual convention in physics consists in calling the temporal direction of the positive energy flow 'future'. In this case, at any point x of the spacetime $T^{0\alpha}(x)V^{\mu}_{(\alpha)}(x)$ belongs to the future light semicone $C_{+}(x)$: the

[11] Of course, this can be done only in the regions of spacetime where $T_{\mu\nu} \neq 0$. In the regions where $T_{\mu\nu} = 0$, we are forced to use the non-vanishing vector field γ^{μ}, as explained in subsection 3.2.

[12] There are, of course, strange cosmological 'objects' whose existence would lead to universes where the dominant energy condition is not satisfied in certain regions of spacetime. For instance, in wormhole spacetimes, the dominant energy condition is violated in the vicinity of the wormhole throat since the wormhole is threaded by negative 'exotic' matter (see [25]). Nevertheless, it is plausible to suppose that universes containing such kinds of objects will surely not satisfy the stronger conditions necessary for defining the arrow of time, that is, time-orientability and existence of global time.

At present, a non-negative pressure cosmological term $T_{\mu\nu}$ seems to exist in the universe. Even though this term does not satisfy the dominant energy condition, it can be considered as a background term, irrelevant for the local issues we are considering here. To the extent that the ordinary matter term does satisfy the dominant energy condition, we can include the non-negative pressure term on the lhs of Einstein's equations as $\Lambda g_{\mu\nu}$, and the resulting rhs will satisfy the dominant energy condition.

energy flows towards the future for any observer. But the relevant point is that this statement acquires a non-conventional meaning only when the *substantial difference* between past and future has been previously established by the time-asymmetry of the spacetime (as shown in subsection 3.2).

5.2. From the global arrow to the local arrow

As we have seen, according to the usual terminological convention, the future light semicone $C_+(x)$ at each point x of the spacetime is defined by the positive energy flow $T^{0\alpha}(x)V^{\mu}_{(\alpha)}(x)$ at this point. But, is $T^{0\alpha}$ really the energy flow used in local physics? Let us remember that $T_{\mu\nu}$ satisfies the 'conservation' equation:

$$\nabla_\mu T^{\mu\nu} = 0. \tag{4}$$

However, this is not a true conservation equation since ∇_μ is a covariant derivative. The usual conservation equation with ordinary derivative reads

$$\partial_\mu \tau^{\mu\nu} = 0 \tag{5}$$

where $\tau_{\mu\nu}$, the energy–momentum (or gravitational stress) pseudo-tensor (since it is not a proper tensor), is defined as [19]

$$\tau^{\mu\nu} = T^{\mu\nu}_{\text{eff}} = T^{\mu\nu} + t^{\mu\nu}. \tag{6}$$

The correction term $t_{\mu\nu}$ reads

$$t_{\mu\nu} = \frac{1}{16\pi}\left[\mathcal{L}g_{\mu\nu} - \frac{\partial\mathcal{L}}{\partial g_{\kappa\lambda,}{}^{\mu}} g_{\kappa\lambda,\nu}\right] \tag{7}$$

where $\mathcal{L}$ is the system gravitational Lagrangian[13]. $t_{\mu\nu}$ is also a homogeneous and quadratic function of the connection $\Gamma^{\lambda}_{\nu\mu}$ [26], precisely:

$$\begin{aligned} t^{\mu\nu} = \frac{1}{16\pi k}\Big[&\left(2\Gamma^{\chi}_{\lambda\rho}\Gamma^{\sigma}_{\chi\sigma} - \Gamma^{\chi}_{\lambda\sigma}\Gamma^{\sigma}_{\rho\chi} - \Gamma^{\chi}_{\lambda\chi}\Gamma^{\sigma}_{\rho\sigma}\right)\left(g^{\mu\lambda}g^{\nu\rho} - g^{\mu\nu}g^{\lambda\rho}\right) \\ &+ g^{\mu\lambda}g^{\rho\chi}\left(\Gamma^{\nu}_{\lambda\sigma}\Gamma^{\sigma}_{\rho\chi} + \Gamma^{\nu}_{\rho\chi}\Gamma^{\sigma}_{\lambda\sigma} - \Gamma^{\nu}_{\chi\sigma}\Gamma^{\sigma}_{\lambda\rho} - \Gamma^{\nu}_{\lambda\rho}\Gamma^{\sigma}_{\chi\sigma}\right) \\ &+ g^{\nu\lambda}g^{\rho\chi}\left(\Gamma^{\mu}_{\lambda\sigma}\Gamma^{\sigma}_{\rho\chi} + \Gamma^{\mu}_{\rho\chi}\Gamma^{\sigma}_{\lambda\sigma} - \Gamma^{\mu}_{\chi\sigma}\Gamma^{\sigma}_{\lambda\rho} - \Gamma^{\mu}_{\lambda\rho}\Gamma^{\sigma}_{\chi\sigma}\right) \\ &+ g^{\lambda\rho}g^{\chi\sigma}\left(\Gamma^{\mu}_{\lambda\chi}\Gamma^{\nu}_{\rho\sigma} - \Gamma^{\mu}_{\lambda\rho}\Gamma^{\nu}_{\chi\sigma}\right)\Big]. \end{aligned} \tag{8}$$

Now we can consider the coordinates $\tau^{0\mu}$, which satisfy a usual conservation equation:

$$\partial_\mu \tau^{0\mu} = \partial_0 \tau^{00} + \partial_i \tau^{0i} = 0 \tag{9}$$

where τ^{00} represents the energy density and τ^{0i} are the three components of the spatial energy flow (the Poynting vector).

Analogous to the case of the energy–momentum tensor, in any orthonormal basis $\left\{V^{\mu}_{(\alpha)}\right\}$, $\tau^{0\alpha}V^{\mu}_{(\alpha)}$ is not a vector since it is not transformed as a vector by the Lorentz transformations. But in contrast to that case, the dominant energy condition cannot be expressed in terms of $\tau^{\mu\nu}$. However, this can be done in the particular case of *local inertial frames*. In fact, in a local inertial frame, $\Gamma^{\lambda}_{\nu\mu} = 0$; thus, $t_{\mu\nu} = 0$ (see equation (8)) and

[13] Since both sides of Einstein's equations satisfy the covariant conservation law (4), the same correction term $t_{\mu\nu}$ can be used on both sides of such equations in order to obtain the corresponding usual conservation law (5). Therefore, such a term can be computed by means of the gravitational Lagrangian that defines the lhs of those equations. This means that the correction term $t_{\mu\nu}$ can be adequately defined when matter is present or in vacuum.

$\tau^{\mu\nu} = T^{\mu\nu}$ (see equation (6)). Therefore, in this case the dominant energy condition implies that:

$$\tau^{00} \geqslant |\tau^{\alpha\beta}| \qquad \text{for each } \alpha, \beta. \tag{10}$$

As a consequence, in the basis $\left\{W^{\mu}_{(\alpha)}\right\} = \left(W^{\mu}_{(0)}, W^{\mu}_{(1)}, W^{\mu}_{(2)}, W^{\mu}_{(3)}\right)$ of the local inertial frame, the energy flow, represented by $\tau^{0\alpha} W^{\mu}_{(\alpha)}$ and satisfying the usual conservation equation, is non-spacelike.

Although this result cannot be generalized to all the reference frames of the whole spacetime, it is relevant in local contexts since, in small regions of the spacetime, the metric tends to the Minkowski form in local inertial frames[14]. In fact, any local region of the spacetime can be approximated by the tangent Minkowski space (with orthonormal basis $\{V^{\mu}_{(\alpha)}\}$) at any point of that region. If $\left\{W^{\mu}_{(\alpha)}\right\}$ is the basis of an inertial frame on this flat tangent space, $\tau^{0\alpha} W^{\mu}_{(\alpha)}$ represents the non-spacelike *local energy flow* satisfying the usual conservation equation[15]. Moreover, at each point x of the local region, the local energy flow $\tau^{0\alpha}(x) W^{\mu}_{(\alpha)}(x)$ belongs to the same light semicone as the one to which $T^{0\alpha}(x) V^{\mu}_{(\alpha)}(x)$ belongs. Therefore, if we have adopted the usual physical convention in the global level, we can also meaningfully say that future is the temporal direction of the positive local energy flow: the local flow emitted at x belongs to the light semicone $C_+(x)$.

This result is particularly relevant because local inertial frames are the reference frames in which the non-relativistic local theories of physics are valid. In turn, ordinary quantum field theory in flat spacetime must be considered as locally formulated in a local inertial frame. This means that the local energy flow directed towards the future is the flow of energy as conceived by this kind of theory, where energy satisfies the usual conservation law expressed by means of ordinary derivatives. Summing up, $\tau^{0\mu}$ inherits the time orientation defined at the global level and, to the extent that it has a local physical meaning, it not only transfers the global arrow of time to local contexts, but also translates the global arrow into a usual magnitude of local physical theories.

5.3. The absolute nature of the arrow of time

As we have seen, the vector $\tau^{0\mu}$ introduced above *is always non-spacelike* and, therefore, its direction defines the arrow of time. However, we know that, given the time-reversal invariance of Einstein's field equations, in a time-orientable spacetime where t is the global time, if we can obtain $\tau^{0\mu}$, we can also obtain $-\tau^{0\mu}$. At this point, the ghost of symmetry threatens again: it seems that we are committed to providing a non-conventional criterion for picking out one of both nomologically admissible solutions, one being the temporal mirror image of the other. Nevertheless, as we will see, the threat is not as serious as it seems.

Replacing t by $-t$ amounts to applying a symmetry transformation, in particular, time reversal. The point is to understand the meaning of such a transformation. Under the active interpretation, a symmetry transformation corresponds to a change from one system to another; under the passive interpretation, a symmetry transformation consists in a change of the point of view from which the system is described. The traditional position about symmetries assumes that, in the case of discrete transformations, only the active interpretation makes sense: an idealized observer can rotate himself in space in correspondence with the given spatial rotation,

[14] Near each point x_0, the metric can be approximated with the metric of the local free inertial frame as

$$\mathrm{d}s^2 = (1 - R_{0l0m}x^l x^m)\,\mathrm{d}t^2 - \left(\tfrac{4}{3} R_{0ljm}x^l x^m\right)\mathrm{d}t\,\mathrm{d}x^j + \left(\delta_{ij} - \tfrac{1}{3} R_{iljm}x^l x^m\right)\mathrm{d}x^i\,\mathrm{d}x^j + O(|x^j|^3)\,\mathrm{d}x^\alpha\,\mathrm{d}x^\beta$$

(see [19] equation (13.73)). Therefore, locally in the inertial frame at x_0, we can guarantee that $\tau^{00} \geqslant |\tau^{0i}|$.

[15] In fact, $\partial_\mu\left(\tau^{0\alpha} W^{\mu}_{(\alpha)}\right) = \partial_\mu \tau^{0\mu} = 0$ (see equation (9)).

but it is impossible to 'rotate in time' (see [12, 27]). Of course, this is true when the idealized observer is immersed in the same spacetime as the observed system. But when the system is the universe as a whole, we cannot change our spatial perspective with respect to the universe: it is as impossible to rotate in space as it is to rotate in time. However, this does not mean that the active interpretation is the correct one: the idea of two identical universes, one translated in space or in time with respect to the other, has no meaning. This shows that both interpretations, when applied to the universe as a whole, collapse into conceptual nonsense.

In fact, in cosmology symmetry transformations are given neither an active nor a passive interpretation: two mathematical models for the universe, defined by (M, g) and (M', g'), are taken to be equivalent if they are isometric, that is, if there is a diffeomorphism $\theta : M \to M'$ which carries the metric g into the metric g' (see [16]). Since symmetry transformations are isometries, two models related by a symmetry transformation (in particular, time-reversal) are considered equivalent descriptions of one and the same universe. Therefore, it is not necessary to find a non-conventional criterion for selecting one of two nomologically admissible solutions to the extent that both are descriptions of a single possible universe.

These considerations point to the absolute character of the arrow of time embodied in $\tau^{0\mu}$. This vector (in particular, its direction) is rigidly linked to the spacetime on which it is defined. Changing $\tau^{0\mu}$ by $-\tau^{0\mu}$ amounts to changing the model by its temporal mirror image; but, as we have shown, this step has no physical meaning since both models are merely conventionally different descriptions of the same universe.

Let us note that we have not used the term 'future' in the present argument. If we adopt the usual terminology, we will call the time direction of the energy flow 'future': in this case, we can say that the vector $\tau^{0\mu}$ points to the future. Nevertheless, it is worth remembering that 'past' and 'future' are words that we use in our everyday language but they do not belong to physical theories. Then, the choice of saying that $\tau^{0\mu}$ points to the future is conventional: we can replace 'future' by 'past' and nothing changes. What remains is the absolute and substantial nature of the arrow of time defined by the unchangeable direction of $\tau^{0\mu}$.

5.4. Breaking the symmetry in time-reversal invariant theories

As we have seen in subsection 2.2, the Ehrenfests criticized Gibbs' approach by pointing out that the increase of the entropy towards the future is always matched by a similar one in the past evolution of the system. It is interesting to note that this old discussion can be generalized to the case of any kind of evolution arising from local time-reversal invariant laws. In fact, time-reversal invariant equations always give rise to what we will call '*time-symmetric twins*' (see [2]), that is, two mathematical structures symmetrically related by a time-reversal transformation: each 'twin', which in some cases represents an irreversible evolution, is the temporal mirror image of the other 'twin'. For instance, electromagnetism provides a pair of advanced and retarded solutions, which are usually related to incoming and outgoing states in scattering situations as described, e.g., by the Lax–Phillips scattering theory [28]. In irreversible quantum mechanics, the analytical extension of the energy spectrum of the quantum systems Hamiltonian to the complex plane leads to poles in the lower half-plane (usually related to decaying unstable states) and symmetric poles in the upper half-plane (usually related to growing unstable states) (see [29]). However, at this level the twins are only conventionally different: we cannot distinguish between advanced and retarded solutions or between lower and upper poles without assuming temporally asymmetric notions, as the asymmetry between past and future or between preparation and measurement. Then, the challenge consists in providing a non-conventional criterion for choosing one of the twins as physically relevant: such a criterion must establish a substantial difference between the two members of the pair.

The arrow of time locally defined by $\tau^{0\mu}$ is precisely what provides us the criterion for establishing the desired substantial difference. In fact, $\tau^{0\mu}$ says that at each point of the spacetime the semicones $C_{-}(x)$ receive an incoming flow of energy while the semicones $C_{+}(x)$ emit an outgoing flow of energy. Therefore, in each case the twin that must be retained as physically relevant is the one that describes this kind of energy flow, from $C_{-}(x)$ to $C_{+}(x)$. For instance, in electromagnetism only retarded solutions fulfil this condition since they describe waves propagating into the semicone $C_{+}(x)$. In irreversible quantum mechanics, only decaying unstable states with poles in the lower half-plane have the physical sense of spontaneous evolutions, since they provide an energy flow contained in $C_{+}(x)$. In contrast, the poles in the upper half-plane represent growing unstable states, which are not spontaneous since they must be generated by pumping energy coming from the semicone $C_{-}(x)$, that is, from the past. Summing up, by translating the geometrical global time-asymmetry in terms of energy flow, $\tau^{0\mu}$ can be used locally (at each point of the spacetime) for breaking the symmetry of the set of solutions produced by time-reversal invariant laws (see [2] for more cases, e.g., thermodynamic twins, quantum measurement twins, decoherence twins, etc).

6. The non-time-reversal invariance of quantum field theory: the second role of the energy–momentum tensor

6.1. The non-time-reversal invariance of axiomatic QFT

In any of its versions, axiomatic quantum field theory includes a non-time-reversal invariant postulate (see [30–32]), which states that the spectrum of the energy–momentum operator P^{μ} is confined to a future light semicone, that is, its eigenvalues p^{μ} satisfy

$$p^2 \geqslant 0 \qquad p^0 \geqslant 0.$$

This postulate says that, when we measure the observable P^{μ}, we obtain a *non-spacelike classical* p^{μ} contained in a future semicone, that is, a semicone belonging to C_{+}.

It is clear that the condition $p^0 \geqslant 0$ selects one of the elements of the time-symmetric twins $p^0 \geqslant 0$ and $p^0 \leqslant 0$ that would arise from the theory in the absence of the non-time-reversal invariant postulate: by means of this postulate, QFT becomes a non-time-reversal invariant theory. In turn, since QFT, being both quantum and relativistic, can be considered one of the most basic theories of physics, the choice introduced by the condition $p^0 \geqslant 0$ is transferred to the rest of the physical theories. But such a choice is established from the very beginning, as a postulate of the theory. The challenge is, then, to *justify* the non-time-reversal invariant postulate by means of independent theoretical arguments.

As is well known, in the energy–momentum tensor, $T^{0i}(x)$ represents the matter–energy flow and $T^{i0}(x)$ represents the linear momentum density. Since $T^{\mu\nu}$ is a symmetric tensor, $T^{\mu\nu}(x) = T^{\nu\mu}(x)$ and, therefore, $T^{0i}(x) = T^{i0}(x)$; in other words, the matter–energy flow is equal to the linear momentum density. This means that if the matter energy flow $T^{0\alpha}$ can be used to define the arrow of time under the dominant energy condition, this is also the case for the linear momentum density $T^{\alpha 0}$. On the other hand, we have seen in subsection 5.2 that the local matter–energy flow $\tau^{\mu\nu}$ can be conceived as a conservative version of $T^{\mu\nu}$ in the orthonormal coordinates of a local inertial frame; in this case, the dominant energy condition has the consequence that $\tau^{0\mu}$ is non-spacelike. Now we know that exactly the same conclusion can be drawn for the local linear momentum density $\tau^{\mu 0}$. But the local linear momentum density $\tau^{\mu 0}$ is precisely the local magnitude corresponding to the classical p^{μ} of QFT; thus, at each point x, $p^{\mu} \sim \tau^{0\mu} \in C_{+}(x)$.

In conclusion, the fact that p^{μ} at each point x of the local context and, therefore, for every classical particle, must be contained in the future light semicone $C_{+}(x)$ turns out to be a consequence of the global time-asymmetry of the spacetime when the dominant energy condition holds everywhere. In other words, the non-time-reversal invariant postulate can be justified on cosmological grounds instead of being imposed as a departure point for the axiomatic version of QFT.

6.2. The non-time-reversal invariance of ordinary QFT

In this section we will analyse how time-reversal invariance is introduced in the ordinary version of QFT. The classification of one-particle states according to their transformation under the Lorentz group leads to six classes of 4-momenta (e.g. see [33], vol. I, p 65). Of these classes, it is considered that only three have physical meaning: these are precisely the cases which agree with the non-time-reversal invariant postulate of the axiomatic version of QFT. In other words, the symmetry group of QFT is the orthochronous group (see [21, 34]), where the space inversion $\mathcal{P}$ but not the time inversion $\mathcal{T}$ is included. This is another way of expressing the non-time-reversal invariance of the QFT. In this case, the non-time-reversal invariance is introduced not by means of a postulate but on the basis of empirical arguments that make certain classes of 4-momenta physically meaningless. However, to the extent that special relativity and standard quantum mechanics are time-reversal invariant theories, they give no theoretically grounded justification for such a breaking of time-reversal invariance. Nevertheless, as we have seen in the previous subsection, this justification can be given on cosmological grounds.

Let us make the point in different terms. The quantum field correlates of $\mathcal{P}$ and $\mathcal{T}$, $\mathbf{P}$ and $\mathbf{T}$, are defined as

$$\mathbf{P}\mathrm{i}P^{\nu}\mathbf{P}^{-1} = \mathrm{i}\mathcal{P}^{\nu}_{\mu}P^{\mu} \qquad \mathbf{T}\mathrm{i}P^{\nu}\mathbf{T}^{-1} = \mathrm{i}\mathcal{T}^{\nu}_{\mu}P^{\mu}$$

where $\mathbf{P}$ is a linear and unitary operator and $\mathbf{T}$ is an antilinear and antiunitary operator. In fact, if $\mathbf{T}$ were linear and unitary, we could simply cancel the i and, then, $\mathbf{T}P^{\nu}\mathbf{T}^{-1} = -P^{\mu}$: the action of the operator $\mathbf{T}$ on the operator P^{ν} would invert the sign of P^{μ}, with the consequence that the spectrum of the inverted energy–momentum operator would be contained in a past light semicone. In particular, for $\nu = 0$, $P^{\mu} = H$, where H is the energy operator; then, if $\mathbf{T}$ were linear and unitary, $\mathbf{T}H\mathbf{T}^{-1} = -H$ with the consequence that for any state of energy E there would be another state of energy $-E$. The antilinearity and antiunitarity of $\mathbf{T}$ avoid these 'anomalous' situations in agreement with the conditions imposed by the non-time-reversal invariant postulate and, at the same time, make QFT non-time-reversal invariant. Once again, there are good empirical reasons for making $\mathbf{T}$ antilinear and antiunitary, but there is no theoretical justification for such a move.

Summing up, in ordinary QFT it is always necessary to take a decision about the time direction of the spectrum of the energy–momentum operator P^{μ}. The point that we want to stress here is that, either in the case of the non-time-reversal invariant postulate of the axiomatic version of QFT or in the case of the usual version of QFT, the decision can be justified on cosmological grounds, as a consequence of the global time-asymmetry of the universe and the dominant energy condition.

Finally, it is worth reflecting on the role of weak interactions in the problem of the arrow of time. The CPT theorem states that **CPT** is the only combination of charge conjugation **C**, parity reflection **P** and time-reversal **T** which is a symmetry of QFT. In fact, it is well known that weak interactions break the **T** of the CPT theorem. According to a common opinion, this empirical fact is precisely what gives the clue for the solution of the problem of the arrow of time: since the **T** symmetry is violated by weak interactions, they introduce a non-conventional distinction between the two directions of time (see [25]). The question is: is it the breaking of **T** that distinguishes both directions of time in QFT? As we have seen, the operator **T** was designed precisely to avoid certain tetra-magnitudes, such as the linear momentum p^{μ}, having the 'anomalous' feature of being contained in a past light semicone: the action of the operator **T** on the energy–momentum operator P^{μ} preserves the time direction of P^{μ} and, therefore, of its eigenvalues. It is this fundamental fact that makes QFT non-time-reversal invariant, and not the incidental violation of **T** by weak interactions[16]. This non-time-reversal invariance of QFT, based on the peculiar features of the operator **T**, distinguishes by itself between the two directions of time, without the need for weak interactions. In other words, even if weak interactions did not exist, QFT would be a non-time-reversal invariant theory which would define the arrow of time. The real problem is, then, to justify the non-time-reversal invariance of a theory which is presented as a synthesis of two time-reversal invariant theories such as special relativity and quantum mechanics. But this problem is completely independent of the existence of weak interactions and the breaking of **T** introduced by them. Summing up, weak interactions do not play a role as relevant in the problem of the arrow of time as is usually assumed.

7. Conclusion

In this paper we have defined a physical object $\tau^{0\mu} \sim p^{\mu}$ that can play the role of the arrow of time. Then, the mysterious and phantomlike arrow of Eddington is at last materialized. In particular, we have shown the dual role played by the energy–momentum tensor in the context of our approach to the problem of the arrow of time. When matter–energy flow is considered, the energy–momentum tensor translates the generic geometrical time-asymmetry of the universe in terms of energy flow. When linear momentum density is considered, the energy–momentum tensor provides the means of justifying the time-asymmetry postulate of axiomatic quantum field theory.

Acknowledgments

We are extremely grateful to Alejandro Gangui for his assistance regarding some astronomical subjects. This paper was partially supported by CONICET and the Secretaría de Estado de Educación y Universidades of Spain.

References

[1] Castagnino M, Gaioli F and Gunzig E 1996 *Found. Cosmic Phys.* **16** 221
Castagnino M and Laura R 1997 *Phys. Rev.* A **56** 108
Castagnino M 1998 The global nature of time asymmetry and the Bohm–Reichenbach diagram *Irreversibility and Causality* ed A Bohm, H Doebner and P Kielarnowski *(Proc. G. 21 Goslar 1996)* (Berlin: Springer)
Castagnino M and Gunzig E 1999 *Int. J. Theor. Phys.* **38** 47
Castagnino M and Laciana C 2002 *Class. Quantum Grav.* **19** 2657
Castagnino M, Gueron J and Ordoñez A 2002 *J. Math. Phys.* **43** 705
Castagnino M, Catren G and Ferraro R 2002 *Class. Quantum Grav.* **19** 4729
Castagnino M, Lara L and Lombardi O 2003 *Int. J. Theor. Phys.* **42** 2487
[2] Castagnino M, Lara L and Lombardi O 2003 *Class. Quantum Grav.* **20** 369
[3] Castagnino M, Lombardi O and Lara L 2003 *Found. Phys.* **33** 857

[16] Of course, this leaves open a different problem: an explanation of why, among all the elementary interactions, only weak interactions break **T**.

[4] Sudarshan C G, Chiu C B and Gorini V 1978 *Phys. Rev.* D **18** 2914
Bohm A 1979 *Quantum Mechanics, Foundations and Applications* (Berlin: Springer)
Bohm A and Gadella M 1989 *Dirac Kets, Gamow Vectors, and Gel'fand Triplets* (Berlin: Springer)
Castagnino M and Gunzig E 1997 *Int. J. Theor. Phys.* **36** 2545
[5] Penrose R 1979 Singularities and time asymmetry *General Relativity, an Einstein Centenary Survey* ed S Hawking and W Israel (Cambridge: Cambridge University Press)
[6] Sachs R G 1987 *The Physics of Time-Reversal* (Chicago, IL: University of Chicago Press)
[7] Price H 1996 *Time's Arrow and the Archimedes' Point* (Oxford: Oxford University Press)
[8] Ehrenfest P and Ehrenfest T 1959 *The Conceptual Foundations of the Statistical Approach in Mechanics* (Ithaca, NY: Cornell University Press) (original 1912)
[9] Boltzmann L 1897 *Ann. Phys., Lpz.* **60** 392
[10] Feynman R P, Leighton R B and Sands M 1964 *The Feynman Lectures on Physics* vol 1 (New York: Addison-Wesley)
[11] Davies P C 1994 Stirring up trouble *Physical Origins of Time Asymmetry* ed J J Halliwell, J Perez-Mercader and W H Zurek (Cambridge: Cambridge University Press)
[12] Earman J 1974 *Phil. Sci.* **41** 15
[13] Penrose O and Percival I C 1962 The direction of time *Proc. Phys. Soc.* **79** 605
[14] Mattews G 1979 *Phil. Sci.* **46** 82
[15] Schulman S 1999 *Phys. Rev. Lett.* **83** 5419
[16] Hawking S and Ellis J 1973 *The Large Scale Structure of Space-Time* (Cambridge: Cambridge University Press)
[17] Schutz B F 1980 *Geometrical Methods of Mathematical Physics* (Cambridge: Cambridge University Press)
[18] Savitt S F 1996 *Br. J. Phil. Sci.* **47** 347
[19] Misner C W, Thorne K S and Wheeler J A 1973 *Gravitation* (San Francisco, CA: Freeman)
[20] Choquet-Bruhat Y and Lamoureau-Brousse L 1972 *C.R. Acad. Sci., Paris* **276** 1217
Cordero P and Teitelboim C 1976 *Ann. Phys., NY* **100** 607
Hughes T, Kato T and Marsden J 1977 *Arch. Ration. Mech. Anal.* **63** 273
[21] Lichnerowicz A 1976 *J. Math. Phys.* **17** 2135
[22] Bennet C L *et al* 2003 First year Wilkinson microwave anisotropy probe (WMAP) *Preprint* astro-ph/0302207
Kogut A *et al* 2003 Wilkinson microwave anisotropy probe (WMAP). First year *Preprint* astro-ph/0302213
[23] Perlmutter S and Schmidt B P 2003 Measuring cosmology with supernovae *Preprint* astro-ph/0303428
[24] Castagnino M, Giacomini H and Lara L 2000 *Phys. Rev.* D **61** 107302
Castagnino M, Giacomini H and Lara L 2001 *Phys. Rev.* D **63** 044003
[25] Visser M 1996 *Lorentzian Wormholes* (Berlin: Springer)
[26] Landau L and Lifchitz E 1970 *Théorie des Champs* (Moscow: Editions Mir)
[27] Sklar L 1974 *Space, Time and Spacetime* (Berkeley, CA: University of California Press)
[28] Lax P D and Phillips R S 1979 *Scattering Theory* (New York: Academic)
[29] Castagnino M and Laura R 1997 *Phys. Rev.* A **56** 108
[30] Haag R 1996 *Local Quantum Physics. Fields, Particles, Algebras* (Berlin: Springer)
[31] Bogoliubov N, Logunov A A and Todorov I T 1975 *Axiomatic Quantum Field Theory* (Reading, MA: Benjamin-Cummings)
[32] Roman P 1969 *Introduction to Quantum Field Theory* (New York: Wiley)
[33] Weinberg S 1995 *The Quantum Theory of Fields* (Cambridge: Cambridge University Press)
[34] Roman P 1961 *Elementary Theory of Particles* (Amsterdam: North-Holland)