

Information Entropy in Cosmology

Akio Hosoya,^{1,*} Thomas Buchert,^{2,1,3,†} and Masaaki Morita^{4,5,‡}

¹*Department of Physics, Tokyo Institute of Technology, Oh-Okayama, Meguro-ku, Tokyo 152-0033, Japan*

²*Theoretische Physik, Ludwig-Maximilians-Universität, Theresienstrasse 37, D-80333 München, Germany*

³*Department of Physics and Research Center for the Early Universe (RESCEU), School of Science, The University of Tokyo, Tokyo 113-0033, Japan*

⁴*Department of Physics, Ochanomizu University, Ohtsuka, Bunkyo-ku, Tokyo 112-8610, Japan*

⁵*Advanced Research Institute for Science and Engineering, Waseda University, Ohkubo, Shinjuku-ku, Tokyo 169-8555, Japan*
(Received 17 August 2003; published 9 April 2004)

The effective evolution of an inhomogeneous cosmological model may be described in terms of spatially averaged variables. We point out that in this context, quite naturally, a measure arises which is identical to a fluid model of the Kullback-Leibler relative information entropy, expressing the distinguishability of the local inhomogeneous mass density field from its spatial average on arbitrary compact domains. We discuss the time evolution of “effective information” and explore some implications. We conjecture that the information content of the Universe—measured by relative information entropy of a cosmological model containing dust matter—is increasing.

DOI: 10.1103/PhysRevLett.92.141302

PACS numbers: 98.80.-k, 04.40.-b, 89.70.+c

A measure of inhomogeneity in the Universe.—Cosmology is based on the hypothesis of simplicity called the cosmological principle, i.e., homogeneity and isotropy. The departure of the actual mass distribution from the homogeneous Universe model is quantified in terms of density contrast or a statistical quantity such as the two-point correlation function, which both have been studied either by perturbation theory or numerical simulations. Behind these investigations there is a belief that the Universe is homogeneous on some large enough scale. This belief has to be quantitatively confronted with observation, explicitly introducing a measure of inhomogeneity for a domain of the Universe.

In this Letter, we propose a measure which quantifies the distinguishability of the actual mass distribution from its spatial average, borrowing a well-known concept in standard information theory. Suppose we are told that the probability distribution is $\{q_i\}$ and would like to examine how close this distribution is to the actual one $\{p_i\}$ by carrying out observations or coin tossing; the relevant quantity in information theory is the *relative entropy*,

$$S\{p \parallel q\} = \sum_i p_i \ln \frac{p_i}{q_i}, \quad (1)$$

which is positive for $q_i \neq p_i$, and zero if the actual distribution $\{p_i\}$ agrees with the presumed one $\{q_i\}$. Note that this relative entropy is not symmetric for the two distributions $\{p_i\}$ and $\{q_i\}$. It is known that this measure always decreases or stays the same under Markovian stochastic processes (i.e., a *linear* positive map). Namely, the actual distribution becomes less and less distinguishable from the priorly informed distribution due to the random process. In cosmology we are interested in how the real matter distribution is different from its spatial average. For a continuum, the relevant

quantity would be

$$\frac{S\{\varrho \parallel \langle \varrho \rangle_{\mathcal{D}}\}}{V_{\mathcal{D}}} = \left\langle \varrho \ln \frac{\varrho}{\langle \varrho \rangle_{\mathcal{D}}} \right\rangle_{\mathcal{D}}, \quad (2)$$

where ϱ is the actual distribution and $\langle \cdots \rangle_{\mathcal{D}}$ its spatial average in the volume $V_{\mathcal{D}}$ on the compact domain \mathcal{D} of the manifold Σ . We shall conjecture that the measure $S\{\varrho \parallel \langle \varrho \rangle_{\Sigma}\}$ continues to grow indefinitely, if Σ is compact.

The resolution of the apparent discrepancy between the gravitational system and the ordinary stochastic system will be (i) we are considering in cosmology a nonisolated system defined by a comoving region \mathcal{D} in contrast to an isolated system for an ordinary stochastic process, and (ii) the time evolution dictated by Einstein’s equations induces a negative feedback due to the attractive nature of the gravitational force, which tends to make the matter distribution more and more inhomogeneous.

Deduction of the measure.—To begin with, let us emphasize that the functional (2), known as the Kullback-Leibler relative information entropy (cf. [1–3]) is not assumed as a measure *a priori*, rather it can be *deduced* from a fundamental kinematical relation that refers to the *noncommutativity* of two operations: spatially averaging and evolving the material mass density field. The specific form of the information measure is, thus, inherently determined by the physical problem at hand, and does not need to be justified empirically or axiomatically as is the common status of information measures in the literature.

We define the averaging operation in terms of Riemannian volume integration, restricting attention to scalar functions $\Psi(t, X^i)$,

$$\langle \Psi(t, X^i) \rangle_{\mathcal{D}} := \frac{1}{V_{\mathcal{D}}} \int_{\mathcal{D}} \sqrt{g} d^3 X \Psi(t, X^i), \quad (3)$$

with the Riemannian volume element $d\mu_g := \sqrt{g} d^3 X$, $g := \det(g_{ij})$, and the volume of an arbitrary compact domain, $V_{\mathcal{D}}(t) := \int_{\mathcal{D}} \sqrt{g} d^3 X$; X^i are coordinates in a $t = \text{const}$ hypersurface (with 3-metric g_{ij}) that are comoving with fluid elements of dust:

$$ds^2 = -dt^2 + g_{ij} dX^i dX^j. \quad (4)$$

It is evident from the above setting that we predefine a simple time-orthogonal foliation (which restricts the matter to an irrotational dust continuum) in order to simplify the framework in which we discuss our measure as a concept of a *spatial* average. We wish to emphasize that the formalism below could be carried over to more general settings, e.g., to perfect fluids or scalar fields (cf. [4]) with possibly further interesting implications.

The above-mentioned “noncommutativity” has been fruitfully exploited in previous work on the averaging problem of inhomogeneous cosmologies [5,4,6–8], and can be compactly written in terms of a *commutation rule* for the averaging of a scalar field Ψ :

$$\begin{aligned} \langle \dot{\Psi} \rangle_{\mathcal{D}} - \langle \dot{\Psi} \rangle_{\mathcal{D}} &= \langle \Psi \theta \rangle_{\mathcal{D}} - \langle \Psi \rangle_{\mathcal{D}} \langle \theta \rangle_{\mathcal{D}} = \langle \Psi \delta \theta \rangle_{\mathcal{D}} \\ &= \langle \theta \delta \Psi \rangle_{\mathcal{D}} = \langle \delta \Psi \delta \theta \rangle_{\mathcal{D}}, \end{aligned} \quad (5)$$

where θ denotes the local expansion rate (as minus the trace of the extrinsic curvature of the hypersurfaces $t = \text{const}$). We have rewritten the right-hand side (rhs) of the first equality in terms of the deviations of the local fields from their spatial averages, $\delta \Psi := \Psi - \langle \Psi \rangle_{\mathcal{D}}$ and $\delta \theta := \theta - \langle \theta \rangle_{\mathcal{D}}$.

The key statement of the commutation rule (5) is that the operations *spatial averaging* and *time evolution* do not commute. In cosmology, we may think of initial conditions at the epoch of last scattering, when the fluctuations imprinted on the cosmic microwave background are considered to be averaged out on a rest frame of a standard Friedmann-Lemaître-Robertson-Walker (FLRW) cosmology. In this picture, the evolution of the Universe is described by first averaging out (or ignoring) inhomogeneities and then evolving the average distribution by a homogeneous (in the above case homogeneous isotropic) Universe model. A realistic model would first evolve the inhomogeneous fields and, at the present epoch, the resulting fields would have to be evaluated by spatial averaging to obtain the final values of, e.g., the averaged density field. In particular, this comment applies to all cosmological parameters (see, e.g., [5,7]). Let us illustrate this statement for the mass density field. Setting $\Psi = \varrho$, Eq. (5) reads

$$\langle \dot{\varrho} \rangle_{\mathcal{D}} + \langle \theta \rangle_{\mathcal{D}} \langle \varrho \rangle_{\mathcal{D}} = \langle \dot{\varrho} + \theta \varrho \rangle_{\mathcal{D}}. \quad (6)$$

Since the rhs vanishes due to the continuity equation, we also have a continuity equation for the averages:

$$\langle \dot{\varrho} \rangle_{\mathcal{D}} + \langle \theta \rangle_{\mathcal{D}} \langle \varrho \rangle_{\mathcal{D}} = 0, \quad (7)$$

which simply expresses the conservation of the total material mass, $M_{\mathcal{D}} = \int_{\mathcal{D}} \sqrt{g} d^3 X \varrho$, in our comoving and synchronous gauge setting.

A fairly general insight that, in principle, will not depend on some specialized setting can be obtained by rewriting Eq. (6): The notion of noncommutativity mentioned above comes into the fore by observing that the time evolution of the average density does not coincide with the average of the locally evolved density:

$$\langle \dot{\varrho} \rangle_{\mathcal{D}} - \langle \dot{\varrho} \rangle_{\mathcal{D}} = \langle \varrho \theta \rangle_{\mathcal{D}} - \langle \varrho \rangle_{\mathcal{D}} \langle \theta \rangle_{\mathcal{D}} = \langle \delta \varrho \delta \theta \rangle_{\mathcal{D}}. \quad (8)$$

For the fluctuation terms on the rhs, which would vanish in the FLRW model without any perturbation, we can give a deeper interpretation. For this end, let us ask which functional will reproduce these terms upon performing the time derivative. First, note that for the averaged expansion rate $\langle \theta \rangle_{\mathcal{D}}$ the corresponding functional is the volume according to

$$\langle \theta \rangle_{\mathcal{D}} = \frac{\dot{V}_{\mathcal{D}}}{V_{\mathcal{D}}} =: 3H_{\mathcal{D}}. \quad (9)$$

The latter equality demonstrates that this quantity may be regarded as an *effective Hubble function*, which will show up in our discussion later.

Interestingly, the answer is provided, for $\varrho > 0$, by the functional $S\{\varrho \parallel \langle \varrho \rangle_{\mathcal{D}}\}$ [Eq. (2)] so that the source of noncommutativity in Eq. (8) is given (up to the sign) by the production of *relative information entropy*, defined as to measure the deviations from the average mass density due to the development of inhomogeneities:

$$\langle \dot{\varrho} \rangle_{\mathcal{D}} - \langle \dot{\varrho} \rangle_{\mathcal{D}} = - \frac{\dot{S}\{\varrho \parallel \langle \varrho \rangle_{\mathcal{D}}\}}{V_{\mathcal{D}}}. \quad (10)$$

This measure can actually be inferred from its definition in phase space in terms of the one-particle distribution function for dust matter; i.e., the matter density multiplied by a delta-function distribution in velocity space [9]. It is here, where generalizations of the matter model, e.g., supported by pressure, vorticity, and/or velocity dispersion could be implemented, resulting in more general entropies after taking velocity moments in phase space.

The reader may ask whether this measure is superior to the density fluctuation measure, which also provides a generally growing and positive-definite valuation of the density distribution. Let us give some answers to this question before we proceed.

A standard index of inhomogeneity in cosmology is the density contrast $\delta := (\delta \varrho) / \langle \varrho \rangle_{\mathcal{D}}$ and the derived positive measure $(\Delta \varrho)^2 := \langle \varrho^2 \rangle_{\mathcal{D}} - \langle \varrho \rangle_{\mathcal{D}}^2$. The relative information entropy or the distinguishability [Eq. (2)] may have further implications by exploiting results from information theory. At the present stage, we do not claim that this measure is superior to the density fluctuation, but rather it

is *complementary*. This can be illustrated by pointing out that both measures are “cousins” in a 1-parameter family of inhomogeneity measures defined by

$$\mathcal{F}_\alpha\{\varrho \parallel \langle \varrho \rangle_{\mathcal{D}}\} := \frac{\langle \varrho \rangle_{\mathcal{D}}}{\alpha} \left[\left\langle \left(\frac{\varrho}{\langle \varrho \rangle_{\mathcal{D}}} \right)^{\alpha+1} \right\rangle_{\mathcal{D}} - 1 \right], \quad (11)$$

with α being a real parameter. In the limit $\alpha \rightarrow 0$, the formula reproduces the relative entropy, $\mathcal{F}_{\alpha \rightarrow 0} \rightarrow S/V_{\mathcal{D}}$, whereas $\alpha = 1$ reproduces the density fluctuation, $\mathcal{F}_{\alpha=1} = (\Delta\varrho)^2/\langle \varrho \rangle_{\mathcal{D}}$. This interpolating formula is known as the *Tsallis relative entropy*. It should be emphasized that the limit $\alpha \rightarrow 0$ is singled out as the only measure that exactly provides the source of noncommutativity with regard to the density evolution.

Properties of the measure.—The measure $S\{\varrho \parallel \langle \varrho \rangle_{\mathcal{D}}\}$ forms one of the central concepts in information theory [1]; $S = 0$ (“zero structure”) is attained by the homogeneous mass distribution, $\varrho = \langle \varrho \rangle_{\mathcal{D}}$.

First, for strictly positive mass density, $\varrho > 0$, $S\{\varrho \parallel \langle \varrho \rangle_{\mathcal{D}}\}$ is positive definite, which can be readily confirmed, i.e., it is indeed a *measure*.

Let us have a closer look at the total time derivative of our measure. Following from what has been said above, we may write the total relative information entropy production as follows:

$$\frac{\dot{S}\{\varrho \parallel \langle \varrho \rangle_{\mathcal{D}}\}}{V_{\mathcal{D}}} = -\langle \delta\varrho\dot{\theta} \rangle_{\mathcal{D}} = -\langle \varrho\delta\dot{\theta} \rangle_{\mathcal{D}} = -\langle \delta\varrho\delta\dot{\theta} \rangle_{\mathcal{D}}. \quad (12)$$

The last quantity is bounded according to Schwarz’s inequality, so that we obtain

$$\left| \frac{\dot{S}\{\varrho \parallel \langle \varrho \rangle_{\mathcal{D}}\}}{V_{\mathcal{D}}} \right| = |\langle \delta\varrho\delta\dot{\theta} \rangle_{\mathcal{D}}| \leq \Delta\varrho\Delta\dot{\theta}, \quad (13)$$

with the positive-definite fluctuation amplitudes

$$\Delta\varrho := \sqrt{\langle (\delta\varrho)^2 \rangle}, \quad \Delta\dot{\theta} := \sqrt{\langle (\delta\dot{\theta})^2 \rangle}. \quad (14)$$

This inequality states that the temporal change of the ratio between the *distinguishability* of the density distribution from the homogeneous distribution and the volume is bounded by the density and expansion fluctuation amplitudes. We may say that the production of information in the Universe and its volume expansion are competing.

We may look more closely at bounds as well as kinematical and dynamical conditions for the total second time derivative of $S\{\varrho \parallel \langle \varrho \rangle_{\mathcal{D}}\}$. In [9], we give sufficient conditions for the *time convexity* of our measure. Let us put one of them into perspective. We consider the question under which condition the time derivative of the relative information entropy production is positive. A straightforward calculation provides

$$\frac{\ddot{S}}{V_{\mathcal{D}}} = -\langle \delta\varrho\delta\ddot{\theta} \rangle_{\mathcal{D}} + \langle \varrho \rangle_{\mathcal{D}}(\Delta\dot{\theta})^2. \quad (15)$$

Raychaudhuri’s equation,

$$\dot{\theta} = \Lambda - 4\pi G\varrho - \frac{1}{3}\theta^2 - 2\sigma^2, \quad (16)$$

with the rate of shear $\sigma := \sqrt{\frac{1}{2}\sigma^i{}_j\sigma^j{}_i}$, the shear tensor σ_{ij} being minus the trace-free part of the extrinsic curvature, together with the commutation rule (5) yields

$$\begin{aligned} \frac{\ddot{S}}{V_{\mathcal{D}}} &= 4\pi G(\Delta\varrho)^2 + \langle \varrho \rangle_{\mathcal{D}}(\Delta\dot{\theta})^2 + \frac{1}{3}\langle \delta\varrho\delta\dot{\theta}^2 \rangle_{\mathcal{D}} \\ &\quad + 2\langle \delta\varrho\delta\sigma^2 \rangle_{\mathcal{D}} \geq 4\pi G(\Delta\varrho)^2 \\ &\quad - \Delta\varrho \left[\frac{1}{3}\Delta\dot{\theta}^2 + 2\Delta\sigma^2 \right] + \langle \varrho \rangle_{\mathcal{D}}(\Delta\dot{\theta})^2. \end{aligned}$$

The rhs is positive if

$$\frac{1}{2} \frac{\Delta(\frac{1}{3}\dot{\theta}^2) + \Delta(2\sigma^2)}{\Delta\dot{\theta}} \leq \sqrt{4\pi G\langle \varrho \rangle_{\mathcal{D}}} = \frac{1}{t_{F_{\mathcal{D}}}}, \quad (17)$$

where $t_{F_{\mathcal{D}}}$ denotes the *effective free-fall time* on \mathcal{D} .

Equation (17) provides a sufficient condition for the time convexity of the relative information entropy, which can be met, if gravity dominates over expansion and shear fluctuations. Time convexity implies that entropy production eventually becomes positive, i.e., the structure eventually surfaces and its rate of formation increases.

Discussion and conjecture.—Looking at Eq. (12), we appreciate that the source, i.e., the averaged relative information entropy production density, can be positive or negative. In cosmology, the processes of a relative accumulation of matter (cluster formation) and a relative dilution of matter (void formation) create *structure* compared with the average distribution. Following from Eq. (12), information entropy is produced if, on average, there are overdense fluid elements ($\delta\varrho > 0$) which are contracting ($\theta < 0$), or underdense elements ($\delta\varrho < 0$) which are expanding ($\theta > 0$), respectively. With regard to cosmological structure formation, these two states are generically encountered in a self-gravitating system, i.e., for large enough times and looking at some regional scale, an asymmetry of states is created due to the coupling of the expansion rate to the rate of change of the density through the continuity equation. We know from a calculation of the measure in linear perturbation theory that the growing-mode solution supports states with $\{\delta\varrho > 0, \theta < 0\}$ (contracting clusters) and $\{\delta\varrho < 0, \theta > 0\}$ (expanding voids). Thus, *for sufficiently large times*, i.e., when the decaying mode disappears, our measure will increase.

Looking at Eq. (17), we conclude that also in the case of the second time derivative we have the possibility of time concavity of the relative information entropy. However, we have evidence that, at least for large enough times and on sufficiently large scales, time convexity *always* holds for a self-gravitating continuum of dust. In particular, in the linear perturbation theory, our measure is *always* time convex [9].

We can illustrate roughly the physical content of the sufficient condition (17) as follows. Concentrating on the linear regime by considering the case in which fluctuations of a quantity are small compared with their average values, we may expand the quadratic expressions and keep only the leading terms:

$$\Delta(\theta^2) \approx 2|\langle\theta\rangle_{\mathcal{D}}|\Delta\theta, \quad \Delta(\sigma^2) \approx 2|\langle\sigma\rangle_{\mathcal{D}}|\Delta\sigma. \quad (18)$$

In this limit, if we additionally think of a large domain featuring approximately vanishing average shear, $\langle\sigma\rangle_{\mathcal{D}} \approx 0$, the sufficient condition (17) reduces to the inequality

$$t_{F_{\mathcal{D}}} \leq t_{H_{\mathcal{D}}} =: |H_{\mathcal{D}}^{-1}|, \quad (19)$$

i.e., if the effective free-fall time on \mathcal{D} is smaller than the effective Hubble time $t_{H_{\mathcal{D}}}$, with $H_{\mathcal{D}}$ defined in Eq. (9), then time convexity of our measure is ensured under the given assumptions. The expectation that both positivity of the relative information entropy production and time convexity, which are supported by the linear perturbation results, will hold in the dust continuum *generically*, at least for large enough times and on sufficiently large scales of averaging, establishes the following.

Conjecture.—The relative information entropy of a dust matter model $S\{\varrho \parallel \langle\varrho\rangle_{\Sigma}\}$ is, for sufficiently large times, globally (i.e., averaged over the whole compact manifold Σ) an increasing function of time.

We are currently investigating nonlinear exact solutions for spherically symmetric domains [9], which may provide further support for our conjecture.

A note is in order as for the relation to observational constraints. In our context, a generalized form of Friedmann's differential equation governs the averaged expansion (9), and a set of four effective cosmological parameters can be defined [5,7]. Assuming that, on sufficiently large scales of averaging, kinematical fluctuations and the averaged 3-Ricci curvature have negligible contributions, then the sum of the cosmological parameters for the matter content and the cosmological term have to add up to 1; the former is indeed given by the fraction of the two competing times:

$$\Omega_{\mathcal{D}}^m := \frac{8\pi G\langle\varrho\rangle_{\mathcal{D}}}{3H_{\mathcal{D}}^2} = \frac{2t_{H_{\mathcal{D}}}^2}{3t_{F_{\mathcal{D}}}^2}. \quad (20)$$

Referring to observational results, e.g., by the Wilkinson Microwave Anisotropy Probe [10], its contribution is $\Omega_{\mathcal{D}}^m \approx 0.3$ and, thus, $t_{F_{\mathcal{D}}}$ is slightly larger than $t_{H_{\mathcal{D}}}$. Note that this does not immediately imply that our mea-

sure is not time convex, because the condition (19) derives from the sufficient condition (17), which provides only a rough estimation and is not very stringent. On cosmological scales, both times are indeed very similar, so that we should make the estimation tighter to see whether or not time convexity holds; this we postpone to the future work [9].

We contemplate that the measure that we propose in the present Letter not only incorporates an assessment of structure, but may turn out to be a fundamental quantity in many other respects, e.g., for the study of black holes and the early Universe.

We would like to thank Alvaro Domínguez, José Gaite, and Atsushi Taruya for constant interest and discussions. T.B. acknowledges hospitality during a visit in 2001 to the Hosoya Laboratory of the Tokyo Institute of Technology, where the main work on this subject was done with support by The Japanese Society for the Promotion of Science (JSPS). He also acknowledges hospitality from the University of Tokyo with support by the Research Center for the Early Universe (RESCEU, Tokyo), COE Monkasho Grant, where this Letter was written. This work is also partially supported by the Sonderforschungsbereich SFB 375 "Astroparticle Physics" by the German science foundation DFG.

*Electronic address: ahosoya@th.phys.titech.ac.jp

†Electronic address: buchert@theorie.physik.uni-muenchen.de

‡Electronic address: morita@okinawa-ct.ac.jp

- [1] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
- [2] S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959) (republished by Dover, New York, 1997).
- [3] S. Kullback and R. A. Leibler, *Ann. Math. Stat.* **22**, 79 (1951).
- [4] T. Buchert, *Gen. Relativ. Gravit.* **33**, 1381 (2001).
- [5] T. Buchert, *Gen. Relativ. Gravit.* **32**, 105 (2000).
- [6] T. Buchert, in *9th JGRG Meeting, Hiroshima 1999*, edited by Y. Eriguchi *et al.* (2000), pp. 306–321.
- [7] T. Buchert and M. Carfora, *Phys. Rev. Lett.* **90**, 031101 (2003).
- [8] T. Buchert and J. Ehlers, *Astron. Astrophys.* **320**, 1 (1997).
- [9] A. Hosoya, T. Buchert, and M. Morita (to be published).
- [10] D. N. Spergel *et al.*, *Astrophys. J. Suppl. Ser.* **148**, 175 (2003).