It is shown that a first-order cosmological perturbation theory for the open, flat and closed Friedmann-Lemaître-Robertson-Walker universes admits one, and only one, gauge-invariant variable which describes the perturbation to the energy density and which becomes equal to the usual Newtonian energy density in the non-relativistic limit. The same holds true for the perturbation to the particle number density. Using these two new variables, a new manifestly gauge-invariant cosmological perturbation theory has been developed.

Density perturbations evolve diabatically. Perturbations in the total energy density are gravitationally coupled to perturbations in the particle number density, irrespective of the nature of the particles. There is, in first-order, no back-reaction of perturbations to the global expansion of the universe.

Small-scale perturbations in the radiation-dominated era oscillate with an increasing amplitude, whereas in older, less precise treatments, oscillating perturbations are found with a decreasing amplitude. This is a completely new and, obviously, important result, since it makes it possible to explain and understand the formation of massive stars after decoupling of matter and radiation.

PACS numbers: 04.25.Nx, 97.10.Bt, 98.62.Ai, 98.80.Jk

I. INTRODUCTION

Since measurements of the fundamental parameters of our universe are today very precise, cosmology is nowadays a mature branch of astrophysics. Despite advances in observational as well as theoretical cosmology, there is, as yet, no manifestly covariant cosmological perturbation theory. It is the purpose of this article to fill up this gap.

We redo the calculations of Lifshitz [1] and Lifshitz and Khalatnikov [2], and we use new results of the literature, which were not known at the time when Lifshitz and Khalatnikov developed their theory. Combining these results with new insights obtained by us and described in detail in our report [3], we are able to develop a perturbation theory with the properties that both the evolution equations and their solutions are invariant under general infinitesimal coordinate transformations \( x^\mu \rightarrow x^\mu - \xi^\mu(x'^\nu) \). We refer to such a theory as a manifestly covariant gauge-invariant perturbation theory.

A. Former Insights

Firstly, it is mandatory [4–9] to use gauge-invariant variables to construct a perturbation theory. Secondly, although Lifshitz and Khalatnikov were aware of the fact that the system of perturbation equations can be divided into three independent systems of equations, namely for gravitational waves, vortices, and scalar perturbations, they did not use explicitly the decompositions (11) and (13) found by York [10] and Stewart [11]. This decomposition makes the computations much more tractable. Finally, an important result [10, 11] is that the perturbed metric tensor for scalar perturbations can be written in terms of two potentials (12). This facilitates the derivation of the Newtonian results (56) and (57) in the non-relativistic limit.

B. Our New Insights

In addition to the results found in the literature, we use the new insights obtained by us [3]. Firstly, we have found, using the combined First and Second Law of Thermodynamics, that, in general, density perturbations do not evolve adiabatically, as we will demonstrate in Section VII. Therefore, it is essential that in a cosmological perturbation theory one uses an equation of state for the pressure \( p = p(n, \varepsilon) \) rather than \( p = p(\varepsilon) \), where \( n \) is the particle number density and \( \varepsilon \) is the energy density of the universe. This enables us to show that a small negative internal pressure fluctuation may have a favorable effect on the growth rate of a density perturbation. Secondly, we have found that for scalar perturbations, the perturbed momentum constraint equations (8b) can be rewritten as an evolution equation (14b) for the local perturbation...
\[ 3R_{(1)} = \] to the global spatial curvature \[ 3R_{(0)} \] (6). Using this new insight we can rewrite the evolution equations for scalar perturbations as simple extensions (14) of the background equations (5). The sets of equations (5) and (14) are written with respect to the same system of reference. From these two sets of equations it follows that only three independent scalars, namely the energy density \( \epsilon \), the particle number density \( n \), and the expansion scalar \( \theta \), play a role in a perturbation theory. This reduces the number of possible gauge-invariant variables considerably. Finally, in a first-order perturbation theory the gravitational field is already weak. Therefore, we only need to take the limit that all particle velocities are negligible with respect to the speed of light, \( v/c \rightarrow 0 \), the so-called non-relativistic limit, see (51), in order to show that our treatment reduces to the Newtonian theory of gravity. In fact, we show that there exist two, and only two, gauge-invariant variables defined by (17a) with the property that in the non-relativistic limit the well-known Newtonian results (56) and (57) show up. Consequently, \( \epsilon_{(1)}^{\beta} \) and \( n_{(1)}^{\beta} \) are the real, physical, energy density perturbation and the particle number density perturbation.

II. RESULTS

The results of our approach are as follows. Perturbations in the total energy density are gravitationally coupled to perturbations in the particle number density (21b), irrespective of the nature of the particles [i.e., ordinary matter or Cold Dark Matter (CDM)] and independent of the scale of the perturbations. Local density perturbations do not affect the global expansion of the universe: in first-order there is no back-reaction.

For large-scale perturbations, our new theory corroborates the outcomes of the standard theory, with the exception of the fact that we do not find the non-physical gauge-modes: see our comments below expressions (34) and (47).

For small-scale perturbations, however, the matter is quite different. In the literature it is assumed that in the small-scale limit the Newtonian theory of gravity applies and that gauge dependent variables become in the small-scale limit automatically gauge-independent [4]. This assumption is not true. Indeed, we show in Section IX that a gauge dependent quantity is still gauge dependent in the non-relativistic limit. Consequently, a manifestly gauge-invariant theory is of the utmost importance for the study of the evolution of small-scale perturbations. For small-scale perturbations our perturbation theory yields results that differ substantially from the standard theory. Firstly, in the radiation-dominated era of a flat FLRW universe we find that small-scale perturbations oscillate with an increasing amplitude according to (35). The outcomes of the standard theory, namely decaying perturbations, are found to be a coordinate artefact. Secondly, after decoupling of matter and radiation, small-scale density perturbations evolve also very differently from the behavior predicted by the standard theory. For small-scale perturbations, our new evolution equation (37a) contains an entropy term which is of the same order of magnitude as the pressure term. This will shed new light on the problem of structure formation in the universe.

III. BASIC EQUATIONS IN SYNCHRONOUS COORDINATES

Due to the general covariance of the Einstein equations and conservation laws, Einstein’s gravitational theory is invariant under a general coordinate transformation \( x^\mu \rightarrow \tilde{x}^\mu(\chi) \), implying that preferred coordinate systems do not exist. In particular, the linearized Einstein equations and conservation laws are invariant under a general linear coordinate transformation

\[ x^0 \rightarrow x^0 - \psi(t, \mathbf{x}), \quad x^i \rightarrow x^i - \xi^i(t, \mathbf{x}), \] (1)

where \( \psi(t, \mathbf{x}) \) and \( \xi^i(t, \mathbf{x}) \) are four arbitrary, first-order (‘infinitesimal’) functions of the time and space coordinates. Since preferred systems of reference do not exist and since our result (21) is manifestly gauge-invariant, one may use any suitable and convenient coordinate system to perform the calculations. In the Newtonian theory of gravity all systems of reference are synchronous because space and time are decoupled in this theory, see (54). In order to facilitate the interpretation of our new gauge-invariant quantities, we use, like Lifshitz and Khalatnikov do, synchronous coordinates. In this coordinate system the metric of FLRW universes has the form

\[ g_{00} = 1, \quad g_{0i} = 0, \quad g_{ij} = -a^2(t)\tilde{g}_{ij}(\mathbf{x}), \] (2)

where \( a(t) \) is the scale factor of the universe, and \( \tilde{g}_{ij} \) is the metric of the three-dimensional maximally symmetric subspaces of constant time. The functions \( \psi \) and \( \xi^i \) of the general infinitesimal transformation (1) reduce to

\[ \psi = \psi(\mathbf{x}), \quad \xi^i = \tilde{g}^{ik}\partial_k \psi(\mathbf{x}) \int^\tau_0 d\tau \frac{a^2(\tau)}{a^2(\tau)} + \chi^i(\mathbf{x}), \] (3)

if only transformations between synchronous coordinates are allowed. In (3), \( \psi(\mathbf{x}) \) and \( \chi^i(\mathbf{x}) \) are four arbitrary, infinitesimal, functions of the spatial coordinates.

A. Background Equations

The complete set of zeroth-order Einstein equations and conservation laws for an open, flat or closed FLRW universe filled with a perfect fluid with energy-momentum tensor

\[ T^{\mu\nu} = (\varepsilon + p)u^\mu u^\nu - pg^{\mu\nu}, \quad p = p(n, \varepsilon), \] (4)
is, in synchronous coordinates, given by
\begin{align}
\text{Constraint: } & \ 3H^2 = 1 \ 3R_{(o)} + \kappa \varepsilon_{(o)} + \Lambda, \quad (5a) \\
\text{Evolution: } & \ 3R_{(o)} = -2H \ 3R_{(o)}, \quad (5b) \\
\text{Conservation: } & \ \dot{\varepsilon}_{(o)} = -3H \varepsilon_{(o)}(1 + w), \quad (5c) \\
\dot{\vartheta}_{(o)} = 0, \quad (5d) \\
\dot{n}_{(o)} = -3H n_{(o)}, \quad (5e)
\end{align}

where the equation of state for the pressure \( p \) is supposed to be a given function of the energy density \( \varepsilon \) and the particle number density \( n \). The \( G_{0i} \) constraint equations and the \( G_{ij}, i \neq j \), dynamical equations are identically satisfied. The \( G_{\mu \nu} \) dynamical equations are equivalent to the time-derivative of \((5a)\). Therefore, the \( G_{ij} \) dynamical equations need not be taken into account. In equations \((5)\) \( \Lambda \) is the cosmological constant, \( \kappa = 8\pi G/c^4 \), and \( w \equiv p_{(o)}/\varepsilon_{(o)} \). An overdot denotes differentiation with respect to \( ct \) and the sub-index \((0)\) refers to the background, i.e., unperturbed, quantities. Furthermore, \( H \equiv \dot{a}/a \) is the Hubble function which is equal to \( H = \frac{1}{3} \theta_{(o)} \), where \( \theta_{(o)} \) is the background value of the expansion scalar \( \theta \equiv u^\mu u_\mu \) with \( u^\mu \) the four-velocity, normalized to unity \((u^\mu u_\mu = 1)\). A semicolon denotes covariant differentiation with respect to the background metric \( g_{(o)\mu \nu} \). The spatial curvature \( 3R_{(o)} \), is given by
\[ 3R_{(o)} = -\frac{6k}{a^2}, \quad k = -1, 0, +1. \quad (6) \]

The quantity \( \vartheta_{(o)} \) is the three-divergence of the spatial part of the four-vector \( u_{(o)i}^\mu \). For an isotropically expanding universe we have \( u_{(o)i}^\mu = \delta_{(o)i}^\mu \), so that \( \vartheta_{(o)} = 0 \).

From the system \((5)\) one may infer that the evolution of an unperturbed FLRW universe is determined by exactly three independent scalars, namely
\[ \varepsilon = T^{\mu \nu} u_{\mu} u_{\nu}, \quad n = N^{\mu} u_{\mu}, \quad \theta = u^{\mu \cdot \mu}, \quad (7) \]

where \( N^{\mu} \equiv n u^{\mu} \) is the cosmological particle current four-vector, which satisfies the particle number conservation law \( N^{\mu \cdot \mu} = 0 \).

### B. Perturbation Equations

The complete set of first-order Einstein equations and conservation laws for the open, flat and closed FLRW universe is, in synchronous coordinates, given by
\begin{align}
\text{H} & \ H^{k}_{i} + \frac{1}{2} 3R_{(1)i} = -\kappa \varepsilon_{(1)i}, \quad (8a) \\
\dot{h}^{k}_{i|j} & = -2k(\varepsilon_{(0)} + p_{(0)}) u_{(1)i}, \quad (8b) \\
\dot{h}^{ij}_{(1)} & + 3H h^{ij}_{(1)} + \delta^{ij} H h^{k}_{k} + 2 3R^{i}_{(1)jk} = -\kappa \delta^{ij} (\varepsilon_{(1)} - p_{(1)}), \quad (8c) \\
\dot{\varepsilon}_{(1)} & + 3H (\varepsilon_{(1)} + p_{(1)}) + (\varepsilon_{(0)} + p_{(0)}) \theta_{(1)} = 0, \quad (8d) \\
\frac{1}{c} \frac{d}{dt} [ & \ (\varepsilon_{(0)} + p_{(0)}) u^{i}_{(1)} ] - u_{(0)i}^{j} \ (\varepsilon_{(0)} + p_{(0)}) n_{(0)} u^{j}_{(1)} = 0, \quad (8e) \\
\dot{n}_{(1)} & + 3H n_{(1)} + n_{(0)} \theta_{(1)} = 0, \quad (8f)
\end{align}

where \( h^{i}_{j} = g^{ij}_{(0)} h_{kj} \) is the perturbed metric, and \( g^{ij}_{(0)} \) the background metric \((2)\) for an open, closed or flat FLRW universe. Quantities with a sub-index \((1)\) are the first-order counterparts of the corresponding background quantities with a sub-index \((0)\). The first-order perturbation to the spatial part of the Ricci tensor is \( 3R^{(1)ij}_{(1)} \) and its trace is given by
\[ 3R_{(1)i} = g^{ij}_{(0)} (h^{k}_{kij} - h^{k}_{ijj}) + \frac{1}{3} 3R_{(0)k} h^{k}_{i}. \quad (9) \]

A vertical bar denotes covariant differentiation with respect to the background metric \( g_{(o)ij} \). Expression \((9)\) is the local perturbation to the global spatial curvature \( 3R_{(0)} \) due to a local density perturbation. Finally, \( \theta_{(1)} \) is the first-order perturbation to the expansion scalar \( \theta \). It is found that
\[ \theta_{(1)} = \vartheta_{(1)} - \frac{1}{2} h^{k}_{i} k_{(1)k}, \quad \vartheta_{(1)} = u^{k}_{(1)k}, \quad (10) \]

where \( \vartheta_{(1)} \) is the divergence of the spatial part of the perturbed four-vector \( u^{i}_{(1)} \). The quantities \((9)\) and \((10)\) play an important role in the derivation of our manifestly gauge-invariant perturbation theory.

### IV. FIRST-ORDER EQUATIONS FOR SCALAR PERTURBATIONS

Since the work of Lifshitz and Khalatnikov, it is known that for FLRW universes, the set \((8)\) can be broken down into three independent systems of equations, namely a system for gravitational waves (tensor perturbations), for vortices (vector perturbations), and for scalar perturbations. In addition they showed that only scalar perturbations are coupled to density perturbations. Lifshitz and Khalatnikov used spherical harmonics to classify the three types of perturbations. A disadvantage of their approach is that using spherical harmonics is computationally demanding. A new insight, gained in 1974 by York [10], and in 1990 elucidated by Stewart [11], makes the computations much more tractable. York and Stewart showed that the symmetric perturbation tensor \( h^{i}_{j} \) can uniquely be decomposed into three parts, i.e.,
\[ h^{i}_{j} = h^{i}_{||j} + h^{i}_{\perp j} + h^{i}_{* j}, \quad (11a) \]
\[ h^{k}_{||k} = 0, \quad h^{k}_{\perp k} = 0, \quad h^{k}_{* |k} = 0, \quad (11b) \]

where the scalar, vectorial and tensorial perturbations are denoted by \( ||, \perp \) and \(*\), respectively. Moreover, they demonstrated that the component \( h^{i}_{||j} \) can be written in terms of two independent potentials \( \phi(t, x) \) and \( \zeta(t, x) \):
\[ h^{i}_{||j} = \frac{2}{c^2} (\phi \delta^{ij} + \zeta^{ij}). \quad (12) \]

Finally, Stewart also proved that the spatial part of the perturbed four-vector \( u^{i}_{(1)} \) can uniquely be decomposed into two parts
\[ u^{i}_{(1)} = u^{i}_{(1)||} + u^{i}_{(1)i} \perp, \quad (13a) \]
\[ \nabla^{i} u^{i}_{(1)} = \nabla^{i} u^{i}_{(1)||}, \quad \nabla^{i} u^{i}_{(1)} = \nabla^{i} u^{i}_{(1)i} \perp. \quad (13b) \]
where \((\tilde{\nabla} f)^i \equiv \tilde{g}^{ij} \partial_j f\).

The perturbed Ricci tensor, \(3R^i_{(1)j}\), being a symmetric tensor, should also obey the decomposition (11a) with the properties (11b), i.e., \(3R^k_{(1)\perp k} = 0\). As a consequence, \(h^{ij}_{\perp}\) must obey \(k^{ij}_{\perp k|l} = 0\), in addition to (11b), as follows from (9). This additional property is a necessary and sufficient condition for the decomposition of the perturbation equations into three distinct types of perturbations, with \(u_{(1)i}\) coupled to scalar perturbations and \(u_{(1)\perp}\) coupled to vortices, according to their properties (13b).

Since only the scalar part of the perturbations is coupled to density perturbations, we may replace in (8)–(10) \(h^{ij}_{\perp}\) by \(h^{ij}_{\perp}\) and \(u_{(1)i}\) by \(u_{(1)i}\), to obtain perturbation equations which exclusively describe the evolution of scalar perturbations. Using the decompositions (11a) and (13a) and the properties (11b) and (13b), we can rewrite the evolution equations for scalar perturbations in the form

\[
2H(\theta_{(1)} - \eta_{(1)}) - \frac{3}{2} \tilde{R}\|_{(1)} = \kappa \varepsilon_{(1)},
\]

\[
3\tilde{R}\|_{(1)} + 2H \tilde{R}\|_{(1)} - 2\kappa \varepsilon_0 (1 + w)\theta_{(1)} + \frac{3}{2} \tilde{R}_0 \varepsilon_{(1)} = 0,
\]

\[
\dot{\varepsilon}_{(1)} + 3H(\varepsilon_{(1)} + p_{(1)}(1 + w)\theta_{(1)} = 0,
\]

\[
\dot{\theta}_{(1)} + H(2 - 3\beta^2)\theta_{(1)} + \frac{1}{\varepsilon_{(1)}(1 + w)} \tilde{\nabla}^2 p_{(1)} = 0,
\]

\[
\dot{\eta}_{(1)} + 3Hn_{(1)} + n_0 \theta_{(1)} = 0.
\]

The set (14), which is the perturbed counterpart of the set (5), consists of five equations for the five unknown functions \(\varepsilon_{(1)}, n_{(1)}, \theta_{(1)}, \tilde{R}\|_{(1)}\) and \(\theta_{(1)}\). The first-order perturbation to the pressure is given by the perturbed evolution of state \(p_{(1)} = p_n n_{(1)} + p_z \varepsilon_{(1)}\), where \(p_n \equiv (\partial p/\partial n)_\varepsilon\) and \(p_z \equiv (\partial p/\partial \varepsilon)_n\) are the partial derivatives of the equation of state \(p(m, \varepsilon)\), and \(\beta^2 = p_\varepsilon(p_\varepsilon/\varepsilon_0)\). Finally, the symbol \(\tilde{\nabla}^2\) denotes the generalized Laplace operator with respect to the three-space metric \(\tilde{g}_{ij}\).

We now sketch the derivation of the basic equations (14) for scalar perturbations. Eliminating \(h^{ij}_{\perp}\) from (8a) with the help of (10) yields (14a). Multiplying both sides of equation (8b) by \(\tilde{g}_{ij}^{\prime}\), and taking the covariant derivative with respect to the index \(j\) and using (9), (10) and the fact that the background connection coefficients \(\Gamma^k_{(0)ij}\) are for FLRW metrics independent of time, we arrive at equation (14b). In the derivation of (14b) we have used that \(g^{ij}_{(1)0} h^{kij}_{(1)|l} = g^{ij}_{(1)0} h^{kij}_{(1)l}\), which follows from \(g^{ij}_{(0)k} = 0\) and the symmetry of \(h^{ij}_{\perp}\). Equations (8c) need not be considered, since for \(i \neq j\) they are not coupled to scalar perturbations, whereas the trace of (8c) is, just as in the background case, equivalent to the time-derivative of the constraint equation (14a). Finally, equation (14d) can be obtained from (8c) by taking the covariant derivative with respect to the metric \(g^{ij}_{(1)}\). This concludes the derivation of the system (14). As follows from its derivation, this system is, for scalar perturbations, equivalent to the full set of first-order Einstein equations and conservation laws (8).

V. UNIQUE COSMOLOGICAL DENSITY PERTRUBATIONS

The background equations (5) and the perturbation equations (14) are both written with respect to the same system of reference. Therefore, these two sets can be combined to describe the evolution of the five background quantities \(\theta_{(0)} = 3H, \tilde{R}_{(0)}, \varepsilon_{(0)}, \vartheta_{(0)} = 0\) and \(n_0\), and their first-order counterparts \(\theta_{(1)}, \tilde{R}_{(1)}\|, \varepsilon_{(1)}, \vartheta_{(1)}\) and \(n_{(1)}\). Just as in the background case, we again come across the three independent scalars (7). Consequently, the evolution of cosmological density perturbations is described by the three independent scalars (7). A complicating factor is that the first-order quantities \(\varepsilon_{(1)}\) and \(n_{(1)}\), which are supposed to describe the energy density and the particle number density perturbations, have no physical significance, as we will now establish.

A first-order perturbation to one of the scalars \((\varepsilon)\) transforms under a general (not necessarily between synchronous coordinates) infinitesimal coordinate transformation (1) as

\[
S_{(1)}(t, x) \rightarrow S_{(1)}(t, x) + \psi(t, x) \tilde{S}_{(0)}(t),
\]

where \(S_{(0)}\) and \(S_{(1)}\) are the background and first-order perturbation of one of the three scalars \(S = \varepsilon, n, \vartheta\). In (15) the term \(\tilde{S} \equiv \psi S_{(0)}\) is the so-called gauge mode. The complete set of gauge modes is given by

\[
\tilde{\varepsilon}_{(1)} = \psi \varepsilon_{(0)}, \quad \tilde{n}_{(1)} = \psi n_{(0)}, \quad \tilde{\theta}_{(1)} = \psi \theta_{(0)},
\]

\[
\tilde{\vartheta}_{(1)} = -\frac{\nabla^2 \psi}{a^2}, \quad \tilde{\tilde{R}}_{(1)\|} = 4H \left[ \frac{\nabla^2 \psi}{a^2} - \frac{1}{2} \tilde{R}_{(0)} \psi \right],
\]

where expressions (16b) hold true only in synchronous coordinates. The quantities (16) are mere coordinate artifacts, which have no physical meaning, since the gauge function \(\psi(x)\) is an arbitrary (infinitesimal) function. Equations (14) are invariant under coordinate transformations (1) with (3). This property combined with the linearity of the perturbation equations, implies that a solution set \((\varepsilon_{(1)}, n_{(1)}, \theta_{(1)}, \vartheta_{(1)}, \tilde{R}_{(1)\|})\) can be augmented with the corresponding gauge modes (16) to obtain a new solution set. Therefore, the solution set \((\varepsilon_{(1)}, n_{(1)}, \theta_{(1)}, \vartheta_{(1)}, \tilde{R}_{(1)\|})\) has no physical significance, since the general solution of the set (14) can be modified by an infinitesimal coordinate transformation.

It is possible to eliminate the gauge modes for the scalars \(\varepsilon, n\) and \(\theta\), by constructing so-called gauge-invariant variables, i.e., variables that do not change under general infinitesimal coordinate transformations (1). All attempts [4–9] to construct a gauge-invariant cosmological perturbation theory yield different outcomes. For instance, Bardeen [4] defines three different gauge-invariant variables to describe energy density perturbations, which become equal to each other only in the small-scale limit. The variables of Bardeen differ from the ones used by Mukhanov et al. [5, 9], and the variables used by Mukhanov et al. differ, in turn, from the ones used by
Ellis et al. [6–8]. Apparently, the number of possibilities to construct gauge-invariant variables is large. This is why there is no consensus about which variable exactly describes the evolution of energy density perturbations in the universe. This is the notorious gauge problem of cosmology.

To unravel this problem, we first have rewritten the perturbation equations (8) into the form (14) in order to isolate the scalar perturbations from the vortices and gravitational waves. This reduces the number of possible gauge-invariant variables substantially, since we need only consider the three independent scalars (7). All scalar perturbations transform in the same way (15). This implies that we can combine two independent scalars to eliminate the gauge function \( \psi(t, x) \). With the three independent scalars (7), we can make three different sets of three gauge-invariant variables. In each of these sets exactly one gauge-invariant quantity vanishes. As we will show in Section IX, the only set for which the corresponding perturbation theory yields in the non-relativistic limit \( v/c \to 0 \) the well-known Newtonian results (56) and (57) is given by

\[
\begin{align*}
\varepsilon^{(1)} &= \varepsilon^{(t)} - \frac{\dot{\varepsilon}^{(0)}}{\theta^{(t)}}, \quad n^{(1)} = n^{(t)} - \frac{\dot{n}^{(0)}}{\theta^{(t)}}, \quad (17a) \\
\theta^{(1)} &= \theta^{(t)} - \frac{\dot{\theta}^{(0)}}{\theta^{(t)}} \equiv 0. \quad (17b)
\end{align*}
\]

It follows from the general transformation rule (15) that the quantities (17) are invariant under the infinitesimal transformation (1), i.e., they are gauge-invariant, hence the superscript ‘gi’. The physical interpretation of (17b) is that, in first-order, the gauge function \( \psi(t, x) \). This implies that we can combine two independent scalars to eliminate the gauge function \( \psi(t, x) \).

The gauge-invariant quantities (17a) are completely determined by the background equations (5) and their perturbed counterparts (14). In principle, we can use these two sets to study the evolution of density perturbations in FLRW universes. The set (14) is still too complicated, since it also admits the non-physical solutions (16).

Our aim will be a system of equations for \( \varepsilon^{(1)} \) and \( n^{(1)} \) that do not have the gauge modes (16) as solution. In other words, we are looking for a perturbation theory for which not only the differential equations are invariant under general infinitesimal coordinate transformations (1), but also their solutions. We refer to such a theory as a manifestly covariant gauge-invariant perturbation theory. The construction of such a theory will be the subject of the next section.

VI. PERTURBATION EQUATIONS FOR DENSITY PERTURBATIONS

In this section we will outline the derivation of our new perturbation theory. Firstly, we observe that the gauge dependent variable \( \theta^{(1)} \) is not needed in our calculations, since its gauge-invariant counterpart \( \theta^{(1)}_{gi} \), (17b), vanishes identically. Eliminating \( \theta^{(1)} \) from the differential equations (14b)–(14e) with the help of the algebraic equation (14a) yields the set of four first-order differential equations

\[
\begin{align*}
3\dot{R}_{(1)} + 2H 3R_{(1)} &= 0, \quad (18a) \\
3\dot{\varepsilon}^{(0)}(1 + w)\dot{\theta}^{(1)} + 3\frac{R^{(0)}}{3H} (\kappa\varepsilon^{(1)} + \frac{1}{2} 3R^{(1)}) &= 0, \quad (18a) \\
\dot{\varepsilon}^{(1)} + 3H (\varepsilon^{(1)} + \rho^{(1)}) + \varepsilon^{(0)}(1 + w)\theta^{(1)} + \frac{1}{2H} (\kappa\varepsilon^{(1)} + \frac{1}{2} 3R^{(1)}) &= 0, \quad (18b) \\
\dot{\varepsilon}^{(1)} + 3H n^{(1)} + n^{(0)}\theta^{(1)} + \frac{1}{H} (\kappa\varepsilon^{(1)} + \frac{1}{2} 3R^{(1)}) &= 0, \quad (18c)
\end{align*}
\]

for the four quantities \( \varepsilon^{(1)}, n^{(1)}, \theta^{(1)}, \) and \( 3R^{(1)} \).

Using the background equations (5) to eliminate all time-derivatives and the first-order constraint equation (14a) to eliminate \( \theta^{(1)} \), we can rewrite the gauge-invariant quantities (17a) as

\[
\begin{align*}
\varepsilon^{(1)} &= \varepsilon^{(0)} - 3\varepsilon^{(0)} (1 + w) (2H \theta^{(1)} + \frac{1}{2} 3R^{(1)}) \frac{3R^{(0)} + 3n^{(0)}\varepsilon^{(1)} + 2H \theta^{(1)} + \frac{1}{2} 3R^{(1)}}{3R^{(0)} + 3\kappa\varepsilon^{(0)}(1 + w)}, \quad (19a) \\
n^{(1)} &= n^{(0)} - \frac{3n^{(0)}\varepsilon^{(1)} + 2H \theta^{(1)} + \frac{1}{2} 3R^{(1)}}{3R^{(0)} + 3\kappa\varepsilon^{(0)}(1 + w)}. \quad (19b)
\end{align*}
\]

These quantities are now completely determined by the background equations (5) and the first-order equations (18). In the study of the evolution of density perturbations, it is convenient not to use \( \varepsilon^{(1)} \) and \( n^{(1)} \) directly, but instead their corresponding contrast functions \( \delta_c \) and \( \delta_n \) defined by

\[
\delta_c(t, x) \equiv \frac{\varepsilon^{(1)}(t, x)}{\varepsilon^{(0)}(t)}, \quad \delta_n(t, x) \equiv \frac{n^{(1)}(t, x)}{n^{(0)}(t)}. \quad (20)
\]

We now rewrite the system of equations (18) for the four independent quantities \( \varepsilon^{(1)}, n^{(1)}, \theta^{(1)}, \) and \( 3R^{(1)} \) into a new system of equations for the two independent quantities \( \delta_c \) and \( \delta_n \). The final result is our new manifestly covariant gauge-invariant perturbation theory:

\[
\begin{align*}
\ddot{\delta}_c + b_1 \dot{\delta}_c + b_2 \delta_c &= \frac{b_1}{b_3} \left[ \delta_n - \frac{\delta_c}{1 + w} \right], \quad (21a) \\
\frac{1}{c^2} \frac{d}{dt} \left[ \delta_n - \frac{\delta_c}{1 + w} \right] &= \frac{3H n^{(0)} p_n}{\varepsilon^{(0)}(1 + w)} \left[ \delta_n - \frac{\delta_c}{1 + w} \right]. \quad (21b)
\end{align*}
\]

These are two differential equations for the two independent and gauge-invariant quantities \( \delta_c \) and \( \delta_n \). It follows from equation (21b) that perturbations in the total energy density are gravitationally coupled to perturbations in the particle number density if \( p_n \equiv (\partial p/\partial n)_{\varepsilon} \leq 0 \). This is the case in a FLRW universe in the radiation-dominated era and after decoupling of matter and radiation. This coupling is independent of the nature of the
particles, i.e., it holds true for ordinary matter as well as CDM.

The system of equations (21) is equivalent to a system of three first-order differential equations, whereas the original set (18) is a fourth-order system. This difference is due to the fact that the gauge modes, which are solutions of the set (18), are completely removed from the solution set of (21): one degree of freedom, namely the gauge function \( \psi \), has disappeared.

The coefficients \( b_1 \), \( b_2 \) and \( b_3 \) occurring in equation (21a) are given by

\[
\begin{align*}
    b_1 &= \frac{\kappa \varepsilon_{(0)}(1 + w)}{H} - \frac{2}{3} b \beta - H(2 + 6w + 3\beta^2) + 3R_{(0)} \left( \frac{1}{3H} + \frac{2H(1 + 3\beta^2)}{3R_{(0)} + 3\kappa \varepsilon_{(0)}(1 + w)} \right), \\
    b_2 &= -\frac{1}{2} \kappa \varepsilon_{(0)}(1 + w)(1 + 2w) + H^2 (1 - 3w + 6\beta^2(2 + 3w)) + 6H \frac{\beta}{\beta} \left( w + \frac{\kappa \varepsilon_{(0)}(1 + w)}{3R_{(0)} + 3\kappa \varepsilon_{(0)}(1 + w)} \right) - 3R_{(0)} \left( \frac{1}{2} w + \frac{H^2(1 + 6w)(1 + 3\beta^2)}{3R_{(0)} + 3\kappa \varepsilon_{(0)}(1 + w)} \right) - \beta^2 \left( \frac{\nabla^2}{a^2} - \frac{1}{2} \frac{3R_{(0)}}{3R_{(0)} + 3\kappa \varepsilon_{(0)}(1 + w)} \right), \\
    b_3 &= \left\{ -18H^2 \left[ \varepsilon_{(0)}p_{\text{cn}}(1 + w) + \frac{2p_n \beta}{3H \beta} \right] - \beta^2 p_n + \varepsilon_{(0)} \beta(1 + 3\beta^2) \right\} + \nabla \left( \frac{\nabla^2}{a^2} - \frac{1}{2} \frac{3R_{(0)}}{3R_{(0)} + 3\kappa \varepsilon_{(0)}(1 + w)} \right),
\end{align*}
\]

where \( p_{\text{cn}} \equiv \partial^2 p / \partial n^2 \) and \( p_{\text{cn}} \equiv \partial^2 p / \partial \varepsilon \, \partial n \).

The background equations (5) and the new perturbation equations (21) constitute a set of equations which enables us to study the evolution of small fluctuations in an open, flat or closed FLRW universe filled with a perfect fluid with an equation of state \( p(n, \varepsilon) \).

VII. DIABOTIC DENSITY PERTURBATIONS

In this section we will relate equations (21) to thermodynamics and we will show that, in general, density perturbations do not evolve adiabatically, as is usually assumed, but diabatically, i.e., they exchange heat with their environment during their evolution.

The combined first and second law of thermodynamics is given by

\[
T ds = d\left( \frac{\varepsilon}{n} \right) + pd\left( \frac{1}{n} \right),
\]

where \( s \) is the entropy per particle. Using the background equations (5) it follows that \( \dot{s}_{(0)} = 0 \). This implies with (15), that \( s_{(1)} \equiv s_{(1)}^{\text{gi}} \) is automatically gauge-invariant. The thermodynamic relation (23) can, using (20), be rewritten in the form

\[
T_{(0)} \varepsilon_{(1)}^{\text{gi}} = -\frac{\varepsilon_{(0)}(1 + w)}{n_{(0)}} \left[ \dot{\varepsilon}_{(0)} - \frac{\delta \varepsilon}{1 + w} \right].
\]

Thus, the right-hand side of (21a) is related to local perturbations in the entropy.

Adiabatic perturbations do not exchange heat with their surroundings, so that \( T_{(0)} s_{(1)}^{\text{gi}} = 0 \). Using (20) and the background conservation laws (5c) and (5e), we find from \( s_{(1)}^{\text{gi}} = 0 \) the adiabatic condition

\[
\dot{n}_{(0)} \varepsilon_{(1)}^{\text{gi}} = \dot{\varepsilon}_{(0)} n_{(0)}^{\text{gi}} = 0.
\]

In a non-static universe (i.e., \( \dot{\varepsilon}_{(0)} \neq 0 \) and \( \dot{n}_{(0)} \neq 0 \)) this equation is fulfilled if, and only if, the energy density is a function of the particle number density only, i.e., if and only if \( \varepsilon = \varepsilon(n, T) \) and \( p = p(n, T) \), where the particle number density \( n \) and the temperature \( T \) are independent quantities. Substituting \( \varepsilon = \varepsilon(n, T) \) into (25) yields

\[
\left( \frac{\partial \varepsilon}{\partial T} \right) \left[ \dot{n}_{(0)} T_{(1)}^{\text{gi}} - n_{(1)}^{\text{gi}} \dot{T}_{(0)} \right] = 0.
\]

Since \( n \) and \( T \) are independent quantities, the adiabatic condition (26) is satisfied if, and only if,

\[
\left( \frac{\partial \varepsilon}{\partial T} \right) = 0,
\]

implying that \( \varepsilon = \varepsilon(n) \). In particular, in the non-relativistic limit, where \( \varepsilon = \rho m c^2 \) and \( p = 0 \), density perturbations are adiabatic. In all other cases, \( \varepsilon = \varepsilon(n, T) \) and \( p = p(n, T) \) [hence \( p = p(n, \varepsilon) \)], local density perturbations evolve diabatically.

VIII. THE FLAT FLRW UNIVERSE

We consider a flat \( k = 0 \), (6) FLRW universe in its radiation-dominated phase and in the era after decoupling of matter and radiation. In this section only, we take \( \Lambda = 0 \).

A. Radiation-dominated Era

In this epoch we have \( \varepsilon = a_B T_R^4 \), where \( a_B \) is the black body constant and \( T_R \) the radiation temperature. The
pressure is \( p = \frac{1}{3} \varepsilon \), so that \( p_n = 0 \), \( p_c = \frac{1}{3} \) and \( \beta^2 = \frac{1}{3} \). In this case, the perturbation equations (21) reduce to

\[
\dot{\delta}_\varepsilon - H \delta_\varepsilon - \left[ \frac{1}{3} \nabla^2 - \frac{2}{3} \kappa \varepsilon_0 \right] \delta_\varepsilon = 0, \quad (28a)
\]

\[
\frac{1}{c} \frac{d}{dt} \left( \delta_n - \frac{3}{4} \delta_\varepsilon \right) = 0. \quad (28b)
\]

Equation (28b) implies that the difference \( \delta_n - \frac{3}{4} \delta_\varepsilon \) depends only on the spatial coordinates. Since the universe is non-static and \( (\partial \delta_\varepsilon / \partial T)_n \neq 0 \), it follows that (26) cannot be satisfied, implying that the perturbations are not adiabatic, so that \( \delta_n(t_0, \mathbf{x}) - \frac{3}{4} \delta_\varepsilon(t_0, \mathbf{x}) \neq 0 \).

Substituting \( \delta_\varepsilon(t, \mathbf{x}) = \delta_\varepsilon(t, \mathbf{q}) \exp(i \mathbf{q} \cdot \mathbf{x}) \) into equation (28a) and using the well-known solutions of the background equations (5)

\[
H \propto t^{-1}, \quad \varepsilon_{(0)} \propto t^{-2}, \quad n_{(0)} \propto t^{-3/2}, \quad a \propto t^{1/2}, \quad (29)
\]
equation (28a) reduces to

\[
\delta_\varepsilon'' - \frac{1}{2r} \delta_\varepsilon' + \left[ \frac{\mu^2}{4r^2} + \frac{1}{2r^2} \right] \delta_\varepsilon = 0, \quad (30)
\]

where \( \mu_r \) is given by

\[
\mu_r = \frac{2\pi}{\lambda_0} \frac{1}{H(t_0)} \frac{1}{\sqrt{3}}, \quad \lambda_0 \equiv \lambda a(t_0), \quad (31)
\]

with \( \lambda a(t_0) \) the physical scale of a perturbation at time \( t_0 \), and \( |\mathbf{q}| = 2\pi/\lambda \). The exact solution of (30) is

\[
\delta_\varepsilon(\tau, \mathbf{q}) = \left[ A_1(\mathbf{q}) \sin(\mu_r \sqrt{\tau}) + A_2(\mathbf{q}) \cos(\mu_r \sqrt{\tau}) \right] \sqrt{\tau},
\]

where \( \tau \equiv t/t_0 \). The ‘constants’ of integration \( A_1(\mathbf{q}) \) and \( A_2(\mathbf{q}) \) are given by

\[
A_{1,2}(\mathbf{q}) = \delta_\varepsilon(t_0, \mathbf{q}) \frac{\sin \mu_r}{\cos \mu_r}
\]

\[
\mp \frac{1}{\mu_r} \cos \mu_r \left[ \delta_\varepsilon(t_0, \mathbf{q}) - \frac{\dot{\delta}_\varepsilon(t_0, \mathbf{q}) - \dot{\delta}_n(t_0, \mathbf{q})}{H(t_0)} \right]. \quad (33)
\]

For large-scale perturbations (\( \lambda \to \infty \)), we arrive at

\[
\delta_\varepsilon(t) = - \left[ \delta_\varepsilon(t_0) - \frac{\delta_n(t_0)}{H(t_0)} \right] \frac{t}{t_0} + \left[ 2 \delta_\varepsilon(t_0) - \frac{\delta_n(t_0)}{H(t_0)} \right] \left( \frac{t}{t_0} \right)^{1/2}. \quad (34)
\]

The energy density contrast has two contributions to the growth rate, one proportional to \( t \) and one proportional to \( t^{1/2} \). These two solutions have been found, with the exception of the precise factors of proportionality, by a large number of authors. See Lifshitz and Khalatnikov [2], (8.11), Adams and Canuto [12], (4.5b), Olson [13], page 329, Peebles [14], (86.20), Kolb and Turner [15], (9.121) and Press and Vishniac [16], (33).

A new result of our perturbation theory is that small-scale perturbations oscillate with an *increasing* amplitude, proportional to \( t^{1/2} \), according to

\[
\delta_\varepsilon(t, \mathbf{q}) \approx \delta_\varepsilon(t_0, \mathbf{q}) \left( \frac{t}{t_0} \right)^{1/2} \cos \left[ \mu_r - \mu_l \left( \frac{t}{t_0} \right)^{1/2} \right]. \quad (35)
\]

These small-scale perturbations manifest themselves as small temperature fluctuations in the cosmic background radiation.

In contrast to our theory, the standard perturbation theory predicts oscillating density perturbations with a *decreasing* amplitude. As is explained in our report [3], a decreasing amplitude is due to the spurious gauge modes present in the standard theory. This is why a manifestly gauge-invariant perturbation theory is of the utmost importance.

### B. Era after Decoupling

In this section it is assumed that the CDM particle mass is larger than or equal to the proton mass, \( m_{\text{CDM}} \geq m_\text{n} \), implying that for the mean particle mass \( m \) we have \( mc^2 \gg k_B T \).

Once protons and electrons combine to yield hydrogen, the radiation pressure becomes negligible, and the equations of state become those of a non-relativistic monatomic perfect gas

\[
\varepsilon(n, T) = nmc^2 + \frac{3}{2} nk_B T, \quad p(n, T) = nk_B T, \quad (36)
\]

where \( k_B \) is Boltzmann’s constant, \( m \) the mean particle mass, and \( T \) the temperature of the matter. We have \( p_\varepsilon = \frac{2}{3} \) and \( p_n = -\frac{2}{3} mc^2 \), so that the system (21) reduces to

\[
\delta_\varepsilon + 3H \delta_\varepsilon - \left[ \frac{\beta^2 \nabla^2}{a^2} + \frac{2}{3} \kappa \varepsilon_0 \right] \delta_\varepsilon = \left[ \frac{2}{3} \nabla^2 \left( \delta_n - \delta_\varepsilon \right) \right], \quad (37a)
\]

\[
\frac{1}{c} \frac{d}{dt} (\delta_n - \delta_\varepsilon) = -2H (\delta_n - \delta_\varepsilon), \quad (37b)
\]

where \( \beta \equiv \sqrt{n_{(0)}/\varepsilon_{(0)}} \) is, to a good approximation, given by

\[
\beta(t) \approx \frac{v_s(t)}{c} = \sqrt{\frac{5}{3} \frac{k_B T(n,t)}{mc^2}}, \quad T(n) \propto a^{-2}, \quad (38)
\]

with \( v_s \) the adiabatic speed of sound and \( T(n) \) the matter temperature. Equation (37b) implies

\[
\delta_n - \delta_\varepsilon \propto a^{-2}, \quad (39)
\]

where we have used that \( H \equiv \dot{a}/a \). From the expressions (36) it follows that the relative pressure perturbation is given by

\[
\delta_p = \delta_n + \delta_T, \quad (40)
\]
where $\delta_T$ is the relative matter temperature perturbation defined by $\delta_T \equiv T_{(0)}^m/T_{(0)}^e$ and $T_{(0)}^{\text{ref}}$ is defined as $\varepsilon_{\text{ref}}^m$ in (17a) with $\varepsilon$ replaced by $T$. Using that after decoupling $k_B T_{(0)}/(mc^2) \ll 1$, we find from (36) the relation
\begin{equation}
\delta_n - \delta_c \approx - \frac{3 k_B T_{(0)}}{2 mc^2} \delta_T. \tag{41}
\end{equation}
Combining (38) and (39) we find from (41) that
\begin{equation}
\delta_T(t, x) \approx \delta_T(t_0, x), \tag{42}
\end{equation}
to a very good approximation. Using the well-known solutions of the background equations (5)
\[ H \propto t^{-1}, \quad \varepsilon_{(0)} \propto t^{-2}, \quad n_{(0)} \propto t^{-2}, \quad a \propto t^{2/3}, \tag{43}\]
and $\delta(t, x) = \delta(t, q) \exp(iq \cdot x)$, equations (37) can be combined into one equation
\begin{equation}
\delta''_\tau + \frac{2}{\tau} \delta' + \left[ \frac{4}{3} \frac{\mu_m^2}{g} - \frac{10}{9 \tau^2} \right] \delta_c = - \frac{4}{15} \frac{\mu_m^2}{g} \delta_T(t_0, q), \tag{44}
\end{equation}
where $\tau \equiv t/t_0$ and a prime denotes differentiation with respect to $\tau$. The parameter $\mu_m$ is given by
\begin{equation}
\mu_m \equiv \frac{2 \pi}{\lambda_0} \frac{1}{H(t_0)} \frac{v_k(t_0)}{c}, \quad \lambda_0 \equiv \lambda a(t_0). \tag{45}
\end{equation}
The general solution of equation (44) is found to be
\begin{equation}
\delta_c(\tau, q) = \left[ B_1(q) J_{+} \left( \frac{2 \mu_m \tau^{-1/3}}{3} \right) \right] + B_2(q) J_{-} \left( \frac{2 \mu_m \tau^{-1/3}}{3} \right) \tau^{-1/2}
- \frac{3}{5} \left( 1 + \frac{5 \tau^{2/3}}{2 \mu_m^2} \right) \delta_T(t_0, q), \tag{46}
\end{equation}
where $J_{\pm}(\tau x)$ are Bessel functions of the first kind and $B_1(q)$ and $B_2(q)$ are the ‘constants’ of integration.

In the large-scale limit $\lambda \rightarrow \infty$ terms with $\nabla^2$ vanish, so that the general solution of equation (44) is
\begin{align*}
\delta_c(t) &= \frac{1}{\lambda_0} \left[ 5\delta_c(t_0) + \frac{2 \delta_c(t_0)}{H(t_0)} \right] \left( \frac{t}{t_0} \right)^2 \frac{5}{6}
+ \frac{2}{\lambda_0} \left[ \delta_c(t_0) - \frac{\delta_c(t_0)}{H(t_0)} \right] \left( \frac{t}{t_0} \right)^{2/3} \frac{5}{3}. \tag{47}
\end{align*}
Thus, for large-scale perturbations the initial value $\delta_T(t_0, q)$ does not play a role during the evolution: large-scale perturbations evolve only under the influence of gravity: they are so large that neither heat exchange nor pressure perturbations do play a role during their evolution in the linear phase. For perturbations much larger than the Jeans scale, gravity alone is insufficient to account for structure formation within 13.75 Gyr. The solution proportional to $t^{2/3}$ is a standard result. Since $\delta_c$ is gauge-invariant, the standard non-physical gauge mode proportional to $t^{-1}$ is absent from our theory. Instead, a physical mode proportional to $t^{-5/3}$ is found. This mode has also been found by Bardeen [4], Table I, and by Mukhanov et al. [5], expression (5.33). In order to arrive at the $t^{-5/3}$ mode, Bardeen has to use the ‘uniform Hubble constant gauge.’ In our treatment the Hubble function is automatically uniform, without any additional gauge condition, see (17b).

In the small-scale limit $\lambda \rightarrow 0$, we find
\begin{align*}
\delta_c(t, q) &\approx - \frac{3}{5} \delta_T(t_0, q)
+ \left( \frac{t}{t_0} \right)^{-1/3} \left[ \frac{2}{5} \delta_T(t_0, q) + \delta_c(t_0, q) \right]
\times \cos \left[ 2 \mu_m - 2 \mu_m \left( \frac{t}{t_0} \right)^{-2/3} \right]. \tag{48}
\end{align*}
Thus, density perturbations with scales much smaller than the Jeans scale oscillate with a decaying amplitude which is smaller than unity: they are so small that gravity is insufficient to let perturbations grow. Pressure perturbations alone cannot make these perturbations grow. Consequently, small-scale perturbations will never reach the non-linear regime.

The study of the evolution of perturbations with intermediate scales will be conducted in a forthcoming paper by solving equation (44) numerically. It is demonstrated that for density perturbations with scales somewhat larger than the Jeans scale, the action of both gravity and an initially small negative pressure perturbation together may result in massive stars several hundred million years after decoupling of matter and radiation.

**IX. NON-RELATIVISTIC LIMIT**

In Section V we have shown that only two gauge-invariant quantities $\varepsilon_{\text{ref}}^m$ and $n_{\text{ref}}^e$ exist, which could be the real energy density and particle number density perturbations. In this section we will demonstrate that in the non-relativistic limit $v/c \rightarrow 0$ our theory reduces to the well-known results (56) and (57) of the Newtonian theory of gravity and that the quantities $\varepsilon_{\text{ref}}^m$ and $n_{\text{ref}}^e$ become equal to their Newtonian counterparts. This implies that $\varepsilon_{\text{ref}}^m$ and $n_{\text{ref}}^e$ are indeed the local perturbations to the energy density and particle number density perturbations.

We use the result (12) found by Stewart [11], which implies that, in general, cosmological density perturbations are described by two potentials $\phi$ and $\zeta$. In the non-relativistic limit of a flat ($k = 0$) FLRW universe, the potential $\zeta$ disappears from our theory, thus paving the way towards the Newtonian theory of gravity.

As follows from (9) and (10), both potentials occur only in the quantities $\theta_{R_{11}}$ and $\theta_{\phi\psi}$. A first step to get rid of the potential $\zeta$ is to consider a flat ($k = 0$) FLRW universe. For a flat universe, we find that, using (12), the
local perturbation to the spatial curvature (9) reduces to

$$3R_{(1)}^{\parallel} = \frac{4}{c^2} \dot{\phi}^{(1)k} [k] = - \frac{4}{c^2} \nabla^2 \phi,$$

(49)

where $\nabla^2$ is the usual Laplace operator. Using this expression, the perturbation equations (14) read, for a flat FLRW universe,

$$H(\theta_{(1)} - \vartheta_{(1)}) + \frac{1}{a^2} \nabla^2 \phi = \frac{4\pi G}{c^2} \left[ \varepsilon^{(1)}_i + \frac{\dot{\varepsilon}^{(1)}}{\theta_{(0)}} i \right],$$

(50a)

$$\nabla^2 \dot{\phi} + \frac{4\pi G}{c^2} \varepsilon^{(0)} (1 + w) \dot{\theta}_{(1)} = 0,$$

(50b)

$$\dot{\varepsilon}^{(1)} + 3H(\varepsilon^{(1)} + p^{(1)}_i) + \varepsilon^{(0)} (1 + w) \theta_{(1)} = 0,$$

(50c)

$$\dot{\theta}_{(1)} + H(2 - 3\beta^2) \theta_{(1)} + \frac{1}{a^2} \nabla^2 p^{(1)} = 0,$$

(50d)

$$\dot{n}_{(1)} + 3H n_{(1)} + n_{(0)} \theta_{(1)} = 0,$$

(50e)

where we have used (17a) to eliminate $\varepsilon_{(1)}$ from the constraint equation (14a). In these equations, the potential $\zeta$ occurs only in the quantity $\theta_{(1)}$, see (10) and (12). This potential disappears from our equations in the non-relativistic limit, as we will now show. The spatial part $u_{(1)}^{i\parallel}$ of the fluid four-velocity is gauge dependent with a physical component and a non-physical gauge part. We define the non-relativistic limit $v/c \to 0$ by

$$u_{(1)}^{i\parallel}_{\text{physical}} \equiv \frac{U^{i\parallel}_{(1)}_{\text{physical}}}{c} \to 0,$$

(51)

i.e., the physical part of the spatial fluid velocity is negligible with respect to the speed of light. In this limit, the kinetic energy per particle $\frac{1}{2}m(v^2) = \frac{1}{2}mc^2 \to 0$ is very small compared to the rest energy $mc^2$ per particle, implying that the pressure $p = nk_B T \to 0 \ (n \neq 0)$ is also vanishingly small. Substituting $p \to 0$ into the momentum conservation law (8c) yields, using also the background equation (5c) with $\equiv p^{(0)} / \varepsilon^{(0)} = 0$,

$$\dot{u}_{(1)}^{i\parallel} = -2Hu_{(1)}^{i\parallel}.$$

(52)

Since the physical part of $u_{(1)}^{i\parallel}$ vanishes in the non-relativistic limit, the general solution of equation (52) is exactly equal to the gauge mode

$$u_{(1)}^{i\parallel}_{\text{gauge}} = -\frac{1}{a^2(i)} \tilde{g}^{ik} \partial_k \psi(x),$$

(53)

where we have used that $H \equiv \dot{a}/a$. Thus, in the non-relativistic limit (51) we are left with the gauge mode (53) only. Consequently, we may, without losing any physical information, put $u_{(1)}^{i\parallel}_{\text{gauge}}$ equal to zero, implying that $\partial_k \psi = 0$, so that $\psi$ is a constant in the non-relativistic limit. Therefore, the relativistic transformation (1) with (3) between synchronous coordinates reduces in the non-relativistic limit to the most general transformation

$$x^0 \to x^0 - \psi, \ x^i \to x^i - \chi^i(x),$$

(54)

which is possible in the Newtonian theory of gravity. In (54), $\psi$ is an arbitrary constant and $\chi^i(x)$ are three arbitrary functions of the spatial coordinates.

Substituting $\vartheta_{(1)} = (u_{(1)}^k) [k] = 0$ and $p = 0$ into the system (50), we get in the non-relativistic limit

$$\text{Constraint:} \quad \nabla^2 \phi = \frac{4\pi G}{c^2} a^2 \varepsilon^{(1)}_i,$$

(55a)

$$\text{Evolution:} \quad \nabla^2 \dot{\phi} = 0,$$

(55b)

$$\text{Conservation:} \quad \dot{\varepsilon}^{(1)} + 3H \varepsilon^{(1)} + \varepsilon^{(0)} \theta_{(1)} = 0,$$

(55c)

$$\dot{n}_{(1)} + 3H n_{(1)} + n_{(0)} \theta_{(1)} = 0.$$

(55d)

The constraint equation (55a) can be found by subtracting $\frac{1}{3} \theta_{(1)} / \dot{H}$ times the time-derivative of the background constraint equation (5a) with $3R_{(0)} = 0$ from the constraint equation (5a) and using that $\theta_{(0)} = 3H$.

Defining the potential $\varphi(x) \equiv \phi(x) / a^2 (t_0, x)$, equations (55a) and (55b) can be combined to yield

$$\nabla^2 \varphi(x) = 4\pi G \rho_{(1)}(x), \quad \rho_{(1)}(x) \equiv \frac{\varepsilon^{(1)}_i(t_0, x)}{c^2},$$

(56)

which is the well-known Poisson equation of the Newtonian theory of gravity.

Equations (55c) and (55d) have no physical significance since $\varepsilon_{(1)}$, $n_{(1)}$ and $\theta_{(1)}$ are gauge dependent also in the non-relativistic limit, as follows from (5), (16a) and (54). Moreover, the physical equations (55a) and (55b) are decoupled from the non-physical equations (55c) and (55d). Therefore, the latter two equations are not part of the Newtonian theory of gravity and need not be considered. Thus, in the non-relativistic limit, the potential $\zeta$ occurring in $\theta_{(1)}$ drops from our perturbation theory and we are left with one potential $\varphi(x) \equiv \phi(x) / a^2 (t_0)$ only.

The expression (19a) reduces in the non-relativistic limit to $\varepsilon_n^{(1)} = -3R_{(1)}^{\parallel}/(2c)$, which is, with (49), equivalent to the Poisson equation (56). Expression (19b) reduces in the non-relativistic limit to the familiar result

$$\varepsilon_n^{(1)} = \frac{\varepsilon_i^{(1)}}{mc^2},$$

(57)

where we have used that $\varepsilon_{(1)} = n_{(1)} mc^2$ and $\varepsilon_{(0)} = n_{(0)} mc^2$.

Finally, (55a) and (55b) imply that $\varepsilon_n^{(1)} \propto a^{-2}$. From (5c) with $w = 0$ we have $\varepsilon_{(0)} \propto a^{-3}$, so that with (20) and (57) we arrive at the well-known result $\delta_n \propto \delta_n \propto a$.

We have shown that our theory based on the gauge-invariant quantities (17) reduces in the non-relativistic limit $v/c \to 0$ to the well-known Newtonian results (56) and (57). Consequently, $\varepsilon_n^{(1)}$ and $n_n^{(1)}$ are the real, physical, perturbations to the energy density and particle number density, respectively.

X. STANDARD THEORY

In order to compare the standard theory with the theory developed in this article, we consider a flat $k = 0$, see
(6)] FLRW universe after decoupling of matter and radiation. Since $k_2 T_{(0)} \ll mc^2$ the standard theory does not distinguish between perturbations in the particle number density and the total energy density. Therefore, we take in our perturbation theory $\delta_n = \delta_\epsilon$, implying that $\delta_T = 0$ and $\delta_\rho = \delta_n$, see (40) and (41). Equation (37a) reads

$$\ddot{\epsilon} + 3H\dot{\epsilon} - \left[ \frac{\beta^2 \nabla^2}{a^2} + \frac{2}{3}\kappa\epsilon\dot{(\epsilon)} \right] \epsilon = 0,$$

(58)

whereas (37b) is identically satisfied. This equation should be compared with the standard equation

$$\ddot{\delta} + 2H\dot{\delta} - \left[ \frac{\nu_s^2 \nabla^2}{c^2} + \frac{2}{3}\kappa\epsilon\dot{(\epsilon)} \right] \delta = 0.$$

(59)

Apart from the immediately obvious differences, namely the coefficients of $H$ and $\kappa\epsilon\dot{(\epsilon)}$ and the very small but unimportant difference between $\beta$ and $\nu_s/c$, there are two major differences between (58) and (59). Firstly, (58) is derived from the full set of perturbed Einstein equations and conservation laws, without any assumptions or approximations. In contrast to (58), the standard equation (59) can not be derived from the General Theory of Relativity; instead, it is derived from the Newtonian theory of gravity, using dubious mathematics [17]. Secondly, the general solution of our equation (58) is gauge-invariant, whereas the general solution of the standard equation (59) depends on the gauge. This can be seen as follows. For large-scale perturbations, $\nabla^2 \delta \to 0$, (59) reduces to

$$\ddot{\delta} + 2H\dot{\delta} - \frac{2}{3}\kappa\epsilon\dot{(\epsilon)} \delta = 0.$$

(60)

In contrast to (59), equation (60) can be derived from Einstein’s gravitation theory: substituting $\delta = \epsilon_{(1)}/\epsilon\dot{(\epsilon)}$, $\dot{\varepsilon}_{(1)} = 0$, $n_{(0)} = n_{(1)} = 0$, $p_{(0)} = p_{(1)} = 0$, $w = 0$ and $\delta R_{(0)} = 0$ into the set (18) and using the background equations (5) yields (60). As a consequence, (60) is a relativistic equation. Using (43), we find for its general solution

$$\delta(\tau) \propto C\tau^{2/3} - 3H(t_0)|\psi\tau^{-1},$$

(61)

where $C$ is an arbitrary constant, and $\tau \equiv t/t_0$. Since the solutions of the set (18) are gauge dependent, the solution (61) of equation (60) is also gauge dependent and the mode proportional to $\tau^{-1}$ is the gauge mode as follows from (16a) and $\epsilon_{(0)} \propto t^{-2}$. Note that $\varepsilon_{(1)} = 0$ implies with (16b) that the gauge function $\psi$ is constant. Since the solution (61) can be obtained from the general solution of (59) by taking the large-scale limit, it follows that also the general solution of (59) contains the gauge function $\psi$. Consequently, one cannot impose initial conditions, since the gauge function $\psi$ cannot be fixed. We have to conclude that also the standard equation (59) has no physical significance and should, therefore, not be used anymore.

ACKNOWLEDGMENTS

The authors should like to take the opportunity of thanking Marie-Louise Pennings MA (lecturer of English at the NLDA) for her assistance in discovering the intricacies of the English language.
restart;

# Consistency check of the perturbation equations (21).
# To check Eqs.(21) with coefficients (22), substitute (20) into Eqs.(21)
# and use the background equations (5) and the perturbation equations (18)
# to evaluate all time derivatives.

# Coefficients b_1, b_2, b_3, (22):

b1 := \frac{\kappa e_0(t)(1+w(t))}{H(t)} - 2\frac{d}{dt}\frac{\beta(t)}{\beta(t)} - H(t)(2 + 6w(t) + 3\beta(t)^2) + R_0(t)\left(\frac{1}{3H(t)} + \frac{2H(t)(1 + 3\beta(t)^2)}{R_0(t)} + 5\kappa e_0(t)(1 + w(t))\right)

b2 := -(1/2)\kappa e_0(t)\left(1 + \frac{3\beta(t)}{R_0(t)}\right) - H(t)^2\left(1 - \frac{3w(t)}{R_0(t)} + 6\beta(t)^2\right) - \frac{R_0(t)}{2} - \frac{H(t)^2}{R_0(t)}\left(1 + 3\frac{\beta(t)}{R_0(t)}\right)

b3 := -\frac{18H(t)^2}{R_0(t)}\left(1 + \frac{w(t)}{R_0(t)}\right) - 3\frac{\kappa e_0(t)(1 + w(t))}{R_0(t)} - \frac{H(t)^2}{R_0(t)}\left(1 + 3\frac{\beta(t)}{R_0(t)}\right)

# Background equations

w(t) := p_0(t)/e_0(t);

w(t) := \frac{p_0(t)}{e_0(t)}

H(t) := \frac{\dot{a}(t)}{a(t)}

H(t) := \frac{\dot{a}(t)}{a(t)}

# The time derivative of Eq.(5a) combined with Eqs.(5b) and (5c) yields the time derivative of the Hubble function H:

\dot{H}(t) := -\frac{1}{6}R_0(t) - \frac{1}{2}\kappaappa(e_0(t) + p_0(t))
Either define the three-space curvature by Eq. (5b):
\[
\frac{1}{6} \left( R_{0}(t) - \frac{1}{2} \kappa \left( e_{0}(t) + p_{0}(t) \right) \right) = -2 H(t) R_{0}(t)
\]

or by Eq. (6) with \( k = \pm 1, \ 0, \ -1 \):
\[
\begin{align*}
\text{if } k = +1: & \quad R_{0}(t) := -6 k / a(t)^2; \\
\text{if } k = -1: & \quad R_{0}(t) := -6 k / a(t)^2; \\
\text{if } k = 0: & \quad R_{0}(t) := -6 k / a(t)^2; \\
\end{align*}
\]

Energy conservation law (5c):
\[
\text{define}(e_{0}, \text{diff}(e_{0}(t), t) = -3 H(t) * e_{0}(t) * (1 + w(t))); \\
\text{diff}(e_{0}(t), t);
\]

Particle number conservation law (5e):
\[
\text{define}(n_{0}, \text{diff}(n_{0}(t), t) = -3 H(t) * n_{0}(t)); \\
\text{diff}(n_{0}(t), t);
\]

Time derivative of \( p_{0}(t) \):
\[
\text{define}(p_{0}, \text{diff}(p_{0}(t), t) = p_{e}(t) * \text{diff}(e_{0}(t), t) + p_{n}(t) * \text{diff}(n_{0}(t), t)); \\
\text{diff}(p_{0}(t), t);
\]

Partial derivative \( p_{e}(t) \) of the pressure \( p \):
\[
\begin{align*}
\text{define}(p_{e}, \text{diff}(p_{e}(t), t) = p_{ee}(t) * \text{diff}(e_{0}(t), t) + p_{en}(t) * \text{diff}(n_{0}(t), t)); \\
\text{diff}(p_{e}(t), t);
\end{align*}
\]

Partial derivative \( p_{n}(t) \) of the pressure \( p \):
\[
\begin{align*}
\text{define}(p_{n}, \text{diff}(p_{n}(t), t) = p_{en}(t) * \text{diff}(e_{0}(t), t) + p_{nn}(t) * \text{diff}(n_{0}(t), t)); \\
\text{diff}(p_{n}(t), t);
\end{align*}
\]

Quantity \( \beta(t) \):
\[
\beta(t) := \sqrt{\frac{-9 p_{e}(t) H(t) e_{0}(t) \left( 1 + \frac{p_{0}(t)}{e_{0}(t)} \right) - 9 p_{n}(t) H(t) n_{0}(t)}{H(t) e_{0}(t) \left( 1 + \frac{p_{0}(t)}{e_{0}(t)} \right)}}
\]

First order perturbation equations:
\[
\begin{align*}
\text{if } k = +1: & \quad R_{0}(t) := -6 k / a(t)^2; \\
\text{if } k = -1: & \quad R_{0}(t) := -6 k / a(t)^2; \\
\text{if } k = 0: & \quad R_{0}(t) := -6 k / a(t)^2; \\
\end{align*}
\]

Energy conservation law (18b):
\[
\begin{align*}
\text{define}(e_{1}, \text{diff}(e_{1}(t), t) = -3 H(t) * (e_{1}(t) + p_{1}(t)) - e_{0}(t) * (1 + w(t)) * (\theta(t) + (\kappa e_{1}(t) + (1/2) * R_{1}(t)) / (2 * H(t)))); \\
\text{diff}(e_{1}(t), t);
\end{align*}
\]

Particle number conservation law (18d):
\[
\begin{align*}
\text{define}(n_{1}, \text{diff}(n_{1}(t), t) = -3 H(t) * n_{1}(t) - n_{0}(t) * (\theta(t) + (\kappa e_{1}(t) + (1/2) * R_{1}(t)) / (2 * H(t)))); \\
\text{diff}(n_{1}(t), t);
\end{align*}
\]
\begin{align}
-3 \frac{d}{dt} n_{\text{i}(t)} - n_{\text{o}(t)} \left( \frac{\kappa e_{\text{f}(t)}}{H(t)} + \frac{1}{2} \frac{R_l(t)}{H(t)} \right) \tag{16}
\end{align}

> # Momentum conservation law (18c):
> define(theta,diff(theta(t),t)=-H(t)*(2-3*beta(t)^2)*theta(t)-Delta/a(t)^2*p_1(t)/(e_0(t)*(1+w(t))))
> diff(theta(t),t);

\begin{align}
-9 p_1(t) H(t) e_{\text{f}(t)} & \left( 1 + \frac{p_1(t)}{e_{\text{f}(t)}} \right) - 9 p_n(t) H(t) n_{\text{o}(t)} H(t) e_{\text{f}(t)} \left( 1 + \frac{p_n(t)}{e_{\text{f}(t)}} \right) \right) 0(t) = \frac{\Delta \left( p_1(t) e_{\text{f}(t)} + p_n(t) n_{\text{o}(t)} \right)}{a(t)^2 e_{\text{f}(t)} \left( 1 + \frac{p_n(t)}{e_{\text{f}(t)}} \right)} \tag{17}
\end{align}

> # Evolution equation for the local perturbation to the
> # global spatial curvature, (18a):
> define(R_1,diff(R_1(t),t)=-2*(H(t)*R_1(t)-kappa*e_0(t)*(1+w(t))*theta(t))-R_0(t)/(3*H(t))*(kappa* e_1(t)+(1/2)*R_1(t)));
> diff(R_1(t),t);

\begin{align}
-2 H(t) R_l(t) + 2 \frac{\kappa e_{\text{f}(t)}}{e_{\text{f}(t)}} \left( \frac{p_1(t)}{e_{\text{f}(t)}} \right) \theta_{\text{f}(t)} - \frac{1}{3} \frac{R_{\text{o}(t)}}{H(t)} \left( \frac{\kappa e_{\text{f}(t)}}{H(t)} + \frac{1}{2} \frac{R_l(t)}{H(t)} \right) \tag{18}
\end{align}

> # Gauge-invariant density perturbations
> #

\begin{align}
e_{\text{gi}(t)} := \frac{e_{\text{f}(t)} R_0(t) - 3 e_{\text{o}(t)} (1+w(t))}{R_0(t) + 3 \kappa e_{\text{f}(t)} \left( 1 + \frac{p_1(t)}{e_{\text{f}(t)}} \right)} \tag{19}
\end{align}

> # Gauge-invariant perturbation to the energy density, (19a):
> e_{\text{gi}(t)} := \frac{(e_{\text{f}(t)} R_0(t) - 3 e_{\text{o}(t)} (1+w(t)) (2H(t)*theta(t)+(1/2)*R_1(t))) / (R_0(t)+3*kappa*e_0(t)*(1+w(t)))};

\begin{align}
n_{\text{gi}(t)} := n_{\text{i}(t)} - 3 n_{\text{o}(t)} \left( \frac{\kappa e_{\text{f}(t)}}{H(t)} + \frac{1}{2} \frac{R_l(t)}{H(t)} \right) \tag{20}
\end{align}

> # Gauge-invariant contrast functions, (20):
> delta_e(t) := e_{\text{gi}(t)} / e_{\text{o}(t)}; delta_n(t) := n_{\text{gi}(t)} / n_{\text{o}(t)};

\begin{align}
delta_e(t) := \frac{e_{\text{f}(t)} R_0(t) - 3 e_{\text{o}(t)} \left( 1 + \frac{p_1(t)}{e_{\text{f}(t)}} \right) \left( 2 H(t) \theta(t) + \frac{1}{2} R_l(t) \right) \left( R_0(t) + 3 \kappa e_{\text{f}(t)} \left( 1 + \frac{p_1(t)}{e_{\text{f}(t)}} \right) \right)}{e_{\text{f}(t)}} \tag{21}
\end{align}

> # Consistency check for Eq. (21b) [left-hand side minus right-hand side]:
> simplify(diff(delta_n(t)-delta_e(t)/(1+w(t))),t-3*H(t)*n_{\text{o}(t)}*p_{\text{n}(t)}/(e_{\text{o}(t)}*(1+w(t)))))*(delta_n(t) - delta_e(t)/(1+w(t)))));

0 \tag{22}

> # Consistency check for Eq. (21a) [left-hand side minus right-hand side]:
> simplify(diff(delta_e(t),t$2)+b1*diff(delta_e(t),t)+b2*delta_e(t)-b3*(delta_n(t)-delta_e(t)/(1+w(t)))));

0 \tag{23}
