PHYSICAL REVIEW D VOLUME 22, NUMBER 8 15 OCTOBER 1980

Gauge-invariant cosmological perturbations

James M. Bardeen*
Institute for Advanced Study, Princeton, New Jersey 08540
(Received 7 April 1980)

The physical interpretation of perturbations of homogeneous, isotropic cosmological models in the early Universe, when the perturbation is larger than the particle horizon, is clarified by defining a complete set of gauge-invariant variables. The linearized perturbation equations written in these variables are simpler than the usual versions, and easily accommodate an arbitrary background equation of state, entropy perturbations, and anisotropic pressure perturbations. Particular attention is paid to how a scalar (density) perturbation might be generated by stress perturbations at very early times, when the non-gauge-invariant perturbation in the density itself is ill-defined. The amplitude of the fractional energy density perturbation at the particle horizon cannot be larger, in order of magnitude, than the maximum ratio of the stress perturbation to the background energy density at any earlier time, unless the perturbation is inherent in the initial singularity.

I. INTRODUCTION

The mathematical theory of perturbations in inhomogeneous, isotropic cosmological models has been worked over many times in the literature. References 1–7 are a selection of the more important comprehensive treatments. Nevertheless, troubling questions still remain about the physical interpretation of density perturbations at early times when the perturbation is larger than the particle horizon, which will here mean when the time for light to travel a characteristic wavelength of the perturbation is larger than the instantaneous expansion (Hubble) time. These questions are particularly relevant to attempts to explain the origin of perturbations which eventually give rise to galaxies through processes occurring at times when temperatures exceed a few hundred MeV and/or densities exceed nuclear densities, times when there is considerable latitude to speculate about the microscopic physics.[8,9]

The problem has to do with the freedom of making gauge transformations. In discussing perturbations one is dealing with two spacetimes—the physical, perturbed spacetime and a fictitious background spacetime, here described by a Robertson-Walker metric. Points in the background are labeled by coordinates x^k (Latin indices will range from 0 to 3, Greek indices from 1 to 3). A one-to-one correspondence between points in the background and points in the physical spacetime carries these coordinates over into the physical spacetime and defines a choice of gauge. A change in the *correspondence*, keeping the background coordinates fixed, is called a *gauge transformation*, to be distinguished from a coordinate transformation which changes the labeling of points in the background and physical spacetime together.

The perturbation in some quantity is the difference between the value it has at a point in the physical spacetime and the value at the *corresponding point* in the background spacetime. A gauge transformation induces a coordinate transformation in the physical spacetime, but it also changes the point in the background spacetime corresponding to a given point in the physical spacetime. Thus, even if a quantity is a scalar under coordinate transformations, the value of the *perturbation* in the quantity will *not* be invariant under gauge transformations if the quantity is nonzero and position dependent in the background.

The prime example is the perturbation in density (energy density or baryon density). Because the density is time dependent in the cosmological background, the value of the density perturbation is altered by any gauge transformation which changes the correspondence between hypersurfaces of simultaneity in the physical spacetime and the background spacetime. When the perturbation is well within the particle horizon, hypersurfaces of simultaneity are physically unambiguous (clocks can be synchronized by exchanging light signals) and the change in density perturbation between "reasonable" gauge choices is negligible. However, at early times there is no compelling physical reason to choose between gauges which give very different results for the time dependence of density perturbations, and several different values for the exponent in the power-law time dependence of what is physically the same mode of density perturbation can be found in the literature, each mathematically correct. Furthermore, if the gauge condition imposed to simplify the form of the metric leaves a residual gauge freedom, the perturbation equations will have spurious "gauge mode" solutions which can be completely annulled by a gauge transformation and have no physical reality.[2,10-12]

One interesting attempt to avoid these problems was made by Hawking,[3] who formulated the perturbation equations in a completely covariant form, without any mention of metric tensor perturbations as such. This did not totally circumvent the problem of gauge ambiguity, since a choice of time slicing in the perturbed spacetime must still be made to define a density perturbation. The paper is flawed by a failure to recognize that hypersurfaces of constant proper time along the fluid-element world lines cannot be orthogonal to these world lines when pressure perturbations as well as density perturbations are present.

The Hawking line of analysis was completed successfully by Olson[13] for the special case of an isentropic perfect fluid and a background with zero spatial curvature. An equation was derived for a *gauge-invariant* variable, proportional to the intrinsic spatial curvature of hypersurfaces orthogonal to the fluid four-velocity everywhere. The density perturbation as such was defined relative to hypersurfaces of constant fluid-element proper time, and thus was subject to an ambiguity associated with the freedom of adjusting the origin of proper time differently for different fluid elements. The Hawking-Olson approach can perhaps be extended to allow for nonzero background curvature, entropy perturbations, and anisotropic pressure, but only with great difficulty.

This paper presents a complete gauge-invariant framework for studying the time development of physically general perturbations at early times, when the wavelength of the perturbation is larger than the particle horizon. The geometrical quantities are defined from the metric perturbations alone, without reference to the matter perturbations. A gauge-invariant formalism in some respects mathematically equivalent to this one has been developed by Gerlach and Sengupta,[14] but they did not consider any cosmological applications.

There are two independent gauge-invariant gravitational "potentials" for scalar (density) perturbations and one for vector (vorticity) perturbations. The appropriate combinations of Einstein equations give all these potentials directly from the matter perturbations through purely algebraic equations once the perturbations have been separated into spatial harmonics.

The gauge-invariant variables which give the mathematically simplest description of the matter dynamics are, for *scalar* perturbations, (1) the velocity amplitude which, when divided by the reduced wavelength of the perturbation, gives the time dependence of the rate of shear of the matter velocity field; (2) the perturbation in the "entropy" of the matter, specifically the excess of the actual fractional pressure perturbation over the adiabatic one; (3) the amplitude of the anisotropic stress associated with the perturbation, if any, and (4) the fractional energy density perturbation on the hypersurfaces orthogonal to the world lines of the matter. When written completely in terms of these gauge-invariant amplitudes, the equations of motion for the matter simplify to a form rather closely analogous to the corresponding *Newtonian* equations for perturbations in an expanding background, even when the wavelength of the perturbation is much larger than the particle horizon. The shear velocity and the comoving energy density perturbation formally respond to input from the entropy perturbation and the anisotropic stress, in the spirit of recent work by Press and Vishniac.[12] Analytic solution of these equations is straightforward when the background spatial curvature is negligible and the ratio of pressure to energy density in the background is independent of time.

The gauge invariance in itself does not resolve the physical ambiguity of what one means by an energy density perturbation or a spatial curvature perturbation when the perturbation wavelength is larger than the particle horizon. The physical perturbations relative to any well-defined set of hypersurfaces of simultaneity can be represented by appropriate combinations of the above gauge-invariant amplitudes. I will consider three such families of hypersurfaces. The *comoving hypersurfaces*, orthogonal to the world lines of the matter, have already been mentioned. The mathematically simplest gauge-invariant gravitational potentials represent directly the perturbations in the spatial curvature and the lapse function relative to *zero-shear hypersurfaces*, hypersurfaces for which the congruence of normal timelike world lines has zero shear; i.e., the traceless part of the extrinsic curvature tensor vanishes. A spatially uniform rate of expansion of the normal world lines characterizes *uniform-Hubble-constant hypersurfaces*. Only the last hypersurface condition carries over without modification to general perturbations, involving divergenceless vector fields and transverse traceless tensors. As long as the perturbations are linear the scalar perturbations can be treated separately.

Requiring the hypersurfaces of simultaneity to satisfy specific physical or geometric criteria is contrary to the usual approach to cosmological perturbations through a synchronous gauge.[2,7,15] In a synchronous gauge the physical properties of only one of the hypersurfaces, at one particular time, can be specified in advance; all the other hypersurfaces depend on the global solution to the perturbation equations. As a result the density perturbation in a synchronous gauge cannot be characterized by gauge-invariant amplitudes, and its value is in principle unrelated to the current physical state of the perturbations.

An important question is whether the perturbations that eventually give rise to galaxies need to be present in the singularity at $t=0$ or can somehow be produced at a later time in an initially homogeneous and isotropic universe. New physical mechanisms for generating inhomogeneities through gauge-theory symmetry breaking have been proposed in recent years.[8,9,16,17] The difficulty is that, with a conventional background equation of state, the comoving volume which produces a galaxy is far larger than the particle horizon at the times the symmetry breaking (or any other exotic process) is likely to occur. No causal process can then produce coherence on a galactic scale. There have been suggestions[18,19] that initially small *statistical* fluctuations on a galactic scale associated with large amplitude perturbations within the particle horizon at some early time would grow to a significant amplitude by the time the galactic scale comes inside the particle horizon. The conventional mythology is that the fractional energy density perturbation increases as the proper time t in a radiation-dominated background as long as the perturbation wavelength is larger than the particle horizon.

Local conservation of energy and momentum requires that any energy density perturbation arise as the result of a stress perturbation. Press and Vishniac[12] point out that for an isotropic stress (entropy) perturbation the fractional energy density perturbation *at the particle horizon* can be no larger than the maximum value of the ratio of stress perturbation to background energy density at earlier times. In this paper I generalize the result to allow for anisotropic stress perturbations and nonlinear excitation of density perturbations. The conclusion still holds that a given amplitude fractional energy density perturbation at the particle horizon can *only* arise from a stress perturbation with a comparable amplitude relative to the background energy density unless the perturbation is inherent in the initial singularity or unless the background equation state abolishes particle horizons at early times. While an anisotropic stress perturbation can produce an energy density perturbation of comparable amplitude (on a comoving hypersurface) at some early time, the fractional energy density perturbation belongs predominantly to the "decaying mode" and is no larger at the particle horizon.

The plan of the paper is as follows. The notation and description of the background and perturbed metric tensor and energy-momentum tensor in conventional terms, without any gauge restrictions, is established in Sec. II. In Sec. III I show how these quantities are affected by arbitrary gauge transformations and identify the gauge-invariant amplitudes describing the perturbation. The equations for the gauge-invariant amplitudes are derived in Sec. IV without any restrictions on the physical nature of the matter other than local conservation of energy and momentum, which is required in any case as an integrability condition on the Einstein equations. In Sec. V general analytic solutions in a background with constant ratio of pressure to energy density and effectively zero spatial curvature are used to discuss the generation of energy density perturbations from stress perturbations from the point of view of each of the hypersurface conditions mentioned above. Contact with the nonperturbative evolution of the geometry and the matter is made in Sec. VI, in order to understand the limits of validity of the linear perturbation analysis and nonlinear effects on the energy density perturbation. The results are summarized in Sec. VII.

The focus in this paper is on the physics. Mathematical questions regarding the existence of linear perturbations and the expansion of the perturbations in spatial harmonics have been largely answered by D'Eath.[20]

II. STANDARD FORMALISM

The background spacetime is described by some version of the Robertson-Walker metric

$$ds^2 = S^2(\tau)(-d\tau^2 + {}^3g_{\alpha\beta}dx^\alpha dx^\beta)\,. \tag{2.1}$$

The tensor ${}^3g_{\alpha\beta}$ is the metric tensor for a three-space of uniform spatial curvature K, with Riemann tensor

$${}^3R_{\alpha\beta\gamma\delta} = K({}^3g_{\alpha\gamma}{}^3g_{\beta\delta} - {}^3g_{\alpha\delta}{}^3g_{\beta\gamma})\,, \tag{2.2}$$

where K is independent of time. The choice of coordinates in the background three-space is left arbitrary. Let a slash denote the covariant derivative of a three-tensor with respect to ${}^3g_{\alpha\beta}$ and a semicolon the covariant derivative in the physical spacetime. The scale factor $S(\tau)$ describes the expansion of the background as a function of the conformal time τ.

The unperturbed energy-momentum tensor must be formally that of a perfect fluid at rest relative to the above coordinates. The only nonzero components are

$$T^0_0 = -E_0\,, \quad T^\alpha_\beta = P_0\delta^\alpha_\beta\,, \tag{2.3}$$

where $E_0(\tau)$ is the background energy density and $P_0(\tau)$ is the background pressure. Let

$$w=P_0/E_0,\quad c_s{}^2=dP_0/dE_0\,. \tag{2.4}$$

A nonzero cosmological constant can be considered part of this background energy-momentum tensor, contributing $+\Lambda$ to E_0 and $-\Lambda$ to P_0.

The time evolution of the background is governed by the equations[15]

$$(\dot S/S)^{\cdot}=-\tfrac{1}{6}(E_0+3P_0)S^2\,, \tag{2.5a}$$

$$(\dot S/S)^2=\tfrac{1}{3}E_0S^2-K\,, \tag{2.5b}$$

$$\dot E_0/(E_0+P_0)=-3\dot S/S\,, \tag{2.6}$$

where $\dot S\equiv dS/d\tau$, the derivative with respect to the *conformal* time, and units have been chosen so $c=8\pi G=1$.

Perturbations in various quantities can be classified, according to how they transform under *spatial* coordinate transformations in the *background* spacetime, as spatial scalars, vectors, and tensors. Furthermore, the homogeneity and isotropy of the background allows a separation of the time dependence and the spatial dependence, with the spatial dependence related to solutions of a generalized Helmholtz equation.[2] The representation in spatial harmonics is not always unique.[20]

Scalar harmonics are solutions of the scalar Helmholtz equation

$$Q^{(0)|\alpha}{}_{|\alpha}+k^2Q^{(0)}=0\,. \tag{2.7}$$

The wave number k sets the spatial scale of the perturbation relative to the comoving background coordinates. For zero background curvature the $Q^{(0)}$ can be taken as plane waves; solutions for nonzero spatial curvature are described by Harrison.[4] *Scalar perturbations* have a spatial dependence derived from one of the $Q^{(0)}$. A vector or tensor quantity, such as the three-velocity of the matter or the perturbation in the spatial metric tensor, which is associated with a scalar perturbation must be constructed from covariant derivatives of $Q^{(0)}$ and the metric tensor. The construction is unique, within a normalization, for any traceless, symmetric tensor. Define the vector

$$Q_\alpha^{(0)}=-(1/k)Q^{(0)}{}_{|\alpha}\,, \tag{2.8}$$

and the traceless, symmetric, second-rank tensor

$$Q_{\alpha\beta}^{(0)}=k^{-2}Q^{(0)}{}_{|\alpha\beta}+\tfrac{1}{3}\,{}^3g_{\alpha\beta}Q^{(0)}\,. \tag{2.9}$$

Higher-rank tensors can be useful in representing moments of the specific intensity of, say, the microwave background radiation field, but are not needed in this paper. All equations governing scalar perturbations are reducible to scalar equations by taking divergences, e.g.,

$$Q^{(0)\alpha}{}_{|\alpha}=kQ^{(0)},\quad Q^{(0)\alpha\beta}{}_{|\alpha\beta}=\tfrac{2}{3}(k^2-3K)Q^{(0)}\,. \tag{2.10}$$

The divergenceless part of a vector field cannot be related to scalar harmonics, but instead must be proportional to a *vector harmonic* $Q^{(1)\alpha}$, a divergenceless vector field which is a solution of the vector Helmholtz equation

$$Q^{(1)\alpha|\beta}{}_{|\beta}+k^2Q^{(1)\alpha}=0\,. \tag{2.11}$$

The corresponding second-rank symmetric tensor, necessarily traceless but not divergenceless, is

$$Q^{(1)\alpha\beta}=-\tfrac{1}{2}k^{-1}(Q^{(1)\alpha|\beta}+Q^{(1)\beta|\alpha})\,. \tag{2.12}$$

A second-rank antisymmetric tensor can also be constructed from $Q_\alpha^{(1)}$, in contrast to $Q_\alpha^{(0)}$, since $Q_\alpha^{(1)}$ is not the gradient of a scalar.

Gravitational waves are described by a traceless, divergenceless tensor $Q_{\alpha\beta}^{(2)}$ which is a solution of

$$Q^{(2)\alpha\beta|\gamma}{}_{|\gamma}+k^2Q^{(2)\alpha\beta}=0\,, \tag{2.13}$$

and in linear perturbation theory are completely decoupled from the scalar and vector perturbations.

A completely general perturbation of the gravitational field can be written as a linear combination of perturbations associated with individual spatial harmonics as defined above, with no coupling between different harmonics. The gravitational wave tensor perturbations only couple to the anisotropic part of the stress tensor, the vector perturbations couple, in addition, to the divergenceless, vortical part of the velocity field of the matter, and scalar perturbations couple to perturbations in the density and isotropic pressure as well as perturbations in the irrotational part of the velocity field and the anisotropic stress.

From now on we assume the separation into individual harmonics has been made. Then a given quantity can be written as a linear combination of all the independent appropriate rank spatial tensors constructable from the fundamental harmonic, with the coefficients functions of time. The Einstein equations and matter evolution equations become ordinary differential equations in time for these coefficients.

Specific representations of the perturbations in the metric tensor and the energy-momentum tensor will now be defined separately for scalar, vector, and tensor perturbations.

A. Scalar perturbations

The conformal factor S^2 is removed from the metric tensor components before defining the perturbations. Let

$$g_{00}=-S^2(\tau)[1+2A(\tau)Q^{(0)}(x^\mu)]\,, \tag{2.14a}$$

$$g_{0\alpha}=-S^2B^{(0)}(\tau)Q_\alpha^{(0)}(x^\mu)\,, \tag{2.14b}$$

$$g_{\alpha\beta}=S^2\{[1+2H_L(\tau)Q^{(0)}(x^\mu)]^3g_{\alpha\beta}(x^\mu)+2H_T^{(0)}(\tau)Q_{\alpha\beta}^{(0)}(x^\mu)\}\,. \tag{2.14c}$$

The representation of the energy-momentum tensor will also be completely general in the context of first-order perturbations. *Define* the rest frame for the matter to be the frame in which the energy flux vanishes. Let u^a be the four-velocity of this frame relative to the coordinate frame. The three-velocity associated with u^a is represented by

$$u^\alpha/u^0=v^{(0)}(\tau)Q^{(0)\alpha}(x^\mu)\,. \tag{2.15}$$

To first order the normalization $u_au^a=-1$ gives

$$u^0=S^{-1}[1-AQ^{(0)}]\,. \tag{2.16}$$

In the rest frame of the matter the energy density is

$$E=-T_0^0=E_0(\tau)[1+\delta(\tau)Q^{(0)}(x^\mu)]\,. \tag{2.17}$$

In transforming back to the coordinate frame the mixed components T_b^a are unchanged to first order except for T_α^0 and T_0^α. The stress tensor T_β^α is represented by an isotropic pressure

$$P\equiv\tfrac{1}{3}T_\alpha^\alpha=P_0(\tau)+P_0(\tau)\pi_L(\tau)Q^{(0)}(x^\mu) \tag{2.18}$$

and a traceless anisotropic stress, with

$$T_\beta^\alpha=P_0[1+\pi_LQ^{(0)}]\delta_\beta^\alpha+P_0\pi_T^{(0)}(\tau)Q_\beta^{(0)\alpha}\,. \tag{2.19}$$

The remaining components are

$$T_\alpha^0=(E_0+P_0)(v^{(0)}-B^{(0)})Q_\alpha^{(0)}\,,\quad T_0^\alpha=-(E_0+P_0)v^{(0)}Q^{(0)\alpha}\,. \tag{2.20}$$

The perturbed isotropic pressure need not be related to the energy density in the same way as the background. The difference between the fractional pressure perturbation and that expected from the background pressure-energy density relation will be called the entropy perturbation,

$$\eta(\tau)Q^{(0)}=\left(\pi_L-\frac{E_0}{P_0}\frac{dP_0}{dE_0}\,\delta\right)Q^{(0)}=\frac{1}{w}\,(w\pi_L-c_s{}^2\delta)Q^{(0)}\,, \tag{2.21}$$

even though it may bear no relation to the true physical entropy, when this is a meaningful concept.

Perturbations in auxiliary quantities associated with the matter, such as a rest-mass density or specific intensity of radiation, can be defined in a similar fashion, and may be required to treat the internal dynamics of the matter, e.g., the interaction of matter in the narrow sense with electromagnetic or neutrino radiation and the propagation of this radiation.[21] However, in this paper we will only consider the overall dynamics of the matter as reflected in the equations $T^b_{a;b}=0$.

B. Vector perturbations

Now all quantities which are scalars under spatial coordinate transformations in the background must be unperturbed, e.g., g_{00}, T_0^0, and u^0. The description of vector and tensor quantities is similar to the scalar case, but the spatial dependence is generated from a fundamental vector harmonic $Q_\alpha^{(1)}$. In the metric tensor,

$$g_{0\alpha}=-S^2(\tau)B^{(1)}(\tau)Q^{(1)}{}_\alpha(x^\mu)\,, \tag{2.22a}$$

$$g_{\alpha\beta}=S^2[{}^3g_{\alpha\beta}+2H_T^{(1)}(\tau)Q_{\alpha\beta}^{(1)}(x^\mu)]\,. \tag{2.22b}$$

In the energy-momentum tensor,

$$T_\alpha^0=(E_0+P_0)(v^{(1)}-B^{(1)})Q_\alpha^{(1)}\,, \tag{2.23a}$$

$$T_\beta^\alpha=P_0\delta_\beta^\alpha+P_0\pi_T^{(1)}(\tau)Q_\beta^{(1)\alpha}\,, \tag{2.23b}$$

where

$$u^\alpha/u^0=v^{(1)}(\tau)Q^{(1)\alpha}(x^\mu)\,. \tag{2.24}$$

C. Tensor perturbations

The intrinsically tensor perturbations affect only the traceless part of the spatial metric and the traceless part of the stress tensor. For a particular tensor harmonic $Q_{\alpha\beta}^{(2)}$,

$$g_{\alpha\beta}=S^2[{}^3g_{\alpha\beta}+2H_T^{(2)}(\tau)Q_{\alpha\beta}^{(2)}]\,, \tag{2.25}$$

$$T_\beta^\alpha=P_0\delta_\beta^\alpha+P_0\pi_T^{(2)}(\tau)Q_\beta^{(2)\alpha}\,. \tag{2.26}$$

III. GAUGE TRANSFORMATIONS AND GAUGE-INVARIANT VARIABLES

As explained in Sec. I, a gauge transformation corresponds to a change of coordinates in the physical spacetime while the background coordinates are held fixed. Consistent with the perturbation analysis, only first-order effects of the coordinate transformation need be considered, and the spatial dependence of the transformation should correspond to the same harmonic that generates perturbations in the metric tensor and energy-momentum tensor.

A. Scalar perturbations

The most general possible gauge transformation associated with a scalar perturbation is the result of the coordinate transformation

$$\tilde\tau=\tau+T(\tau)Q^{(0)}(x^\mu)\,, \tag{3.1a}$$

$$\tilde x^\alpha=x^\alpha+L^{(0)}(\tau)Q^{(0)\alpha}(x^\mu) \tag{3.1b}$$

with T and $L^{(0)}$ arbitrary functions of τ.

The changes in the metric tensor are computed from

$$g_{ab}(x^c) = \frac{\partial \tilde{x}^k}{\partial x^a}\frac{\partial \tilde{x}^l}{\partial x^b}\tilde{g}_{kl}(\tilde{x}^m) . \quad (3.2)$$

In $\tilde{g}_{kl}$ and g_{ab} the scale factors are related by

$$S(\tilde{\tau}) \simeq S(\tau)[1+(\dot{S}/S)TQ^{(0)}] \quad (3.3)$$

and

$$^3g_{\alpha\beta}(\tilde{x}^\mu) \simeq {}^3g_{\alpha\beta}(x^\mu) + L^{(0)}Q^{(0)\mu}\frac{\partial}{\partial x^\mu}{}^3g_{\alpha\beta} . \quad (3.4)$$

The metric derivatives in Eq. (3.4) combine with the coordinate derivative of $Q^{(0)\alpha}$ in $\partial\tilde{x}^\alpha/\partial x^\beta$ to give covariant derivatives of $Q^{(0)\alpha}$. The final result for the changes in the amplitudes of the metric perturbations defined by Eqs. (2.14) is

$$\tilde{A} = A - \dot{T} - (\dot{S}/S)T , \quad (3.5a)$$

$$\tilde{B}^{(0)} = B^{(0)} + \dot{L}^{(0)} + kT , \quad (3.5b)$$

$$\tilde{H}_L = H_L - (k/3)L^{(0)} - (\dot{S}/S)T , \quad (3.5c)$$

$$\tilde{H}_T^{(0)} = H_T^{(0)} + kL^{(0)} . \quad (3.5d)$$

Now consider the matter perturbations. The new matter three-velocity is, by definition,

$$\tilde{v}^{(0)}Q^{(0)\alpha} = d\tilde{x}^\alpha/d\tilde{\tau} \simeq dx^\alpha/d\tau + \dot{L}^{(0)}Q^{(0)\alpha} ,$$

so

$$\tilde{v}^{(0)} = v^{(0)} + \dot{L}^{(0)} . \quad (3.6)$$

The energy density E is a coordinate scalar, but

$$E(\tilde{\tau}) = E_0(\tilde{\tau})[1+\tilde{\delta}Q^{(0)}]$$
$$\cong E_0(\tau)[1+(\tilde{\delta}+T\dot{E}_0/E_0)Q^{(0)}] .$$

Thus the energy density perturbation does change by

$$\tilde{\delta} = \delta + 3(1+w)(\dot{S}/S)T . \quad (3.7)$$

Equation (2.6) is used to eliminate $\dot{E}_0$. Similarly,

$$\tilde{\pi}_L = \pi_L - T\dot{P}_0/P_0 = \pi_L + 3(1+w)\frac{c_s^2}{w}\frac{\dot{S}}{S}T . \quad (3.8)$$

The amplitude of the traceless part of the stress tensor $\pi_T^{(0)}$ is gauge invariant.

The usual way of dealing with this gauge freedom is to impose conditions on the form of the metric tensor and/or matter perturbations. Examples are the synchronous gauge $A = B^{(0)} = 0$,[2,7,15] the longitudinal gauge $H_T^{(0)} = B^{(0)} = 0$,[4] the comoving proper time gauge $A = v^{(0)} = 0$,[5] and the comoving time-orthogonal gauge $B^{(0)} = v^{(0)} = 0$.[22] Such a condition may or may not specify the gauge uniquely. For instance, the transformation from some other gauge to the synchronous gauge contains two free constants of integration in T and $L^{(0)}$. Any ambiguity in the gauge condition implies the existence of extra, unphysical gauge modes when the Einstein equations are solved.[2,11]

The general covariance of Einstein's theory of gravity guarantees complete freedom in the choice of gauge as long as one can demonstrate the existence of a gauge transformation to that gauge from arbitrary metric tensor components and/or matter variables. The corollary of this principle is that *only gauge-invariant quantities have any inherent physical meaning.* Gauge-dependent quantities, such as the energy density perturbation, have physical meaning only to the extent that, in a particular gauge, they can be identified with a gauge-independent quantity either exactly or approximately.

To have genuine physical significance, gauge-independent quantities should be constructed from the variables naturally present in the problem, here the perturbations in the metric tensor and energy-momentum tensor, without reliance on artificially introduced variables, such as the four-velocity of an *ad hoc* congruence of "observers."

First consider the amplitudes of the metric tensor perturbations. Only two independent gauge-independent quantities can be constructed from the metric tensor amplitudes alone, since there are two gauge functions and four metric tensor amplitudes. By inspection of Eqs. (3.5), these are conveniently taken as

$$\Phi_A \equiv A + \frac{1}{k}\dot{B}^{(0)} + \frac{1}{k}\frac{\dot{S}}{S}B^{(0)} - \frac{1}{k^2}\left(\ddot{H}_T^{(0)} + \frac{\dot{S}}{S}\dot{H}_T^{(0)}\right) \quad (3.9)$$

and

$$\Phi_H \equiv H_L + \tfrac{1}{3}H_T^{(0)} + \frac{1}{k}\frac{\dot{S}}{S}B^{(0)} - \frac{1}{k^2}\frac{\dot{S}}{S}\dot{H}_T^{(0)} . \quad (3.10)$$

The physical interpretation of these gauge-invariant potentials will be postponed until after we have considered the matter perturbations.

The simplest gauge-invariant matter "velocity" amplitude is obviously, from Eqs. (3.6) and (3.5d),

$$v_s^{(0)} \equiv v^{(0)} - \frac{1}{k}\dot{H}_T^{(0)} . \quad (3.11)$$

This has a direct physical interpretation in terms of the shear of the matter velocity field. The shear tensor is[15]

$$\sigma_{ab} = \tfrac{1}{2}P_a^k(u_{k;l} + u_{l;k})P_b^l - \tfrac{1}{3}P_{ab}u^k_{;k} ,$$

with $P_{ab} = g_{ab} + u_a u_b$. This vanishes in the background and to first order the only nonzero components are

$$\sigma_{\alpha\beta} = S(\dot{H}_T^{(0)} - kv^{(0)})Q_{\alpha\beta}^{(0)} .$$

The magnitude of the shear is then

$$\sigma \equiv (\tfrac{1}{2}\sigma_{kl}\sigma^{kl})^{1/2} = \left|\frac{k}{S}v_s^{(0)}\right| (\tfrac{1}{2}Q^{(0)\alpha\beta}Q_{\alpha\beta}^{(0)})^{1/2} . \quad (3.12)$$

The velocity amplitude $v_s^{(0)}$ is the velocity which, when divided by the proper reduced wavelength S/k, gives the time dependence of the rate of shear associated with the perturbation.

The energy density perturbation amplitude δ must be combined with other quantities to produce a gauge-invariant measure of the density perturbation, and one obvious criterion is that the gauge-invariant quantity reduce to δ as soon as the perturbation comes inside the particle horizon, $k^{-1}\dot{S}/S \ll 1$. There are two obvious possibilities. First, consider

$$\epsilon_m \equiv \delta + 3(1+w)\frac{1}{k}\frac{\dot{S}}{S}(v^{(0)} - B^{(0)}), \tag{3.13}$$

which is gauge invariant by Eqs. (3.7), (3.6), and (3.5b). The amplitude ϵ_m is equal to δ in any gauge in which $v^{(0)} = B^{(0)}$, but this is just the condition that the matter world lines be orthogonal to the τ = constant spacelike hypersurface. Thus, *ϵ_m is the natural choice of gauge-invariant energy density perturbation amplitude from the point of view of the matter.* It is the density perturbation relative to the spacelike hypersurface which represents everywhere the matter local rest frame.

This is *not* the same as the density perturbation defined by Olson,[13] which compares energy densities at the same proper time calculated as an integral along each matter world line from the initial singularity. The Olson density perturbation definition has direct operational meaning for individual comoving observers only after the perturbation comes within the particle horizon. It has no gauge-invariant meaning which is local in time and can give a nonzero value for the density perturbation at times when the Universe is actually exactly homogeneous and isotropic, if a real perturbation was present temporarily at an earlier time.[23]

An alternative gauge-invariant density perturbation amplitude is

$$\epsilon_g \equiv \delta - 3(1+w)\frac{1}{k}\frac{\dot{S}}{S}\left(B^{(0)} - \frac{1}{k}\dot{H}_T^{(0)}\right). \tag{3.14}$$

From the discussion of shear following Eq. (3.11) and the fact that $B^{(0)}$ is the three-velocity amplitude of world lines normal to the τ = constant hypersurface, one sees that ϵ_g measures the energy density perturbation relative to the hypersurface whose normal unit vectors have *zero shear*. This geometrically selected hypersurface is as close as possible to a "Newtonian" time slicing.

Of course, any linear combination of ϵ_m and ϵ_g is gauge invariant, and the physical significance of one such linear combination will be discussed in Sec. V [see Eq. (5.26)]. Here I focus on ϵ_m because first, it acts in the Einstein equations as the source for the gauge-invariant potential Φ_H [see Eq. (4.3)], and second, the equations governing the dynamics of the matter are more transparent physically when written in terms of ϵ_m. The difference

$$\epsilon_m - \epsilon_g = 3(1+w)\frac{1}{k}\frac{\dot{S}}{S}v_s^{(0)} \tag{3.15}$$

is small once the perturbation is well inside the particle horizon, but is large at early times. Typically, $\epsilon_m \propto (k\tau)^2$ and ϵ_g is constant at $k\tau \ll 1$ for perturbations regular as $S \to 0$.

The zero-shear hypersurface can be invoked to give physical (geometrical) meaning to the gauge-invariant potentials Φ_A and Φ_H. In a gauge where each constant-τ hypersurface has normals with zero shear, i.e.,

$$B^{(0)} - \frac{1}{k}\dot{H}_T^{(0)} = 0,$$

Eqs. (3.9) and (3.10) greatly simplify to give

$$\Phi_A = A, \quad \Phi_H = H_L + \tfrac{1}{3}H_T^{(0)}.$$

But then Φ_A is the amplitude for the spatial dependence of the proper time intervals along the normals between two neighboring such zero-shear hypersurfaces (the lapse function), while the intrinsic scalar curvature of a zero-shear hypersurface is [see Eq. (A6)], to first order in the perturbation,

$$\mathcal{R}_{\text{zero shear}} = [6K + 4(k^2 - 3K)\Phi_H Q^{(0)}]/S^2. \tag{3.16}$$

In this sense, Φ_H physically represents a "curvature perturbation." Sufficient conditions for the global perturbations of the spacetime geometry to be small are $\Phi_A Q^{(0)} \ll 1$, $\Phi_H Q^{(0)} \ll 1$, but these are not necessary conditions since other hypersurfaces may be less strongly warped by the perturbation (see Sec. V).

To complete the gauge-invariant description of the matter perturbations, note that the fractional isotropic pressure perturbation can be expressed in terms of the energy density perturbation and the entropy perturbation $\eta Q^{(0)}$ through Eq. (2.21). But from Eqs. (2.21), (3.7), and (3.8) η is gauge invariant.

Certain gauges greatly simplify the representation of the gauge-invariant variables. For instance, in the longitudinal gauge $H_T^{(0)} = B^{(0)} = 0$, Eqs. (3.9)–(3.11) and Eq. (3.14) become $\Phi_A = A$, $\Phi_H = H_L$, $v_s^{(0)} = v^{(0)}$, $\epsilon_g = \delta$. A gauge in which $\epsilon_m = \delta$ and the other amplitudes are relatively simple is $v^{(0)} - B^{(0)} = 0$, $H_T^{(0)} = 0$. Both these gauge conditions uniquely specify the gauge.

B. Vector perturbations

The gauge-transformation properties of vector perturbations are much simpler than for scalar perturbations, since now there is no gauge ambiguity about the time coordinate. The most general gauge transformation is

$$\tilde{x}^{\alpha}=x^{\alpha}+L^{(1)}(\tau)Q^{(1)\alpha}(x^{\mu}), \tag{3.17}$$

so

$$\tilde{B}^{(1)}=B^{(1)}+\dot{L}^{(1)}, \tag{3.18a}$$

$$\tilde{H}_T^{(1)}=H_T^{(1)}+kL^{(1)}, \tag{3.18b}$$

and the matter velocity transforms as

$$\tilde{v}^{(1)}=v^{(1)}+\dot{L}^{(1)}. \tag{3.19}$$

The only gauge-invariant combination of metric tensor amplitudes is

$$\Psi\equiv B^{(1)}-\frac{1}{k}\dot{H}_T^{(1)}. \tag{3.20}$$

This, times k/S, is the amplitude of the shear of the normals to the constant-τ hypersurface. What was naturally constrained to vanish in dealing with scalar perturbations is now gauge invariant and necessarily nonvanishing if vector perturbations are present.

As for the energy density perturbations in the scalar case, there are two alternative choices for gauge-invariant forms of the matter velocity perturbation. The one related to the *shear* by the vector perturbation analog of Eq. (3.12) is

$$v_s^{(1)}\equiv v^{(1)}-\frac{1}{k}\dot{H}_T^{(1)}. \tag{3.21}$$

The other is related to the vorticity tensor

$$\omega_{ab}=\tfrac{1}{2}P_a^k(u_{k;l}-u_{l;k})P_b^l,$$

the only nonzero components of which are

$$\omega_{\alpha\beta}=S(v^{(1)}-B^{(1)})kW_{\alpha\beta}$$
$$=S(v^{(1)}-B^{(1)})(Q^{(1)}_{\alpha|\beta}-Q^{(1)}_{\beta|\alpha}).$$

The magnitude, the intrinsic angular velocity of an individual fluid element

$$\omega=[\tfrac{1}{2}\omega_{\alpha\beta}\omega^{\alpha\beta}]^{1/2}$$
$$=\left|\frac{k}{S}(v^{(1)}-B^{(1)})\right|[\tfrac{1}{2}W_{\alpha\beta}W^{\alpha\beta}]^{1/2} \tag{3.22}$$

is, like the shear, directly measurable from the local behavior of the matter. Let

$$v_c\equiv v^{(1)}-B^{(1)}=v_s^{(1)}-\Psi. \tag{3.23}$$

It is v_c, the velocity relative to the normal to the constant-τ hypersurface, that is the source for Ψ in the Einstein equations, rather than $v_s^{(1)}$.

C. Tensor perturbations

Since no three-vector or scalar can be formed from a tensor harmonic $Q^{(2)}_{\alpha\beta}$, the amplitudes $H_T^{(2)}$ and $\pi_T^{(2)}$ as defined in Eqs. (2.25) and (2.26) are automatically gauge invariant.

IV. PERTURBATION EQUATIONS IN GAUGE-INVARIANT VARIABLES

The usual approach to the derivation of the equations governing linearized perturbations in cosmology has been to impose at the beginning a gauge condition to simplify the form of the metric and/or matter perturbations and then work directly with the metric tensor components and matter variables. However, a complete set of equations can be obtained directly in terms of the gauge-invariant variables defined in Sec. III. These equations are mathematically simpler and physically more transparent than the usual ones, particularly in comparison with the commonly used synchronous gauge. Spurious gauge modes are automatically excluded.

In the Appendix we give the perturbations in the Ricci tensor components in a general gauge. The perturbation in the Einstein tensor is

$$\delta G^a_b=\delta R^a_b-\tfrac{1}{2}\delta^a_b\delta R.$$

A. Scalar perturbations

For scalar perturbations, one gauge-invariant combination is

$$\delta G^0_0-\frac{3}{k^2}\frac{\dot{S}}{S}(\delta G^0_\alpha)^{|\alpha}=-2\frac{(k^2-3K)}{S^2}\Phi_H Q^{(0)}, \tag{4.1}$$

the other

$$\delta G^\alpha_\beta-\tfrac{1}{3}\delta^\alpha_\beta\delta G^\mu_\mu=-\frac{k^2}{S^2}(\Phi_A+\Phi_H)Q^{(0)\alpha}_\beta. \tag{4.2}$$

Upon equating δG^a_b with δT^a_b through the Einstein equations one finds from Eqs. (4.1), (2.17), (2.20), and (3.13)

$$2\frac{(k^2-3K)}{S^2}\Phi_H=E_0\epsilon_m \tag{4.3}$$

and from Eqs. (4.2) and (2.19)

$$-\frac{k^2}{S^2}(\Phi_A+\Phi_H)=P_0\pi_T^{(0)}. \tag{4.4}$$

Thus, both gauge-invariant amplitudes for the metric tensor perturbation are related algebraically in a very simple way to gauge-invariant amplitudes of the matter perturbations. Equation (4.3), in spite of appearances, really derives from the energy initial-value equation on the zero-shear hypersurface (see Sec. VI). For a perfect fluid, Eq. (4.4) gives $\Phi_A=-\Phi_H$.

All of the dynamics for scalar (and also vector) perturbations resides in the equations of motion of the matter, $T^k_{0;k}=0$ and $T^k_{\alpha;k}=0$. These are also written out in the Appendix for a general gauge. Straightforward manipulation of the momentum equation, with the help of Eqs. (3.9), (3.11), (3.13), and (2.21), gives the explicitly gauge-invariant form

$$\dot{v}_s^{(0)}+\frac{\dot{S}}{S}v_s^{(0)}=k\Phi_A+k(1+w)^{-1}(c_s^2\epsilon_m+w\eta)-\tfrac{2}{3}k(1-3K/k^2)(1+w)^{-1}w\pi_T^{(0)}. \quad (4.5)$$

The second term on the right-hand side of Eq. (4.5) is S times the acceleration in the rest frame of the matter due to the pressure gradient force, and the third term is S times the acceleration due to the divergence of the anisotropic part of the stress tensor. The inertial mass per unit volume is $E_0(1+w)$, and the *proper* wave number is k/S. However, the first term is *not* S times the gravitational acceleration in the matter rest frame. If it were, the left-hand side of Eq. (4.5) would be zero. Rather, it is the gravitational acceleration in the "shear-free frame" associated with the geometry perturbations, as discussed in Sec. III. Equation (4.5) has *exactly* the same form [except for the factor $(1-3K/k^2)$ in the third term on the right-hand side] as the corresponding *Newtonian* equation in an expanding background, with $v_s^{(0)}$ the analog of the Newtonian peculiar velocity and Φ_A the analog of the Newtonian gravitational potential.

The energy equation is less transparent because it is sensitive to first-order changes in the frame of reference in ways the momentum equation is not. Begin by considering Eq. (A4a) in a particular gauge, the comoving, time-orthogonal gauge $v^{(0)}=B^{(0)}=0$. In this gauge $\epsilon_m=\delta$, $v_s^{(0)}=-(1/k)\dot{H}_T^{(0)}$, and from Eq. (3.10)

$$\dot{H}_L=\dot{\Phi}_H+\tfrac{1}{3}kv_s^{(0)}-\frac{1}{k}\left(\frac{\dot{S}}{S}v_s^{(0)}\right)^{\cdot}. \quad (4.6)$$

The obvious substitutions and elimination of $\dot{v}_s^{(0)}$ through Eq. (4.5) give a gauge-invariant equation for $\dot{\Phi}_H$, but a rather messy one. Now use Eqs. (4.3) and (4.4) to eliminate ϵ_m and Φ_A in favor of Φ_H and simplify with the help of the background Eqs. (2.5) and (2.6). After elimination of a common factor $2[k^2-3K+\tfrac{3}{2}(E_0+P_0)S^2]/E_0S^2$,

$$\dot{\Phi}_H+\frac{\dot{S}}{S}\Phi_H=-\frac{1}{2}\,\frac{(E_0+P_0)S^2}{k}v_s^{(0)}-\frac{P_0S^2}{k^2}\,\frac{\dot{S}}{S}\pi_T^{(0)}. \quad (4.7)$$

Alternatively, write Eq. (4.7) in a more familiar form

$$[E_0S^3\epsilon_m]^{\cdot}=-(1-3K/k^2)(E_0+P_0)S^3kv_s^{(0)}-2(1-3K/k^2)P_0S^2\dot{S}\pi_T^{(0)}. \quad (4.8)$$

This has some resemblance to a *special* relativistic energy equation, in that $(k/S)(E_0+P_0)v_s^{(0)}$ is the divergence of the energy flux. However, ϵ_m is not the energy density perturbation in the appropriate frame, and the special relativistic equation would have $3\pi_L$ in place of $2\pi_T^{(0)}$.

Equations (4.3)–(4.5) and Eq. (4.8) combine into a single second-order equation for ϵ_m,

$$(E_0S^3\epsilon_m)^{\cdot\cdot}+(1+3c_s^2)\frac{\dot{S}}{S}(E_0S^3\epsilon_m)^{\cdot}+[(k^2-3K)c_s^2-\tfrac{1}{2}(E_0+P_0)S^2](E_0S^3\epsilon_m)$$
$$=(1-3K/k^2)\{-k^2(P_0S^3\eta)+\tfrac{2}{3}[k^2+3(1+3c_s^2)K](P_0S^3\pi_T^{(0)})+2(w-c_s^2)(E_0S^2)(P_0S^3\pi_T^{(0)})-2\dot{S}(P_0S^2\pi_T^{(0)})^{\cdot}\}, \quad (4.9)$$

or a corresponding equation for Φ_H. A spatially homogeneous perturbation or the lowest inhomogeneous mode $k^2=3K$ in a closed universe require special treatment in that $Q_\alpha^{(0)}$ and/or $Q_{\alpha\beta}^{(0)}$ vanish identically, Φ_H, Φ_A, and $v_s^{(0)}$ are no longer gauge-invariant, and some of the above equations, including Eq. (4.9), are not applicable. See the discussion around Eq. (6.27). A homogeneous scalar perturbation is really no perturbation at all, but an inappropriate choice of background.

Formally, the entropy perturbation amplitude and the anisotropic stress amplitude are free functions of time which act as "sources" for the comoving energy density perturbation in Eq. (4.9). However, if the anisotropic stress is from a shear viscosity, $\pi_T^{(0)}$ should be proportional to $v_s^{(0)}$. In a simple kinetic-theory model the coefficient of viscosity is a density times a mean free path times a velocity. An upper limit to the density is E_0, a characteristic velocity is $w^{1/2}$, and a maximum effective mean free path is the *lesser* of the reduced wavelength of the perturbation S/k and the distance a particle can travel in a Hubble time, $w^{1/2}/(\dot{S}/S^2)$. At the upper limit the simple kinetic-theory picture breaks down and the relation between shear and anisotropic stress becomes nonlocal in time and space. Nevertheless, a reasonable estimate is

$$P_0\pi_T^{(0)} \sim \alpha E_0 \left\{\frac{S}{k}, w^{1/2}\frac{S^2}{\dot{S}}\right\}_{\min} w^{1/2}\left(\frac{k}{S}v_s^{(0)}\right), \quad (4.10)$$

where α is a dimensionless fudge factor unlikely to be larger than one. At early times, $k\tau < 1$,

$$\pi_T^{(0)} \lesssim \frac{kS}{\dot{S}} v_s^{(0)} . \quad (4.11)$$

With this restriction, the shear-stress terms in Eqs. (4.8) and (4.9) can never dominate the early evolution.

An anisotropic stress as well as an entropy perturbation might also arise *ab initio* out of some speculative nonkinetic origin, perhaps inhomogeneous gauge-theory phase transitions of the vacuum.[8,9,16,17] This possibility will be discussed further in connection with specific solutions of Eq. (4.9) in Sec. V.

B. Vector perturbations

The equations for vector perturbations (see the Appendix) can be put in gauge-invariant form by inspection. The initial-value equation $\delta G_\alpha^0 = \delta T_\alpha^0$ for the "frame-dragging potential" Ψ is [see Eq. (A2b)]

$$\frac{1}{2}\frac{(k^2-2K)}{S^2}\Psi Q_\alpha^{(1)} = (E_0+P_0)v_c Q_\alpha^{(1)} . \quad (4.12)$$

The equation of motion for the matter is

$$\dot{v}_c = \frac{\dot{S}}{S}(3c_s^2-1)v_c - k\frac{w}{1+w}\pi_T^{(1)} . \quad (4.13)$$

C. Tensor perturbations

There is just one equation for tensor perturbations, i.e.,

$$\frac{1}{S^2}\left(\ddot{H}_T^{(2)} + 2\frac{\dot{S}}{S}\dot{H}_T^{(2)} + (k^2+2K)H_T^{(2)}\right) = P_0\pi_T^{(2)} . \quad (4.14)$$

V. SOLUTIONS OF THE PERTURBATION EQUATIONS

A. Scalar perturbations

Explicit solutions of the gauge-dependent equations governing scalar perturbations have been obtained in the past[11] for the case of a perfect fluid ($\pi_T^{(0)}=0$), usually only for adiabatic perturbations ($\eta=0$), when the background equation of state satisfies $w=c_s^2=$const and the background spatial curvature $K=0$. The assumption $K=0$ is well justified for perturbations on the scale of clusters of galaxies or smaller, since then $k^2 \gg K$ and at least prior to recombination $(\dot{S}/S)^2 \gg K$ for any K in the range allowed by observation. The assumption $w=c_s^2=$const is made largely for mathematical convenience, though in the radiation-dominated phase of the universe following electron-position recombination and neutrino decoupling $w=\frac{1}{3}$ is a good approximation. Here the solutions will be extended to allow entropy and anisotropic pressure perturbations with arbitrary time dependence, and the physical significance of the solutions will be analyzed through their expression in gauge-invariant variables.

With $w=c_s^2=$const and $K=0$ the background equations (2.5) and (2.6) have the well-known solution

$$S \propto \tau^\beta, \quad \beta \equiv 2/(3w+1) \quad (5.1)$$

and

$$E_0 S^2 = 3(\dot{S}/S)^2 = 3\beta^2\tau^{-2} . \quad (5.2)$$

The parameter β ranges from 2 (for $w=0$) to 1 (for $w=\frac{1}{3}$) to $\frac{1}{2}$ (for $w=1$).

Define a new independent variable

$$x \equiv k\tau \quad (5.3)$$

and in Eq. (4.9) a new independent variable

$$f \equiv x^{\beta-2}\epsilon_m = \tfrac{2}{3}\beta^{-2}x^\beta\Phi_H . \quad (5.4)$$

Let a prime denote d/dx. Then Eq. (4.9) becomes

$$f'' + 2x^{-1}f' + [c_s^2 - \beta(\beta+1)x^{-2}]f = -x^{\beta-2}[w\eta - \tfrac{2}{3}w\pi_T^{(0)} + 2\beta x(x^{-2}w\pi_T^{(0)})'] . \quad (5.5)$$

Given a solution for f, Eq. (4.8) determines $v_s^{(0)}$ as

$$v_s^{(0)} = -\frac{3}{2}\frac{\beta}{\beta+1}x^{2-\beta}f' - 3\frac{\beta^2}{\beta+1}x^{-1}w\pi_T^{(0)} . \quad (5.6)$$

First consider the homogeneous version of Eq. (5.5), with $w\eta = w\pi_T^{(0)} = 0$. The solutions are obviously spherical Bessel functions, either j_β or the spherical Neumann function n_β, and

$$f = a j_\beta(c_s x) + b n_\beta(c_s x) . \quad (5.7)$$

The simple form of the homogeneous solution, with nothing in it to mark the particle horizon at $x \sim 1$ as distant from the sound horizon at $c_s x \sim 1$, confirms the choice of ϵ_m, Φ_H, and $v_s^{(0)}$ as mathematically, and in some sense physically, natural gauge-invariant amplitudes to describe the perturbation.

Now take the limit $c_s x \ll 1$, but not necessarily $x \ll 1$. The oscillatory behavior at $c_s x > 1$ becomes power-law behavior. Renormalize the coefficients so

$$f \simeq cx^\beta + dx^{-\beta-1} , \quad (5.8)$$

with

$$c = \frac{\sqrt{\pi}\,c_s^\beta}{2^{\beta+1}\Gamma(\beta+3/2)}a, \quad d = -\frac{2^\beta\Gamma(\beta+1/2)}{\sqrt{\pi}\,c_s^{\beta+1}}b . \quad (5.9)$$

The standard gauge-invariant amplitudes at $c_s x \ll 1$ are

$$\epsilon_m \simeq cx^2 + dx^{-2\beta+1}, \quad \Phi_H \simeq \tfrac{3}{2}\beta^2(c + dx^{-2\beta-1}),$$
$$v_s^{(0)} \simeq \tfrac{3}{2}\beta\left(-c\,\frac{\beta}{\beta+1}x + dx^{-2\beta}\right). \tag{5.10}$$

The two independent modes may, at $c_s x \ll 1$, be characterized as a "growing mode" and a "decaying" mode by looking at the amplitude of the fractional energy density perturbation on comoving hypersurfaces. The amplitude Φ_H measures the distortion of the zero-shear hypersurfaces by the perturbation. The spatial curvature $\sim (k^2/S^2)\Phi_H$ is coherent over the reduced wavelength $\sim S/k$. For the growing mode this distortion is nonzero at $x=0$ and remains constant until the sound wave begins to oscillate. We shall see that this behavior is also characteristic of the distortion of the comoving hypersurfaces and uniform-Hubble-constant hypersurfaces, so geometrically the growing mode is a "constant" mode at $x \ll 1$. However, the apparent relative amplitude of the energy density perturbation is strongly hypersurface dependent at $x \ll 1$. Consider the amplitude ϵ_g, which measures the fractional energy density perturbation on zero-shear hypersurfaces. From Eq. (3.15),

$$\epsilon_g \simeq c(x^2 + 3\beta^2) + dx^{-2\beta-1}[x^2 - 3\beta(\beta+1)], \tag{5.11}$$

still assuming $(c_s x)^2 \ll 1$. For both modes ϵ_g is of order x^{-2} times ϵ_m at $x \ll 1$, though ϵ_g and ϵ_m coincide for all $x \gg 1$ even if $c_s \ll 1$.

About the only measure of the relative amplitude of the perturbation which is strictly hypersurface independent is the ratio of the rate of shear σ to the background expansion rate, because the shear vanishes in the background. The amplitudes of the shear and expansion are, respectively, $(k/S)\, v_s^{(0)}$ and $\dot{S}/S^2$. The ratio is

$$\xi \equiv (kS/\dot{S})v_s^{(0)} \simeq \frac{3}{2}\left(-c\,\frac{\beta}{\beta+1}x^2 + dx^{-2\beta+1}\right) \tag{5.12}$$

at $x \ll 1$, which has the same time dependence for each mode as ϵ_m. For $w<1$, or $\beta > \frac{1}{2}$, the decaying homogeneous mode is unambiguously singular at $x=0$. The gauge-invariant amplitude Φ_H becomes singular as $x^{-(2\beta+1)}$ in the decaying mode, but we shall see that this overstates the physical strength of the singularity. On the comoving and the uniform-Hubble-constant hypersurfaces the distortion only becomes singular as $x^{-(2\beta-1)}$, similar to ϵ_m and the relative shear in Eq. (5.12).

The results so far are familiar ones,[4,11,22] though the explicit gauge-invariant form is new. However, relatively little attention has been paid to the generation of density perturbations at early times through stress perturbations. As mentioned in Sec. I and as is obvious from Eq. (4.9), an energy density perturbation can only arise from a nonadiabatic pressure perturbation or an anisotropic stress perturbation if the Universe began perfectly homogeneous. Local conservation of energy and momentum prevents any action directly on the energy and momentum densities.

Entropy perturbations have been considered before,[12] but it appears from Eq. (5.5) that anisotropic stress perturbations might be more interesting. At $x \ll 1$ the second anisotropic stress term is of order x^{-2} times the entropy perturbation term for comparable values of $w\eta$ and $w\pi_T^{(0)}$.

The solution of Eq. (5.5) for an arbitrary source is accomplished by variation of parameters or (equivalently) by constructing the Green's function. I will apply initial conditions that the perturbations vanish at $x=0$, but some of the homogeneous solution, Eq. (5.7), can always be added if desired. The derivative on the right-hand side of Eq. (5.5) can be integrated by parts. The identity

$$\frac{d}{dy}[y^{\beta+1}n_\beta(c_s y)] = c_s y^{\beta+1} n_{\beta-1}(c_s y), \tag{5.13}$$

and similarly for j_β, simplifies the result, which is

$$f(x) = c_s \int_0^x dy\, y^\beta\{(w\eta - \tfrac{2}{3}w\pi_T^{(0)})[j_\beta(c_s x)n_\beta(c_s y) - n_\beta(c_s x)j_\beta(c_s y)]$$
$$- 2\beta c_s y^{-1} w\pi_T^{(0)}[j_\beta(c_s x)n_{\beta-1}(c_s y) - n_\beta(c_s x)j_{\beta-1}(c_s y)]\}. \tag{5.14}$$

If $c_s x \ll 1$, the spherical Bessel functions may be replaced by

$$j_\beta(c_s x)n_\beta(c_s y) \simeq -(2\beta+1)^{-1}(c_s y)^{-1}(x/y)^\beta \text{ and } x \leftrightarrow y,$$
$$j_\beta(c_s x)n_{\beta-1}(c_s y) \simeq -(4\beta^2-1)^{-1}(x/y)^\beta, \quad \beta > \tfrac{1}{2},$$
$$n_\beta(c_s x)j_{\beta-1}(c_s y) \simeq -(c_s x)^{-2}(y/x)^{\beta-1}.$$

Keeping only the dominant terms, Eq. (5.14) simplifies to

$$f(x)\simeq (2\beta+1)^{-1}\left\{x^{\beta}\int_0^x dy\, y^{-1}\left[\frac{2}{3}\,\frac{\beta+1}{2\beta-1}w\pi_T^{(0)}-w\eta\right]+x^{-\beta-1}\int_0^x dy\, y^{2\beta}[-2\beta(2\beta+1)y^{-2}w\pi_T^{(0)}+w\eta]\right\}. \tag{5.15}$$

If $\beta=\frac{1}{2}$ the contribution of $w\pi_T^{(0)}$ to the growing mode is not infinite; rather, $n_{-1/2}$ contains a logarithm, and

$$\frac{2}{3}\,\frac{\beta+1}{2\beta-1}\to\tfrac{1}{2}\ln(1/y). \tag{5.16}$$

In Eq. (5.15), the entropy perturbation contributes comparable amounts of growing mode and decaying mode to f and therefore to $\epsilon_m=x^{2-\beta}f$. The contribution to ϵ_m while the entropy perturbation is present is of order $x^2w\eta$, once the entropy perturbation has been on at roughly constant amplitude for at least one expansion time. Press and Vishniac[12] claim that an entropy perturbation couples *only* to the growing mode at $x\ll 1$. The apparent contradiction arises because their definition of the "physical" energy density perturbation is based on uniform-Hubble-constant hypersurfaces; the relative strength of the growing and decaying modes is hypersurface dependent, as we shall soon see in detail. Still, if the entropy perturbation turns off at $x\ll 1$, by the time the perturbation comes within the particle horizon at $x\sim 1$ the decaying mode is insignificant compared with the growing mode. At $x\sim 1$ the latter has an amplitude in ϵ_m which is roughly the maximum previous amplitude of $w\eta$, the ratio of the nonadiabatic pressure perturbation to the background energy density, averaged over one e-folding in the conformal time τ.

The anisotropic stress source term of order x^{-2} relative to the entropy perturbation source term contributes *only* to the decaying mode at $x\ll 1$ in Eq. (5.15). Except for a modest enhancement if β is close to $\frac{1}{2}$, as indicated in Eq. (5.16), anisotropic stress and isotropic stress perturbations of the same amplitude generate comparable amounts of the growing mode in f, ϵ_m, and [see Eq. (5.4)] Φ_H. After an anisotropic stress perturbation has been on at roughly constant amplitude for one expansion time, Eq. (5.15) shows that the part of ϵ_m associated with the decaying mode is of order $w\pi_T^{(0)}$ and the part of ϵ_m associated with the growing mode is of order $x^2w\pi_T^{(0)}$. If the anisotropic stress then disappears at $x=x_1$, by the time $x=1$ the decaying-mode contribution to ϵ_m is of order $x_1^{+(2\beta-1)}w\pi_T^{(0)}$, small (unless $\beta\simeq\frac{1}{2}$) compared with the growing-mode contribution of order $w\pi_T^{(0)}$, if $x_1\ll 1$.

The perturbation in ϵ_m is always small if $w\pi_T^{(0)}\ll 1$. On the other hand, the amplitude Φ_H, which measures the global distortion of the zero-shear hypersurfaces, is of order $x^{-2}\epsilon_m$ and while the anisotropic stress is present can be larger than one even if the anisotropic stress is small compared with the background energy density, as long as the anisotropic stress is present at a very early time, with $x^2<w\pi_T^{(0)}$. Some gauge-theory symmetry breaking might take place near the Planck time.[8] At the Planck time the value of $x=k\tau$ corresponding to a mass M in the present universe is $x\sim 10^{-26}(10^{12}M_\odot/M)^{1/3}$, so even though the comoving volume of a galactic mass is large compared with the particle horizon at this value of x, it is not inconceivable that a small statistical residual anisotropy on a galactic scale could make in a sense the perturbations nonlinear.

Does $\Phi_H>1$ really imply a physically significant nonlinearity, one that could perhaps couple the large amplitude decaying mode to the growing mode and give a value of ϵ_m at the particle horizon large compared with the maximum previous value of $w\pi_T^{(0)}$? To answer this question, at least in part, one should look carefully at the complete description of the perturbation on various types of hypersurfaces while the anisotropic stress is present or just after it turns off.

The description definitely is potentially nonlinear on the zero-shear hypersurfaces. The distortion amplitude $\phi_g\equiv\Phi_H$ is a measure of the amplitude of the spatial metric perturbations on the zero-shear hypersurfaces, and the amplitude $\alpha_g\equiv\Phi_A$ measures the fractional perturbation in the lapse function. Both these are of order $x^{-2}w\pi_T^{(0)}$, as is the fractional energy density perturbation amplitude ϵ_g from Eq. (5.11). The matter velocity amplitude on zero-shear hypersurfaces is $v_g\equiv v_s^{(0)}$, and from Eq. (5.6) is of order $x^{-1}w\pi_T^{(0)}$, though the shear to expansion rate ratio is of order $w\pi_T^{(0)}$ [see Eq. (5.12)].

Now consider the comoving hypersurfaces. We have already seen that the fractional energy density perturbation amplitude is only of order $w\pi_T^{(0)}$. The matter velocity relative to the comoving hypersurface is zero by definition. To get the geometrical amplitudes, consider a gauge transformation from the longitudinal gauge $H_T^{(0)}=B^{(0)}=0$, where $A=\Phi_A$, $H_L=\Phi_H$, $v^{(0)}=v_s^{(0)}$, to a comoving gauge with $\tilde{v}^{(0)}-\tilde{B}^{(0)}=0$. Equations (3.5b) and (3.6) give

$$T=k^{-1}(v^{(0)}-B^{(0)})=k^{-1}v_s^{(0)}. \tag{5.17}$$

Then from Eq. (3.5a) the fractional perturbation in the lapse function has an amplitude

$$\tilde{A}=\Phi_A-k^{-1}\left(\dot{v}_s^{(0)}+\frac{\dot{S}}{S}v_s^{(0)}\right)\equiv\alpha_m\,, \tag{5.18}$$

and the measure of the distortion of the intrinsic geometry is

$$\phi_m=\tilde{H}_L+\tfrac{1}{3}\tilde{H}_T=\Phi_H-\frac{\dot{S}}{S}k^{-1}v_s^{(0)}\,. \tag{5.19}$$

Equation (4.5) gives

$$\alpha_m=-(1+w)^{-1}(c_s^{\,2}\epsilon_m+w\eta)+\tfrac{2}{3}(1-3K/k^2)\frac{w}{1+w}\pi_T^{(0)}\,, \tag{5.20}$$

so α_m is of order $w\pi_T^{(0)}$ and $w\eta$. An equation for $\dot{\phi}_m$ can be derived from Eqs. (4.5), (4.7), and (2.5), with the result

$$\dot{\phi}_m=-Kk^{-1}v_s^{(0)} - (1+w)^{-1}\frac{\dot{S}}{S}[c_s^{\,2}\epsilon_m+w\eta-\tfrac{2}{3}(1-3K/k^2)w\pi_T^{(0)}]\,. \tag{5.21}$$

Each term on the right-hand side of Eq. (5.21) is at most of order $\tau^{-1}(w\eta, w\pi_T^{(0)})$ and therefore ϕ_m is also of order $w\eta$, $w\pi_T^{(0)}$, without any special assumption that w is constant.

On comoving hypersurfaces, then, all relative perturbation amplitudes are small up until the time the perturbation enters the particle horizon if the perturbation vanishes initially and subsequent stress perturbations are small ($w\eta\ll 1$, $w\pi_T^{(0)}\ll 1$), no matter how early the stress perturbations turn on, as long as the stress perturbations are on in full strength for only a reasonably finite number of e-foldings in τ. The larger amplitudes ϕ_g, α_g, ϵ_g, and $v_s^{(0)}$ for zero-shear hypersurfaces then are *only* due to a large warping of the zero-shear hypersurfaces relative to comoving hypersurfaces.

Finally, consider the perturbations relative to uniform-Hubble-constant hypersurfaces. Again, the simplest route to the gauge-invariant amplitudes describing the matter and geometry from the point of view of this hypersurface is the gauge transformation from the longitudinal gauge to the gauge satisfying the hypersurface condition [see Eq. (A7)]

$$\dot{\tilde{H}}_L+\frac{k}{3}\tilde{B}^{(0)}-\frac{\dot{S}}{S}\tilde{A}=0\,. \tag{5.22}$$

The amplitude of change in the time coordinate is

$$T=-3[k^2-3K+\tfrac{3}{2}(E_0+P_0)S^2]^{-1}\left(\dot{\Phi}_H-\frac{\dot{S}}{S}\Phi_A\right) =\tfrac{3}{2}(E_0+P_0)S^2[k^2-3K+\tfrac{3}{2}(E_0+P_0)S^2]^{-1}k^{-1}v_s^{(0)}\,, \tag{5.23}$$

after simplifying with the help of Eqs. (4.4) and (4.7). The intrinsic geometry of the uniform-Hubble-constant hypersurface is governed by the amplitude

$$\phi_h=\phi_m+[1+\tfrac{3}{2}(k^2-3K)^{-1}(E_0+P_0)S^2]^{-1}\frac{\dot{S}}{S}k^{-1}v_s^{(0)}\,. \tag{5.24}$$

Similarly, the lapse function perturbation amplitude is

$$\alpha_h=\alpha_m+k^{-1}S^{-1}\{[1+\tfrac{3}{2}(k^2-3K)^{-1}(E_0+P_0)S^2]^{-1}Sv_s^{(0)}\}^{\boldsymbol{\cdot}}\,. \tag{5.25}$$

The amplitude of the fractional energy density perturbation can be written as

$$\epsilon_h=[1+\tfrac{3}{2}(k^2-3K)^{-1}(E_0+P_0)S^2]^{-1}[\epsilon_m+3(1+w)\Phi_m]\,. \tag{5.26}$$

After simplification using the Einstein equations, the amplitude of the matter velocity relative to the hypersurface is

$$v_h=\tilde{v}^{(0)}-\tilde{B}^{(0)} =[1+\tfrac{3}{2}(k^2-3K)^{-1}(E_0+P_0)S^2]^{-1}v_s^{(0)}\,. \tag{5.27}$$

In a constant-w background, and with $k^2\gg K$, the ubiquitous factor in Eqs. (5.24)–(5.27) becomes

$$[1+\tfrac{3}{2}(k^2-3K)(E_0+P_0)S^2]^{-1}=x^2/[3\beta(\beta+1)+x^2]. \tag{5.28}$$

The description of both the matter and the geometry changes character depending on whether the perturbation is larger or smaller than the particle horizon at $x\sim 1$. On the other hand, the description of the matter and geometry relative to the comoving hypersurface is in all respects oblivious to the particle horizon. I have already remarked on this for ϵ_m, but it also holds for ϕ_m and α_m as is obvious from Eqs. (5.20) and (5.21). Relative to the zero-shear hypersurfaces, the description of the *geometry* is unaffected by the particle horizon, since Φ_H acts as a quasi-Newtonian potential foe ϵ_m, but ϵ_g does change character at the particle horizon [see Eq. (5.11)].

There are several measures of the relative amplitude of the perturbation whose physical interpretation is hypersurface dependent, but only one, the ratio ξ of matter shear rate to expansion rate [Eq. (5.12)], which has a hypersurface-independent *physical* significance. Since all are formally gauge invariant, the *mathematical* values of the amplitudes are independent of the gauge/hypersurface choice.

Table I gives the comparison of the hypersurface-dependent amplitudes with ξ for each of the above hypersurface conditions and for each of the two *homogeneous* modes of scalar perturbations,

TABLE I. Amplitudes of physical perturbations at $x \ll 1$ compared with amplitude ξ of the ratio of the shear rate of the matter to the background expansion rate.

Hypersurface	ϵ/ξ	v/ξ	ϕ/ξ	α/ξ
		(a) "Growing" mode, $\xi \sim x^2$		
Uniform-Hubble-constant	$-\frac{2}{3}\frac{2\beta+1}{\beta}$	$\frac{1}{3}\frac{\beta}{\beta+1}\left(\frac{x}{\beta}\right)$	$-\frac{2\beta+1}{\beta}\left(\frac{x}{\beta}\right)^{-2}$	$\frac{2}{3}\frac{2\beta+1}{\beta(\beta+1)}$
Comoving	$-\frac{2}{3}\frac{\beta+1}{\beta}$	0	$-\frac{2\beta+1}{\beta}\left(\frac{x}{\beta}\right)^{-2}$	$\frac{1}{3}\frac{2-\beta}{\beta}$
Zero-shear	$-2\frac{\beta+1}{\beta}\left(\frac{x}{\beta}\right)^{-2}$	$\left(\frac{x}{\beta}\right)^{-1}$	$-\frac{\beta+1}{\beta}\left(\frac{x}{\beta}\right)^{-2}$	$\frac{\beta+1}{\beta}\left(\frac{x}{\beta}\right)^{-2}$
		(b) "Decaying" mode, $\xi \sim x^{-(2\beta-1)}$		
Hypersurface	ϵ/ξ	v/ξ	ϕ/ξ	α/ξ
Uniform-Hubble-constant	$\frac{2}{9}\frac{\beta}{2\beta-1}\left(\frac{x}{\beta}\right)^2$	$\frac{1}{3}\frac{\beta}{\beta+1}\left(\frac{x}{\beta}\right)$	$\frac{1}{3}\frac{\beta}{2\beta-1}$	$-\frac{2}{9}\frac{\beta}{(\beta+1)(2\beta-1)}\left(\frac{x}{\beta}\right)^2$
Comoving	$\frac{2}{3}$	0	$\frac{1}{3}\frac{\beta(2-\beta)}{(\beta+1)(2\beta-1)}$	$-\frac{1}{3}\frac{2-\beta}{\beta+1}$
Zero-shear	$-2\frac{\beta+1}{\beta}\left(\frac{x}{\beta}\right)^{-2}$	$\left(\frac{x}{\beta}\right)^{-1}$	$\left(\frac{x}{\beta}\right)^{-2}$	$-\left(\frac{x}{\beta}\right)^{-2}$

in the limit that $x \ll 1$. Of course, I also assume w is constant and $k^2 \gg K$. A couple of points are particularly noteworthy. First, for the growing mode the amplitude of the distortion in the intrinsic geometry (ϕ) is independent of time and roughly the same in all three hypersurfaces. The values of ϕ in all three hypersurfaces then predicts the value of the fractional energy density perturbation at the particle horizon, where ϵ_m, ϵ_g, and ϵ_h converge. Second, for the *decaying* mode, but not the growing mode, the amplitudes ϵ_h and α_h are down by a factor of x^2 relative to ϵ_m and α_m. The relatively small value of α_h means that the uniform-Hubble-constant hypersurfaces are close to being synchronous when $x^2 \ll 1$. The relatively small value of ϵ_h is the basis for the claim of Press and Vishniac[12] that an entropy perturbation couples only to the growing mode, but this claim is valid just for the one hypersurface condition. The value of ϕ_h is *not* suppressed relative to ϕ_m, so the perturbation as a whole has the same time dependence, $x^{-(2\beta-1)}$, from the point of view of the uniform-Hubble-constant hypersurfaces as from the point of view of the comoving hypersurfaces.

In conjunction with Table I, it may be helpful to give explicit results for how the entropy perturbation and the anisotropic stress contribute to the amplitude ξ. Assume that the stress perturbations are each on for n e-foldings in the scale factor S, i.e., $n\beta$ e-foldings in the conformal time τ, at a time when $kS/\dot{S} = x/\beta \ll 1$. During this time the relative stress amplitudes $w\eta$ and $w\pi_T^{(0)}$ are constant. The inhomogeneous stress perturbations disappear at $x = x_1$. The growing mode in ξ at $x > x_1$ is then

$$\xi = n\left[\frac{3}{2}\,\frac{\beta^2}{(\beta+1)(2\beta+1)}w\eta - \frac{\beta^2}{4\beta^2-1}w\pi_T^{(0)}\right]\left(\frac{x}{\beta}\right)^2 . \tag{5.29}$$

If $n \gg 1$, and $-\frac{1}{3} < w < 1$ $(\infty > \beta > \frac{1}{2})$, the ratio of the entropy perturbation part of the decaying mode in ξ to the corresponding part of the growing mode in ξ is $n^{-1}\,[(\beta+1)/(2\beta+1)](x_1/x)^{2\beta+1}$. The same ratio associated with the anisotropic stress is $3(n\beta)^{-1} \times (2\beta+1)(\beta/x_1)^2(x_1/x)^{2\beta+1}$.

While all relative amplitudes on the comoving and uniform-Hubble-constant hypersurfaces are small if excited by small stress perturbations with $w\eta \ll 1$ and $w\pi_T^{(0)} \ll 1$, it is perhaps conceivable that explicit nonlinear terms in the Einstein equations could generate enough growing mode at early times to give a large fractional energy density perturbation at the particle horizon. This possibility will be laid to rest in Sec. VI. Still, the strictly linear perturbation analysis will be shown to fail on uniform-Hubble-constant hypersurfaces at the same point it fails on zero-shear hypersurfaces, when $w\pi_T^{(0)} > x^2$, though only because the linear perturbation in fractional energy density is anomalously small, not because the hypersurface is strongly warped. The nonperturbative formulation of the

Einstein equations in Sec. VI also gives directly the perturbation equations in terms of ϵ_h, v_h, ϕ_h, and α_h.

So far, in discussing the solutions to the perturbation equations I have implicitly assumed that the background equation of state has some relation to that of a fluid, with $0 \leq w \leq 1$, so β is always of order unity. However, it is conceivable that the net pressure could have been negative in the early Universe. One possibility is that at early times the *vacuum* could have had a large energy density and pressure, i.e., a contribution to the energy-momentum tensor like that of a cosmological constant, with pressure equal to minus the energy density.[24] A phase change of the vacuum would be required at some point to reduce the effective cosmological constant to zero or to a small value consistent with the present Universe. Another possibility is that quantum fluctuations of nongravitational and/or gravitational fields could result in an effective negative pressure around the Planck time or before.[25] Zel'dovich *et al.*[16] propose that domain boundaries associated with gauge theories of weak interactions could produce an average $w = -\frac{2}{3}$.

In idealized models with w constant $\dot{S}/S$ is constant or decreasing as $S \to 0$ if $w \leq -\frac{1}{3}$, i.e., if the strong energy condition is violated.[26] Thus as $S \to 0$ the conformal time $\tau \to -\infty$, even though for $w > -1$ the proper time from the initial singularity is finite. If $w \leq -\frac{1}{3}$ at the beginning of the Universe, light signals can propagate arbitrarily far relative to the comoving background coordinates, and particle horizons in the strict sense do not exist. Nevertheless, it makes sense to talk of a perturbation being larger than the *effective* particle horizon while $k^{-1}\dot{S}/S > 1$. It is $k^{-1}\dot{S}/S$, equal to β/x if w is constant, that governs which terms in the perturbation equations are dominant. If $w < -\frac{1}{3}$ initially, $k^{-1}S/\dot{S}$ for a protogalaxy begins less than one, increases to a value much greater than one, and then, once $w > -\frac{1}{3}$, decreases and eventually becomes less then one again at relatively recent times.

A density perturbation established during an early epoch when communication over a wavelength is possible will persist through the period when the perturbation is larger than the effective particle horizon. Assuming a homogeneous solution, the equation

$$(E_0S^2\epsilon_m)^{\cdot\cdot} + 3(1+c_s{}^2)\frac{\dot{S}}{S}(E_0S^2\epsilon_m)^{\cdot} + [c_s{}^2k^2 - 2(1+3c_s{}^2)K + (c_s{}^2 - w)E_0S^2](E_0S^2\epsilon_m) = 0 \tag{5.30}$$

shows that $E_0S^2\epsilon_m$ is roughly constant for the growing mode while $k^{-1}\dot{S}/S \gg 1$. The change in the background equation of state from $w \leq -\frac{1}{3}$ to $w \sim \frac{1}{3}$ changes $E_0S^2\epsilon_m$ by only a factor of 2 or so. Since E_0S^2 increases and decreases with $(\dot{S}/S)^2$, ϵ_m is very small when $k^{-1}\dot{S}/S \gg 1$ relative to its value when $k^{-1}\dot{S}/S \sim 1$. The "growth" of ϵ_m at early times in conventional backgrounds is a purely kinematic consequence of the time dependence of E_0S^2, with no dynamic significance.

B. Vector perturbations

The solution of Eq. (4.13) for the vortical velocity amplitude v_c can be put in the form

$$S^4(E_0+P_0)v_c = -\int_0^{k\tau} dy\, S^4P_0\pi_T^{(0)}, \tag{5.31}$$

assuming no perturbation initially. Since $S^4P_0 \propto y^{2(\beta-1)}$, once the anisotropic stress has been on for more than a few expansion times,

$$v_c \sim -\frac{x}{2\beta-1}\,\frac{w}{1+w}\,\pi_T^{(0)}. \tag{5.32}$$

After the anisotropic stress turns off,

$$v_c \propto [S^4(E_0+P_0)]^{-1} \sim x^{-2(\beta-1)}. \tag{5.33}$$

At the particle horizon v_c is at best roughly comparable to the maximum previous value of $w\pi_T^{(1)}$, and then only if $\pi_T^{(1)}$ is at full strength at the particle horizon or if β is close to $\frac{1}{2}$.

The "frame-dragging potential" Ψ is

$$\Psi = 2(k^2-2K)^{-1}S^2(E_0+P_0)v_c \sim 4\beta(\beta+1)x^{-2}v_c. \tag{5.34}$$

Even if $w\pi_T^{(1)} \ll 1$ and $v_c \ll 1$ at all times, $\Psi \gtrsim 1$ is quite possible if the anisotropic stress is present when $kS/\dot{S} \lesssim w\pi_T^{(1)}$. While $\Psi > 1$ does mean that there is no gauge in which all the metric perturbations are small, there is a formulation of the Einstein equations (see Sec. VI) in which the equations remain approximately linear. The physical significance is that, as in the ergoregion around a rotating black hole, a timelike observer cannot be at rest relative to nonshearing spatial coordinates (with $\dot{H}_T^{(1)} = 0$).

C. Tensor perturbations

The homogeneous solution of Eq. (4.14) for tensor perturbations is well known (see Ref. 15), and the inhomogeneous solution raises no new physical questions. The gravitational wave amplitude $H_T^{(2)}$ can never exceed in order of magnitude the maximum previous value of $w\pi_T^{(2)}$, if $H_T^{(2)}$ vanishes at the initial singularity, as long as $E_0S^2/k^2 \gg 1$.

VI. RELATIONSHIP TO NONLINEAR DYNAMICS

The standard Arnowitt-Deser-Misner (ADM) approach[27] to the nonlinear dynamics of the gravitational field, with or without the presence of matter, begins with the initial-value problem[28] on a spacelike hypersurface characterized by an *intrinsic metric tensor* $h_{\alpha\beta}$ and an *extrinsic curvature tensor* $\mathcal{K}_{\alpha\beta}$. The proper-time spacing between successive hypersurfaces is characterized by the *lapse function* N, a spatial scalar determined from the geometric condition specifying the spacelike hypersurfaces by, typically, an elliptic equation.[29] The dynamic equations evolve $\mathcal{K}_{\alpha\beta}$ and $h_{\alpha\beta}$ forward in time from one hypersurface to the next. It is instructive to consider the cosmological perturbation problem from this point of view, to understand better the physical meaning of the gauge-invariant variables, to derive the equations governing the amplitudes ϕ_h, α_h, ϵ_h, and v_h appropriate to uniform-Hubble-constant hypersurfaces, and to investigate the possible effect of nonlinearities on the development of the perturbations.

Let each hypersurface be labeled by a single value of the time coordinate τ, and let n_a be the unit future-directed four-vector normal to the hypersurface. Then

$$n_0=-N,\ n_\alpha=0,\ n^0=N^{-1},\ n^\alpha=N^{-1}N^\alpha . \tag{6.1}$$

The *shift vector* N^α, the coordinate three-velocity of the normal world line, describes how spatial coordinates are propagated from one hypersurface to the next. In this section a three-vector is a three-vector relative to the exact spatial metric $h_{\alpha\beta}$. The metric tensor of the spacetime g_{ab} is given by

$$g_{00}=-N^2+N^\alpha N_\alpha,\quad g_{0\alpha}=-N_\alpha,\quad g_{\alpha\beta}=h_{\alpha\beta}, \tag{6.2}$$

and the inverse is

$$g_{00}=-N^{-2},\quad g^{0\alpha}=-N^{-2}N^\alpha,\quad g^{\alpha\beta}=h^{\alpha\beta}-N^{-2}N^\alpha N^\beta. \tag{6.3}$$

Note $h^{\alpha\beta}$ is the inverse of $h_{\alpha\beta}$ and $N_\alpha \equiv h_{\alpha\beta}N^\beta$.

From the three-metric $h_{\alpha\beta}$ one calculates the spatial Ricci tensor $\mathcal{R}_{\alpha\beta}$ and the scalar intrinsic curvature $\mathcal{R}\equiv h^{\alpha\beta}\mathcal{R}_{\alpha\beta}$. The extrinsic curvature tensor describes the embedding of the hypersurface in the spacetime and is the natural focus of a geometrical condition picking out a particular family of hypersurfaces. Mathematically, it is given by[27]

$$\begin{aligned}\mathcal{K}_{\alpha\beta}&=-n_{\alpha;\beta}=-N\Gamma^0_{\alpha\beta}\\&=-\tfrac{1}{2}N^{-1}[h_{\alpha\beta,0}+N_{\alpha|\beta}+N_{\beta|\alpha}].\end{aligned} \tag{6.4}$$

The semicolon denotes a covariant derivative in the spacetime, the slash a covariant derivative (with respect to $h_{\alpha\beta}$) in the spacelike hypersurface. The comma indicates an ordinary partial derivative.

The correspondence with my representation of the cosmological perturbations is made more transparent by defining the *conformal metric tensor* $\tilde{h}_{\alpha\beta}$,

$$\tilde{h}_{\alpha\beta}\equiv h^{-1/3}h_{\alpha\beta}, \tag{6.5}$$

where h is the determinant of $h_{\alpha\beta}$. The background scale factor is not present in $\tilde{h}_{\alpha\beta}$, and the perturbation in $\tilde{h}_{\alpha\beta}$ is entirely due to the traceless part of the metric tensor perturbation. In a similar spirit, split $\mathcal{K}_{\alpha\beta}$ into the trace $\mathcal{K}$ and the traceless part

$$\bar{\mathcal{K}}_{\alpha\beta}\equiv\mathcal{K}_{\alpha\beta}-\tfrac{1}{3}h_{\alpha\beta}\mathcal{K}. \tag{6.6}$$

Equation (6.4) then becomes the two equations

$$h_{,0}/h=-2N\mathcal{K}-2N^\alpha{}_{|\alpha} \tag{6.7}$$

and

$$\tilde{h}_{\alpha\beta,0}+\tilde{h}_{\alpha\beta,\mu}N^\mu+\tilde{h}_{\alpha\beta}N^\mu{}_{,\beta}+\tilde{h}_{\mu\beta}N^\mu{}_{,\alpha}-\tfrac{2}{3}\tilde{h}_{\alpha\beta}N^\mu{}_{,\mu}=-2Nh^{-1/3}\bar{\mathcal{K}}_{\alpha\beta}, \tag{6.8}$$

which correspond to decomposition of $\mathcal{K}^\alpha_\beta$ in terms of perturbation amplitudes given in Eq. (A7).

The Einstein equations separate into the initial-value equations relating the extrinsic and intrinsic geometry of the constant-τ hypersurface to the matter energy density

$$\mathcal{E}\equiv n_aT^{ab}n_b=N^2T^{00} \tag{6.9}$$

and the momentum density

$$\mathcal{J}_\alpha\equiv-n_aT^a_{\ \alpha}=NT^0_{\ \alpha}, \tag{6.10}$$

and into the dynamic equations for the evolution of the extrinsic curvature in time.[29] The momentum initial-value equation

$$\bar{\mathcal{K}}^\beta_{\alpha|\beta}=\mathcal{J}_\alpha+\tfrac{2}{3}\mathcal{K}_{|\alpha} \tag{6.11}$$

is conventionally viewed[28] as an equation for the longitudinal part of $\bar{\mathcal{K}}^\beta_\alpha$ and the energy equation

$$\mathcal{R}=2\mathcal{E}+\bar{\mathcal{K}}_{\alpha\beta}\bar{\mathcal{K}}^{\alpha\beta}-\tfrac{2}{3}\mathcal{K}^2 \tag{6.12}$$

as an equation for the determinant of the spatial metric tensor h, given the conformal metric $\tilde{h}_{\alpha\beta}$, the transverse part of $h^{1/2}\bar{\mathcal{K}}^\beta_\alpha$ relative to $\tilde{h}_{\alpha\beta}$, and, as a hypersurface condition, $\mathcal{K}$.

The dynamic equations[29] involve the matter stress tensor

$$\mathcal{S}_{\alpha\beta}\equiv T_{\alpha\beta}. \tag{6.13}$$

Let the traceless part of $\mathcal{S}_{\alpha\beta}$ be $\bar{\mathcal{S}}_{\alpha\beta}$, and denote the trace by $\mathcal{S}$. Let $\bar{\mathcal{R}}_{\alpha\beta}$ be the traceless part of the spatial Ricci tensor. Then

$$\mathcal{K}_{,0}+N^\alpha\mathcal{K}_{,\alpha}=-\Delta N+N(\mathcal{R}+\mathcal{K}^2+\tfrac{1}{2}\mathcal{S}-\tfrac{3}{2}\mathcal{E}) \tag{6.14}$$

and

$$\bar{\mathcal{K}}^{\alpha}_{\beta,0}+N^{\mu}\bar{\mathcal{K}}^{\alpha}_{\beta,\mu}-N^{\alpha}_{,\mu}\bar{\mathcal{K}}^{\mu}_{\beta}+N^{\mu}_{,\beta}\mathcal{K}^{\alpha}_{\mu}$$
$$=-N^{|\alpha}{}_{|\beta}+\tfrac{1}{3}\delta^{\alpha}_{\beta}\Delta N+N(\bar{\mathcal{R}}^{\alpha}_{\beta}+\mathcal{K}\bar{\mathcal{K}}^{\alpha}_{\beta}-\bar{\mathcal{S}}^{\alpha}_{\beta}). \quad (6.15)$$

In both equations the left-hand side is the time derivative along the normal to the hypersurface, and Δ is the Laplacian operator in the hypersurface.

The physical dynamics of scalar and vector cosmological perturbations lies in the matter evolution equations, rather then Eqs. (6.14) and (6.15). The equation governing local energy conservation, $T^{0a}{}_{;a}=0$, gives the rate of change of the energy density $\mathcal{E}$ along the normal to the hypersurface,

$$\mathcal{E}_{,0}+N^{\alpha}\mathcal{E}_{,\alpha}=N\mathcal{K}(\mathcal{E}+\tfrac{1}{3}\mathcal{S})+N\bar{\mathcal{K}}^{\alpha\beta}\bar{\mathcal{S}}_{\alpha\beta}+N^{-1}(N^2\mathcal{J}^{\alpha})_{|\alpha}. \quad (6.16)$$

Similarly, the evolution of the momentum density $\mathcal{J}_{\alpha}$ is given by $T^{b}_{\alpha;b}=0$, or

$$\mathcal{J}_{\alpha,0}+N^{\beta}\mathcal{J}_{\alpha,\beta}+N^{\beta}_{,\alpha}\mathcal{J}_{\beta}=N\mathcal{K}\mathcal{J}_{\alpha}-(\mathcal{E}\delta^{\beta}_{\alpha}+\mathcal{S}^{\beta}_{\alpha})N_{|\beta}-N\mathcal{S}^{\beta}_{\alpha|\beta}. \quad (6.17)$$

Although redundant, it will prove useful to have an equation directly for the evolution of the spatial scalar curvature. Take a convective time derivative of Eq. (6.12) along the normal to the hypersurface and eliminate the time derivatives of $\mathcal{K}$, $\bar{\mathcal{K}}^{\alpha}_{\beta}$, and $\mathcal{E}$ with Eqs. (6.14)–(6.16). After simplification with the help of Eq. (6.12) again, the result is

$$\mathcal{R}_{,0}+N^{\alpha}\mathcal{R}_{,\alpha}=\tfrac{2}{3}N\mathcal{K}\mathcal{R}+2N^{-1}(N^2\mathcal{J}^{\alpha})_{|\alpha}+\tfrac{4}{3}\mathcal{K}\Delta N+2\bar{\mathcal{K}}^{\alpha}_{\beta}(N\bar{\mathcal{R}}^{\beta}_{\alpha}-N^{|\beta}{}_{|\alpha}). \quad (6.18)$$

There is considerable cancellation between the time derivatives of the separate terms on the right-hand side of Eq. (6.12).

The solution of the above equations requires a hypersurface condition to determine N and, of secondary importance in the present circumstances, some prescription for N^{α}. The most straightforward hypersurface condition in the nonlinear context is the uniform-Hubble-constant condition that $\mathcal{K}$ be spatially uniform on each hypersurface.[29] The choice of the time dependence of $\mathcal{K}$ is the choice of time coordinate τ; at least in spatially closed cosmologies each value of $\mathcal{K}$ picks out a more or less unique hypersurface in the spacetime. The value of $\mathcal{K}$ may also pick out a unique hypersurface in spatially open cosmologies.[30] On uniform-Hubble-constant hypersurfaces, then $\mathcal{K}_{,\alpha}=0$, $\mathcal{K}_{,0}$ can be specified, and Eq. (6.14) becomes an elliptic equation for N whose mathematical properties, including existence and uniqueness theorems, have been explored rather extensively (see Ref. 30).

The zero-shear hypersurface condition, in the present language $\bar{\mathcal{K}}_{\alpha\beta}=0$, is applicable only to scalar perturbations, since it is compatible with Eq. (6.11) only if $\mathcal{J}_{\alpha}$ is the gradient of a scalar. The obvious generalization, which I will call the *minimal shear hypersurface condition*, is

$$\bar{\mathcal{K}}^{\alpha\beta}{}_{|\alpha\beta}=0. \quad (6.19)$$

The longitudinal part of Eq. (6.11) becomes an equation for $\mathcal{K}$,

$$\Delta\mathcal{K}=-\tfrac{3}{2}\mathcal{J}^{\alpha}{}_{|\alpha}. \quad (6.20)$$

Then the right-hand side of Eq. (6.11) is the transverse part of the vector field $\mathcal{J}_{\alpha}$. An equation for N is more difficult to come by. The best procedure seems to be to combine a time derivative of Eq. (6.20) with a Laplacian of Eq. (6.14). Eliminate $\Delta\dot{\mathcal{K}}$ between the two equations and eliminate $\dot{\mathcal{J}}_{\alpha}$ using Eq. (6.17). The result is a complicated fourth-order equation for N. Acceptable solutions may not always exist in an open universe if highly nonlinear regions such as black holes are present, and may not be unique in closed universes. The linear scalar perturbation with $k^2=3K$ satisfies the minimal-shear condition for any amplitude of warping of the hypersurface since, as mentioned in Sec. IV, $Q^{(0)}_{\alpha\beta}$ vanishes identically.

The complexity of the minimal-shear hypersurface condition magically disappears for linear scalar perturbations. Then, with $\bar{\mathcal{K}}_{\alpha\beta}=0$, Eq. (6.15) reduces to Eq. (4.4) for the perturbation in N. Also, note that Eq. (6.12) reduces to Eq. (4.3) relating Φ_H to ϵ_m after the perturbation in $\mathcal{K}$ is eliminated in favor of $\mathcal{J}_{\alpha}$ through Eq. (6.20). The simplicity of Eq. (4.3) is somewhat deceptive.

The comoving hypersurface condition generalizes from $\mathcal{J}_{\alpha}=0$ for linear scalar perturbations to

$$\mathcal{J}^{\alpha}{}_{|\alpha}=0. \quad (6.21)$$

The equation for N, obtained by taking a divergence of Eq. (6.17), is only second order, simpler than in the minimal-shear case, but is still considerably more complicated than for uniform-Hubble-constant hypersurfaces.

The uniform-Hubble-constant hypersurface condition will be used in our analysis of nonlinear effects on the time development of perturbations. Somewhat different questions arise in the consideration of nonlinearities associated with scalar and vector perturbations, so these will be considered separately.

A. Scalar perturbations

The gauge-invariant scalar perturbation amplitudes appropriate to uniform-Hubble-constant hypersurfaces were defined in Eqs. (5.24)–(5.27). The perturbation equations directly in terms of these amplitudes are most easily derived from the equations of this section, rather than by reworking the equations of Sec. IV, and their derivation will clarify the later consideration of nonlinear effects.

On a uniform-Hubble-constant hypersurface $\phi_h = \tilde{H}_L + \frac{1}{3}\tilde{H}_T^{(0)}$ and $\alpha_h = \tilde{A}$. The hypersurface condition, Eq. (5.22), gives a gauge-invariant amplitude for the traceless part of the extrinsic curvature tensor

$$\bar{\mathcal{K}}^{\alpha}_{\beta} = -\frac{1}{S}(\dot{\tilde{H}}_T^{(0)} - k\tilde{B}^{(0)})Q^{(0)\alpha}_{\beta}$$
$$= -\frac{3}{S}\left(\dot{\phi}_h - \frac{\dot{S}}{S}\alpha_h\right)Q^{(0)\alpha}_{\beta}. \qquad (6.22)$$

The two initial-value equations (6.11) and (6.12) become

$$-\frac{2}{S}k(1-3K/k^2)\left(\dot{\phi}_h - \frac{\dot{S}}{S}\alpha_h\right) = S(E_0+P_0)v_h \qquad (6.23)$$

and

$$4S^{-2}(k^2-3K)\phi_h = 2E_0\epsilon_h. \qquad (6.24)$$

Equations (6.14) and (6.24) combine to determine α_h in terms of ϕ_h and η,

$$[k^2 - 3K + \tfrac{3}{2}(E_0+P_0)S^2]\alpha_h$$
$$= -(1+3c_s^{\,2})(k^2-3K)\phi_h - \tfrac{3}{2}P_0S^2\eta. \qquad (6.25)$$

The geometric evolution equation (6.15) reduces to

$$\left(\dot{\phi}_h - \frac{\dot{S}}{S}\alpha_h\right)^{\cdot} + 2\frac{\dot{S}}{S}\left(\dot{\phi}_h - \frac{\dot{S}}{S}\alpha_h\right) - \tfrac{1}{3}k^2(\phi_h+\alpha_h) = \tfrac{1}{3}P_0S^2\pi_T^{(0)}. \qquad (6.26)$$

Elimination of α_h with Eq. (6.25) gives a single equation in a single unknown ϕ_h, but the equation is much more complicated than Eq. (4.9) for ϵ_m. The solution is also not nearly as simple a function of time as the solution for ϵ_m/Φ_H given in Eq. (5.14); consider Eqs. (5.24), (5.19), and (5.6).

The uniform-Hubble-constant hypersurfaces do deal successfully with the spatial harmonics $k^2 = 3K$ in a closed universe. Since $Q^{(0)}_{\alpha\beta}$ vanishes identically,[2] Eq. (6.23) no longer applies. Equation (6.24) gives $\epsilon_h = 0$, and from Eqs. (6.25) and (6.17)

$$\alpha_h = -\frac{w}{1+w}\eta\,, \quad [S^4(E_0+P_0)v_h]^{\cdot} = 0. \qquad (6.27)$$

The amplitude ϕ_h now depends on the way spatial coordinates are propagated from one hypersurface to the next through the hypersurface condition Eq. (5.22). The traceless part of the metric tensor perturbation and the spatial curvature perturbation vanish. The absence of any physical *adiabatic* mode when $k^2 = 3K$ was first recognized by Lifshitz and Khalatnikov.[2]

A relatively simple direct solution of Eqs. (6.23)–(6.26) is possible under the assumptions of Sec. V when the perturbation is large compared with the particle horizon, $x/\beta \ll 1$. To lowest order,

$$\epsilon_h \simeq \tfrac{2}{3}\beta^{-2}x^2\phi_h\,, \qquad (6.28)$$

$$\alpha_h \simeq -\frac{w}{1+w}\eta\,, \qquad (6.29)$$

and with

$$g \equiv \phi_h' - (\beta/x)\alpha_h \simeq \phi_h' + \frac{\beta w}{1+w}x^{-1}\eta\,, \qquad (6.30)$$

Eq. (6.26) becomes

$$g' + 2(\beta/x)g \simeq (\beta/x)^2 w\pi_T^{(0)}. \qquad (6.31)$$

With no perturbation at $x=0$, the solution is

$$g = \beta^2 x^{-2\beta}\int_0^x w\pi_T^{(0)}y^{2\beta-2}dy\,, \qquad (6.32)$$

$$\phi_h = \int_0^x g(y)dy - \frac{\beta}{1+w}\int_0^x w\eta y^{-1}dy\,. \qquad (6.33)$$

An entropy perturbation present only at $y \ll 1$ couples predominantly to the growing mode of energy density perturbation on uniform-Hubble-constant hypersurfaces, while anisotropic stress at $y \ll 1$ couples with comparable strength to both modes. The discussion in Sec. V showed rather different behavior for the energy density perturbation on comoving or zero-shear hypersurfaces. These and the hypersurface-independent ratio of shear rate to expansion rate displayed a coupling of the entropy perturbation to both growing and decaying modes and a coupling of the anisotropic stress predominantly to the decaying mode at early times. The situation is not as simple as presented by Press and Vishniac.[12]

The linear perturbations relative to the uniform-Hubble-constant hypersurfaces are small as long as $w\eta \ll 1$ and $w\pi_T^{(0)} \ll 1$. If the stress perturbations are present for at least one expansion time at $x \ll 1$, Eq. (6.33) gives $\phi_h \sim w\eta,\ w\pi_T^{(0)}$; Eq. (6.28) gives $\epsilon_h \sim x^2 w\eta,\ x^2 w\pi_T^{(0)}$; Eq. (6.25) gives $\alpha_h \sim w\eta,\ x^2 w\pi_T^{(0)}$; and from Eqs. (6.23) and (6.32) $v_h \sim x^2 g \sim x^3 w\eta,\ xw\pi_T^{(0)}$. Nevertheless, some second-order terms in the Einstein equations are larger than the linear perturbations whenever $w\pi_T^{(0)} > x^2$. Consider, for instance, the initial value Eq. (6.12). The linear perturbation in $\mathcal{R}$ is of order $(k^2/S^2)\phi_h \sim (k/S)^2[w\eta,\ w\pi_T^{(0)}]$, as is the linear perturbation in $\mathcal{E}$. The term $\bar{\mathcal{K}}_{\alpha\beta}\bar{\mathcal{K}}^{\alpha\beta}$, since $\bar{\mathcal{K}}^{\alpha}_{\beta} \sim (k/S)g \sim (k/S)[xw\eta,\ x^{-1}w\pi_T^{(0)}]$, is of order $(k/S)^2[x^2(w\eta)^2,\ x^{-2}(w\pi_T^{(0)})^2]$ times the first order perturbations. When $1 \gg w\pi_T^{(0)} > x^2$ the second-order term is not large in an absolute sense, since $\overline{\mathcal{K}}_{\alpha\beta}\overline{\mathcal{K}}^{\alpha\beta}/\mathcal{K}^2$ is of order $(w\pi_T^{(0)})^2$; the first-order perturbation in $\mathcal{E}$ is anomalously small. The key physical question is whether this breakdown of linearity on uniform-Hubble-constant hypersurfaces has an effect on the time development of the energy density perturbation, so that the fractional energy density perturbation as the perturbation enters the particle horizon is not predicted correctly by the linear theory.

A partial answer is that on *comoving* hypersurfaces the initial second-order perturbation is small compared with the corresponding first-order perturbation as long as $w\pi_T^{(0)} \ll 1$, independent of the value of x. The term $\overline{\mathcal{K}}_{\alpha\beta}\overline{\mathcal{K}}^{\alpha\beta}$ in Eq. (6.12), now the square of the shear of the matter, is still of order $(k/S)^2 x^{-2}(w\pi_T^{(0)})^2$. However, the linear perturbations in $\mathcal{E}$ and $\mathcal{K}^2$ are now of order $(k/S)^2 x^{-2}(w\pi_T^{(0)})^2$.

Can nonlinear effects couple the decaying mode of fractional energy density perturbation on comoving hypersurfaces to the growing mode so that at the particle horizon the fractional energy density is large compared with the original value of $w\pi_T^{(0)}$? At the particle horizon the fractional energy density perturbation is comparable to the distortion of the spacelike hypersurface, which is $(S/k)^2$ times the perturbation in scalar curvature $\mathcal{R}$. The time development of $\mathcal{R}$ is given by Eq. (6.18). From Eq. (6.14) the fractional perturbation in N is at most the order of the fractional perturbation in $\mathcal{E}$ or $\mathcal{K}$, which is the order of $w\pi_T^{(0)}$. The explicit second-order terms in Eq. (6.18) are of order $\tau^{-1}(k/S)^2(w\pi_T^{(0)})^2$, down by a factor of order $w\pi_T^{(0)}$ relative to the first-order terms. The nonlinearities are unable to alter appreciably the linear theory prediction for $\mathcal{R}$ as long as $w\pi_T^{(0)} \ll 1$.

The linear perturbation equations are really the same equations, but with variables regrouped, no matter what hypersurface condition is used to interpret the variables physically. The validity of the linear equations on the comoving hypersurfaces implies their validity on all sets of hypersurfaces as long as the physical interpretation of the linear perturbation amplitudes is qualified appropriately. For instance, if $w\pi_T^{(0)} > x^2$ in a linear perturbation calculation on uniform-Hubble-constant hypersurfaces (or zero-shear hypersurfaces), ϵ_h (or ϵ_g) is at first not the amplitude of the actual fractional energy density perturbation, but by the time the perturbation comes inside the particle horizon the discrepancy has disappeared. The dynamical origin of the nonlinear corrections to the fractional energy density perturbation is the work done by the anisotropic stress on the shearing volume element [see Eq. (6.16)].

B. Vector perturbations

The amplitude Ψ of the frame-dragging potential is a gauge-invariant measure of the amplitude of metric tensor perturbations associated with vortical motions of the matter. It is generated by anisotropic stress through Eqs. (4.12) and (4.13). In a standard background $\Psi \sim x^{-1}(w\pi_T^{(1)})$ from Eqs. (5.32) and (5.34). The possibility that $\Psi \gtrsim 1$ even if $w\pi_T^{(1)} \ll 1$ at all times was pointed out in Sec. V. The physical significance of this can best be understood in the context of the full Einstein equations as presented in this section.

One can verify that $w\pi_T^{(1)} \ll 1$ at all times does guarantee the validity of the linear equations in a gauge such that $\dot{H}_T^{(1)}$ vanishes. Then $\Psi = B^{(1)}$, and

$$N^\alpha = \Psi Q^{(1)\alpha} \quad . \tag{6.34}$$

The lapse function N, as a spatial scalar, is unperturbed to first order, and from Eq. (6.2)

$$g_{00} \simeq -S^2(1 - \Psi^2 Q^{(1)\alpha} Q_\alpha^{(1)}) \,. \tag{6.35}$$

Indices on the spatial harmonics are always raised and lowered with the background metric.

If $\Psi > 1$ there are likely, depending on the precise normalization of $Q_\alpha^{(1)}$, to be regions where a physical observer or particle cannot be at rest relative to nonshearing spatial coordinates, analogous to an ergoregion in a stationary spacetime. However, g_{00} as such does not appear in the nonlinear equations (6.11)–(6.18), and these equations are not in any way singular when $g_{00} \geqslant 0$. That the second-order perturbation in N is small on uniform-Hubble-constant hypersurfaces can be verified from Eq. (6.14).

The dynamically significant nonlinearities arise in the same way as in the scalar case. From Eq. (A7)

$$\overline{\mathcal{K}}{}^\alpha_\beta = (k/S)\Psi Q_\beta^{(1)\alpha} \sim (k/S)x^{-1} w\pi_T^{(1)} \,, \tag{6.36}$$

which in Eq. (6.16) implies a nonlinear contribution to the fractional energy density perturbation of order $(w\pi_T^{(1)})^2$. Again, this belongs predominantly to the decaying mode at $x \ll 1$, and Eq. (6.18) shows that the nonlinear correction to $\mathcal{R}$ is small compared to the linear scalar perturbation for mixed scalar-vector perturbations. If the only linear perturbations are pure vector harmonics, the dominant nonlinearities in $\mathcal{R}$ come from the $\frac{4}{3}\mathcal{K}\Delta N$ term in Eq. (6.18), since $\overline{\mathcal{R}}{}^\alpha_\beta$ vanishes to first order for vector perturbations, and give $\phi_h \sim (S/k)^2\mathcal{R} \sim (w\pi_T^{(1)})^2$ for the induced scalar perturbation. This is roughly what one expects for the fractional energy density perturbation at $x \sim 1$.

Even though N^{α} may be large at $x \ll 1$, the convective terms involving N^{α} and a spatial gradient are of order $x\Psi \sim w\pi_T^{(1)}$ times the time-derivative term in Eqs. (6.14)–(6.18) as the perturbation is being generated.

C. Tensor perturbations

The nonlinear perturbation of the energy density follows the same pattern for tensor perturbations as for vector and scalar perturbations. If the completely transverse anisotropic stress is on for the order of one expansion time, Eq. (4.14) integrates to give $H_T^{(2)\prime} \sim x^{-1} w \pi_T^{(2)}$, so

$$\overline{\mathcal{K}}^{\alpha}_{\beta} \sim (k/S) x^{-1} w \pi_T^{(2)} \,. \qquad (6.37)$$

The potentially largest nonlinear contribution to the fractional energy density perturbation is the order of the square of the ratio of anisotropic stress to background energy density, both when the anisotropic stress is present (presumably at $x \ll 1$) and at the particle horizon, regardless of whether the anisotropic stress is associated with scalar, vector, or tensor harmonics or a mixture of all three.

VII. SUMMARY AND CONCLUSIONS

The primary gauge-invariant amplitudes for scalar (density) perturbations were chosen in Sec. III on the grounds of mathematical simplicity, based on a description of the geometry through the metric tensor and a description of the matter through the energy-momentum tensor. Three of these amplitudes have, to first order in the deviation from the homogeneous and isotropic background, a universal physical significance. The amplitude $v_s^{(0)}$, or $\xi \equiv (k/S) v_s^{(0)}$, is a measure of the shear of the curl-free part of matter velocity field. The amplitude η is a measure of the nonadiabatic, relative to the background pressure-energy-density relation, part of the perturbation in the isotropic pressure, while $\pi_T^{(0)}$ measures the completely longitudinal part of the anisotropic stress. In each case the physical quantity vanishes in the background.

On the other hand, the amplitude ϵ_m, while mathematically gauge invariant, has the physical significance of measuring the fractional energy density perturbation only on spacelike hypersurfaces orthogonal to the four-velocity of the frames in which the matter energy flux vanishes. The purely geometrical amplitudes Φ_H and Φ_A also have physical meaning with respect to a particular set of spacelike hypersurfaces, the hypersurfaces for which the normal unit vectors have zero shear. Specifically, Φ_H measures the amount of warping of the zero-shear hypersurfaces due to the presence of the perturbation, and Φ_A is the amplitude of the fractional perturbation in the lapse function.

The mathematical simplicity of the definition of these amplitudes carries over into the simple quasi-Newtonian structure of the equations which govern them. The nonadiabatic stress amplitudes η and $\pi_T^{(0)}$ are formally regarded as known functions of time which through the source terms in Eq. (4.9) for ϵ_m generate the perturbations in energy density shear, and spatial curvature. A more detailed treatment of the matter would require additional equations governing the evolution of individual components such as the microwave background radiation, neutrinos, and at very early times quarks and various gauge bosons, particularly if these components are thermally decoupled. It is straightforward to write these additional equations in gauge-invariant form, and presumably a complete theory of the matter would determine the time dependence of η and $\pi_T^{(0)}$, rather than leaving them as free functions. In the optically thick limit, $\eta Q^{(0)}$ is roughly the fractional perturbation in the photon-to-baryon ratio times the ratio of rest-mass energy density to total energy density in a standard radiation-plus-matter model.

The gauge-invariant variables and equations are closely related to the variables and perturbation equations in certain specific gauges. For instance, Φ_H and Φ_A are the metric perturbation amplitudes in the longitudinal gauge of Harrison, and ϵ_m is the density perturbation amplitude in the comoving gauge of Sakai.[11] Both Harrison[4] and Sakai[11] arrived at equations equivalent to the homogeneous version of Eq. (4.9). Unfortunately, most of the standard references, including the textbooks of Weinberg[15] and Peebles,[7] rely on the synchronous gauge. This gauge is unnecessarily complicated mathematically, since the presence of purely gauge modes means that the analog of Eq. (4.9) is a fourth-order equation, and the physical interpretation of the results is not straightforward. Of course, a calculation can be done in any gauge if done consistently and completely, and in practice difficulties of interpretation disappear once the perturbation is well within the particle horizon.

The physical interpretation of such hypersurface-dependent quantities as the fractional energy density perturbation is somewhat ambiguous even in the gauge-invariant formalism when the perturbation wavelength is larger than the particle horizon. In Secs. III and V I define in addition to ϵ_m gauge-invariant amplitudes ϵ_g and ϵ_h which measure the fractional energy density perturbation on zero-shear and uniform-Hubble-constant hypersurfaces, respectively. Which is considered the physically most appropriate is a matter of taste. A calculation done in a particular gauge or with a particular set of gauge-invariant variables can always be physically interpreted by computing the gauge-invariant amplitude appropriate to the preferred hypersurface.

The same is true of the geometrical amplitudes. While $\phi_g \equiv \Phi_H$ measures the distortion of zero-shear hypersurfaces and $\alpha_g \equiv \Phi_A$ the perturbation in the lapse function between zero-shear hypersurfaces, appropriate combinations of Φ_H and Φ_A as defined in Sec. V give the amplitudes of the distortion and the perturbation in the lapse function for comoving (ϕ_m and α_m) and uniform-Hubble-constant (ϕ_h and α_h) hypersurfaces.

The comparison between the values of the gauge-invariant amplitudes appropriate to these three hypersurfaces for each of the two independent homogeneous modes in a background with negligible spatial curvature and $w = P_0/E_0$ independent of time is given in Table I. The fractional energy density perturbation is particularly sensitive to the choice of hypersurface when $kS/\dot{S} = k\tau/\beta \ll 1$.

The hypersurface conditions discussed in this paper have the property that *each* hypersurface is picked out by a physical criterion based either on the extrinsic geometry of the hypersurface or on its relation to the matter. The geometric criteria are spatially global in character, since the lapse function relating one hypersurface to the next satisfies an elliptic equation. Even the comoving hypersurface condition is fundamentally global, since the separation of the longitudinal and vortical parts of the velocity field also requires the solution of an elliptic equation.

In contrast, the hypersurfaces might be chosen as surfaces of constant *proper* time along the world lines of a particular family of observers. If these observers are comoving with the matter one has the sort of hypersurface condition used by Sachs and Wolfe[5] and Olson[13] to define a density perturbation, even though Olson's dynamical calculation was performed with a variable (essentially equivalent to ϕ_m) defined on hypersurfaces orthogonal to the matter world lines. In a synchronous gauge the observers are freely falling and their world lines are orthogonal to the hypersurfaces of constant propertime.

The problem with these proper-time hypersurface conditions is that at any *one* time the choice of the hypersurface is completely arbitrary. There is a corresponding arbitrariness in the value of, say, the fractional energy density perturbation, and the description is intrinsically non-gauge-invariant. Press and Vishniac[12] avoid this problem, even though they calculate in a synchronous gauge, by continuously transforming to a uniform-Hubble-constant hypersurface to interpret their results. However, it would seem better to work directly with variables whose physical meaning is unambiguous.

A proper-time description may be useful in considering local physics in the presence of given inhomogeneities, e.g., in studying element formation in the early Universe. However, it is in appropriate when the dynamics of the inhomogeneities themselves are the focus of interest.

Is there any one best measure of the true amplitude of the perturbation? Once the perturbation is within the particle horizon the value of the fractional energy density perturbation on any of the standard hypersurfaces is a reasonable choice. At early times, when $kS/\dot{S} \ll 1$, the true amplitude should indicate the limits of validity of the linear perturbation analysis for both growing and decaying modes. All of the amplitudes compared in Table I are relative amplitudes in the sense some sort of breakdown of linearity is implied if one of them is larger than unity. A "good" hypersurface condition should not introduce any apparent nonlinearities which are *only* due to a large warping of the hypersurface. In this sense the zero-shear hypersurface condition is "bad," since for a decaying mode ϕ_g and α_g can exceed one even though all the relative amplitudes are small on comoving and uniform-Hubble-constant hypersurfaces.

On no hypersurface does the fractional energy density perturbation indicate the true amplitude for both homogeneous scalar modes at $kS/\dot{S} \ll 1$. The one hypersurface-invariant amplitude ξ is unsuitable because ξ is down by a factor $(kS/\dot{S})^2$ compared with the irreducible relative amplitude of the geometry perturbations for the growing mode. For the decaying mode ϕ_m is anomalously small compared with, say, ξ when $P_0/E_0 \ll 1$ (β close to 2). The one generally suitable choice is ϕ_h, the amplitude of warping of the uniform-Hubble-constant hypersurface. Both comoving and uniform-Hubble-constant hypersurfaces minimize (within a factor of order unity) the maximum relative amplitude, and ϕ_h is always comparable with, if not identical to, this maximum relative amplitude.

A further advantage of ϕ_h can be seen from Eq. (6.26). If no nonadiabatic stress perturbations are present ($w\eta = w\pi_T^{(0)} = 0$) at a time when $kS/\dot{S} \ll 1$, Eq. (6.25) shows that $\alpha_h = O[(kS/\dot{S})^2\phi_h]$, and Eq. (6.26) reduces to

$$\ddot{\phi}_h + 2\frac{\dot{S}}{S}\dot{\phi}_h - \tfrac{1}{3}k^2\phi_h \simeq 0 . \qquad (7.1)$$

In a time $\Delta\tau \ll k^{-1}$, so light can travel only a small fraction of a wavelength, the growing mode in ϕ_h

is to a good approximation *constant* in amplitude *regardless of any change in the background equation of state*, even a change from $P_0/E_0 < -\frac{1}{3}$ to $P_0/E_0 > 0$ as discussed in connection with Eq. (5.30). Meanwhile, the fractional energy density perturbation as measured by ϵ_h or ϵ_m varies inversely with E_0S^2 or $(\dot{S}/S)^2$ and can increase or decrease by many orders of magnitude. It is ϕ_h, rather than ϵ_m or ϵ_h, that gives the correct physical picture of the dynamics of the perturbation while the perturbation is large compared with the effective particle horizon. The perturbation amplitude can decay adiabatically but cannot grow except in *direct* response to a nonadiabatic stress perturbation. This conclusion may seem an obvious consequence of causality, but is obscured in many standard references which emphasize power-law growth of the fractional energy density perturbation at early times.

While the above considerations give preference to the uniform-Hubble-constant hypersurface for dealing with the nonlinear Einstein equations, we did point out in Sec. VI that the linear perturbation treatment is strictly valid to the fullest possible extent only on comoving hypersurfaces. On uniform-Hubble-constant hypersurfaces the second-order correction to the fractional energy density perturbation can exceed the linear term at $kS/\dot{S} \ll 1$ even though the overall perturbation amplitude is small. In this circumstance $\epsilon_h Q$ is not equal to the actual perturbation in the fractional energy density on uniform-Hubble-constant hypersurfaces, but the gauge-invariant linear perturbation equations still give the correct evolution and by the time $kS/\dot{S} \sim 1$, the fractional energy density perturbation is equal to $\epsilon_h Q$.

The exact inhomogeneous solutions to the perturbation equations obtained in Sec. V, assuming negligible spatial curvature and P_0/E_0 independent of time in the background, are new, though Press and Vishniac[12] did consider the effect of entropy perturbations at $kS/\dot{S} \ll 1$. The exact solution displayed in Eq. (5.14) reduces to Eq. (5.15) in this limit, assuming $-\frac{1}{3} < P_0/E_0 < 1$, so $\infty > \beta > \frac{1}{2}$. With $-\frac{1}{3} < w < 1$ the strong energy condition is satisfied and $kS/\dot{S}$ increases toward the future; the particle horizon expands rather than shrinks relative to the wavelength of the perturbation. While the adiabatic speed of sound is imaginary for $w < 0$, a single-component treatment of the matter is inappropriate when the net pressure is negative.

Over the whole range $-\frac{1}{3} < w < 1$ the analytic solution shows that the fractional energy density at the particle horizon cannot greatly exceed the maximum previous ratio of stress perturbation to background energy density, averaged over one e-folding in S, no matter how early the stress perturbation occurs. For w close to one ($\beta \sim \frac{1}{2}$), $(2\beta - 1)^{-1}$ should be replaced by $\ln(\dot{S}/kS)$ in the integral over the anisotropic stress. This logarithmic enhancement of the perturbation amplitude over the amplitude of the stress perturbation is at most a factor of order 10^2 even if the stress perturbation was excited as early as the Planck time.

As measured by ϕ_h an entropy or isotropic stress perturbation predominantly excites the growing mode (which really has constant amplitude until the perturbation comes within the particle horizon). An anisotropic stress perturbation excites comparable amounts of the growing mode and the decaying mode. Of course, at the particle horizon all that is left is the growing mode.

From the point of view of the fractional energy density perturbation it seems that small nonlinear coupling of the decaying mode excited by anisotropic stress to the growing mode could perhaps significantly enhance the amount of growing mode beyond what is expected from the linear theory. I showed rather conclusively in Sec. VI that as long as the amplitude of the stress perturbation relative to the background energy density and therefore ϕ_h is small, nothing of the sort can happen. The nonlinear effects of anisotropic stress associated with scalar (purely longitudinal), vector (semilongitudinal), and tensor (purely transverse) perturbations on the fractional energy density are similar even though the frame dragging associated with a vector perturbation can generate large metric perturbations at $kS/\dot{S} \ll 1$.

In summary, then, I conclude that there is no possibility of explaining the origin of galaxies through the dynamical evolution of perturbations which arise from genuinely small or "statistical" causes within the context of general relativity and the strong energy condition. Either rather large perturbations, with a relative amplitude of at least 10^{-3} or so, are present initially or correspondingly large stress perturbations appear at a later time. This conclusion is hardly new, but has been strengthened by consideration of completely general energy-momentum tensor perturbations and the nonlinear interaction of modes. Of course, it also requires consideration of the physical processes which act on the perturbation once it is inside the particle horizon, but for anything like a galaxy scale these are fairly well understood.[7]

The one real hope for a dynamical explanation of the origin of structure in the Universe is the abolition of particle horizons at early times, perhaps through quantum modifications to the energy-momentum tensor and/or the gravitational field equations which in effect violate the strong energy condition that $E_0 + 3P_0 > 0$.

ACKNOWLEDGMENTS

This research was supported in part by National Science Foundation Grant No. PHY 79-19884 at the Institute for Advanced Study and in part by the National Science Foundation Grant No. GP-15267 at the University of Washington. A draft discussing some aspects of the cosmological perturbation problem and which was a foundation for this paper was written while the author was a Senior Fellow at the Center for Theoretical Physics, University of Maryland. I would like to thank W. H. Press and P. J. E. Peebles for stimulating discussions.

APPENDIX

The general expressions are written down for the perturbed Ricci tensor components and the perturbed matter equations of motion in terms of the metric tensor perturbations and energy-momentum tensor perturbations defined in Eqs. (2.14)–(2.20). Also included are the intrinsic and extrinsic curvature tensors for the constant-τ spacelike hypersurface.

The Ricci tensor for scalar perturbations is

$$\delta R^0_0 = \frac{3}{S^2}\left[\ddot{H}_L + \frac{\dot S}{S}\dot{H}_L - \frac{\dot S}{S}\dot{A} + [k^2/3 - 2(\dot S/S)^{\cdot}]A + \frac{k}{3}\left(\dot{B}^{(0)} + \frac{\dot S}{S}B^{(0)}\right)\right]Q^{(0)} , \tag{A1a}$$

$$\delta R^0_\alpha = \frac{2}{S^2}\left[-k\dot{H}_L - \tfrac{1}{3}k(1-3K/k^2)\dot{H}^{(0)}_T + k\frac{\dot S}{S}A - KB^{(0)}\right]Q^{(0)}_\alpha , \tag{A1b}$$

$$\begin{aligned}\delta R^\alpha_\beta = \frac{1}{S^2}&\left[\tfrac{4}{3}(k^2-3K)(H_L+\tfrac{1}{3}H^{(0)}_T) + \ddot{H}_L + 5\frac{\dot S}{S}\dot{H}_L - \frac{\dot S}{S}\dot{A} + (k^2/3 - 2S^{-2}(S\dot S)^{\cdot})A + \frac{k}{3}\left(\dot{B}^{(0)} + 5\frac{\dot S}{S}B^{(0)}\right)\right]\delta^\alpha_\beta Q^{(0)}\\ &+\frac{1}{S^2}\left[\ddot{H}^{(0)}_T + 2\frac{\dot S}{S}\dot{H}^{(0)}_T - k\left(\dot{B}^{(0)} + 2\frac{\dot S}{S}B^{(0)}\right) - k^2(H_L + \tfrac{1}{3}H^{(0)}_T + A)\right]Q^{(0)\alpha}_\beta .\end{aligned} \tag{A1c}$$

The Ricci tensor for vector perturbations:

$$\delta R^0_0 = 0 , \tag{A2a}$$

$$\delta R^0_\alpha = \frac{k^2-2K}{2S^2}\left(B^{(1)} - \frac{1}{k}\dot{H}^{(1)}_T\right)Q^{(1)}_\alpha , \tag{A2b}$$

$$\delta R^\alpha_\beta = \frac{1}{S^2}\left[\ddot{H}^{(1)}_T + 2\frac{\dot S}{S}\dot{H}^{(1)}_T - \frac{k}{5}\left(\dot{B}^{(1)} + 2\frac{\dot S}{S}B^{(1)}\right)\right]Q^{(1)\alpha}_\beta . \tag{A2c}$$

The Ricci tensor for tensor perturbations:

$$\delta R^0_0 = 0 , \tag{A3a}$$

$$\delta R^0_\alpha = 0 , \tag{A3b}$$

$$\delta R^\alpha_\beta = \frac{1}{S^2}\left[\ddot{H}^{(2)}_T + 2\frac{\dot S}{S}\dot{H}^{(2)}_T + (k^2+2K)H^{(2)}_T\right]Q^{(2)\alpha}_\beta . \tag{A3c}$$

Equations of motion for scalar perturbations:

$$[E_0S^3\delta]^{\cdot} + (E_0+P_0)S^3(kv^{(0)} + 3\dot{H}_L) + 3P_0S^2\dot{S}\pi_L = 0 , \tag{A4a}$$

$$[v^{(0)} - B^{(0)}]^{\cdot} + \frac{\dot S}{S}(1-3c_s^2)[v^{(0)} - B^{(0)}] - A - k\frac{w}{1+w}\pi_L + \tfrac{2}{3}k(1-3K/k^2)\frac{w}{1+w}\pi^{(0)}_T = 0 . \tag{A4b}$$

Equation of motion for vector perturbations:

$$[v^{(1)} - B^{(1)}]^{\cdot} + \frac{\dot S}{S}(1-3c_s^2)[v^{(1)} - B^{(1)}] + k\frac{w}{1+w}\pi^{(1)}_T = 0 . \tag{A5}$$

Intrinsic curvature tensor on the constant-τ spacelike hypersurface for general perturbations:

$$\mathcal{R}^\alpha_\beta = \frac{1}{S^2}[2K + \tfrac{4}{3}(k^2-3K)(H_L + \tfrac{1}{3}H^{(0)}_T)Q^{(0)}]\delta^\alpha_\beta - \frac{k^2}{S^2}(H_L + \tfrac{1}{3}H^{(0)}_T)Q^{(0)\alpha}_\beta + \frac{(k^2+2K)}{S^2}H^{(2)}_T Q^{(2)\alpha}_\beta . \tag{A6}$$

Extrinsic curvature tensor for general perturbations:

$$\mathcal{K}^\alpha_\beta = -\frac{1}{S}\left[\frac{\dot S}{S} + \left(\dot{H}_L - \frac{\dot S}{S}A + \frac{k}{3}B^{(0)}\right)Q^{(0)}\right]\delta^\alpha_\beta - \frac{1}{S}[\dot{H}^{(0)}_T - kB^{(0)}]Q^{(0)\alpha}_\beta - \frac{1}{S}[\dot{H}^{(1)}_T - kB^{(1)}]Q^{(1)\alpha}_\beta - \frac{1}{S}\dot{H}^{(2)}_T Q^{(2)\alpha}_\beta . \tag{A7}$$

Neither the intrinsic curvature nor the extrinsic curvature are gauge invariant, but both are invariant under a purely spatial gauge transformation ($T=0$) and depend only on the instantaneous value of T for a time gauge transformation.

*Permanent address: Department of Physics, University of Washington, Seattle, Washington 98195.

[1]E. M. Lifschitz, J. Phys. (Moscow) 10, 116 (1946).
[2]E. M. Lifschitz and I. M. Khalatnikov, Adv. Phys. 12, 185 (1963).
[3]S. W. Hawking, Astrophys. J. 145, 544 (1966).
[4]E. R. Harrison, Rev. Mod. Phys. 39, 862 (1967).
[5]R. K. Sachs and A. M. Wolfe, Astrophys. J. 147, 73 (1967).
[6]G. B. Field, in *Galaxies and the Universe*, edited by A. Sandage, M. Sandage, and J. Kristian (Univeristy of Chicago Press, Chicago, 1975).
[7]P. J. E. Peebles, *Cosmology: The Physics of Large Scale Structure* (Princeton University Press, Princeton, 1980).
[8]A. H. Guth and S.-H. H. Tye, Phys. Rev. Lett. 44, 631 (1980).
[9]W. H. Press, Phys. Scr. 21, 702 (1980).
[10]G. B. Field and L. C. Shepley, Astrophys. Space Sci. 1, 309 (1968).
[11]K. Sakai, Prog. Theor. Phys. 41, 1461 (1969).
[12]W. H. Press and E. T. Vishniac, Astrophys. J. 239, 1 (1980).
[13]D. W. Olson, Phys. Rev. D 14, 327 (1976).
[14]U. H. Gerlach and U. K. Sengupta, Phys. Rev. D 18, 1789 (1978).
[15]S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).
[16]Ya. B. Zel'dovich, L. B. Okun', and I. Yu. Kabzarev, Zh. Eksp. Teor. Fiz. 67, 3 (1974) [Sov. Phys.-JETP 40, 1 (1975)].
[17]T. W. B. Kibble, J. Phys. A 9, 1387 (1976).
[18]I. D. Novikov, Zh. Eksp. Teor. Fiz. 46, 686 (1964) [Sov. Phys.-JETP 19, 467 (1964)].
[19]P. J. E. Peebles, Nature 220, 237 (1968). Peebles does recognize the unphysical nature of these "statistical fluctuations."
[20]P. D. D'Eath, Ann. Phys. (N.Y.) 98, 237 (1976).
[21]P. J. E. Peebles and J. T. Yu, Astrophys. J. 162, 815 (1970).
[22]J. Silk, Astrophys. J. 151, 459 (1968).
[23]B. Carter, in 1979 Les Houches Summer School Proceedings (unpublished).
[24]E. W. Kolb and S. Wolfram (unpublished).
[25]M. V. Fischetti, J. B. Hartle, and B. L. Hu, Phys. Rev. D 20, 1757 (1979).
[26]S. W. Hawking and G. F. R. Ellis, *Large Scale Structure of Spacetime* (Cambridge University Press, Cambridge, 1973).
[27]R. Arnowitt, S. Deser, and C. W. Misner, in *Gravitation, An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962). See also, C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[28]N. Ó Murchadha and J. W. York, Phys. Rev. D 10. 428 (1974).
[29]L. Smarr and J. W. York, Phys. Rev. D 17, 2529 (1978).
[30]D. M. Eardley and L. Smarr, Phys. Rev. D 19, 2239 (1979).