Observational cosmology. VI. The microwave background and the Sachs-Wolfe effect

W. R. Stoeger
Vatican Observatory Research Group, Steward Observatory, The University of Arizona, Tucson, Arizona 85721

G. F. R. Ellis
Department of Applied Mathematics, Univeristy of Cape Town, Rondebosch 7700, South Africa

Chongming Xu
Lehr und Forschungbereich Theoretische Astrophysik, Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany

(Received 11 June 1993; revised manuscript received 12 October 1993)

The use of observational coordinates allows the formulation of redshift in a general cosmological space-time in a simple form, $(1+z)=A_0 dw/d\tau$, where A_0 is a normalization constant, w is the observational time coordinate, and τ is the proper time along the fundamental flow lines. This in turn allows easy calculation of the anisotropy of the cosmic microwave background radiation (CMBR) due to the Sachs-Wolfe (SW) effect. We reproduce the usual dominant first-order effect $\delta T_R/T_R=\frac{1}{3}(\delta_{bB}-\delta_{bA})$, where δ_b is the density contrast of baryons on the last scattering surface; as implied by this equation, the observationally significant result is the difference of δ_b in two different directions A and B on the plane of the sky. In order to obtain the actual result, one also needs to study perturbations of the temperature on the background last-scattering surface and on its first-order counterpart. In addition to the usual dominant term in the SW effect, we obtain a second term when the pressure p is significant at last scattering; this term depends on the difference in the pressure at the points of emission A and B on the last-scattering surface, and enters the CMBR anisotropy with a sign opposite to that of the usual term. For the adiabatic case, this pressure term can reduce the SW effect by up to 87% in a low-density universe. When $p=0$ at the last scattering surface, the usual SW result is obtained. Our results are gauge invariant to first order. Other explicit contributions to the Sachs-Wolfe anisotropy in the observational-coordinate calculation are clearly higher order. We discuss the interpretation of these results and compare them to other calculations of the large-scale cosmic microwave background anisotropy.

PACS number(s): 98.80.Es, 04.20.Cv, 95.30.Sf, 98.70.Vc

I. INTRODUCTION

The cosmic microwave background radiation (CMBR) is perhaps the observational cornerstone of cosmology. Precise measurement of its temperature, spectrum, and anisotropy has given us access to the origin, evolution, and development of the Universe itself, as well as to the structure within it. In particular, the anisotropy of the CMBR has often been used as the key to placing limits on the amplitude and size of the primordial seeds of superclusters, cluster of galaxies, and galaxies themselves at the time of last scattering. The preliminary positive Cosmic Background Explorer (COBE) measurement of the temperature anisotropy by Smoot and his collaborators [1,2] at the level of $(\delta T/T)_{\rm observ}=1.2\times10^{-5}$ on angular scales greater than 7 degrees, with a quadrupole term $(\delta T/T)_{\rm quad}=6\times10^{-6}$, has been used to strongly constrain competing theories of galaxy formation. The basis for the comparison with these measurements is the Scahs-Wolfe (SW) effect, [3] whereby photons traveling from the last-scattering surface to the observer are redshifted slightly more if they have to climb from an increased gravitational potential due to a density enhancement over part of the surface. Calculation of the SW effect, which will give the predominant contribution to the anisotropy of the CMBR for angular scales larger than a degree, which correspond to scales larger than the horizon at the time of last scattering ($z\approx1200$), is simple in principle. But in practice it has turned out to be quite difficult, and fraught both with gauge problems and calculational complications, which have necessitated various approximations (see [4,5] and references therein for a discussion). The calculation has been repeated in detail by a number of different authors, each making improvements and clarifications [6–8]. But ambiguities and complications in the calculation remain; in particular the relation of the measurements to gauge freedom, and the closely related issue of the precise placing of the surface of last scattering in the perturbed space-time remain elusive. In our view the meaning of many of the computer calculations of the SW effect remains obscure, because the way they handle these issues is not made explicit in their discussions.

We here present a calculation of the SW effect which is both simpler and easier to interpret than the usual treatments. It is based on using observational coordinates, [9–13] in which the redshift, the key parameter in calculating the SW effect, takes a particularly simple form. The variation of the observed temperature T_R coming from the last-scattering surface is given generally by

Sachs and Wolfe [3], Ellis [14], and Stoeger *et al.* [4]:

$$\frac{\delta T_R}{T_R}=\frac{\delta T_E}{T_E}-\frac{\delta z}{1+z} , \tag{1}$$

where T_E is the temperature on the surface of emission, in this case the last scattering surface, and $(1+z)$ is the redshift from the observer R to the points of emission E on that surface. Equation (1) should be interpreted as specifying the difference in temperature of the background radiation observed in *different* directions on the plane of the sky, corresponding to different points of emission on the last-scattering surface. The redshift is generally given by

$$(1+z)=\frac{(u^a k_a)_E}{(u^b k_b)_R} , \tag{2}$$

where u_E^a and u_R^a are the four-velocities of the emitter and the observer ("receiver"), respectively, and k^a is the vector tangent to the null geodesic connecting the events of emission and observation. In this paper we use these two relationships, specified in observational coordinates, to calculate the temperature anisotropy $\delta T_R/T_R$ for the microwave background.

We first describe and specify observational coordinates and show that, in terms of them, the redshift in a general space-time is given by Eq. (8) below. Then we write down the Friedmann-Lemaître-Robertson-Walker (FLRW) metric in observational coordinates, as well as the most general perturbed FLRW metric. Using this, we give $(1+z)$ to first order in terms of these metric components. The next step will be to give the conservation equations in observational coordinates—these will relate the contributing metric components to the density and pressure on the last-scattering surface. Following this, we calculate the variation of the redshift as specified in Eq. (1), interpreted as giving the difference between two directions on the plane of the sky corresponding to two different points on the last-scattering surface. The last stage in the calculation is the determination of the emission temperature T_E on the last-scattering surface itself; we calculate this, assuming that the surface is given by constant free-electron density [7]. Combining these steps we then have $\delta T_R/T_R$ to first order in a simple form, and give it for the pressure-free case ($p=0$), and the adiabatic ($p=\rho_\gamma/3$) and nonadiabatic radiation-pressure cases. In the pressure-free case we recover the usual dominant SW contribution; in the $p\neq 0$ cases we find another contribution which partially compensates for the dominant density-contrast contribution, and can possibly reduce the effect to 13% of that predicted by simple use of the standard SW formula. We discuss and interpret our results in the final two sections of the paper, comparing and contrasting certain of its key features with those of other calculations of the CMBR anisotropy, and relating it to the way in which the observations are made.

This is the sixth in a series of papers setting up a theoretical gravitational physics framework in which cosmologically significant observations can be optimally incorporated into a direct study of the geometry of space-time. Papers I and II are by Ellis *et al.* [11]. Papers III, IV, and V are by Stoeger *et al.*, Refs. [15], [16], and [17], respectively. In this approach, we determine what data tell us directly about space-time geometry, without imposing geometrical presuppositions (e.g. a FLRW geometry) on the analysis. In the case of the CMBR, examined in this paper, this approach links the observational data directly to conditions on the last-scattering surface, rather than through a model of the growth of irregularities in the Universe.

II. THE REDSHIFT IN OBSERVATIONAL COORDINATES

Observational coordinates $\{x^a\}=\{w,y,\theta,\phi\}$ [9,10, 18,11–13] are characterized in the following way. The "timelike" coordinate $w\equiv x^0$ is defined so that $\{w=\text{const}\}$ gives the past light cone $C^-(p)$ for each event p on the observer's world line C. Often we normalize w by requiring that w measure proper time along C, i.e., $w|_C=\tau|_C$, where τ is the proper time. Here we shall not impose that requirement. Instead we require that

$$\left.\frac{dw}{d\tau}\right|_C=\left.\frac{1}{A}\right|_C , \tag{3}$$

implying that A^2 is the g_{00} component of the metric. We choose $w=w_0$ to correspond to the event p_0 ("here and now").

The null geodesic vector field generating the ruling geodesics of these light cones $w=\text{const}$ can be written

$$\mathbf{k}=\partial/\partial v \Longrightarrow k^i=dx^i/dv , \tag{4}$$

where v is an affine parameter along the null geodesics and where (Friedlander [19], pp. 79–80)

$$k_i\equiv w_{,i}=\delta_i^0, \quad k^i k_i=0 . \tag{5}$$

This definition implies that $\mathbf{k}$ is the hypersurface-orthogonal ($k_{a;b}=k_{b;a}$) null geodesic ($k^a{}_{;b}k^b=0$) vector field lying in (and orthogonal to) the null surfaces on which w is constant ($w_{,i}k^i=k^0=0$).

The angular coordinates $\theta\equiv x^2$ and $\phi\equiv x^3$ will be observed angles on the plane of the sky. More technically, we let the null geodesics generating a given null cone be given by $\{\theta,\phi\}=\text{const}$ in the surface $\{w=\text{const}\}$. Then $\theta_{,i}k^i=\phi_{,i}k^i=0 \Longleftrightarrow k^2=k^3=0$. Furthermore, we normalize θ and ϕ by the central condition that they are standard spherical coordinates based on a parallely propagated orthonormal tetrad along the observer's world line C. Taken together, the above results show

$$k^j=dx^j/dv=(1/\beta)\delta_1^j ,$$

where $k^1=(1/\beta)=dy/dv$; it follows from (5) and $k_i=g_{ij}k^j$ that β is the g_{01} component of the metric [11].

Now $y\equiv x^1$ is the radial coordinate which measures "distance" down these null geodescis. There are various possible choices for y. It can be the affine parameter v itself, redshift, observer area distance, or something else. In order to mesh this choice easily with a simple formulation of the FLRW metric in observational coordinates, we choose $y=\tilde{r}$, where (a) the initial value of $\tilde{r}$ on our past light cone $w=w_0$ is determined by setting

$g_{00}=-g_{01}$ there; (b) the coordinate is then "dragged off" this light cone by the condition that $\tilde{r}$ is comoving with the fluid four-velocity **u**, so its Lie derivative with respect to **u** is zero: that is,

$$d\tilde{r}/d\tau=0 \Longleftrightarrow u^1=0 \; . \tag{6}$$

In the background FLRW space-time, this choice makes $\tilde{r}$ a standard comoving distance parameter in the surfaces $t=$const [15]. It should be mentioned that choosing $g_{00}=-g_{01}$ on $w=w_0$ does does not mean that this relationship will hold in general as $\tilde{r}$ is dragged off onto other past light cones: g_{00} may not equal $-g_{01}$ at other values of the retarded time w,[1] although this will indeed be true in the FLRW case, and so will be true in the background model.

Now, from Eq. (5) and from the fact that the components of the four-velocity **u** are just $u^i=dx^i/d\tau$, we immediately see that

$$k_a u^a = dw/d\tau \; . \tag{7}$$

Then, from Eqs. (2) and (7) and the central condition (3) we have that the redshift in observational coordinates is simply

$$(1+z)=A_0 dw/d\tau \; , \tag{8}$$

where A_0 is A evaluated at the position of the observer, that is at $w=w_0$ and $\tilde{r}=0$ ($\tilde{r}$ is zero on the observer's world line C). Equation (8) is the key result which simplifies the calculation of the SW CMBR temperature anisotropy in observational coordinates.

III. THE METRIC IN OBSERVATIONAL COORDINATES

As has been shown elsewhere [11,20], in observational coordinates the background FLRW metric can be written as

$$d\tau^2=R^2(w-\tilde{r})\{-dw^2+2\,dw\,d\tilde{r}+\kappa^{-2}\sin^2\kappa\tilde{r}\,d\Omega^2\} \; , \tag{9}$$

where $\kappa=1,0,-1$, respectively, for the elliptic, flat, and hyperbolic cases, and $d\Omega^2$ is the metric on the unit two-sphere. The most general metric in observational coordinates, written in the "FLRW-based" form, is [11,20,16]

$$\begin{aligned} d\tau^2=&-(R^2-Z^2)dw^2+2(R^2+B^2)dw\,d\tilde{r}+2v_2 dw\,d\theta \\ &+2v_3 dw\,d\phi+(R^2\hat{f}^2+h_{22})d\theta^2+2h_{23}d\theta d\phi \\ &+(R^2\hat{f}^2\sin^2\theta+h_{33})d\phi^2 \; , \end{aligned} \tag{10}$$

where the functions $R=R(w-\tilde{r})$ and $\hat{f}=\kappa^{-1}\sin\kappa\tilde{r}$ characterize the FLRW part of the metric, and the functions $Z^2, B^2, v_2, v_3, h_{22}, h_{23}$, and h_{33} characterize the deviations in the metric from the FLRW form. In general, they will be functions of all four coordinates. When these are small, they represent perturbations from the background FLRW space-time.

The quantity A of the previous section [see Eq. (3)] is related to these metric components by $A^2=R^2-Z^2$. Now as we have already indicated, by suitable coordinate choice we can put $Z^2=-B^2$ on our past light cone $w=w_0$ without loss of generality. In fact, this defines our choice of null radial coordinate $\tilde{r}$ uniquely, and so fixes the "fitting" of the FLRW model to the real (lumpy) Universe [21]; this in turn defines what is meant by the perturbation quantities such as δ_b [22]. Essentially we have chosen a unique "gauge," which defines the perturbed quantities.

IV. THE REDSHIFT TO FIRST ORDER

If we now write down explicitly the equation $u^a u_a=-1$ on $w=w_0$ from the general metric (10), we obtain

$$\begin{aligned} -1=&-(R^2-Z^2)(dw/d\tau)^2+2(R^2+B^2)(dw/d\tau)(d\tilde{r}/d\tau)+2v_2(d\theta/d\tau)(dw/d\tau)+2v_3(d\phi/d\tau)(dw/d\tau) \\ &+(R^2\hat{f}^2+h_{22})(d\theta/d\tau)^2+2h_{23}(d\theta/d\tau)(d\phi/d\tau)+(R^2\hat{f}^2\sin^2\theta+h_{33})(d\phi/d\tau)^2 \; , \end{aligned} \tag{11}$$

where we mean here and hereafter that the equation is evaluated at $w=w_0$ unless otherwise indicated. We want to solve this equation for $dw/d\tau$, which will give us $1+z$ modulo the normalization.

In our observational coordinates $\tilde{r}$ is comoving. So $d\tilde{r}/d\tau=0$ and the g_{01} component is eliminated. If we write $d\theta/d\tau=(d\theta/dw)(dw/d\tau)$ and $d\phi/d\tau=(d\phi/dw)(dw/d\tau)$ and solve Eq. (10) for $dw/d\tau$, we obtain

$$\begin{aligned} dw/d\tau=&[R^2-Z^2-2v_2(d\theta/dw)-2v_3(d\phi/dw)-(R^2\hat{f}^2+h_{22})(d\theta/dw)^2-2h_{23}(d\theta/dw)(d\phi/dw) \\ &-(R^2\hat{f}^2\sin^2\theta+h_{33})(d\phi/dw)^2]^{-1/2} \; . \end{aligned} \tag{12}$$

Here $d\theta/dw$ and $d\phi/dw$ are proper motions. Then, if we expand the brackets in Eq. (12), using only the first two terms of the binomial series for the right-hand side, we obtain

[1]J. Ehlers (private communication) has pointed out that it is only possible to set $g_{00}=-g_{01}$ everywhere, with y a comoving coordinate, in restricted cases. This affects some details, but not the overall argument, of Stoeger *et al.* [15,17] where it is erroneously claimed this specialization is possible generally. Details of the required amendments to those papers will be given elsewhere.

$$dw/d\tau=\frac{1}{R}+\frac{Z^2}{2R^3}+\frac{v_2 d\theta/dw}{R^3}+\frac{v_3 d\phi/dw}{R^3}+\left[\frac{R^2\hat{f}^2}{2R^3}+\frac{h_{22}}{2R^3}\right](d\theta/dw)^2+\frac{h_{23}}{2R^3}(d\theta/dw)(d\phi/dw)+\left[\frac{R^2\hat{f}^2\sin^2\theta}{2R^3}+\frac{h_{33}}{2R^3}\right](d\phi/dw)^2 , \tag{13}$$

where (considering the case of a perturbed FLRW universe) the first term is zeroth order, and only the second term is first order. The remaining terms are second or third order ($d\theta/dw$ and $d\phi/dw$ are each first order, as are the non-FLRW metric components v_2, v_3, h_{22}, h_{23}, and h_{33}). To obtain all the second- and third-order contributions to $dw/d\tau$ we would have to include the third and fourth terms of the binomial expansion of Eq. (12) as well. We can easily see that the second- and third-order terms already present in Eq. (13) contain direct contributions from velocities of the emitters on the last-scattering surface and from gravitational radiation. Later we shall study whether the first-order term contains a gravitational wave contribution.

In what follows we shall be concerned only with the first-order SW contributions. From Eqs. (13) and (8), we see that the redshift to first order is

$$(1+z)=A_0\left[\frac{1}{R}+\frac{Z^2}{2R^3}\right] , \tag{14}$$

where A_0 is just the quantity $(R^2-Z^2)^{1/2}$ evaluated at the point of observation, and R and Z in the brackets are evaluated at the point of emission. It is obvious that when $Z=0$ this reduces to the usual expression for redshift in the FLRW case, which is just

$$(1+z)_{\rm FLRW}=(R_0/R) . \tag{15}$$

Here $R_0 \equiv R(w=w_0,\ \tilde{r}=0)$. It is important to notice that in these observational coordinates, the points of emission and observation have the same value of the "time" coordinate, $w=w_0$, because they are on the same light cone. One parameter serves to distinguish these events, their $\tilde{r}$-coordinate values, and the point of observation will always be at $\tilde{r}=0$ because it is on the observer's world line C. This feature of observational coordinates is one of the reasons for the simplicity of the SW calculation using them.

V. THE CONSERVATION EQUATIONS

As we can see from Eq. (14), in observational coordinates the redshift to first order depends only on the metric variable Z^2, since $R=R(w,\tilde{r})$ is known from the FLRW background. If we can determine Z^2 and link it to the matter variables in some way, we will have gone a long way towards solving our problem.

It turns out, fortunately, that if we write down the conservation equations $T^{ab}{}_{;b}=0$ in the observational coordinates we can easily establish this connection. The easiest way of doing this is through the Fluid-Ray (FR) tetrad formalism, in which the four components of the conservation equations in the case of a single perfect fluid take the form (Stoeger *et al.* [13], Eqs. (29))

$$\Delta\mu+(l+\bar{l}+b+\bar{b}-n-\bar{n})(\mu+p)=0 , \tag{16a}$$

$$(l+\bar{l})(\mu+p)+\Delta p+Dp=0 , \tag{16b}$$

$$\bar{a}(\mu+p)+\bar{\delta}p=0 , \tag{16c}$$

$$a(\mu+p)+\delta p=0 . \tag{16d}$$

Here μ is the total relativistic energy density, p is the pressure, l, n, b, and a are FR-tetrad spin coefficients, and Δ, D, and δ the tetrad derivative operators in the **u**, **k**, and angular directions, respectively. Using the differential relationships between the spin coefficients and the metric components [16], the purely angular conservation Eqs. (16c) and (16d), which are relevant to this problem, can be written to first order (note that as these equations are entirely in the past light cone $w=w_0$, we can without loss of generality take $Z^2=-B^2$ in evaluating them):

$$-\frac{\mu_0+p_0}{2R^2}\frac{\partial Z^2}{\partial\theta}+\frac{\partial p_1}{\partial\theta}=0 , \tag{17a}$$

$$-\frac{\mu_0+p_0}{2R^2}\frac{\partial Z^2}{\partial\phi}+\frac{\partial p_1}{\partial\phi}=0 . \tag{17b}$$

Here the subscripts 0 and 1 refer to zeroth-order and first-order quantities, respectively. Further, as is clear, R is zeroth order, and Z^2 is first order.

Now it is clear that (to first order) Eqs. (17a) and (17b) just give Z^2 in terms of the pressure p, plus a first-order function $F(w,\tilde{r})$, that is

$$Z^2=2R^2\frac{p_1}{\mu_0+p_0}+F(w,\tilde{r}) . \tag{18}$$

This is the relationship we were looking for. We do not, of course, know $F(w,\tilde{r})$. This must be determined from the field equations and the other two conservation equations (16a) and (16b). But it turns out that for calculating the SW effect to first order we shall not need it; F will only contribute beginning at second order.

VI. THE SACHS-WOLFE ANISOTROPY FOR ADIABATIC PERTURBATIONS

From Eq. (1) we see that in general the SW anisotropy will be made up of two contributions, one from the difference in the redshift from the observer to two different points on the last-scattering surface, and the other from the difference of emission temperatures at those points on the last-scattering surface itself. If we define the last-scattering surface as a surface of constant free electron density η_e [7,5] then it can be easily shown that for adiabatic perturbations ("adiabatic" in the cosmological sense—all components of the mass-energy participate equally in the perturbation), the temperature of the perturbed last-scattering surface will be that of the background last-scattering surface and therefore will be a constant [5]. This *does not* mean, however, that the perturbed last-scattering surface will be at the same distance $\tilde{r}$ down our past light cone in every direction. Thus, for adiabatic perturbations we can focus just on the first contribution, which we have been concentrating on calculating above.

From (14), in this case the SW anisotropy measured between two different directions A and B on the plane of the sky will be [using "difference notation" for Eq. (1)]

$$\frac{\Delta T_R}{T_R}=-\frac{\Delta z}{1+z}=-R_A\left[\frac{1}{R_B}-\frac{1}{R_A}+\frac{Z_B^2}{2R_B^3}-\frac{Z_A^2}{2R_A^3}\right] \tag{19}$$

where the last-scattering surface at points A and B will be at two slightly different values of $\tilde{r}$ on our past light cone $w=w_0$. In the last two terms the (first-order) difference between R_A and R_B can be neglected, because Z^2 is already first order. This same difference in the first two terms in Eq. (19) can be written in a differential form; we rewrite Eq. (19) therefore as

$$\frac{\Delta T_R}{T_R}=\frac{1}{R_A}\frac{\partial R}{\partial \tilde{r}}\bigg|_A \Delta\tilde{r}-\frac{1}{2R_A^2}[Z_B^2-Z_A^2]\ , \tag{20}$$

where Z^2 at points A and B will be given by Eq. (18) and $\Delta\tilde{r}=\tilde{r}_B-\tilde{r}_A$. The difference between $F(\omega_0,\tilde{r}_B)$ and $F(\omega_0,\tilde{r}_A)$ contained in the Z^2 terms can be rewritten $F'|_A\Delta\tilde{r}$, which will be second order and can be neglected [F is already first order, and $\Delta\tilde{r}$ is first order]. The reason the purely angular difference between Z^2 at two points at the last-scattering surface cannot be neglected is because the difference in angle can be large and is essentially zeroth order, so the difference in the values of the purely spatial part of Z^2 will still be first order.

We now must evaluate the first term on the right-hand side of Eq. (20). This is the contribution to the temperature anisotropy due to the slightly different distances at which we encounter the last-scattering surface as we look down our past light cone in different directions. As we mentioned above, this is clearly a first-order effect. We take "last scattering" to be defined by the place where the optical depth becomes unity; this in turn occurs when matter becomes ionized, because the main effect here is Thomson scattering, arising when the CMBR temperature exceeds the ionization level of matter in the Universe. That is, we characterize the last-scattering surface in the real space-time as a surface of constant free-electron density, this density being equal to the zero-order (background model) electron density on the background last scattering surface.

Characterizing the region of decoupling as a surface of constant free-electron density is still only an idealization, but a very helpful and appropriate one. The transition from collisional to free photons takes the order of one Hubble time *at* the epoch of coupling. So the decoupling region is actually a layer or shell of some thickness in redshift $\Delta_d z$. However, as Hogan *et al.* [23] point out, if $z_d>1000$, $\Delta_d z\approx\frac{1}{15}z_d$, that is, the thickness of the decoupling shell will be relatively narrow, from our point of view as observers. And we recall that we are assuming that $1+z_d=1200$. Thus, it makes sense to idealize it as a surface, and particularly as one of constant free-electron density. A more precise treatment of the last-scattering region would demand a more sophisticated description of the photons and their interaction with the free electrons.

Further justification for this idealization of the last-scattering region as a surface derives from "the compensation effect" [5]. Because the redshift in the radiation after last scattering goes like T and the change in the temperature of the matter scattering the radiation goes like $1/T$, since the baryonic matter is still coupled to the radiation until the time of last scattering, the two effects cancel out or compensate. Thus, within the last-scattering layer itself, one may calculate the Sachs-Wolfe anisotropy from any location, as long as one uses the appropriate temperature at that location: The difference in redshift due to placing the point of emission at a different null radial coordinate position will be exactly compensated by the different temperature of the matter at that position. This in itself enables one to collapse the last-scattering shell to a surface

With this idealized description of the last-scattering region as a surface of constant free-electron density, Panek's result [7] holds:[2] calculating the displacement of the real last-scattering surface from the background one, we find

$$(\dot{R}/R)\Delta_0\eta=(R'/R)\Delta_0\tilde{r}=\frac{1}{3+D}\delta_b+\frac{D}{4(3+D)}\delta_\gamma\ . \tag{21}$$

Here the overdot signifies partial differentiation with respect to the conformal time η, and prime with respect to $\tilde{r}$. D is a parameter containing the functional dependence of the free-electron density on baryon density and on temperature along with the first derivatives of this dependence [7]. δ_b and δ_γ are the perturbations in the baryons and the radiation, respectively, as measured relative to the background last-scattering surface. The Δ_0 operator refers to differences between values on the perturbed last-scattering surface and on the background last-scattering surface. This will be different from our $\Delta\tilde{r}$, which is the difference in $\tilde{r}$, or equivalently in conformal time, between two points on the perturbed last-scattering surface itself. But obviously, for our situation, $\Delta\tilde{r}$ can be related to $\Delta_0\tilde{r}$.

Because the position of the last-scattering surface in the FLRW background is at constant $\tilde{r}$, the $\Delta\tilde{r}$ will equal the variation of $\Delta_0\tilde{r}$ with direction. That is, we can write

$$\left[\frac{R'}{R}\right]\bigg|_A\Delta\tilde{r}=\left[\frac{R'}{R}\right]\bigg|_A[(\Delta_0\tilde{r})_B-(\Delta_0\tilde{r})_A]\ . \tag{22}$$

Now, for adiabatic perturbations and a Planckian radia-

[2]Detailed derivations of all Panek's equations are given by Katz [24].

tion spectrum, $\delta_\gamma = \frac{4}{3}\delta_b$ and Eq. (21) reduces to [7,5]

$$\left[\frac{R'}{R}\right]\Delta_0 \bar{r} = \frac{1}{3}\delta_b \ . \tag{23}$$

Now we can write down our final result, from Eqs. (18), (20), and (23):

$$\frac{\Delta T_R}{T_R} = \frac{1}{3}(\delta_{bB} - \delta_{bA}) - \frac{1}{\mu_0 + p_0}(p_{1B} - p_{1A}) \ . \tag{24}$$

This equation can be read either from right to left (the conventional sense), predicting the CMBR anisotropy directly from conditions on the last-scattering surface; or from left to right (the sense of the "observational cosmology" program mentioned in Sec. I), determining relations on the last-scattering surface from observational quantities.

The first, nonpressure, term in (24) is just the usual SW result arrived at in the standard calculation and often quoted in the literature. For the pressure-free case, this is the result—the pressure terms going to zero. But for $p \neq 0$, we have the second contribution, which is traceable to the nongeodesic flow of the emitters (due to pressure). This term is not recovered in the standard calculations (see, for example, Panek [7]) because they usually do not treat the $p \neq 0$ case, which will be important only for low-density Universes ($\Omega_0 < 1$). For the two-component situation, where baryons and photons are present Eq. (24) leads to

$$\frac{\Delta T_R}{T_R} = \frac{(\delta_{bB} - \delta_{bA})(\mu_b)}{3(\mu_b + 4\mu_\gamma/3)} \ . \tag{25}$$

Here the baryonic pressure p_b is neglected and the adiabatic condition $\delta_\gamma = \frac{4}{3}\delta_b$ is used. In a situation where radiation dominates completely at last scattering, Eq. (25) has the limit

$$\frac{\Delta T_R}{T_R} = 0 \ , \tag{25a}$$

for in that case the pressure contribution exactly cancels the density-perturbation contribution in Eq. (24). On the other hand when μ_γ is much less than μ_b, then (25) becomes

$$\frac{\Delta T_R}{T_R} = \frac{1}{3}(\delta_{bB} - \delta_{bA}) \ , \tag{25b}$$

which is the normal Sachs-Wolfe effect. For a general mixture of baryons and radiation, in which pressure is important, the result (25) will be somewhere in between the standard result [Eq. (25b)], and that for the radiation dominated case [Eq. (25a)], valid for a pure radiation fluid, or when the universe is so hot that all particles are essentially relativistic. The exact result depends on the value of Ω_0. Specifically, because μ_γ goes as $1/R^4$ but μ_b goes as $1/R^3$, the ratio of densities at last scattering is

$$\left.\frac{\mu_\gamma}{\mu_b}\right|_d = \frac{\mu_{\gamma 0}}{\mu_{b0}}\frac{R_0}{R_d} = \frac{\Omega_{\gamma 0}}{\Omega_{b0}}(1+z_d) \ . \tag{26}$$

But $\Omega_{\gamma 0}h^2 = 4.18 \times 10^{-5}h^2$ [25,26] where h is the Hubble constant in units of 100 km/sec Mpc. Taking $1+z_d = 1200$ we find

$$\left.\frac{\mu_\gamma}{\mu_b}\right|_d = \frac{5\times 10^{-2}}{h^2\Omega_{b0}} \ . \tag{27}$$

This tells us the ratio of the terms in Eq. (25). Let

$$f \equiv \frac{\mu_b}{\mu_b + 4\mu_\gamma/3} \tag{28}$$

which is the fraction of the usual SW result (25a) that holds in the Universe with pressure [given by (25)]. Then we see that

$$f = \frac{1}{1 + 0.067/h^2\Omega_{b0}} \ . \tag{29}$$

At the low density end, $\Omega_0 = 0.04$ gives $f = 0.13$ ($h = 0.5$) to $0.37 (h = 1.0)$; while at the high density end, $\Omega_0 = 1$ gives $f = 0.79 (h = 0.5)$ to $f = 0.94$ ($h = 1.0$). Thus the maximum suppression of anisotropy that can be attained through the pressure term, in practice, is to about 13% of the SW result—a very significant reduction. If we take the observationally preferred middle range of Ω_0, we find for $\Omega_0 = 0.1$ the factor f is in the range 0.27 ($h = 0.5$) to 0.6 ($h = 1$) and for $\Omega_0 = 0.3$, the factor f is in the range 0.53 ($h = 0.5$) to $f = 0.82$ ($h = 1$).

This reduction in CMBR anisotropy due to the pressure term could be significant in assessing the microwave background anisotropy limits for various galaxy formation scenarios, and in estimating the amplitude of density perturbations at the epoch of recombination, in low-density universes (while these are unfashionable because of the prevailing inflationary universe dogma, they are certainly a possibility which must be taken into account if we want to determine which of all possible scenarios are compatible with observations, and which are not). One should notice also the sensitivity of the results to the value of the Hubble constant, basically because the radiation energy density μ_r, is accurately known independent of the value of H_0, but (for given μ_r) Ω_r scales with h^{-2}.

In particular, the reduction of the SW anisotropy in the radiation-dominated case has the interesting consequence that if the matter density is low, the magnitude of the SW effect for a given-density perturbation $\delta\rho$ can be considerably less than we anticipate if we ignore the effect of the pressure term, and simply use the usual SW formula. This means that the low measured value of the CMBR anisotropy, sometimes taken as evidence that Ω must be very close to 1, could be compatible with lower values of Ω (and still allow formation of galaxies and large scale structures).

However, one needs to know the equation of state on the last-scattering surface before we can use the CMBR anisotropy to set limits on Ω_0; the result is quite sensitive to this equation of state.

VII. THE SACHS-WOLFE EFFECT IN THE NONADIABATIC CASE

In the nonadiabatic case, in which the equation of state is perturbed, we will not be able to find such a simple result as that we have just calculated in the adiabatic-perturbation case. Equations (18), (19), and (20) will still hold, but we will have to use the perturbed equation of state to put the redshift contribution to the anisotropy in a form such as that of Eq. (25). Furthermore, now the contribution from the variation of the temperature along the surface of last scattering itself will be nonzero.

In this case we can write the temperature on the last-scattering surface as [5]

$$T_E=(T_E)_0\left[1-\chi\frac{3}{3+D}\right],$$

where $(T_E)_0$ is the temperature on the FLRW background last-scattering surface, and $\chi\equiv\frac{1}{3}\delta_b-\frac{1}{4}\delta_\gamma$. Using this in Eq. (1) and in calculating $(R'/R)\Delta\bar{r}$, we obtain

$$\frac{\Delta T_R}{T_R}=\tfrac{1}{3}(\delta_{bB}-\delta_{bA})-\chi_B+\chi_A-\frac{p_{1B}-p_{1A}}{\mu_0+p_0}, \quad (30)$$

where the pressure term will depend on χ and on the equation of state, assuming that radiation dominates over the baryons (but not necessarily the total matter).

VIII. DISCUSSION

Comparing our calculation of the CMBR anisotropy with those in the literature, for example those of Sachs and Wolfe [3], Peebles [6], Panek [7], and Abbott and Schaefer [8], we see a number of features which distinguish it and which demand explanation.

First of all, some more recent calculations [7,8] have been carried out in terms of Bardeen's [27] gauge-invariant variables. It is important that the result be gauge-invariant, that is, that it be immune to small changes in the coordinate system in the perturbed spacetime and thus free from purely coordinate perturbation modes, which have no physical significance. One way of handling this is to write the result in such gauge-invariant variables. But if a scalar, vector, or tensor vanishes in the background, then it is automatically gauge invariant [22], even if it is not written in explicitly gauge-invariant quantities. This is the case with our result, even though the observations we deal with are essentially two-point relations and thus not directly covered by the Stewart-Walker theorem [28]. Both sides of Eqs. (24), (25), and (30) are given in terms of differences in physically determined quantities (temperature, density, pressure) at points on either the real last-scattering surface (left-hand side) or the background last-scattering surface (right-hand side) corresponding to two different directions on the plane of the sky. In the FLRW background these differences vanish. They are thus gauge-invariant, as may be explicitly checked in the case of scalar perturbations by using the Bardeen equations (3.3) and (3.7) [27]. Thus the fact we have determined the result in a particular gauge does not matter; the result will be the same in any gauge.

The theoretical results usually given are not often explicitly in terms of such differences, which are the observationally relevant quantities. Instead, they are given in terms of the differences between the background quantity and the perturbed quantity on one null ray—that is, in one direction; but this is gauge dependent. The difference of this quantity in two directions will generally be equivalent to our results, and will be gauge invariant also. But that is not what most workers have written down. Furthermore, (a) they often discard small terms (see, for instance Panek [7]) in the single-direction anisotropy they calculate. If what is observationally significant is the difference in such quantities in two different directions, this could be the source of significant error; what is discardable in one direction may not be negligible when we are taking the difference of a quantity in two directions. (b) Even if gauge-invariant variables are used, the result is not gauge invariant unless the position of the last-scattering surface is fixed in a gauge-invariant (physical) way [4]; but this is not always done (in many calculations it is assumed that the last-scattering surface is given by the same equation in the real space as in the background space; but this is a gauge-dependent, nonphysical prescription).

With some effort we can construct gauge-invariant quantities relative to observational coordinates and have done so (Stoeger, unpublished)—these will be different from those of Bardeen, because the observational coordinate system is "far" from the coordinate system Bardeen and others employ. We can write our result in terms of these observational-coordinate gauge-invariant quantities. But, since the result is already automatically gauge invariant, there is no need to do so. In fact, in the case of the background radiation anisotropy, the gauge-invariant variables complicate the result and make it very hard to interpret physically, except in simple cases and after the use of the evolution and conservation equations. This is seen clearly in Panek [7], and Abbott and Schaefer [8].

Second, we notice that in our calculations and results there are no time integrations, and no use of the evolution equations for the perturbations along timelike world lines, as there are in other calculations and results [we do, however, use the conservation equations, but along the light cone, not along timelike world lines]. Time integrations are not necessary—we do not need to solve the null geodesic equation through an explicit integration, because the redshift $(1+z)$ is already given in a simple form fitted to null cone data, the geodesic equation having been implicitly integrated through use of suitably adapted coordinates. Furthermore, in observational coordinates, everything in the calculation is done at one specific observational time, $w_0=\text{const}$. The result is given in terms of the differences in density, pressure, and χ, at pairs of points on the last-scattering surface which intersect our past light cone. The history of those values is not needed in the calculation, however important and interesting it might otherwise be. As we have seen, this significantly facilitates the calculation of the anisotropy.

Third, the observational coordinate system is "privileged" from an observational point of view. Our past light cone is defined by our observational context. The angles on the plane of the sky give the direction of the null rays generating the light cone and connecting us with the sources we are observing. The only coordinate we use that is not "privileged" in this sense is $\bar{r}$, which measures distance down the generators of the light cone. In other words, our observational situation gives us automatically an observational time w_0, and observational angles, but it does not give us a natural measure of distance down the light cone which is easily usable. This partially "privileged" character of the coordinates is another way of describing why the calculation of the anisotropy is relatively simple and is done all at one "time" $w=w_0=\text{const}$.

Fourth, a related difference is that we do not have to decompose the perturbation modes into scalar, vector, and tensor contributions, as is usually done. Nor have we had to separate the growing and decaying modes, and argue the neglect of the decaying mode. We simply get a first-order result for the cosmic microwave anisotropy, without having to worry about the detailed dynamics of the key contributions. In the usual calculations, one ends up with three separate contributions to the anisotropy—a scalar, vector, and tensor contribution (see Panek [7], and Abbott and Schaefer [8]). The total anisotropy is the sum of these. These can be difficult to interpret in the general case, when all three occur.

Fifth, in the Panek [7] and Abbott and Schaefer [8] calculations, there are vector and tensor contributions, which seem to have no parallel in our first-order result. The vector contribution in the simplest cases will have only a decaying mode, which people usually argue they can neglect; the tensor contribution will have both a growing and a decaying mode [8]. The growing mode will represent the contribution of gravitational waves to the anisotropy. There are transverse, traceless metric tensor contributions in the scalar and vector parts of the anisotropy also, but these are not in themselves gauge-invariant and do not represent real gravitational radiation. However gravitational waves can change the distance from us to the source, i.e., reposition last-scattering surfaces relative to the point of observation.

The velocity contributions are found both in the scalar and vector components of the decomposition. If we look at the details of our calculation, especially at Eq. (12) above, we see that the transverse velocities (proper motions) begin to enter our results only at second order, and the terms involving h_{22},h_{23}, and h_{33}, which will carry some of the gravitational wave contribution, enter only at third order. Furthermore, the first-order term Z^2 cannot contain angular gravitational-radiation effects; for the zero-pressure case, Z^2 must be independent of θ and ϕ to first order, and, for the pressure case, the angular dependence is completely determined to first order by the conservation equations. There will in general be higher-order angular gravitational contributions from Z^2, however. In our calculation where are the first-order gravitational-radiation contributions and first-order velocity contributions analogous to those of Panek and others?

This question requires further study. Certainly, the first-order velocity and gravitational-radiation terms in Panek, for instance [7] are due to velocity components in the radial direction and to gravitational-radiation fields also in the radial direction, respectively. These latter must be due to gravitational waves traveling transverse to the generators of our past light cone, not along them, according to the transverse traceless condition. It turns out that gravitational waves are already incorporated implicitly in our first-order result. Why do we not see them explicitly in our calculation? The answer is because the splitting into scalar, vector, and tensor contributions is purely heuristic, and depends on the gauge, as does the splitting of redshift into velocity and gravitational effects, as remarked by Sachs and Wolfe [3]. As we have remarked elsewhere [4,5], we can change this scalar-vector-tensor splitting by changing the gauge; it is only preserved under a very restricted set of gauge transformations. (Those considered by Panek [7] and Bardeen [27], whose "gauge-invariant" formalism Panek employs, constitute such a very restricted set—essentially the infinitesimal coordinate transformations from FLRW expressed in the usual $3+1$ coordinates with a conformal time coordinate. Relative to our observational coordinates, those coordinates represent a very "large" gauge transformation.) Thus the scalar-vector-tensor splitting is not present in any truly covariant results. It may be useful to make such a splitting for heuristic or interpretational purposes, but the resulting analysis will not be completely gauge invariant. One of the strengths of our observational-coordinate calculation here is that we do have to introduce this splitting. We have automatically included all three types of contribution (scalar, vector, and tensor) precisely because we do not use such a splitting.

Finally, one might wonder why we do not recover a dipole-anisotropy contribution in our formulation. Since we have set up our problem with the observer moving with the cosmological fluid flow, it does not appear. This is reflected, in particular, in the form of the FLRW metric in observational coordinates, given in Eq. (9). If we had freed the observer from following the cosmological fluid flow, allowing him or her to have a velocity with respect to it, then we would have recovered the dipole-velocity contribution. Then, too, the background FLRW metric in observational coordinates with respect to our observer would have been "tilted," with time-angle cross terms [20,16].

IX. CONCLUSIONS

We have calculated both the redshift and the SW temperature anisotropy of the CMBR to first order in observational coordinates, recovering the standard results plus a pressure contribution. In doing so we have demonstrated that this formulation has a number of advantages over the usual formulations. The results are relatively simple and straightforward, and easy to interpret. There is no need to harmonically analyze the equations or to separate them into their scalar, vector, and tensor components—nor is there a need, to first order, of knowing the solutions to the field equations. The results are given simply in terms of the differences in the density contrasts and the pressures (if $p\neq 0$) at the two points on the last-scattering surface being compared. Furthermore, it is clear that the results are gauge invariant, that is, the values of the temperature anisotropy given by Eqs. (24), (25), or (30) will be independent of the coordinate system in which they are calculated. Though the underlying formalism is not gauge invariant (everything is worked out in a particular coordinate system—observational coordinates), the differences in the quantities contributing to $\Delta T_R / T_R$ all vanish in the FLRW background, and are gauge invariant, as can been seen in our formulation.

There are essentially three contributions to the temperature anisotropy. The first is the variation of the emission temperature on the last-scattering surface itself. This is zero in the adiabatic-perturbation case (see, e.g., Ellis *et al.* [5]). The second contribution is due to the first-order variation in the distance $\tilde{r}$ the last-scattering surface is found in different directions on the plane of the sky. This is the dominant contribution and is independent of pressure. It is determined by examining the consequences of demanding that the last-scattering surface is one of constant free-electron density, as Panek [7] has done. The third first-order contribution is due to the difference in the pressures on the last-scattering surface. This is essentially the result of the nongeodesic flow of the emitters.

We find that the additional pressure term can significantly depress the SW effect when Ω is low, or when radiation pressure dominates matter pressure at last scattering. This shows that CMBR anisotropies are depressed in adiabatic low-density universes where the radiation pressure can dominate at the surface of last scattering.

In the near future, we hope to employ this formulation of the SW effect to investigate more fully first-order velocity and gravitational-wave contributions and the second-order contributions to the anisotropy of the microwave background.

ACKNOWLEDGMENTS

We thank the FRD (South Africa) and MURST (Italy) for financial support. W.R.S. thanks Shoba V. Veeraraghavan for helpful discussions.

[1] G. E. Smoot *et al.*, Astrophys. J. Lett. **396**, L1 (1992).

[2] C. L. Bennett *et al.*, Astrophys. J. Lett. **396**, L7 (1992).

[3] R. K. Sachs and A. M. Wolfe, Astrophys. J. **147**, 73 (1967).

[4] W. R. Stoeger, G. F. R. Ellis, and B. G. Schmidt, Gen. Relativ. Gravit. **23**, 1169 (1991).

[5] W. R. Stoeger, C.-M. Xu, G. F. R. Ellis, and M. Katz (in preparation).

[6] P. J. E. Peebles, *The Large-Scale Structure of the Universe* (Princeton University Press, Princeton, NJ, 1980), pp. 363–372.

[7] M. Panek, Phys. Rev. D **34**, 416 (1986).

[8] L. F. Abbott and R. K. Schaefer, Astrophys. J. **308**, 546 (1986).

[9] G. F. R. Ellis, Ann. N. Y. Acad. Sci. **336**, 130 (1980).

[10] G. F. R. Ellis, South African J. Sci. **76**, 504 (1980).

[11] G. F. R. Ellis, S. D. Nel, R. Maartens, W. R. Stoeger, and A. P. Whitman, Phys. Rep. **124**, 315 (1985); **124**, 348 (1985).

[12] S. D. Nel, in *Theory and Observational Limits in Cosmology*, edited by W. Stoeger (Vatican Observatory, Rome, 1987), pp. 255–274.

[13] W. R. Stoeger, S. D. Nel, R. Maartens, and G. F. R. Ellis, Class. Quantum Grav. **9**, 493 (1992).

[14] G. F. R. Ellis, in *General Relativity and Gravitation*, edited by R. K. Sachs (Academic, New York, 1971), pp. 104–182.

[15] W. R. Stoeger, G. F. R. Ellis, and S. D. Nel, Class. Quantum Grav. **9**, 509 (1992) [paper III].

[16] W. R. Stoeger, S. D. Nel and G. F. R. Ellis, Class. Quantum Grav. **9**, 1711 (1992) [paper IV].

[17] W. R. Stoeger, S. D. Nel, and G. F. R. Ellis, Class. Quantum Grav. **9**, 1725 (1992) [paper V].

[18] G. F. R. Ellis, in *General Relativity and Gravitation*, edited by B. Bertotti, F. de Felice, and A. Pascolini (Reidel, Dordrecht, 1984), pp. 215–288.

[19] F. G. Friedlander, *The Wave Equation on a Curved Space-Time* (Cambridge University Press, Cambridge, England, 1975).

[20] W. R. Stoeger, in *Theory and Observational Limits in Cosmology*, [12], pp. 275–305.

[21] G. F. R. Ellis and W. R. Stoeger, Class. Quantum Grav. **4**, 1679 (1987).

[22] G. F. R. Ellis and M. Bruni, Phys. Rev. D **40**, 1804 (1989).

[23] C. J. Hogan, N. Kaiser, and M. J. Rees, Philos. Trans. R. Soc. London **A307**, 97 (1982).

[24] M. Katz, M. Sc. thesis, University of Cape Town, 1993.

[25] M. Turner, "Inflation after COBE" report, 1993 (unpublished).

[26] E. W. Kolb and M. S. Turner, *The Early Universe* (Addison-Wesley, Reading, MA, 1990), pp. 76–77.

[27] J. M. Bardeen, Phys. Rev. D **22**, 1882 (1980).

[28] J. M. Stewart and M. Walker, Proc. R. Soc. London **341**, 49 (1974).