New J. Phys. 9 (2007) 377, Focus Issue on Dark Energy

arXiv:gr-qc/0702082v4 1 Nov 2007

Cosmic clocks, cosmic variance and cosmic averages

David L. Wiltshire

Department of Physics and Astronomy, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand

E-mail: David.Wiltshire@canterbury.ac.nz
http://www2.phys.canterbury.ac.nz/∼dlw24/

Abstract. Cosmic acceleration is explained quantitatively, purely in general relativity with matter obeying the strong energy condition, as an *apparent* effect due to quasilocal gravitational energy differences that arise in the decoupling of bound systems from the global expansion of the universe. "Dark energy" is recognised as a misidentification of those aspects of gravitational energy which by virtue of the equivalence principle cannot be localised. Matter is modelled as an inhomogeneous distribution of clusters of galaxies in bubble walls surrounding voids, as we observe. Gravitational energy differences between observers in bound systems, such as galaxies, and volume–averaged comoving locations in freely expanding space can be so large that the time dilation between the two significantly affects the parameters of any effective homogeneous isotropic model one fits to the universe. A new approach to cosmological averaging is presented, which implicitly solves the Sandage-de Vaucouleurs paradox. Comoving test particles in freely expanding space, which observe an isotropic cosmic microwave background (CMB), possess a quasilocal "rest" energy $E = \langle\gamma(\tau,\mathbf{x})\rangle mc^2$ on the spatial hypersurfaces of homogeneity. Here $1 \le \gamma < \frac{3}{2}$: the lower bound refers to fiducial reference observers at "finite infinity", which is defined technically in relation to the demarcation scale between bound systems and expanding space. Within voids $\gamma > 1$, representing the quasilocal gravitational energy of expansion and spatial curvature variations. Since all our cosmological measurements apart from the CMB involve photons exchanged between objects in bound systems, and since clocks in bound systems are largely unaffected, this is entirely consistent with observation. When combined with a non–linear scheme for cosmological evolution with back–reaction via the Buchert equations, a new observationally viable model of the universe is obtained, without "dark energy". A quantitative scheme is presented for the recalibration of average cosmological parameters. It uses boundary conditions at the time of last scattering consistent with primordial inflation. The expansion age is increased, allowing more time for structure formation. The baryon density fraction obtained from primordial nucleosynthesis bounds can be significantly larger, yet consistent with primordial lithium abundance measurements. The angular scale of the first Doppler peak in the CMB anisotropy spectrum fits the new model despite an average negative spatial curvature at late epochs, resolving the anomaly associated with ellipticity in the CMB anisotropies. Non–baryonic dark matter to baryonic matter ratios of about 3:1 are typically favoured by observational tests. A number of other testable consequences are discussed, with the potential to profoundly change the whole of theoretical and observational cosmology.

Keywords: Cosmology, dark energy, dark matter, cosmological parameters

1. Introduction

It is a cornerstone of cosmology that the observed near exact isotropy of the cosmic microwave background radiation (CMBR), together with the assumption that our spatial location is not special – the *Copernican* or *Cosmological Principle* – leads to the conclusion that, to a reliable degree of approximation, we live in a homogeneous isotropic universe characterised by a Friedmann–Lemaître–Robertson–Walker (FLRW) geometry. However, although the matter distribution was certainly very homogeneous at the epoch of last–scattering when the CMBR was laid down, in the intervening aeons the matter distribution has become very inhomogeneous through the growth of structure, and the problem of fitting a smooth geometry to a universe with a lumpy matter distribution [1, 2] is central to relating observations to the numerical values of the averaged parameters which describe the Universe and its evolution as a whole.

Our conventional interpretation of observations within the FLRW models has yielded a standard model of cosmology in broad agreement with observations. However, this model requires that most of the matter in the universe is in forms of clumped non–baryonic dark matter and smooth dark energy, the nature of which has been described by many commentators as the greatest challenge to science. Furthermore, even if a mysterious dark energy is accepted, the parameters which provide the best fit to observation also give a model with a number of puzzling anomalies. These include: the apparent very early formation of galaxies [3]; the low power of the quadrupole in the spectrum of the CMB radiation anisotropies and the unexplained alignments of low multipoles which theory says should be random [4]; ellipticity in the CMB anisotropies consistent with the geodesic mixing expected from average negative spatial curvature [5]; and the fact that the baryon to photon ratio which best fits CMBR anisotropy data [6] gives, via big bang nucleosynthesis, a predicted lithium abundance at variance with what is observed [7]. Recent astronomical observations of one globular cluster suggest a possible resolution of the lithium anomaly in terms of stellar astrophysics [8], but this remains to be confirmed by modelling.

Another puzzle, serious enough to be described as a "crisis" [9], is that our present universe is dominated by voids which are far emptier than predicted by models of structure formation. The present clumped matter distribution is inhomogeneous, with an observed hierarchical structure, the largest structures being clusters and superclusters of galaxies bound in walls and filaments surrounding voids. Some 40–50% of the volume [10] of the universe at the present epoch is in voids of order $30h^{-1}$Mpc in diameter, h being the dimensionless Hubble parameter, $H_0 = 100h\,\mathrm{km\ sec^{-1}\,Mpc^{-1}}$, and there is much evidence for voids of 3–5 times this size [11], as well as local voids on smaller scales [12]. When considering the volume fraction of voids on all scales, it appears that our observable universe is "void–dominated" at the present epoch.

In this paper I propose that the resolution of various cosmological anomalies is related to understanding why our universe is dominated by voids, and to taking the correct average of such an inhomogeneous matter distribution to obtain the smoothed

Hubble flow. This involves careful examination of the operational understanding of the relationship between our measurements and the average FLRW geometry. I will do so by addressing central issues which are ambiguous in general relativity and often ignored in cosmology: the definition of gravitational energy, and the question of what is meant by the expansion of space and its influence on bound systems.

While the definition of gravitational energy in general is difficult and possibly unresolvable in general relativity on a completely arbitrary background, I propose that in model universes which are inhomogeneous but began from very close to an homogeneous state with scale–invariant density perturbations, as ours did, there is a clear physical answer operationally. The answer put forward here has the consequence that all average cosmological parameters must be recalibrated, as we have systematically ignored the variation in quasilocal gravitational energy in the universe at late epochs in so far as it affects the rest–energy of ideal comoving observers, and the synchronisation of their clocks with respect to ours. Understanding this point may make it possible to obtain a viable model of the universe, without a substantial fraction of dark energy at the present epoch. Furthermore, the new paradigm suggests a framework in which other cosmological puzzles could conceivably be resolved, as shall be discussed.

There are two key elements to the new solution. The first is the fact that in an inhomogeneous universe the appropriate averaged Einstein equations which describe the dynamical evolution of the universe are not the Friedmann equation, but modified equations obtained by a suitable non–linear averaging scheme, for which a number of possible alternatives exist [13, 14]. Such schemes cannot succeed, however, unless the average quantities are understood operationally rather than by convention. In this paper I argue that the second key element to obtaining a viable cosmology involves a careful examination of the physical meaning of the time parameter of the averaging scheme and its operational relationship to the proper time of observers in galaxies. Since time parameterisations in an arbitrary inhomogeneous universe are also inhomogeneous, this is not a trivial issue.

In this paper I will propose a new relationship between our measurements and those of volume–averaged observers. Since the proposal overturns a standard simplifying assumption that has been made for 80 years, but is not demanded by general relativity or observation, its intuition – already briefly outlined in a limiting case in a previous paper [15] – may not be obvious. The purpose of this paper therefore is to provide an expository clarification of the physical basis of the new solution to the fitting problem at a conceptual level. Sufficient quantitative details and numerical examples will be provided to demonstrate that it is likely that a new concordance cosmology can be found purely with matter obeying the strong energy condition. However, the numerical derivation of best–fit cosmological parameters is left to other papers [16, 17].

Firstly let me briefly outline what is already well–known about cosmic averages in an inhomogeneous universe. Conventionally we assume that the universe is homogeneous and isotropic on a suitably large scale, take smoothed averages of the curvature and energy–momentum and substitute these in the Einstein equations, giving the standard

Friedmann and Raychaudhuri equations. However, since the Einstein equations are local and non–linear, strictly speaking we should start with an inhomogeneous curvature and energy–momentum, evolve these via the Einstein equations and then take the average. This yields equations corrected by a *back–reaction*.

There are many alternatives to averaging, depending on whether one chooses to begin by averaging tensor quantities [14], or scalars [13], for example, and whether one chooses spacelike volume averages or null cone averages. As one example, for irrotational dust cosmologies, characterised by an energy density, $\rho(t,\mathbf{x})$, expansion, $\theta(t,\mathbf{x})$, and shear, $\sigma(t,\mathbf{x})$, on a compact domain, $\mathcal{D}$, of a suitably defined spatial hypersurface of constant average time, t, and spatial 3–metric, ${}^3g_{ij}(t,\mathbf{x})$, average cosmic evolution in Buchert's scheme [13] is described by the exact equations

$$3\frac{\dot{\bar{a}}^2}{\bar{a}^2} = 8\pi G\langle\rho\rangle - \tfrac{1}{2}\langle\mathcal{R}\rangle - \tfrac{1}{2}\mathcal{Q}, \tag{1}$$

$$3\frac{\ddot{\bar{a}}}{\bar{a}} = -\,4\pi G\langle\rho\rangle + \mathcal{Q}, \tag{2}$$

$$\partial_t\langle\rho\rangle + 3\frac{\dot{\bar{a}}}{\bar{a}}\langle\rho\rangle = 0, \tag{3}$$

where $\bar{a}(t) \equiv \left[\mathcal{V}(t)/\mathcal{V}(t_0)\right]^{1/3}$ with $\mathcal{V}(t) \equiv \int_{\mathcal{D}} \mathrm{d}^3x\sqrt{\det {}^3\!g}$, overdot denotes a t–derivative,

$$\begin{aligned}\mathcal{Q} &\equiv \frac{2}{3}\left\langle(\theta - \langle\theta\rangle)^2\right\rangle - 2\langle\sigma\rangle^2 \\ &= \frac{2}{3}\left(\langle\theta^2\rangle - \langle\theta\rangle^2\right) - 2\langle\sigma\rangle^2\,,\end{aligned} \tag{4}$$

and angle brackets denote the spatial volume average of a quantity, so that $\langle\mathcal{R}\rangle \equiv \left(\int_{\mathcal{D}} \mathrm{d}^3x\sqrt{\det {}^3\!g}\,\mathcal{R}(t,\mathbf{x})\right)/\mathcal{V}(t)$ is the average spatial curvature, for example. The following integrability condition follows from (1)–(4):

$$\partial_t\left(\bar{a}^6\mathcal{Q}\right) + \bar{a}^4\partial_t\left(\bar{a}^2\langle\mathcal{R}\rangle\right) = 0. \tag{5}$$

The extent to which the back–reaction, $\mathcal{Q}$, can lead to apparent cosmic acceleration or not has been the subject of much debate [18]–[20].

Despite the common misrepresentation in the literature that "we measure acceleration" since we actually measure apparent magnitudes of type Ia supernovae (SneIa), we are in reality determining luminosity *distances* as a function of redshift. The deduction of acceleration requires two time derivatives. The issue of how any time parameter, t, such as that used in (1)–(5), is related to our own clocks must therefore be central to the debate about whether cosmic acceleration can be obtained from back–reaction, and if so what its magnitude is. While the relationship between any two average cosmic time parameters can be expected to be monotonic on physical grounds, in the context of inhomogeneous cosmology it cannot simply be assumed that any time parameter we write down is necessarily the time on our own clocks. The operational meaning of average cosmic time parameters is central and cannot be ignored.

The subject of inhomogeneous cosmology is a vast one [21], and in examining the question of back–reaction I shall make particular choices. Firstly, while cosmological perturbation theory is highly important near the epoch of last–scattering, I do not believe that issues concerning back–reaction at the present epoch can be resolved in the context of perturbation theory. While perturbative approaches have naturally led to realisation of the significance of back–reaction [19], to account for "74% dark energy" the effect of back–reaction on the background of the universe would be so great that a viable quantitative model is beyond the domain of applicability of perturbation theory, and so these approaches will not be considered further in this paper. Secondly, while exact inhomogeneous models such as the spherically symmetric Lemaître–Tolman–Bondi (LTB) models [22, 23] are immensely useful, both as exact models for isolated systems in an expanding universe, or as toy models for understanding fundamental concepts, they could only be applied to the universe as a whole if one abandoned the Copernican Principle. I am interested in the problem of obtaining the correct homogeneous average from the inhomogeneous geometry within our present particle horizon volume, and I shall retain the Copernican Principle. Thus I shall not further consider approaches based on the exact LTB models [22, 23] or the exact Szekeres models [24].

The outline of this paper is as follows: an outline of broad conceptual issues and some important preliminary definitions, in particular the notions of finite infinity and the true critical density are first given in §2 and §3. The key physical ideas are then laid out. In §4 an implicit solution of the Sandage–de Vaucouleurs paradox is given. In §5 and §6 a mathematical model is presented, which may offer a simple viable model of the universe that successfully describes observations from the time of last scattering to the present epoch, without dark energy. The proof of principle is demonstrated by a numerical example. In §7 specific quantities associated with the CMB and the early universe are recalibrated, including the baryon–to–photon ratio, the sound horizon and the angular scale associated with the first Doppler peak. It is demonstrated that the broad features of the CMB anisotropy spectrum may be fitted while simultaneously resolving a number of observational anomalies. The scale of homogeneity is physically identified with the "comoving" baryon acoustic oscillation scale and consequences for variance of the Hubble flow are discussed. In §8 the physical implications of the model on relatively small cosmological scales are considered, and the potential for further cosmological tests discussed. In §9 arguments are presented as to why the new model universe is a natural consequence of primordial inflation, and broader questions concerning the primordial perturbation spectrum are discussed. A concluding discussion is presented in §10, to which a reader without much time is referred for a summary of the main results.

2. The Copernican Principle

Much of the debate about inhomogeneous back–reaction has centred on perturbation theory, and has overlooked the key point that in taking averages we must be certain that we are correctly relating observations to parameters of the average geometry. In particular, the assumption that the background geometry is very close to an FLRW geometry, with the metric

$$\mathrm{d}\bar{s}^2 = -\mathrm{d}t^2 + \bar{a}^2(t)\mathrm{d}\bar{\Omega}_k^2 \tag{6}$$

where $\mathrm{d}\bar{\Omega}_k^2$ is the 3–metric of a space of constant curvature, with $k = -1, 0, +1$, is well justified by the Copernican Principle, almost exact isotropy of the CMBR and the average isotropic Hubble flow.

In interpreting (6), it has been assumed since the early days of relativistic cosmology that the time parameter t, is to a good approximation the time on our own clocks. This simplifying assumption, made at an early point in most cosmology texts, is not required by either theory, principle or observation, as I shall now argue. Furthermore, it is not the natural choice in a far–from–equilibrium universe with the inhomogeneous structure that grows from an initially almost scale–free spectrum of density perturbations, as one would expect from primordial inflation.

By the Copernican Principle if we reside at an "average point" on a spatial hypersurface then observers at other "average points" also measure a near to isotropic CMBR, apart from a dipole anisotropy of order 10^{-3} of the mean CMBR temperature due to small peculiar velocities. However, taking our clocks rates to be very nearly the same as the clock rates at other average points on an appropriate spatial hypersurface, involves an implicit assumption over–and–above that implied by the Copernican Principle. In particular, the naïve identification of surfaces of homogeneity with surfaces of synchronicity implies that there is a single class of average "comoving" observers who measure an almost identical mean CMBR temperature. This need not be the case.

In a universe of voids and galaxies confined to bubble walls there are in fact two classes of average points to consider: (1) the mass–averaged observers residing in bound systems, typically in galaxies, where space is *not* expanding; and (2) the volume–averaged observers in freely expanding space, typically in voids which occupy the largest volume of space. Both classes of observers can measure an isotropic CMB in accord with the Copernican Principle, while measuring a *different mean temperature* and a *different angular anisotropy scale*.

Imagine a static geometry such as that of a system of an infinite number of equidistant black holes of equal electric charges and masses held in equilibrium by mutual gravitational attraction and Coulomb repulsion. On a large scale, this geometry is effectively a homogeneous landscape, whether viewed from deep in a gravitational well close to a mass source or in an almost asymptotic region equidistant from the closest sources. By gravitational time dilation, however, clock rates at the two vantage points differ, despite the overall homogeneity. In an expanding universe, the situation is more subtle, but actually quite similar. In particular, total gravitational energy is very important in inhomogeneous universes – a point which was realised long ago by Bondi [23] in his study of the spherically symmetric LTB models [22]. In freely expanding situations, *quasilocal* gravitational energy differences can also be significant.

We reside in a gravitationally bound system: locally space has not been expanding for over 10 billion years. This has the consequence that measurements by our local clocks relate parametrically to solutions of the geodesic equations in the Schwarzschild geometry centred on our sun, with a time parameter related to what is effectively almost a Killing vector, $\partial/\partial\tau$. To match our clock rate, τ, which has been effectively frozen in for billions of years to the clock rate of the idealised comoving observers at *volume–averaged points* of the average geometry (6) involves matching geometry from the scale of stars, to galaxies, to galaxy clusters, to bubble walls, and ultimately a homogeneous scale. Until one solves this fitting problem [1, 2] the assumption that our clock rate, τ, closely matches the parameter, t, of (6) is an ansatz.

The typical justification for assuming the conventional clock ansatz to be reliable comes from considerations such as that of particle motion in the Schwarzschild and Kerr geometries. In the Schwarzschild geometry, a radially moving clock of rest mass, m, with locally measured velocity βc on a geodesic possesses a conserved energy of

$$E = (1 - r_{\rm s}/r)\gamma mc^2, \qquad \gamma = (1-\beta^2)^{-1/2}, \tag{7}$$

as follows directly from Killing's equations. Given values of the ratio of the Schwarzschild radius, $r_{\rm s}$, of galaxies to their radius, r, and similar calculations for objects in closed bound orbits, the effect of gravitational binding energy appears to be negligible‡, and is only significant in the vicinity of highly compact objects such as neutron stars and black holes. However, considerations with regard to gravitational binding energy do not translate directly to the quasilocal kinetic energy of expansion and gravitational energy associated with spatial curvature variations, as I shall argue further in the next section.

The conventional clock ansatz was verified as being consistent by Einstein and Straus [26] in the Swiss cheese model in which spheres are excised from a dust FLRW model and replaced by point sources of mass equal to that of the excised dust in the holes§. Such a model would be accurate if the universe did actually consist of isolated galaxies moving uniformly in the "coins on the balloon" analogy described in undergraduate texts. However, the observed void / bubble wall structure indicates that the actual universe is quite different from this hypothetical situation. In fact, the difference is important enough to be a phenomenological puzzle. This puzzle, the *Sandage–de Vaucouleurs* paradox‖, arises since we expect that the statistical scatter in peculiar velocities of galaxies as a fraction of their "recession velocity" should be large until the scale of homogeneity is approached. In fact, on the scale of 20 Mpc – of order 10% of the scale of homogeneity – the scatter ought to be so large that no linear Hubble flow should be derivable, statistically speaking. Yet, 20 Mpc is the local scale over which

‡ More rigorous discussions of gravitational binding energy for a perfect fluid source in a quasilocal framework in stationary spacetimes have been given by Katz, Lynden-Bell and Bičák [25].

§ The clock rate for an observer in the Einstein-Straus vacuole differs from that of the comoving clock [27]. However, the differences turn out to be observationally negligible, as has been discussed in detail by Harwit [28]. I thank a referee for bringing this work to my attention.

‖ My nomenclature derives from the fact that this problem was originally raised by Sandage and collaborators [29] in objection to de Vaucouleurs' hierarchical cosmology [30] before the evidence for the void structure of the universe was as good as it is now. In the literature it sometimes called the "Hubble–de Vaucouleurs paradox" [31, 32] and sometimes the "Hubble–Sandage paradox" [33].

Hubble originally obtained his famous linear law. By conventional understanding this does not make sense.

3. Where is infinity?

It is well–known that the definition of gravitational energy in general relativity is difficult, since space itself carries energy and momentum. The dynamical nature of space in general relativity is well illustrated by the phenomenon of frame–dragging in the Kerr geometry, whereby an infalling test particle initially on a radial geodesic will rotate more and more with respect to spatial infinity, the closer it gets to the source of the geometry. The best attempts at quasilocal definitions of energy in general relativity involve integrals over 2–surfaces [34]–[36], but have not been widely applied in cosmology.

It is intrinsic to the physical assumptions of general relativity that an expanding universe must possess some sort of gravitational *kinetic energy of expansion*; yet this is an issue which is not considered much beyond loosely identifying the l.h.s. of eq. (1) with such a term when we consider the Friedmann equation (with $\mathcal{Q} = 0$). Our understanding of such a kinetic energy is based largely on Newtonian thinking, and indeed in the ad hoc Newtonian derivation of the Friedmann equation, the correspondence is exact, while the total energy is associated with the spatial curvature term up to an overall sign. To make progress, we must extend these concepts beyond the Newtonian limit.

Our best understanding of gravitational energy is for exact solutions where space is assumed to be asymptotically flat or (anti)-de Sitter. Since the universe is not asymptotically flat, we need an understanding of gravitational energy that extends beyond bound regions. In his pioneering work on the fitting problem [1], Ellis introduced the notion of *finite infinity*, "*fi*" , as being a timelike surface within which the dynamics of an isolated system such as the solar system can be treated without reference to the rest of the universe. No system is truly isolated since we continually receive electromagnetic and gravitational radiation from distant parts of the universe. However, this radiation is so small that the solution to the problem of geodesic motion within the isolated system can be considered without reference to the rest of the universe. Ellis' suggestion provides a notional answer to the question of: "Where is infinity?". Within finite infinity a solution might be considered to be almost asymptotically flat, and governed by "almost" Killing vectors.

I now propose to modify Ellis' suggestion slightly and will identify finite infinity in terms of average expansion and its relationship to the *true critical density.*

3.1. The true critical density

It is a direct consequence of averaging in an inhomogeneous universe that the average internal energy density of a dust universe that we would measure at the present epoch is not the same as the time evolution of the average density that we would have measured at some time in the past. In terms of Buchert's scheme, the non–commutativity of averaging and time evolution is described by the exact relation [13]

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\Psi\rangle - \langle\frac{\mathrm{d}\Psi}{\mathrm{d}t}\rangle = \langle\Psi\theta\rangle - \langle\theta\rangle\langle\Psi\rangle \tag{8}$$

for any scalar, Ψ, such as the internal energy density, ρ.

In Buchert's scheme it is natural to rewrite (1) as

$$\Omega_M^{\mathcal{D}} + \Omega_k^{\mathcal{D}} + \Omega_{\mathcal{Q}}^{\mathcal{D}} = 1, \tag{9}$$

where

$$\Omega_M^{\mathcal{D}} \equiv \frac{8\pi G\langle\rho\rangle}{3\bar{H}^2}, \qquad \Omega_k^{\mathcal{D}} \equiv -\frac{\langle\mathcal{R}\rangle}{6\bar{H}^2}, \qquad \Omega_{\mathcal{Q}}^{\mathcal{D}} \equiv -\frac{\mathcal{Q}}{6\bar{H}^2}, \tag{10}$$

and $\bar{H} \equiv \dot{\bar{a}}/\bar{a}$. It should be noted that density parameters are considered as fractions of the region–dependent quantity $3\bar{H}^2/(8\pi G)$, which does not, however, play the role of a critical density in delineating the critical case between a "closed" and an "open" universe.

If we fit an FLRW model to the averaged geometry then there are further issues associated with matching volumes of a homogeneous isotropic model, which Buchert and Carfora have described by the slogan [37]: "Cosmological parameters are dressed". Quantities such as $\Omega_M^{\mathcal{D}}$ must still be corrected by volume factors to get the equivalent "dressed" parameters of the equivalent FLRW model, which according to an earlier estimate of Hellaby [38] can give corrections of 10–30%. Once again, however, the dressed parameters are still region dependent quantities, and one does not have the notion of a critical density. So is there a notion of critical density, and if so where is it to be found?

I will now make the following crucial physical observation. By the evidence of the CMB, the universe at last scattering was very close to being truly homogeneous and isotropic. Therefore an operational definition of critical density does exist, provided we assume the Copernican Principle and accept that the universe was globally smooth at that epoch and not just in our present past horizon volume. This might also be viewed as a direct consequence of primordial inflation, as is further discussed in §9.

At the epoch of last scattering, t_i, the Hubble expansion was uniform, as the local velocity perturbations were tiny. Given a uniform initial expansion rate there must have existed a uniform critical density of matter required for gravity to be able to eventually bring that expansion to zero. This critical density, $\rho_{\rm cr}(t_i)$, therefore sets *a universal scale* which delineates the boundary between density perturbations which will become bound, as opposed to density perturbations which are unbound.

This may seem a trivial point. However, when one considers an inhomogeneous evolution it is clear that the average Hubble parameter on a given domain does not correspond to the time evolved critical density, whether in a dressed or undressed form. The naïve use of the Friedmann equation in cosmology to date means that we could well be making a gross error in choice of background in structure formation studies. In particular, we estimate the critical density by extrapolating back in time using our present Hubble parameter, H_0, assuming the evolution of the universe is smooth and that

$3H_0^2/(8\pi G)$ is the critical density at the present epoch, when it is not. If the total density of the universe is very close to one, this has the effect that we can *mis-estimate* the background density of the universe at early epochs where structure formation boundary conditions are set. We are in effect perturbing about the wrong background by implicitly assuming that the average density at the present epoch, as determined by the recent past within our past light cone, is identical to the universe as a whole. By cosmic variance there is no reason to expect this to be the case.

3.2. Location of finite infinity

We will define *finite infinity* as the timelike boundaries which demarcate the boundary between bound – or more strictly *potentially* bound – systems and unbound systems, assuming that the observable universe is at present void–dominated. At the technical level we will propose the following working definition:

> With respect to a foliation of spacetime by spacelike hypersurfaces, finite infinity is identified with the set of *timelike boundaries* of (disjoint) compact domains, $\mathcal{F}_I$, within which the *average* expansion vanishes, while being positive outside:

$$\begin{array}{l}\text{(i)}\ \ \langle\theta(p)\rangle_{\mathcal{F}_I} = 0;\\ \text{(ii)}\ \exists\ \mathcal{D}_I \text{ such that } \mathcal{F}_I \subset \mathcal{D}_I \text{ and } \theta(p) > 0 \quad \forall\, p \in \mathcal{D}_I \setminus \mathcal{F}_I\end{array} \tag{11}$$

The index, I, is taken to run over the disjoint domains. Thus finite infinity, $\mathit{fi} \equiv \cup_I \partial\mathcal{F}_I$. Finite infinity becomes operationally defined only once collapsing regions form, and each $\mathcal{F}_I$ is centred on a region which is initially collapsing. Furthermore, the average is a volume average over the *smallest* scales on which (11) applies. Each $\mathcal{D}_I$ must be contained within the particle horizon volume at any epoch.

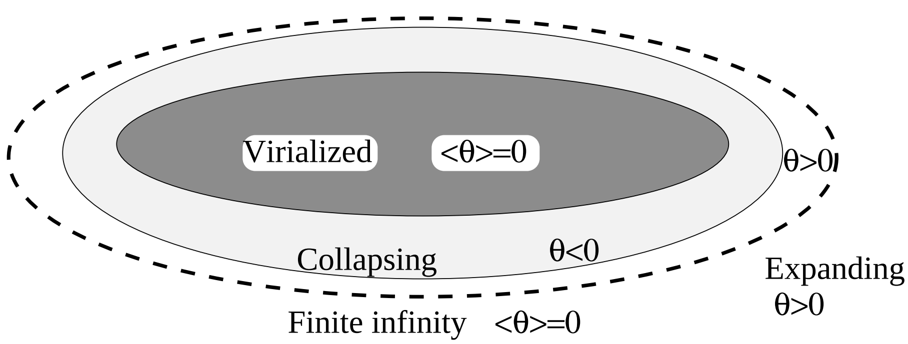

Figure 1. *A schematic illustration of the notion of finite infinity,* fi*: the boundary (dashed line) to a region with average zero expansion inside, and positive expansion outside. It may or may not contain collapsing regions.*

Some comments are in order. At first sight, it may seem more natural to identify finite infinity with the set of timelike boundaries of (disjoint) compact domains, $\mathcal{F}_I$, outside which the local expansion is positive, rather than using an average. However, we must bear in mind that pressure, shear and vorticity are neglected here. Vorticity certainly becomes important when regions collapse, and our present study is within the context of an averaging scheme in which the vorticity is neglected. Therefore it seems that we can only talk about *averaged* expansion. The specification of the spatial hypersurfaces on which domains are averaged is deferred until the next section.

We also note that while each $\mathcal{F}_I$ is centred on an initially collapsing region, to obtain an averaged zero expansion one must extend $\mathcal{F}_I$ outwards to connected regions with positive expansion which will ultimately cease expanding. There is an intrinsic ambiguity in assigning the shape to the boundary; though there are probably ways to do this in terms of a surface of minimal proper distance to regions with $\theta(t,\mathbf{x}) \leq 0$ on the hypersurfaces in question. Thus the *zero expansion boundaries*, at which $\theta(t,\mathbf{x}_0) = 0$, lie within the finite infinity regions. The zero expansion boundary may be thought of as the instantaneous "tipping point" which separates regions where space is not expanding from regions where it is expanding. Finite infinity represents a boundary to the region within which space would *ultimately* stop expanding if its entire future evolution were determined from the local dynamics of the $\mathcal{F}_I$ region alone; a local, rather than a global statement.

The idea behind our definition – which is the crucial property we wish to preserve should the technical definition (11) require further refinement – is that the true critical density may now be *defined* by

$$\rho_{\rm cr}(\tau) = \langle \rho(\tau,\mathbf{x}) \rangle_{\mathcal{F}_I} \tag{12}$$

for each $\mathcal{F}_I$ defined by (11). Our definition encompasses more situations than envisaged in the original suggestion of Ellis [1], since it is our aim that it should apply as soon as regions start collapsing. Since galaxies and clusters of galaxies are still growing by matter infall at the present epoch, it would not be reasonable to try to define finite infinity for groups of galaxies as a static boundary at a fixed proper distance from the barycentre of a mass concentration. Thus while Ellis may have had something closer to the zero expansion surfaces in mind, our version has greater utility as it will correspond to a surface where space is locally expanding as long as the observed universe is, and the true critically density will be directly related to the local expansion, $\theta_{\mathcal{F}_I} = 3\bar{H}$, at finite infinity. Here local expansion refers to a small region straddling the finite infinity boundary in Fig. 1, within which $\theta > 0$.

Physically finite infinity would have to be located well outside the concentrated visible mass and invisible halo mass of galaxies, and clusters of galaxies. Essentially one needs to include enough low density regions surrounding the dense cores of bound structures in order to get an average density which corresponds to the present true critical density. As long as bound structures are growing the proper distance from the barycentre of some $\mathcal{F}_I$ to its finite infinity boundary will increase. The rate of increase of this distance will be commensurate with the locally measured Hubble flow, $\bar{H}$, when the growth of density contrasts is in the linear regime but will differ in general, particularly at late epochs.

By our definition the regions $\mathcal{F}_I$ should be seen as analogues of the spheres cut out in the Einstein–Straus solution, which also possess a non–static boundary. The principal differences are: (i) finite infinity contains an average rather than exact geometry; (ii) no assumptions about homogeneity are made beyond finite infinity. Beyond finite infinity we find regions below the true critical density, which are non–uniformly distributed. In particular, we expect a non–uniform situation in which the expansion within the filamentary structures of the bubble walls is different to the expansion in the voids, if referred to a single set of clocks.

4. Gravitational energy and the definition of homogeneity

Since a broadly isotropic Hubble flow is observed, it is clear that an average sense of homogeneity must exist despite the observed large inhomogeneities in the matter distribution at the present epoch. Reconciling these two facts – which find their quantitative expression in the Sandage–de Vaucouleurs paradox – one must face what critics of inhomogeneous back–reaction have described as the "seemingly impossible burden of explaining why the universe appears to be so well described by a model that has only very small departures from a FLRW metric" [20].

This section contains the essential conceptual arguments of my proposal. The "seemingly impossible burden" can be resolved not by departing from an average FLRW geometry, albeit one that evolves on average with the inclusion of back–reaction, but by a careful reinterpretation of the relationship between the parameters of the average homogeneous geometry and our own observations. My proposal will simultaneously clarify the role of gravitational energy in the averaged cosmology, and resolve the Sandage–de Vaucouleurs paradox, issues which are fundamental but otherwise unanswered.

4.1. The locally synchronous gauge

To set the scene, let us carefully examine the gauge choices that have been made in Buchert's averaging scheme, and the interpretation of the parameters. For a globally hyperbolic manifold, we can make a standard $3+1$-split

$$\mathrm{d}s^2 = -\boldsymbol{\omega}^0 \otimes \boldsymbol{\omega}^0 + g_{ij}(\tau,\mathbf{x})\,\boldsymbol{\omega}^i \otimes \boldsymbol{\omega}^j, \tag{13}$$

where

$$\begin{aligned}\boldsymbol{\omega}^0 &= \gamma(\tau,\mathbf{x})\,\mathrm{d}\tau,\\ \boldsymbol{\omega}^i &= \mathrm{d}x^i + \beta^i(\tau,\mathbf{x})\,\mathrm{d}\tau.\end{aligned} \tag{14}$$

Here $\gamma(\tau,x^k)$ is the *lapse function*, which measures the difference between coordinate time, τ, and proper time, t, on curves normal to hypersurfaces Σ_τ, the unit normal being $n_\alpha = (-\gamma,0,0,0)$, or $n^\alpha = \gamma^{-1}(1,\beta^i)$. Also, $\beta^i(\tau,x^k)$ is the *shift vector*, which measures the difference between a worldline with fixed spatial coordinate, and the point reached by following the worldline along the normal $\mathbf{n}$ from one hypersurface to the next.

On account of the Bianchi identities, the Einstein equations include the Hamiltonian and momentum constraints. As a consequence, four out of the ten algebraically independent components of the metric (13)–(14) can always be removed locally by gauge choices at the dynamical level. The choice made by Buchert is to assume that

(i) $\beta^i(\tau, \mathbf{x}) = 0$, which is to say that the coordinates are *comoving*;

(ii) $\gamma(\tau, \mathbf{x}) = 1$, which is to say that if the unit normal, $\boldsymbol{n}$, is assumed to be the 4–velocity then τ is the local proper time¶.

Buchert shows that these choices can always be made in irrotational dust cosmologies [13]. In the presence of a perfect fluid with non–zero pressure, the choice $n^\alpha n_\alpha = -1$ does not generally lead to $\gamma(\tau, \mathbf{x}) = 1$ [39]. However, our concern here is primarily with dust cosmologies since the universe can be assumed homogeneous and isotropic before last scattering at the level of accuracy we require. We will be averaging over scales larger than the domains bounded by finite infinity, and it will be assumed that at those scales pressures, shear and vorticity can be neglected.

The physical implications of Buchert's gauge choices need to be examined. The choice that $\beta^i(\tau, \mathbf{x}) = 0$ can be understood as the statement that the momentum flux within space is neglected at the level of averaging. In particular, spatial variations in vorticity, i.e., the angular momentum content of space, are neglected, as are non–expanding regions as compared to expanding regions. The choice that $\gamma(\tau, \mathbf{x}) = 1$ can be understood as the statement that variations in the gravitational energy of space can also be neglected at the level of averaging.

It is clear that neither of these gauge choices is appropriate if one actually wishes to consider the worldlines in collapsing or virialised regions as physically distinct from worldlines in expanding regions. Since timelike geodesics converge in collapsing regions comoving coordinates cannot be adopted globally, but only in an averaged sense. With respect to comoving observers in expanding regions, collapsing regions should be physically distinguishable in terms of a momentum flux which differs from the average expansion within the spatial hypersurfaces. Nonetheless, provided the averaging scale is much larger than collapsing or virialised regions then the choice $\beta^i = 0$ is a reasonable one provided all measurements are referred to observers in expanding regions. Furthermore, since angular momentum perturbations decay in the linear regime, neglecting vorticity is reasonable. It is only after overdense regions break away from the Hubble flow that rotation cannot be neglected. As long as we are averaging over suitably large scales, the effect of vorticity will be a correction, but a small one. The magnitude of its correction might be judged from any net circulation of galaxy clusters around voids.

The choice that $\gamma(\tau, \mathbf{x}) = 1$ is less reasonable as a physical assumption, since density contrasts grow even in the linear regime when one is very close to matter homogeneity, suggesting possible large variations in gravitational energy. However, provided all measurements are referred to the same single class of average observers in freely expanding space, then of course Buchert's choice $\gamma(\tau, \mathbf{x}) = 1$ is consistent.

¶ The choice $\gamma = 1$ is usually part and parcel of what are termed "comoving coordinates". However, strictly speaking $\gamma = 1$ is an independent gauge choice from $\beta^i = 0$.

The crucial issue, however, is that all observers in bound systems – such as ourselves and all other galaxies – do not reside in regions of locally expanding space. Therefore they do not follow *average* geodesics tangent to the average fluid 4–velocity, $\bar{n}^\mu$. While Buchert's scheme may be consistent if applied to volume–averaged clocks, variations in gravitational energy and its effect on physical clocks *cannot* be neglected once back–reaction becomes completely non–linear. The average surfaces of homogeneity are not necessarily surfaces of synchronicity of local proper time.

To make this point more transparent we must recall that particle interactions in relativity involve conservation of 4–momentum, not 4–velocity. Similarly, Einstein's equations involve the energy–momentum tensor. Although a dust cosmology has an energy–momentum tensor

$$T^{\mu\nu} = \rho\,\bar{n}^\mu \bar{n}^\nu \tag{15}$$

this merely represents the *internal energy* of the fluid. In cosmology the kinetic energy of the expansion of space is contained in the l.h.s. of Einstein's equations. This is an inevitable consequence of the fact that on account of the equivalence principle there is no local definition of gravitational energy, but at best quasilocal definitions (see, e.g., [34]–[36]). The kinetic energy content of an expanding space involves motion and cannot be localised, and so cannot be expressed in the internal form appropriate to an energy–momentum tensor; instead it is contained in the Einstein tensor. Similarly spatial curvature cannot be measured at a point but only by geodesic deviation. Thus variations in spatial curvature between galaxies and voids are encoded in variations of the total quasilocal gravitational energy within the Einstein tensor. Since the walls containing galaxies are assumed to be near to spatially flat, with large negative spatial curvatures only pertaining to voids, the quasilocal curvature variations can well be the dominant contribution to gravitational energy differences.

By the principles of general relativity, the physical effect of gravitational energy on local clocks must nonetheless be real, even if it cannot be localised, except in an averaged sense over a region. Furthermore, if local spatial curvature differs systematically between bound systems and the volume average, then we cannot naïvely assume that our measured angular positions of the Doppler peaks in the spectrum of CMB anisotropies will be the same as those at the volume average. As soon as one is dealing with a genuinely inhomogeneous cosmology then the operational understanding of our measurements in relation to those of the average geometry of the universe is a nontrivial question which must be confronted.

4.2. The quasilocally uniformly expanding gauge

I propose that the ultimate rigorous understanding of the relationship between local measurements in bound systems and measurements at an average location in freely expanding space, will be through an extension of quasilocal formulations [25],[34]–[36] to cosmology. Such an approach has yet to be developed; what is outlined here is an attempt to come to grips with its essential physical principles. As such the proposal is not yet complete. I will write down a quantitatively viable model universe in §5, while leaving numerous issues to be further refined.

One general feature of many quasilocal approaches is the subtraction of an energy–momentum integral of a fiducial geometry from an energy–momentum integral of some more general system – usually a bound system. Typically spatial infinity in an asymptotically flat spacetime provides the fiducial geometry. There one has a notion of static or zero–angular momentum observers, with tangent vectors appropriately defined in terms of timelike Killing vectors. This leads to an appropriate notion of gravitational binding energy [25]. The idea here, is that in the actual universe similar observers are still the relevant class for defining the reference point with respect to which we must quantify the gravitational energy of expansion and global spatial curvature variations relative to the bound systems within which we live. We no longer have exact Killing vectors, but approximate Killing vectors within virialised regions, and spatial infinity in asymptotically flat geometries is replaced by finite infinity, defined in terms of the scale set by the true critical density.

It is usually assumed implicitly that stars and black holes in distant galaxies are described in terms of asymptotically flat geometries with a stationary Killing vector, ξ^μ, with clock rates synchronised to our own despite the fact that the intervening space is expanding and not described by timelike Killing vectors. The issue of how this is to be achieved mathematically is generally not addressed beyond the approximations used in the Swiss cheese model. I will begin from the premise that the specification of finite infinity – a direct physical consequence of the initial uniform expansion afforded by primordial inflation – does allow a means of "asymptotically" synchronising the clocks corresponding to approximate Killing vectors in disjoint bound systems, even when clocks in the freely expanding space between such regions are not synchronised but vary in a way which reflects the underlying gravitational energy variations.

The global vector field, $\xi^\mu = \left(\frac{\partial}{\partial\tau}\right)^\mu$, of the geometry (13)–(14) is assumed to be an "almost" Killing vector within finite infinity regions, but to differ from the local normal to the spatial hypersurface in general. In the absence of exact asymptotic flatness, it is normalised according to

$$\langle -\xi^\mu n_\mu \rangle_{\mathcal{F}_I} = \langle \gamma(\tau, \mathbf{x}) \rangle_{\mathcal{F}_I} \equiv \gamma_{\mathcal{F}_I}(\tau) = 1 \qquad (16)$$

for all values of *wall time*, τ. Within the virialised regions within finite infinity it is assumed that the average geometry is approximately a flat Minkowski geometry. The timelike direction of this geometry coincides with the "almost" stationary Killing vector ξ^μ. To this extent the time variation of γ can be neglected, and we can define static observers near mass concentrations within the virialised regions in the usual sense as observers at fixed spatial coordinates with 4–velocity $W^\mu = \xi^\mu/(-\xi^\nu\xi_\nu)^{1/2} = \gamma(\mathbf{x})^{-1}\xi^\mu$ and 4–acceleration $a_\mu = \nabla_\mu \ln\gamma(\mathbf{x})$.

Outside finite infinity in freely expanding space it is assumed that when the geometry (13)–(14) is averaged over regions within voids, in which vorticity, shear and pressure can be neglected, the shift vector then vanishes at the level of averaging, $\beta^i = 0$, while for any spatial domain, $\mathcal{D}_I$, entirely contained within a void

$$\gamma_{\mathcal{D}_I}(\tau) \equiv \langle -\xi^\mu n_\mu \rangle_{\mathcal{D}_I} > 1.$$

The fact that the lapse function is generally greater than one is a consequence of the positive gravitational energy associated with negative spatial curvature. The value of $\gamma_{\mathcal{D}_I}$ thus obtained is region–dependent. If we consider a sequence of disjoint spatial regions of equal proper volume within a void then we expect that

$$1 = \gamma_{\mathcal{F}_I}(\tau) < \gamma_{\mathcal{D}_1}(\tau) < \gamma_{\mathcal{D}_2}(\tau) < \cdots < \gamma_{\mathcal{D}_C}(\tau),$$

i.e., that $\gamma_{\mathcal{D}_I}$ is progressively greater for averaging regions closer and closer to the void centre in region $\mathcal{D}_C$. Effectively, with respect to finite infinity observers, a comoving particle of rest mass, m, in freely expanding space possesses a quasilocal "rest" energy $E = \langle \gamma(\tau, \mathbf{x}) \rangle mc^2$, which increases towards a void centre.

For cosmological scales we are interested in the combined average over all such spatial scales contained within our present particle horizon volume, $\mathcal{H}$. We will use an overbar to indicate this "global" average of the lapse function[+]:

$$\bar{\gamma}(\tau) \equiv \langle \gamma(\tau, \mathbf{x}) \rangle_{\mathcal{H}} = -\xi^\mu \bar{n}_\mu = \langle -\xi^\mu n_\mu \rangle_{\mathcal{H}} \tag{17}$$

The volume–average geometry obtained by averaging (13), (14) over the horizon volume is then

$$\mathrm{d}s^2 = -\bar{\gamma}^2(\tau)\,\mathrm{d}\tau^2 + \bar{g}_{ij}(\tau, \mathbf{x})\,\mathrm{d}x^i \mathrm{d}x^j\,, \tag{18}$$

where the average of the spatial metric, $\bar{g}_{ij}$, remains to be discussed.

We note that as is conventional the energy–momentum tensor is given by (15), where the 4–velocity $\bar{n}^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}t}$ is written in terms of local proper time, t, on geodesics of dust "particles" which actually represent averaging regions which are small enough that the variation of gravitational energy is insignificant. The important physical distinction is that the time parameter τ of wall observers in typical galaxies is not locally defined on a volume–average "dust geodesic". Equivalently, whereas the average dual normals, $\bar{n}^\mu$, may be considered to be tangents to the comoving average dust geodesics with affine parameter t; they are geodesics with non–affine parameter, τ, if the geodesic equation is rewritten in terms of $\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}$. Since $\bar{n}^\mu = \bar{\gamma}(\tau)^{-1}\xi^\mu$, there is a sense in which $\bar{n}^\mu$ may be considered to be the volume–averaged extension of the vector field, W^μ, associated with a static observer within finite infinity.

In establishing a *non–locally synchronous gauge* set by the clocks at finite infinity, we still have one gauge freedom remaining, as the condition $\langle \gamma(\tau, \mathbf{x}) \rangle_{\mathcal{F}_I} = 1$ merely fixes a reference normalisation of clocks, which might equally have been chosen in reference to observers at void centres, if that was where typical observers were located. To complete

[+] One must be careful to distinguish this averaged expression from the case of tilted cosmologies in models with perfect fluids with non–zero pressure [40]. There one has a similar non–averaged relation, $U^\mu n_\mu = -\gamma$, which arises, however, as a result of a *local* boost of the unit normal, n^μ, relative to the local fluid 4-velocity, U^μ, which has nontrivial 4-acceleration.

the identification of surfaces of average homogeneity, we will therefore choose the gauge by the condition that when averaged over regions, $\mathcal{D}$, of freely expanding space, the quasilocally measured expansion is a uniform function of local proper time, independent of spatial location:

$$\frac{\mathrm{d}\ell_r(t)}{\mathrm{d}t} = v_r(t) \tag{19}$$

where $\ell_r \equiv \mathcal{V}^{1/3} = \left[\int_{\mathcal{D}_I} \mathrm{d}^3x\sqrt{{}^3\!g}\right]^{1/3}$, for each averaging region $\mathcal{D}_I$. We can rephrase this in terms of a conventional Hubble parameter by choosing the averaging regions to have equal fiducial proper volumes, even though their spatial curvature will typically vary. With this understanding, an equivalent statement is

$$\frac{1}{\ell_r(t)}\frac{\mathrm{d}\ell_r(t)}{\mathrm{d}t} = \frac{1}{3}\langle\theta\rangle_{\mathcal{D}_1} = \frac{1}{3}\langle\theta\rangle_{\mathcal{D}_2} = \cdots = \bar{H}(t), \tag{20}$$

t being the local proper time in each region. It must be assumed that each averaging region, $\mathcal{D}_I$, is freely expanding at its boundary. Any region, $\mathcal{F}_I$, bounded by finite infinity must be entirely encompassed within a $\mathcal{D}_I$: $\mathcal{F}_I \subset \mathcal{D}_I$, $\forall\, I$. No averaging region need necessarily contain a finite infinity – in particular void regions will correspond to regions $\mathcal{D}_J$, with $\theta(t,\mathbf{x}) > 0$ throughout the region as long as the universe is expanding as a whole – which is assumed throughout this paper.

Eq. (20) can be understood as the physical statement that while the locally measured proper times, t, and local spatial curvature may both vary over the hypersurfaces of average homogeneity – the *quasilocally measured* average expansion is homogeneous. This provides an *implicit* resolution of the Sandage–de Vaucouleurs paradox: a universe which appears to have large voids growing more rapidly than environments around virialised and collapsing regions, when measured by one set of clocks, can nonetheless have an almost uniform quasilocal proper expansion. Operationally, this is possible if we demand that ideal comoving observers measure an isotropic CMB with no peculiar velocity dipole, while making no demands on the synchronisation of their clocks. This is consistent with non–uniformity in quasilocal gravitational energy, which translates to non–uniformity in local measurements of the mean CMB temperature and local average spatial curvature, for which we have only one data point.

The proposed uniform proper expansion gauge arises by the assumption that on average the differential increase in local proper volume between voids and filaments is accompanied by a differential increase in the relative positive gravitational energy associated with negative spatial curvature, which feeds back on relative clock rates. We are talking about a circumstance in which total energy is assumed to be conserved*.

* One often sees statements to the effect that in cosmology "energy is not conserved" in a comoving volume in the presence of pressure, e.g., for a comoving photon gas. (See ref. [32] for an overview.) What is true is that *internal energy* is not conserved in such situations. However, since gravitational energy cannot be localised, one should not talk about energy conservation without a suitable quasilocal formulation.

As the universe decelerates much of the kinetic energy of expansion is converted to different forms: thermal energy and localised forms of mechanical energy, such as rotational energy, within bound systems where most of the matter is located, and to the gravitational energy associated with negative spatial curvature within voids. There is of course also a difference in the kinetic energy of the expansion of space between bound systems, where this term vanishes, and the voids where it makes a residual contribution.

To have a more direct physical understanding of the proposal, we need to specify the actual scales relevant to our observed portion of the universe. I will identify these scales as the bubble walls – namely the filamentary structures containing galaxy clusters – and voids which the bubble walls surround. Finite infinity domains are contained well inside bubble walls, where the locally measured expansion♯ defines a local Hubble flow

$$\bar{H}_{\rm w}(\tau_{\rm w}) = \frac{1}{3}\langle\theta\rangle_{\rm w} \equiv \frac{1}{a_{\rm w}}\frac{{\rm d}a_{\rm w}}{{\rm d}\tau_{\rm w}}\,.$$

The parameter $\tau_{\rm w}$ coincides with the coordinate τ of (18). We temporarily (until the end of section §5.2) add the subscript "w" to distinguish it in what follows from the time parameter $\tau_{\rm v}$ that results from normalising γ to be unity in dominant void centres.

In principle, the local Hubble flow could be measured by placing spacecraft with laser–ranging devices near finite infinity. In practice, we must make measurements from galaxies contained within zero expansion surfaces, which lie within the finite infinity regions. Operationally, we need to infer $\bar{H}$ from the expansion between galaxies which have no peculiar velocity with respect to the cosmic rest frame, within two or more disjoint finite infinity regions for which the separation of the finite infinity boundaries is very small. Since average galaxies such as ours do have small peculiar velocities with respect to the cosmic rest frame, as determined by the dipole anisotropy, in practice one would need to average over many finite infinity regions to extract the local Hubble parameter.

Likewise, within the dominant voids the locally measured expansion defines a local Hubble flow

$$\bar{H}_{\rm v}(\tau_{\rm v}) = \frac{1}{3}\langle\theta\rangle_{\rm v} \equiv \frac{1}{a_{\rm v}}\frac{{\rm d}a_{\rm v}}{{\rm d}\tau_{\rm v}}\,.$$

Since observers are found in bound systems, not in voids, a cosmological test to determine $\bar{H}_{\rm v}$ is a considerable challenge, as it requires direct access to the clocks that tick at a rate, $\tau_{\rm v}$. Since voids appear to exist on all scales from a dominant volume fraction with diameters $30\,h^{-1}$Mpc up to larger scales, and down to minivoids with diameters of order 1Mpc [12], the time parameter $\tau_{\rm v}$ must refer to the average extreme clock rate in the centres of the *dominant* voids, as measured by volume fraction. Within

♯ There is always an inherent ambiguity in use of the word "local" until one specifies a scale. In general relativity one strictly uses "local" for a measurement at a point, and other measurements are "quasilocal". However, quasilocal measurements can also be made over a great variety of scales. Since the terminology of "local" Hubble flow is conventionally used, we will assume this applies to the smallest quasilocal scales over which a linear Hubble law can be extracted, and hope that this use of the word "local" does not cause confusion as compared to our earlier stricter usage.

the filamentary bubble walls there are many minivoids, but they contribute to the expansion of the bubble walls, and the amount by which their clock rates vary will be smaller as compared to the dominant void scales.

On account of the differences in quasilocal gravitational energy,

$$\frac{\mathrm{d}\tau_{\mathrm{w}}}{\mathrm{d}\tau_{\mathrm{v}}} \neq 1\,, \tag{21}$$

so even though $\bar{H}_{\mathrm{v}} = \bar{H}_{\mathrm{w}}$, as soon as we try to construct a present horizon volume average over both walls and voids, if we refer the average to a single set of clocks, this "global average"‖ will differ from $\bar{H}$.

To an observer sitting inside a bound galaxy, within an $\mathcal{F}_I$ region, within a filamentary wall, it will appear that the Hubble rate determined from galaxies on the far side of a large local void is somewhat greater than the Hubble rate within her wall. However, if she accounted for the fact that due to quasilocal gravitational energy differences the comoving clocks within the voids are ticking faster than her own clocks, the different Hubble rates become uniform to first approximation. Void "clocks" are not observed directly – rather the photons in distant galaxies originate within disjoint finite infinity regions. Thus different apparent Hubble rates are determined by differing relative path–integrals along the null geodesics.

It must be noted that a larger apparent Hubble rate across voids as opposed to within bubble walls, is the exact opposite of Newtonian intuition, where one thinks of an exactly smooth Hubble flow, with local anisotropies being a consequence of peculiar velocities induced by mass concentrations, with larger Hubble rates expected between an observer and the directions of largest mass concentrations. I remark that in the present instance I am considering variations in the underlying Hubble flow, with peculiar velocities averaged out. This is consistent with the notion that deceleration is more rapid in regions of greater density. The Newtonian logic, which does not appear to give a very consistent picture of actual motion in the local volume [41, 42], needs to be revised – a point to which I will return in §8.

It may come as a surprise that the effect I am proposing can be large, since when thinking about the effects of gravitational energy on clocks we are most familiar with bound systems, where the effects of binding energy give only very small differences between a galactic system and spatial infinity in an asymptotically flat spacetime, similarly to the small differences expected from (7). Similarly, when we consider the integrated Sachs–Wolfe effect, we find it to be small because we consider ourselves to be at the volume average and calculate the effect of small perturbations about that average. The approach here is different because rather than considering differential effects about the volume average, I am considering the position of the observer to be of crucial importance. Since galaxies are bound systems and are *not* at the volume–average, there are potentially nontrivial aspects of general relativity to be taken into account.

‖ It is only global with respect to the entire present particle horizon volume, $\mathcal{H}$, which though representing the entire presently observable universe, is not the entire universe.

In particular, the dynamics of expanding space is intrinsically different to the dynamics of non–expanding space, and there is no inherent reason to expect that when measured with respect to wall clocks that $\gamma(\tau, \mathbf{x})$ should be only infinitesimally larger than unity in the voids. There have been 10 billion years or more for the clock rates to slowly diverge. In the absence of a dark energy component the only absolute upper bound on the difference in clock rates is that within voids $\gamma(\tau, \mathbf{x}) < \frac{3}{2}$, which represents the ratio of the local expansion rate of an empty Milne universe region to an Einstein–de Sitter one.

In short, I am proposing a physical understanding of the expansion of space which attempts to clarify various misconceptions that have plagued efforts to understand this notion. I suggest that within bound systems space really does not expand at all, just as simple model calculations in Newtonian gravity [43] and in general relativity [26] have usually suggested. However, when space does expand its effect on particles "at rest" should be locally indistinguishable from equivalent motion of particles in a static space. At a foundational level this can be understood in terms of the equivalence principle, as I shall discuss in a separate paper [44]. It demands that in considering widely separated particles we must account for the quasilocal gravitational energy variations resulting from the differential kinetic energy of the expanding universe and variations in its spatial curvature. While the average of each of these quantities appears in the Friedmann equation, their *variation* cannot be incorporated in the energy–momentum tensor by virtue of the equivalence principle. Efforts to reparameterise inhomogeneous back–reaction as a smooth dark energy field [45] can therefore never be entirely quantitatively successful, as they can never account for the fact that our measurements differ systematically from those made at the volume average.

On account of a lack of conceptual clarity and the seductive charm of the very simple FLRW models with which we can perform successful calculations while avoiding fundamental issues, we have come to a historical situation in which we misidentify quasilocal cosmological gravitational energy with "dark energy". As bound system observers who perform observations on other bound systems, which are in regions of locally non–expanding space, this circumstance is an unfortunate consequence of an observer selection effect, and failing to account for the fact that mass and volume averages can differ drastically. If nature had provided us with observable freely falling clocks in the depths of voids where space is locally expanding and negatively curved then, if my thesis is correct, observations of such clocks could well have saved us one or more decades of work in the progress of theoretical cosmology.

5. A viable model for the observable universe

The simplest nontrivial approximation is a two–scale model, in which there are two homogeneous average local cosmic times, $\tau_{\rm w}$ and $\tau_{\rm v}$, at finite infinity boundaries within the walls, and at dominant void centres respectively. Since a true critical density exists the evolution of finite infinity boundaries gives a notion of true cosmic time,

$\tau_{\rm w}$. However, this parameter differs in general from the volume–averaged comoving proper time parameter on surfaces of average homogeneity. A suitable "global average" Hubble parameter is then given by the volume weighted average over all walls and voids within the particle horizon volume, $\mathcal{H}$. The two–scale model was first introduced in ref. [15], where the wall and void scales were considered to evolve independently and the only concern was to ascertain an order of magnitude for the effect of the clock rate variations between walls and voids in determining the overall cosmological parameters. The two–scale model will now be generalised to include the coupling of the two scales via a back–reaction term.

Ultimately we should reformulate an averaging scheme formally from explicit regional averages involving quasilocal energy integrals. For the purposes of the two–scale approximation, however, it should be possible to retain Buchert's formalism, while recognising that the time parameter, t, would represent the proper time at a volume average position in freely expanding space. We should be able to obtain the broad features of a viable model universe by solving Buchert's equations, but being careful to relate our own clocks to the volume–averaged ones, and any other quantities which result from such relations. This may be seen as analogous to using the synchronous gauge in perturbation theory while taking care about identifying physical observables [46], even if calculations could be alternatively performed in a more physical gauge, the "uniform Hubble–constant gauge" [47] being the closest to our choice in §4.2.

Since our clocks, and all clocks within bound systems, are expected to differ little from those at finite infinity, we will neglect any small differences and take $\tau_{\rm w}$ to represent the time on our clocks, though ultimately this may also require some correction.

5.1. Average expansion and kinematic back–reaction

We construct the two scale model by assuming that on spatial hypersurface of average homogeneity, our present particle horizon volume can be represented as the disjoint union of regions corresponding to bubble walls containing bound systems, and of regions corresponding to voids. In this fashion, the entire present horizon volume is given by $\mathcal{V} = \mathcal{V}_{\rm i}\bar{a}^3$, where

$$\bar{a}^3 = f_{\rm wi}{a_{\rm w}}^3 + f_{\rm vi}{a_{\rm v}}^3 \tag{22}$$

Here $f_{\rm vi}$ is an initial fraction of the total volume in void regions, and $f_{\rm wi} = 1 - f_{\rm vi}$ an initial fraction of the total volume in wall regions. In other words the volumes of the void and wall regions are respectively $\mathcal{V}_{\rm v} = \mathcal{V}_{\rm vi}{a_{\rm v}}^3$ and $\mathcal{V}_{\rm w} = \mathcal{V}_{\rm wi}{a_{\rm w}}^3$, while $f_{\rm vi} = \mathcal{V}_{\rm vi}/\mathcal{V}_{\rm i}$ and $f_{\rm wi} = \mathcal{V}_{\rm wi}/\mathcal{V}_{\rm i}$. The size of the averaging domains will be taken to be such that the term representing the bubble walls is a union of domains containing finite infinity regions, $\mathcal{F}_I$, which can be assumed to be *spatially flat*. The averaging regions in the void term contain no finite infinity regions, and are assumed to have negative spatial curvature on average.

We will employ Buchert's scheme with vorticity, pressure and shear assumed to be negligible on the scale of averaging. In accord with (20)

$$\bar{H}(t) = \bar{\gamma}_{\rm w} \frac{1}{a_{\rm w}} \frac{{\rm d}a_{\rm w}}{{\rm d}t} = \bar{\gamma}_{\rm v} \frac{1}{a_{\rm v}} \frac{{\rm d}a_{\rm v}}{{\rm d}t}\,, \tag{23}$$

where

$$\bar{\gamma}_{\rm w} = \frac{{\rm d}t}{{\rm d}\tau_{\rm w}}, \qquad \text{and} \qquad \bar{\gamma}_{\rm v} = \frac{{\rm d}t}{{\rm d}\tau_{\rm v}}. \tag{24}$$

The quantity $\bar{\gamma}_{\rm w} \equiv \bar{\gamma}(\tau)$ is defined by (17); we have temporarily added the subscript "w" to distinguish, $\bar{\gamma}_{\rm w}$, the volume–average of γ with respect to wall clocks, from $\bar{\gamma}_{\rm v}$, the volume–average of γ with respect to clocks at dominant void centres.

It is convenient to refer the expansion in the voids and in the walls to one set of clocks. Since Buchert's scheme is given in terms of a volume–average time parameter, t, over both of these scales, that is the convenient parameter to use. Thus we define

$$H_{\rm w} \equiv \frac{1}{a_{\rm w}} \frac{{\rm d}a_{\rm w}}{{\rm d}t}, \qquad \text{and} \qquad H_{\rm v} \equiv \frac{1}{a_{\rm v}} \frac{{\rm d}a_{\rm v}}{{\rm d}t}. \tag{25}$$

We emphasise that neither $H_{\rm w}$ nor $H_{\rm v}$ coincides with the "local" Hubble parameter any longer in the respective wall and void locations, $\bar{H}_{\rm w}$ and $\bar{H}_{\rm v}$, which are both equal to $\bar{H}$ on account of (20) or (23). In fact, we will now have $H_{\rm w} < \bar{H} < H_{\rm v}$. Since $\bar{H}$ is the locally measured Hubble parameter at the horizon volume average location, by (22)

$$\bar{H} = \frac{1}{3} \langle \theta \rangle_{\mathcal{H}} = f_{\rm w} H_{\rm w} + f_{\rm v} H_{\rm v}, \tag{26}$$

where $f_{\rm w}(t) = f_{\rm wi} {a_{\rm w}}^3 / \bar{a}^3$ is the *wall volume fraction*, $f_{\rm v}(t) = f_{\rm vi} {a_{\rm v}}^3 / \bar{a}^3$ is the *void volume fraction*, and $f_{\rm w}(t) + f_{\rm v}(t) = 1$.

It is convenient to define the relative expansion rate

$$h_r(t) \equiv \frac{H_{\rm w}}{H_{\rm v}} < 1. \tag{27}$$

Then $\bar{H} = \bar{\gamma}_{\rm w} H_{\rm w} = \bar{\gamma}_{\rm v} H_{\rm v}$, where

$$\bar{\gamma}_{\rm w} = 1 + \frac{(1-h_r) f_v}{h_r}, \tag{28}$$

and $\bar{\gamma}_{\rm v} = h_r \bar{\gamma}_{\rm w}$. Also, by direct computation of the derivative of $f_{\rm vi} {a_{\rm v}}^3 / \bar{a}^3$,

$$\begin{aligned} \dot{f}_{\rm v} = -\dot{f}_{\rm w} &= 3(1-f_{\rm v}) \left(1 - \bar{\gamma}_{\rm w}^{-1}\right) \bar{H} \\ &= \frac{3 f_{\rm v} (1-f_{\rm v})(1-h_r) \bar{H}}{h_r + (1-h_r) f_{\rm v}} \end{aligned} \tag{29}$$

and many equivalent versions of the same relation exist in terms of $f_{\rm w}$ or $\bar{\gamma}_{\rm v}$. Combining (26) and (29) with

$$\langle \theta^2 \rangle_{\mathcal{H}} = 9 \left(f_{\rm w} H_{\rm w}^2 + f_{\rm v} H_{\rm v}^2 \right) \tag{30}$$

and (4) we find that in the absence of shear the kinematic back–reaction is given by

$$\begin{aligned} \mathcal{Q} &= 6 f_{\rm v} (1-f_{\rm v}) \left(H_{\rm v} - H_{\rm w}\right)^2 \\ &= \frac{6 f_{\rm v} (1-f_{\rm v}) (1-h_r)^2 \bar{H}^2}{\left[h_r + (1-h_r) f_{\rm v}\right]^2} \end{aligned} \tag{31}$$

Räsänen [48] has also considered a 2–scale model in Buchert's scheme¶ in terms of a single time parameter and has written down expressions for the average Hubble expansion (26) and the average deceleration incorporating a kinematic back–reaction term in a form equivalent to (31). Aside from the key fact that the physical interpretation developed in §2 and in eqs. (23), (24), (28) has not been pursued by others, another difference in Räsänen's model is that he considers a combination of voids and collapsing regions for the two scales, modelling the latter by the spherical collapse model, rather than by directly solving Buchert's equations. Since the spherical collapse model only remains a good approximation until the overdense spherical perturbations reach their maximum size, and since this would have occurred of order 10 Gyr ago or more for galaxies, Räsänen's model would appear to be an interesting toy model which exhibits the effects of apparent acceleration, but which cannot be applied to the observed universe as a whole at late epochs+.

Räsänen overlooks the possibility, we introduce here, of taking one of the scales to represent spatially flat average regions which enclose the collapsing regions, and which will have a close to Einstein–de Sitter time–dependence. This may be due to his statement that in a universe whose expansion goes smoothly from Einstein–de Sitter behaviour ($\bar{a} \propto t^{2/3}$) to empty universe behaviour ($\bar{a} \propto t$), as will be the case in the model here, there can be *"no acceleration as the variance of the expansion rate is too small"* [48]. This claim will be disproved in eq. (62) below. The claim is true insofar as it applies to volume–averaged observers in freely expanding space. However, the crucial point which to the best of my knowledge appears to have been completely overlooked by others in the study of back–reaction and averaging is that in an inhomogeneous universe writing down a timelike parameter, t, does not imbue it with physical meaning until one specifies how this parameter is *operationally* related to our own clocks. That is the issue I confront: it will turn out that the back–reaction is indeed too small to register as apparent acceleration in terms of measurements made at a volume–average location in voids. However, when translated to measurements synchronous with the wall time, $\tau_{\rm w}$, which is the parameter most closely related to our own clocks, then *apparent acceleration can be obtained.*

¶ Since Räsänen [48] considers his model to be a toy model only, he pictures the two scales as two disjoint simply connected regions, and is just considering the effects of averaging over the two regions without considering their embedding in the universe as a whole. It should be stressed that here the two scales are summed over many different disjoint regions. The manner in which these disjoint regions are embedded in the universe as a whole arises by the specification of finite infinity.

+ Räsänen claims his approximation is relevant at epochs when the universe is 10 Gyr old [48], but fails to identify what the collapsing regions are observationally. It would appear that such regions should be necessarily larger than the comoving scale on which the local Hubble flow is still observed at the present epoch [42]. While our local flow could be atypical of wall regions, given the actual structure of voids and filaments that we observe, a *spherical dust* approximation for large collapsing regions at late epochs seems hard to justify.

5.2. The dynamical equations and observational interpretation

To implement Buchert's equations we note that (3) is solved by taking $\langle\rho\rangle_{\mathcal{H}} = \rho_0 \bar{a}_0/\bar{a}^3$, as in the standard Friedmann equation, while the average curvature includes a contribution from the void regions diluted by the larger combined volume of walls and voids:

$$\langle\mathcal{R}\rangle_{\mathcal{H}} = \frac{6k_{\rm v} f_{\rm v}}{{a_{\rm v}}^2} + \frac{6k_{\rm w} f_{\rm w}}{{a_{\rm w}}^2} = \frac{6k_{\rm v} {f_{\rm vi}}^{2/3} {f_{\rm v}}^{1/3}}{\bar{a}^2}\,, \tag{32}$$

where we have assumed that the average curvatures, $\langle\mathcal{R}\rangle_{\rm v} = 6k_{\rm v}/{a_{\rm v}}^2$ and $\langle\mathcal{R}\rangle_{\rm w} = 6k_{\rm w}/{a_{\rm w}}^2$ in the voids and walls respectively, and that $k_{\rm w} = 0$ in the final step of (32). Furthermore, the kinematic back–reaction (31) may also be written as

$$\mathcal{Q} = \frac{2{\dot{f}_{\rm v}}^{\,2}}{3f_{\rm v}(1-f_{\rm v})}\,. \tag{33}$$

The independent Buchert equations (1) and (5) are then found to reduce to

$$\frac{\dot{\bar{a}}^2}{\bar{a}^2} + \frac{{\dot{f}_{\rm v}}^{\,2}}{9f_{\rm v}(1-f_{\rm v})} - \frac{\alpha^2 {f_{\rm v}}^{1/3}}{\bar{a}^2} = \frac{8\pi G}{3}\bar{\rho}_0 \frac{\bar{a}_0^3}{\bar{a}^3}\,, \tag{34}$$

$$\ddot{f}_{\rm v} + \frac{{\dot{f}_{\rm v}}^{\,2}(2f_{\rm v}-1)}{2f_{\rm v}(1-f_{\rm v})} + 3\frac{\dot{\bar{a}}}{\bar{a}}\dot{f}_{\rm v} - \frac{3\alpha^2 {f_{\rm v}}^{1/3}(1-f_{\rm v})}{2\bar{a}^2} = 0\,, \tag{35}$$

if $f_{\rm v}(t) \neq$ const. Here we have assumed that the spatial curvature of the void regions is negative, and have thus defined $\alpha^2 = -k_{\rm v} {f_{\rm vi}}^{2/3}$.

To derive cosmological parameters requires more than simply solving the coupled system (34), (35). Firstly, we observed that the time derivative in these equations is t, the time parameter at a volume–averaged location, which will be in voids. Having solved the differential equations, we must convert all quantities involving time to expressions in terms of wall time, $\tau_{\rm w}$, which is very close to the time we measure. Furthermore, the average spatial curvature at the volume average is negative, whereas the locally measured spatial curvature in wall regions is almost zero on average. These differences between the rods and clocks of wall observers and volume–average observers will both contribute to the "dressing" of cosmological parameters.

The interpretation of the solutions to (34), (35) requires particular care in relation to our own measurements. By the construction of finite infinity, when the full geometry (13)–(14) is averaged over an expanding region immediately containing finite infinity, since $\langle -\xi^\mu n_\mu \rangle_{\mathcal{F}_I} = 1$ and since the average spatial curvature within a finite infinity region is close to zero, it follows that the local average geometry near a finite infinity boundary is

$$\begin{aligned} {\rm d}s^2_{\mathcal{F}_I} &= -\,{\rm d}\tau_{\rm w}^2 + {a_{\rm w}}^2(\tau_{\rm w})\left[{\rm d}\eta_{\rm w}^2 + \eta_{\rm w}^2 {\rm d}\Omega^2\right] \\ &= -\,{\rm d}\tau_{\rm w}^2 + \frac{(1-f_{\rm v})^{2/3}\,\bar{a}^2}{{f_{\rm wi}}^{2/3}}\left[{\rm d}\eta_{\rm w}^2 + \eta_{\rm w}^2 {\rm d}\Omega^2\right]. \end{aligned} \tag{36}$$

where ${\rm d}\Omega^2 = {\rm d}\vartheta^2 + \sin^2\vartheta\,{\rm d}\varphi^2$ is the standard metric on a 2–sphere, and the second line follows on account of (22). At a void centre, the average of the local geometry (13)–(14)

over a small domain is similarly given by an effective homogeneous isotropic geometry specified by the void time, $\tau_{\rm v}$, and the local volume scale factor $a_{\rm v}$,

$$\begin{aligned} {\rm d}s^2_{\mathcal{D}_C} &= -\,{\rm d}\tau_{\rm v}^2 + a_{\rm v}^2(\tau_{\rm v})\left[{\rm d}\eta_{\rm v}^2 + \sinh^2(\eta_{\rm v}){\rm d}\Omega^2\right] \\ &= -\,{\rm d}\tau_{\rm v}^2 + \frac{{f_{\rm v}}^{2/3}\bar a^2}{{f_{\rm vi}}^{2/3}}\left[{\rm d}\eta_{\rm v}^2 + \sinh^2(\eta_{\rm v}){\rm d}\Omega^2\right]. \end{aligned} \tag{37}$$

The volume average scale factor, $\bar a$, of the geometry (38) has been constructed as the average of these two geometries via (22). If we take a common centre coinciding with that of a finite infinity region, the volume average geometry may be considered to have the form

$$\begin{aligned} {\rm d}s^2 &= -\,{\rm d}t^2 + \bar a^2(t)\,{\rm d}\bar\eta^2 + A(\bar\eta,t)\,{\rm d}\Omega^2, \\ &= -\,\bar\gamma^2(\tau){\rm d}\tau^2 + \bar a^2(\tau)\,{\rm d}\bar\eta^2 + A(\bar\eta,\tau)\,{\rm d}\Omega^2, \end{aligned} \tag{38}$$

where we have dropped the subscript "w" from both $\tau_{\rm w}$ and $\bar\gamma_{\rm w}$. Since we will no longer make explicit reference to the time measured in void centres, we will assume that τ and $\bar\gamma$ now refer to wall time in all subsequent expressions. We also note that in (38) the area quantity, $A(\bar\eta,t)$, satisfies $\int_0^{\bar\eta_{\mathcal{H}}}{\rm d}\bar\eta\,A(\bar\eta,t) = \bar a^2(t)\mathcal{V}_{\rm i}(\bar\eta_{\mathcal{H}})/(4\pi)$, $\bar\eta_{\mathcal{H}}$ being the conformal distance to the particle horizon relative to an observer at $\bar\eta = 0$, since we have chosen the particle horizon as the scale of averaging.

It must be emphasised that (38) is not locally isometric to the geometry in either the walls or the void centres, and thus it is *not* a metric ansatz that one substitutes into the field equations and then solves. Consequently, it does *not* represent a LTB model, even though the line element formally has the same form [22, 23]. This LTB form of the metric is a consequence of the fact that the effective scale factor of the volume average does not evolve as a simple time scaling of a space of constant Gaussian curvature. Its radial null geodesics nonetheless have a length which is the quantity of physical interest to a finite infinity observer making measurements on cosmological scales, which is why it is written in a spherically symmetric form.

Rather than substituting (38) as an ansatz in field equations, we first solve the averaged Einstein equations, in the form of the Buchert equations, and then reconstruct (38) as the appropriate average geometry. In particular, we can identify radial null geodesics propagating at fixed (ϑ,φ) in (38) and then reconstruct the average spatial curvature of (38) by a suitable average of the equations of geodesic deviation. Comparing (36) and (38), we see that the radial null geodesics of the two geometries coincide provided that

$${\rm d}\eta_{\rm w} = \frac{{f_{\rm wi}}^{1/3}{\rm d}\bar\eta}{\bar\gamma\,(1-f_{\rm v})^{1/3}} \tag{39}$$

along the geodesics. We note that although the volume average parameter, t, differs from wall time, τ, the radial null section of (38) coincides with that of the conformal rescaling of (36), $\bar\gamma^2(\tau)\,{\rm d}s^2_{\mathcal{F}_I}$. This follows since null geodesics are unaffected by conformal transformations. Although it is possible to identify the null sections in this manner, the complete conformally related metrics will not be isometric due to differences in spatial curvature. The conformal factor is physically postulated to encode the gravitational energy variation between wall observers and the volume average. Likewise the fact that the spatial sections are not fully related by a conformal factor has a physical origin: the volume average spatial curvature differs from the average spatially flat curvature within walls. Consequently there is an average focussing of light rays from the volume average in voids relative to observers such as ourselves within walls.

Once the wall geometry (36) is rewritten in terms of the conformal time $\bar{\eta}$ of the volume average geometry (38), the coordinate $\eta_{\rm w}$ can no longer be regarded as constant if related to the position of a volume average comoving observer. Rather we define $\eta_{\rm w}$ in terms of an integral of (39) on the radial null geodesics of (38). The wall geometry (36) therefore becomes

$$\begin{aligned} {\rm d}s^2_{\mathcal{F}_I} &= -\,{\rm d}\tau^2 + \frac{\bar{a}^2}{\bar{\gamma}^2}\,{\rm d}\bar{\eta}^2 + \frac{\bar{a}^2\,(1-f_{\rm v})^{2/3}}{{f_{\rm wi}}^{2/3}}\,\eta_{\rm w}^2(\bar{\eta},\tau)\,{\rm d}\Omega^2 \\ &= -\,{\rm d}\tau^2 + a^2(\tau)\left[{\rm d}\bar{\eta}^2 + r_{\rm w}^2(\bar{\eta},\tau)\,{\rm d}\Omega^2\right] \end{aligned} \tag{40}$$

where $a \equiv \bar{\gamma}^{-1}\bar{a}$ and $r_{\rm w} \equiv \bar{\gamma}\,(1-f_{\rm v})^{1/3}\,{f_{\rm wi}}^{-1/3}\eta_{\rm w}(\bar{\eta},\tau)$.

If any single geometry can be considered to encode relevant cosmological parameters with respect to our own measurements as wall observers, it is the geometry (40) written in terms of the global average conformal time

$$\bar{\eta} = \int_t^{t_0} \frac{{\rm d}t}{\bar{a}} = \int_\tau^{\tau_0} \frac{\bar{\gamma}\,{\rm d}\tau}{\bar{a}} = \int_\tau^{\tau_0} \frac{{\rm d}\tau}{a}\,, \tag{41}$$

which is constructed numerically from the solution to the equations (34), (35). Eq. (40) is in fact the closest thing we have to an FLRW geometry synchronous to our own clocks: it may be used to define the relevant conventional luminosity and angular diameter distances with respect to our measurements. The quantity $\bar{\eta}$ plays the role of a "comoving distance" in the conventional sense. Since we are not at the volume average, we do not need to construct the geometry (38) directly for simple distance measurements, although a knowledge of its average spatial curvature would be directly relevant to the computation of the integrated Sachs–Wolfe effect, cumulative gravitational lensing etc.

5.3. Volume–average and wall–average cosmological parameters

Our identification of the clocks and rods of wall observers relative to the volume average means that there will be differences between the dressing of parameters in the present model and that discussed by Buchert and Carfora [37]. From (40) the global horizon volume average Hubble parameter over walls and voids, as measured by the wall observer is not $\bar{H} = \frac{1}{\bar{a}}\frac{{\rm d}\bar{a}}{{\rm d}t} = \frac{1}{a_{\rm w}}\frac{{\rm d}a_{\rm w}}{{\rm d}\tau}$, but

$$\begin{aligned} H \equiv \frac{1}{a}\frac{{\rm d}a}{{\rm d}\tau} &= \frac{1}{\bar{a}}\frac{{\rm d}\bar{a}}{{\rm d}\tau} - \frac{1}{\bar{\gamma}}\frac{{\rm d}\bar{\gamma}}{{\rm d}\tau}, \\ &= \bar{\gamma}\bar{H} - \dot{\bar{\gamma}}, \\ &= \bar{\gamma}\bar{H} - \bar{\gamma}^{-1}\tfrac{{\rm d}}{{\rm d}\tau}\,\bar{\gamma}, \end{aligned} \tag{42}$$

where the overdot still denotes a derivative w.r.t. the volume average time parameter, t.

Since $\bar{H}$ is a measurable "local" Hubble parameter at any location, eq. (42) defines the relationship between this local Hubble parameter and the conventional globally measured Hubble parameter, H, according to wall observers. Since both $\bar{H}$ and H can be measured, this relationship is ultimately open to observational testing, once the scale of the "local Hubble flow" is empirically determined.

An interesting consequence of our definition of homogeneity is that even though the bare Hubble parameter has no relationship to the true critical density in Buchert's scheme for arbitrary domains of averaging, for our particular choice of domain it does. By definition finite infinity boundaries encompass regions of true critical density, and insofar as $\bar{H} = \bar{H}_{\rm w}$ is the locally measured Hubble parameter at these boundaries, then $3\bar{H}^2/(8\pi G)$ *is indeed the true critical density.* As we shall see shortly, we nonetheless must take care, as the volume average observer will perceive the critical density to take a different numerical value.

Since the horizon volume average Hubble parameter as measured by wall observers, H, differs from $\bar{H}$ according to (42), then clearly the quantity $3H^2/(8\pi G)$ is not the critical density, contrary to our usual naïve assumptions. If we live in an epoch for which the time variation of the lapse function, $\bar{\gamma}$, is small, then since $1 < \bar{\gamma} < \frac{3}{2}$ it follows from (42) that we can over–estimate the value of the critical density by a significant amount, typically by 40–80% rather than the absolute upper bound of 125%. However, a factor 40–80% is already enough to significantly change the matter budget, and has the consequence that our recalibration of cosmological parameters is going to be significant.

The bare cosmological parameters, expressed as fractions of the true critical density, $3\bar{H}^2/(8\pi G)$, are from (10) and (32)–(34),

$$\bar{\Omega}_M = \frac{8\pi G \bar{\rho}_{M0} \bar{a}_0^3}{3\bar{H}^2 \bar{a}^3}, \tag{43}$$

$$\bar{\Omega}_k = \frac{\alpha^2 f_{\rm v}^{\,1/3}}{\bar{a}^2 \bar{H}^2}, \tag{44}$$

$$\bar{\Omega}_{\mathcal{Q}} = \frac{-\dot{f}_{\rm v}^{\,2}}{9 f_{\rm v}(1 - f_{\rm v})\bar{H}^2}, \tag{45}$$

so that eq. (9) and eq. (34) coincide. These are the parameters as measured locally by volume–average observers: the time derivative in (45) refers to their local clocks.

The parameters (43)–(45) are averaged over the present particle horizon volume, $\mathcal{H}$. This does not imply that they are globally measured quantities outside our past light cone, as is implicitly assumed in FLRW models. In fitting cosmological parameters, we will work from the premise that the universe is close to critical density overall, and that our horizon volume average *at the present epoch* is below critical density. The implication is that horizon volume averages in the past would have given values $\bar{\Omega}_M \simeq 1$, $|\bar{\Omega}_k| \ll 1$, $|\bar{\Omega}_{\mathcal{Q}}| \ll 1$. This certainly would have been the case at the time of last scattering.

To determine when the universe undergoes a "void dominance" transition one must integrate the equations (34)–(35) and fit the results to all available data.

Whereas the parameters (43)–(45) are the bare or "true" cosmological parameters, our conventional parameters, based on the geometry (36) or (40), synchronous with our clocks, will differ. The question of how cosmological parameters are to be dressed is an interesting one. Indeed, if our claim in eq. (12) is correct, then there must be a sense in which if we restricted our considerations to the matter purely within walls, ignoring the void contribution, then that would give the "true critical density", with $\Omega_{M\mathrm{w}} = 1$ for some sort of dressed density parameter. This suggests that the equations (34), (35) should possess a simple first integral. Indeed, this is the case, as we now demonstrate.

In the two–scale model, the equation for the second derivative of the volume average scale factor (2), which may also be derived directly from a derivative of (34) in combination with (35), is

$$\frac{\ddot{\bar a}}{\bar a} = \frac{2\dot f_{\mathrm v}^{\,2}}{9f_{\mathrm v}(1-f_{\mathrm v})} - \frac{4\pi G}{3}\bar\rho_0\frac{\bar a_0^3}{\bar a^3}\,. \tag{46}$$

We may combine (34), (35) and (46) to obtain

$$6\frac{\ddot{\bar a}}{\bar a} + 3\frac{\dot{\bar a}^2}{\bar a^2} - \frac{2\ddot f_{\mathrm v}}{1-f_{\mathrm v}} - \frac{6\dot{\bar a}\dot f_{\mathrm v}}{\bar a\,(1-f_{\mathrm v})} - \frac{\dot f_{\mathrm v}^{\,2}}{(1-f_{\mathrm v})^2} = 0\,. \tag{47}$$

If we now multiply (47) by $24\pi G\bar\rho_0\bar a_0^3/[\dot f_{\mathrm v}\bar a - 3\,(1-f_{\mathrm v})\,\dot{\bar a}]^3$, use (43) and also note from (28) and (29) that

$$\bar\gamma = \frac{3\,(1-f_{\mathrm v})\,\dot{\bar a}}{3\,(1-f_{\mathrm v})\,\dot{\bar a} - \dot f_{\mathrm v}\bar a} \tag{48}$$

then the resulting equation may be recognised as

$$\frac{\mathrm d}{\mathrm dt}\left(\frac{\bar\gamma^2\bar\Omega_M}{1-f_{\mathrm v}}\right) = 0\,. \tag{49}$$

Its integral gives us

$$\Omega_{M\mathrm w} \equiv \frac{(1-\epsilon_i)\,\bar\gamma^2\bar\Omega_M}{1-f_{\mathrm v}} = 1, \tag{50}$$

where $\epsilon_i \ll 1$ is a small constant determined by initial conditions, since at early times $\bar\Omega_M \to 1$, $\bar\gamma \to 1$ and $f_{\mathrm v} \to f_{\mathrm{vi}} \ll 1$. Using the matter–dominated approximation back to the surface of last scattering, at $t = t_i$, the small value of ϵ_i is given by

$$\epsilon_i = 1 - \frac{1-f_{\mathrm{vi}}}{\bar\gamma_i^2\,\bar\Omega_i}. \tag{51}$$

Strictly, we should also include radiation in considering the very early time limit, and modify (50) to include $\bar\Omega_R$. However, provided photon–electron decoupling occurs within the matter dominated era – as will occur for best–fit parameter values considered later – then the effect of $\bar\Omega_R$ is negligible, and is omitted for the present considerations.

The fact that wall observers and volume average observers have different clock rates, and that the frames (38) and (40) differ by both a conformal factor and spatial curvature factor, means that there are subtleties in the definitions of densities. Just as in special relativity, where definition of internal energy density is frame dependent, so too is the corresponding definition here, where there is a conformal frame issue relating to gravitational energy and synchronisation of clocks. Systematically differing results will be obtained when averages are referred to different clocks. In the present case, the density $\bar{\rho}$ has been defined in Buchert's scheme at the volume–average position in freely expanding space. In the two scale approximation it may be considered as the sum

$$\bar{\rho} = (1 - f_{\rm v})\,\rho_{\rm w} + f_{\rm v}\rho_{\rm v} \tag{52}$$

where the mean densities of matter within walls and voids

$$\rho_{\rm w} = \frac{\bar{\rho}_0 \bar{a}_0^3}{a_{\rm w}^{\ 3}}, \qquad \rho_{\rm v} = \frac{\bar{\rho}_0 \bar{a}_0^3}{a_{\rm v}^{\ 3}} \tag{53}$$

are *referred to the volume average frame.* This volume average observer perceives that the expansion rate within the finite infinity regions, $H_{\rm w}$, is slower than the locally measured expansion rate, $\bar{H}$. Thus in fact, using (22), (25) and (53) we see that

$$\frac{3H_{\rm w}^2 \rho_{\rm w}}{8\pi G} = \frac{f_{\rm wi}\bar{\gamma}^2 \bar{\Omega}_M}{1 - f_{\rm v}} = \frac{f_{\rm wi}\Omega_{M\rm w}}{1 - \epsilon_i} \simeq \Omega_{M\rm w}\,, \tag{54}$$

consistent with (50), given that $f_{\rm wi} \simeq 1$ and $\epsilon_i \ll 1$.

Neither the volume average density parameter, $\bar{\Omega}_M$, nor the "true" density parameter, $\Omega_{M\rm w}$, will take a numerical value close to those of the "concordance" dark–energy cosmology. It is possible to define a *conventional* density parameter in terms of the effective global average scale factor, $a = \bar{\gamma}^{-1}\bar{a}$ of (40). This gives a matter density parameter "dressed" by a factor $\bar{\gamma}^3$

$$\Omega_M = \bar{\gamma}^3(\tau)\,\bar{\Omega}_M\,. \tag{55}$$

Since (40) is the closest approximation to a FLRW geometry referred to local rulers and clocks, the value of Ω_M thus defined will come the closest numerically to the conventional matter density parameter of the standard dark–energy cosmology, even though it is not a fundamental parameter in the way that both $\bar{\Omega}_M$ and $\Omega_{M\rm w}$ are. We could similarly define other dressed parameters, by introducing appropriate volume factors. We note that even though the bare parameters (43)–(45) sum to unity (9), this will not be true for such dressed parameters.

Eq. (50) has two important consequences for the characterisation of solutions. Firstly, only three of the parameters $\bar{H}$, $\bar{\Omega}_M$, $f_{\rm v}$ and $\bar{\gamma}$ are independent, meaning that other relations can be simplified. For example,

$$\bar{\Omega}_{\mathcal{Q}} = \frac{-(1 - f_{\rm v})(\bar{\gamma} - 1)^2}{f_{\rm v}\bar{\gamma}^2} = \frac{-(\bar{\gamma} - 1)^2 \bar{\Omega}_M}{1 - \bar{\gamma}^2 \bar{\Omega}_M}\,. \tag{56}$$

The global average Hubble parameter (42) as measured by wall observers is

$$\begin{aligned} H &= \bar{\gamma}\bar{H}\left[2 - \frac{3}{2\bar{\gamma}} + \tfrac{1}{2}\bar{\Omega}_M + 2\bar{\Omega}_{\mathcal{Q}}\right] \\ &= \bar{H}\left[2\bar{\gamma} - \frac{3}{2} - \frac{\bar{\gamma}(\bar{\gamma} - 1)(3\bar{\gamma} - 1)\bar{\Omega}_M}{2(1 - \bar{\gamma}^2\bar{\Omega}_M)}\right]. \end{aligned} \tag{57}$$

The second important consequence of (50) is that in place of (34) and (35) we now only need to solve two *first order* ODEs, namely (34) and

$$(2f_{\rm v}-1)\frac{\dot{\bar a}^2}{\bar a^2}+\frac{2}{3}\dot f_{\rm v}\frac{\dot{\bar a}}{\bar a}-\frac{\alpha^2 {f_{\rm v}}^{4/3}}{\bar a^2}+\frac{8\pi G}{3}\bar\rho_0\frac{\bar a_0^3}{\bar a^3}(1-\epsilon_i-f_{\rm v})=0, \qquad (58)$$

as may be found by combining (34), (48) and (50). Numerically, this is a considerable simplification. In fact, the combinations of $[f_{\rm v}(34)-(58)]$, and $[(1-f_{\rm v})\,(34)+(58)]$ can both be directly factored, leading to the even simpler equations*

$$(1-f_{\rm v})\,\frac{\dot{\bar a}}{\bar a}-\tfrac{1}{3}\dot f_{\rm v}=\sqrt{\frac{8\pi G}{3}\bar\rho_0(1-\epsilon_i)\,(1-f_{\rm v})\,\frac{\bar a_0^3}{\bar a^3}}\,, \qquad (59)$$

$$\frac{\dot{\bar a}}{\bar a}+\frac{\dot f_{\rm v}}{3f_{\rm v}}=\frac{\alpha}{{f_{\rm v}}^{1/3}\bar a}\sqrt{1+\frac{8\pi G}{3\alpha^2}\frac{\bar\rho_0\epsilon_i\bar a_0^3}{{f_{\rm v}}^{1/3}\bar a}}\,. \qquad (60)$$

5.4. Apparent cosmic acceleration

From (2), (31), (43) and (45) the bare or volume–average deceleration parameter is

$$\bar q\equiv\frac{-\ddot{\bar a}}{\bar H^2\bar a}=\tfrac12\bar\Omega_M+2\bar\Omega_{\mathcal Q}=\tfrac12\bar\Omega_M-\frac{2f_{\rm v}(1-f_{\rm v})(1-h_r)^2}{[h_r+(1-h_r)f_{\rm v}]^2}\,. \qquad (61)$$

The back–reaction contribution, $2\bar\Omega_{\mathcal Q}$, to (61) is negligible both at early times when $h_r\to1$ independently of the initial small value of $f_{\rm v}$, and at very late times when $f_{\rm v}\to1$. It is certainly possible for the back–reaction term to dominate so that apparent acceleration is obtained; it is the extent to which this is possible for reasonable parameters which is the subject of debate. The maximum value of $|\bar\Omega_{\mathcal Q}|$ is obtained at the epoch when $\ddot f_{\rm v}=0$. One may solve (35), in combination with (34), as a quadratic equation for $\dot f_{\rm v}$ at this epoch to determine whether the local minimum of deceleration is negative for a particular solution. One finds that whether the back–reaction is large enough to give acceleration depends on initial conditions together with cosmological parameters. As a simple example, if $\ddot f_{\rm v}=0$ is attained at the epoch when $f_{\rm v}=4/7$ then $2\bar\Omega_{\mathcal Q}=\frac{-1}{24}\left(1-\bar\Omega_M\right)$ at the epoch in question, so that acceleration is only attained if $\bar\Omega_M<\frac{1}{13}$ by that epoch. This would not be reasonable if $\bar\Omega_M$ were the standard matter density fraction related to our clocks and rulers, since $\bar\Omega_M(t)$ is monotonically decreasing and the value we measure today is believed to be larger.

The crucial point to observe, however, is that the bare deceleration parameter (61) is defined for the volume–averaged rulers and clocks so that debates based on equations such as (61) are misdirected if they ignore the question as to how average parameters are operationally defined in an inhomogeneous universe. To make contact with our rulers and clocks, or indeed of other observers in bound systems, we need to use dressed, or wall–average parameters. The wall–average global deceleration parameter analogous to

* The simple form (59), (60) was only found after this paper was originally submitted. The general analytic solution of eqns. (59), (60) is readily obtained, and will be presented in a separate paper [49].

the global wall–average Hubble parameter (42) is

$$\begin{aligned} q &\equiv \frac{-1}{H^2 a^2}\frac{\mathrm{d}^2 a}{\mathrm{d}\tau^2} \\ &= -\left(\frac{\dot{\bar a}}{\bar a}-\frac{\dot{\bar\gamma}}{\bar\gamma}\right)^{-2}\left[\frac{\ddot{\bar a}}{\bar a}-\frac{\dot{\bar\gamma}\dot{\bar a}}{\bar\gamma\bar a}-\frac{\ddot{\bar\gamma}}{\bar\gamma}+\frac{\dot{\bar\gamma}^2}{\bar\gamma}\right] \\ &= \left(\bar\gamma\bar H-\dot{\bar\gamma}\right)^{-2}\left(\bar\gamma^2\bar H^2\bar q+\bar\gamma\ddot{\bar\gamma}\right)+\left(\bar\gamma\bar H-\dot{\bar\gamma}\right)^{-1}\dot{\bar\gamma}\,, \end{aligned} \tag{62}$$

Realistic solutions exist with the term proportional to $\ddot{\bar\gamma}$ providing a dominant negative contribution at late times. Thus it is quite possible to obtain regimes in which *the wall observers measure apparent acceleration, $q<0$, even though void observers do not.*

For the purposes of comparison of numerical solutions to (34), (58) with observed quantities, a number of useful relations can be derived by combining these equations with (28) and (29). The derivative of the lapse function with respect to void average time, t, is

$$\dot{\bar\gamma} = \bar\gamma^{-1}\tfrac{\mathrm{d}}{\mathrm{d}\tau}\,\bar\gamma = \bar\gamma\bar H\left[\tfrac{3}{2}\bar\gamma^{-1}-1-\tfrac{1}{2}\bar\Omega_M-2\bar\Omega_{\mathcal Q}\right], \tag{63}$$

where $\bar\Omega_M$ is given by (43) while the global average deceleration parameter (62) as measured by wall observers is

$$\begin{aligned} q = {}& \frac{\bar\gamma^2\bar H^2}{H^2}\Bigg[\frac14-\frac1{\bar\gamma}+\frac{2}{f_{\rm v}}\left(1-\frac1{\bar\gamma}\right)+\frac32\bar\Omega_M+\frac34\frac{(16-13f_{\rm v})}{(1-f_{\rm v})}\bar\Omega_{\mathcal Q} \\ &\qquad -\frac{2}{f_{\rm v}}(1-f_{\rm v})\left(1-\frac1{\bar\gamma}\right)(\bar\Omega_M+\bar\Omega_{\mathcal Q})-\frac14\bar\Omega_M^2-2\bar\Omega_M\bar\Omega_{\mathcal Q}-4\bar\Omega_{\mathcal Q}^2\Bigg] \\ = {}& \frac{\bar\gamma^2\bar H^2}{H^2}\Bigg[\frac{3}{2\bar\gamma}-\frac{9}{4\bar\gamma^2}+\frac{2}{f_{\rm v}}(1-f_{\rm v})\left(1-\frac1{\bar\gamma}\right)\left(1-\frac{\bar\Omega_M}{\bar\gamma}\right)-\frac{\bar\Omega_M^2}{4}+\frac{3\bar\Omega_M}{2} \\ &\qquad -\frac{12}{f_{\rm v}}(1-f_{\rm v})\left(1-\frac1{\bar\gamma}\right)^2-\frac{2}{f_{\rm v}^2}(1-f_{\rm v})^2\left(1-\frac1{\bar\gamma}\right)^3\left(1-\frac2{\bar\gamma}\right)\Bigg]. \end{aligned} \tag{64}$$

It must be recalled that in these expressions $\bar\Omega_M$ is a time–varying parameter given by (43), which begins close to unity and decreases monotonically. One of the parameters $f_{\rm v}$, $\bar\gamma$ and $\bar\Omega_M$ can be further eliminated by (50). However, the resulting expressions are no more compact than (64).

5.5. Qualitative behaviour of solutions

The physical boundary conditions we apply to the problem are that at early epochs near the surface of last scattering, the expansion rate is essentially uniform so that $h_r = 1-\epsilon$, $\epsilon \ll 1$. In this limit $\dot f_{\rm v} \simeq 0$. The volume fraction of voids thus grows very slowly from some small initial value and the back–reaction is negligible. The initial void volume fraction is a parameter input to numerical integration; the only constraint is that initial density contrast in voids should be consistent with observed bounds inferred from the CMB. In this initial phase $\bar\gamma \simeq 1$, $\bar\Omega_M \simeq 1$ and $\bar\Omega_{\mathcal Q} \simeq 0$ so that by (63) $\dot{\bar\gamma} \simeq 0$. Thus the universe evolves essentially as an Einstein–de Sitter universe with negligible spatial curvature, and only very small differences between clocks in overdense and underdense regions.

As the universe evolves density contrasts grow, bound systems form and enter the non–linear regime where vorticity and shear cannot be ignored. We have to average over regions larger than finite infinity domains so that these effects can be neglected. The voids remain in the "linear regime" but their volume increases more rapidly than the wall regions where galaxies form and cluster. It is claimed that the differences in gravitational energy grow as spatial curvature differences between the wall and void regions increase. This is implicit in our assumption that the true surfaces of average homogeneity are those with a uniform quasilocally measured expansion. Although some of the differences in quasilocal gravitational energy between bound systems and voids may be attributed to the kinetic energy of expansion, given the present epoch value of the Hubble parameter, spatial curvature variations should in fact be the dominant contribution.

From (63) we see that the rate of growth of $\bar{\gamma}$ is abetted by both the decrease in $\bar{\Omega}_M$ and the initial increase in $-\bar{\Omega}_{\mathcal{Q}}$, although there is an eventual decrease in $-\bar{\Omega}_{\mathcal{Q}}$ as $\bar{H}$ decreases. The rate of growth of $\bar{\gamma}$ as seen by wall clocks is amplified since $\frac{\mathrm{d}}{\mathrm{d}\tau}\,\bar{\gamma} = \bar{\gamma}\dot{\bar{\gamma}}$. Depending upon parameter values it is possible for wall observers to register an apparent acceleration with the deceleration parameter of (64) taking values $q < 0$. Back–reaction and the rate of increase of $\bar{\gamma}$ are largest in an epoch during which the universe appears to undergo a void–dominance transition, or equivalently a transition in which spatial curvature $\bar{\Omega}_k$ becomes significant. The reason for apparent acceleration at such an epoch has been partly described by Räsänen [48]: in the transition epoch the volume of the less rapidly decelerating regions increases dramatically, giving rise to apparent acceleration in the volume average. We must be careful to note that these statements are true, when *referred to one set of clocks*, such as our own. Taking gravitational energy differences into account the quasilocally measured expansion can nonetheless be uniform, resolving the dilemma, stated in ref. [20], of the geometry being close to an FLRW one.

Since $\ddot{\bar{\gamma}}$ also contributes to (62) apparent acceleration can be seen by observers in galaxies in a transition from global Einstein–de Sitter–like evolution to open FLRW–like evolution, in contradiction to the picture one would have if one were at the volume average position in voids. In fact, we will find that the deceleration parameter (64) is small and negative for typical parameter values. One very important feature of the present model is that since we are no longer dealing with a purely Gaussian curvature evolution, the effect of an apparent acceleration on the luminosity distance is greater than would be obtained for a FLRW model based on the numerical value of q. In particular, even if q is written in terms of wall time, cosmological evolution is no longer simply described by any quantity derived from a single scale factor. The deviation of r_{w} in (40) from the conformal scale, $\bar{\eta}$, which physically amounts to the rate of change of spatial curvature differences, will also contribute to the average increase of the luminosity distance.

At very late times eventually $f_{\rm v}\to1$ and by (56) $\bar\Omega_Q\to0$. In this limit (57) and (64) yield

$$H \simeq \bar H\left[(2+\tfrac12\bar\Omega_M)\bar\gamma-\tfrac32\right] \tag{65}$$

and

$$q \simeq \frac{\bar H^2}{H^2}\left[\tfrac32(\bar\gamma-\tfrac32)+(\tfrac32-\tfrac14\bar\Omega_M)\bar\gamma^2\bar\Omega_M\right]. \tag{66}$$

Since we also have $\bar\Omega_M\to0$, by (63) there is a stable limit in which $\bar\gamma\to\frac32$, $\dot{\bar\gamma}=0$ and $q\to0$. Equivalently, $\bar a\sim t$ while $(1-f_{\rm v})\sim t^{-1}$, so that within walls $a_{\rm w}\sim t^{2/3}$, just as at early times. At the same time the volume average evolution is approximately the late–time coasting phase of an open FLRW universe. This justifies the analytic approximation of ref. [15]. Thus as $t\to\infty$, $h_r\to\frac23$, the limiting Hubble rate of the Einstein–de Sitter divided by that of the Milne universe.

Since the averaging scheme is defined relative to a particular scale, this limit will not necessarily ever be reached. In the present case, the relevant scale is the size of our present particle horizon volume: the initial void volume fraction is defined relative to this volume. At a much later epoch we would need to re-perform the analysis, with a new volume fraction corresponding to the largest correlated perturbation within the past horizon volume at the epoch in question. This perturbation might be above critical density, which would require modification to the analysis above. We will return to this discussion in §9.

It must be noted that the late–time behaviour given by (65), (66), when $f_{\rm v}\to1$, is an attractor–like feature in the phase space of the differential equations (34), (58), whereas the early–time Einstein–de Sitter–like behaviour is not an attractor if the equations are time–reversed. The early Einstein–de Sitter–like phase is imposed in our case, by taking a small initial void fraction, as a reasonable initial boundary condition consistent with the initial uniformity of the primordial plasma as evidenced by the isotropy of the CMB, and as consistent with the expectations of primordial inflation.

5.6. The zero back–reaction analytic approximation

It is possible to obtain an analytic approximation which demonstrates the variation of gravitational energy, and clock rates, between the wall and void regions by setting the back–reaction to zero and taking $f_{\rm v}=1$, $\dot f_{\rm v}=0$ identically. In this case (34) is simply the Friedmann equation for an open universe. Since the wall regions and voids are then decoupled, a simple but effective ansatz is to assume that the expansion rate of the spatially flat wall regions remains close to that of an Einstein–de Sitter universe as measured by wall clocks. As discussed above, this is justified by the fact that both initially and at late times the local wall geometry does expand in this manner.

This approximation was discussed in ref. [15]. It should be noted that setting $f_{\rm v}=1$ and $\dot f_{\rm v}=0$ identically does not give expressions which coincide exactly with the $f_{\rm v}\to1$ and $\dot f_{\rm v}\to0$ limit discussed in §5.5 above, since in the latter case we are dealing with a system of equations with nontrivial back–reaction in which the ratio $\dot{f}_v/(1-f_v)$ remains finite as $f_v \to 1$. Thus the Hubble parameter and deceleration parameter given in ref. [15] do not have the precise analytic forms of (65) and (66). Nonetheless, the Hubble parameter in particular is extremely close to (65), and the approximation of ref. [15] may be sufficient for determining the global average Hubble constant.

The approximation of ref. [15] was tested against SneIa data in ref. [16]. Although it is in some ways a crude approximation, it fits the data remarkably well with best fit χ^2 values only 7% larger than for the standard ΛCDM model. It appears that any apparent cosmic acceleration in the SneIa data is in fact marginal and any model in which $q \to 0$ compares favourably with observation. Remarkably, the best-fit values of the Hubble constant derived by this analysis [16], $H_0 = 62.7^{+1.1}_{-1.7}\,\mathrm{km\ sec^{-1}\,Mpc^{-1}}$ using the 2004 gold data set of Riess *et al.* [50] or alternatively $H_0 = 60.5 \pm 1.5\,\mathrm{km\ sec^{-1}\,Mpc^{-1}}$ using the 2006 gold data set [51], agree with the recent measured value published by Sandage *et al.* [52] to within 1–2%.

The parameter fits of ref. [16] indicate that any ratio of non–baryonic dark matter to baryonic matter from 0:1 up to the standard 5:1 is admitted, and do not constrain non–baryonic dark matter themselves. The match to the baryon acoustic oscillation scale does appear to constrain non–baryonic matter, however, as will be discussed in §7. Tests similar to that of ref. [53] may also serve to bound the amounts of non–baryonic dark matter, but as will be emphasised in §8.3 below, all steps in measuring masses of galaxy clusters would need to be carefully reconsidered. The parameter fits of ref. [16] are of course based on the crude approximation of ref. [15]. In ref. [17] they are updated to include back–reaction and its effects on cosmological evolution.

6. Null geodesics and observable quantities

With the possible exception of high energy cosmic rays, most cosmological information comes to us on null geodesics carried by photons. In order to relate the average geometry to observations we have to specify how light propagates on average. The manner in which we have related the wall geometry at finite infinity to global average parameters on null geodesics according to (40) in fact already allows us to determine average quantities which depend only on average geodesic length.

6.1. Cosmological redshift

In the two scale model, the average geometry may be approximated by (38), and it assumed that light follows radial null geodesics of this geometry, when a common origin with a point within a finite infinity region is chosen. Since null geodesics are unaffected by conformal transformations, and since the radial null sections of (40) have been chosen to coincide with those of (38), it is simplest to use the wall geometry (40) in order to determine the redshift we measure, since wall time, τ, is assumed to differ little from that of our own clocks. As discussed in ref. [15] the cosmological redshift can be derived in the standard fashion, but there is a difference in values of redshifts measured between our location and the volume average location in voids.

The cosmological redshift, z, determined by wall observers is related to that determined by volume average comoving observers, $\bar{z}$, by

$$1+z=\frac{a_0}{a}=\frac{\bar{a}_0\bar{\gamma}}{\bar{a}\bar{\gamma}_0}=\frac{\bar{\gamma}}{\bar{\gamma}_0}(1+\bar{z})\,. \tag{67}$$

This is a direct consequence of the quasilocal energy difference, and variation of clock rates, between the two locations. Since freely–falling matter in voids is so diffuse as to be unobservable, the quantity $\bar{z}$, which is larger than z, is not directly relevant to emitters or absorbers in any cosmological test yet conceived.

While eq. (67) seems little more than a reparameterisation which relates observed redshifts to numerical solutions of the volume–average Buchert equations (34), (58), there are more profound ramifications. In particular, the redshift of the CMB as measured at a comoving volume average position is greater than in bound systems. Since the volume average CMB temperature is used to calibrate several parameters associated with the primordial plasma, we need to recalibrate all quantities associated with the early universe. Local physics at that epoch is unchanged but our interpretation is systematically changed. It is these recalibrations, which account for quasilocal energy variations, which will allow us to obtain a viable model of the universe without dark energy. These recalibrations are performed in §7.

6.2. Luminosity and angular diameter distances

The luminosity and angular diameter distances can also be determined in the standard fashion from (40), since for all known cosmological observations apart from the CMB both the emitters and observers are in bound systems within finite infinity regions, which are close to spatially flat. Furthermore, the surface of last scattering is located at an epoch when $\bar{\gamma}\simeq 1$, making the distinction between the two possible classes of observer irrelevant for measurements involving the CMB.

While there are changes in both photon energies and the average focussing of geodesics, as null congruences exit and enter voids and finite infinity regions, the average effect of this depends only on the average comoving distance, $\bar{\eta}$, to a source, as determined from the solution to (34), (58). Consequently, the relative difference in the absolute luminosity of a source in a wall region and the observed flux in another wall region is accounted for by taking the standard luminosity distance derived from (40),

$$d_L=a_0(1+z)\,r_{\rm w}=\bar{\gamma}_0^{\;-1}\bar{a}_0(1+z)\,r_{\rm w} \tag{68}$$

where

$$\begin{aligned} r_{\rm w}(\tau)&=\bar{\gamma}\,(1-f_{\rm v})^{1/3}\int_\tau^{\tau_0}\frac{{\rm d}\tau}{\bar{\gamma}\,(1-f_{\rm v})^{1/3}\,a}\\ &=\bar{\gamma}\,(1-f_{\rm v})^{1/3}\int_t^{t_0}\frac{{\rm d}t}{\bar{\gamma}\,(1-f_{\rm v})^{1/3}\,\bar{a}} \end{aligned} \tag{69}$$

for a source emitting a photon at wall time, τ, or volume average time, t. Note that $r_{\rm w}$ coincides with the conventional conformal scale $\bar\eta=\int_t^{t_0}{\rm d}t/\bar a$ at early times when $f_{\rm v}\to0$ and $\bar\gamma\to1$.

If we multiply (68) by the bare Hubble constant, it follows that

$$\bar H_0 d_L=\bar\gamma_0{}^{-1}\bar a_0\bar H_0(1+z)\,r_{\rm w}\,. \tag{70}$$

Equivalently using (42) and (44) also,

$$H_0 d_L=(1-\bar\gamma_0{}^{-1}\bar\gamma_0')\alpha f_{\rm v0}^{1/6}\bar\Omega_{k0}^{-1/2}(1+z)\,r_{\rm w}\,, \tag{71}$$

in terms of the wall–measured average Hubble constant, H_0, where $\bar\gamma_0'\equiv\bar H_0{}^{-1}\left.\frac{{\rm d}\bar\gamma}{{\rm d}t}\right|_{t_0}$.

The effective angular diameter distance to sources observed by wall observers may now be defined in the standard fashion

$$d_A=\frac{d_L}{(1+z)^2}, \tag{72}$$

where d_L is given by (68).

6.3. Numerical example

As a proof of principle, we will give the results of one particular numerical example in this section. It is has not been determined to be the "best–fit" example, but is one case which fits all known tests well, including the broad features of the CMB as will be demonstrated in §7. We adopt the terminology *Fractal Bubble* (FB) universe to describe our new cosmological model, for ease of reference in what follows.

We numerically evolve the equations (34), (35) forward in volume–average time, in dimensionless units of $\bar H_0 t$, from suitable initial conditions. We simultaneously integrate eqs. (41) and (69), and also wall time with respect to volume–average time $\tau=\int_0^t{\rm d}t/\bar\gamma$, to determine the value of the lapse function at the present epoch. This is then used to recalibrate the solutions in terms of wall–measured times and redshifts.

The initial time is taken to coincide with the epoch of last scattering, when the universe is smooth and we deem a suitable initial condition to be one which is consistent with the amplitude of the observed primordial density perturbations by this epoch. In particular, whereas the density contrast of photons and the baryons coupled to them is of order $\delta\rho/\rho\sim10^{-5}$, the density contrast in non–baryonic dark matter can be of order $\delta\rho/\rho\sim10^{-3}$, and this may also be taken as representative of a typical void perturbation.

By our assumptions, if we take our present horizon volume, $\mathcal{H}$, then at last scattering the greatest proportion of its volume, $f_{\rm wi}$, will be in perturbations whose mean is exactly the true critical density. However, in addition to these perturbations there will be one underdense dust perturbation, of initial density contrast $\left(\frac{\delta\rho}{\rho}\right)_{\rm vi}<0$ occupying a small fraction, $f_{\rm vi}$, of $\mathcal{H}$ so that the overall density contrast of $\mathcal{H}$ at that epoch is

$$\left(\frac{\delta\rho}{\rho}\right)_{\mathcal{H}{\rm i}}=f_{\rm vi}\left(\frac{\delta\rho}{\rho}\right)_{\rm vi}\sim-10^{-6}\,{\rm to}-10^{-5}. \tag{73}$$

This might be achieved in various ways, by having $f_{\rm vi} \sim 10^{-3}$ and $(\delta\rho/\rho)_{\rm vi} \sim -10^{-3}$, or $f_{\rm vi} \sim -10^{-2}$ and $(\delta\rho/\rho)_{\rm vi} \sim -10^{-4}$, or $f_{\rm vi} \sim 3 \times 10^{-3}$ and $(\delta\rho/\rho)_{\rm vi} \sim -3 \times 10^{-3}$ etc. The important point to note is that $f_{\rm vi}$ does not refer to a single smooth underdense region, but a region with other perturbations embedded in it, which is significantly larger than the particle horizon at last scattering, and only becomes correlated within our past light cone on time scales which are a significant fraction of the present age of the universe, as will be further discussed in section 9.

Detailed cosmological parameter fits will be presented elsewhere [17]. Here and in §7, we will present one numerical example, as a demonstration that the new model fits the major current observations which support the ΛCDM paradigm, while in addition resolving some anomalies at odds with the ΛCDM paradigm. Decoupling occurs at a scale $z_{\rm dec} \simeq 1100$ with respect to wall observers to within approximately 2%. We cannot assume a value of $z_{\rm dec}$ to the accuracy quoted for WMAP [6], since the value quoted includes statistical uncertainties specific to the ΛCDM model. The recalibration of the detailed peak fitting algorithms for the CMB anisotropy spectrum in the present model requires an immense undertaking and is left for future work. We will offer a proof of principle, by demonstrating agreement with observations at the few percent level.

The example we have chosen has an initial void fraction of $f_{\rm vi} = 5.5 \times 10^{-4}$ at $z_{\rm dec}$, which with $\left(\frac{\delta\rho}{\rho}\right)_{\rm vi} \sim -2 \times 10^{-3}$ gives $\left(\frac{\delta\rho}{\rho}\right)_{\mathcal{H}\rm i} \sim -10^{-6}$. Integrating the equations we find that by the present epoch, the void fraction has risen to $f_{\rm v0} = 0.759$, with a density parameter $\bar{\Omega}_{M0} = 0.127$ with respect to the volume average. This translates to a conventional density parameter (55) of $\Omega_{M0} = 0.33$, as being the value relative to the wall geometry (40) normalised to global average parameters. The difference in the rates of the volume average clocks with respect to our clocks at the present epoch is $\bar{\gamma}_0 = 1.38$. The global average Hubble constant takes the value $H_0 = 62.0\,{\rm km\ sec^{-1}\,Mpc^{-1}}$, while the bare Hubble constant which would represent the value we should determine from measurements restricted to lie *solely within local filamentary walls*, is $\bar{H}_0 = 48.4\,{\rm km\ sec^{-1}\,Mpc^{-1}}$.

The distance modulus is plotted in Fig. 2, in comparison to the Riess Gold06 data set [51] of 182 supernovae. We find a χ^2 of 0.9 per degree of freedom, making the goodness of fit essentially indistinguishable from the ΛCDM model, since on a statistical basis χ^2 of order 1 per degree of freedom is to be expected. We follow Riess *et al.* [51] in excluding supernovae at extremely low redshifts within the "Hubble bubble" from the analysis. Whereas the reasons for doing so are not theoretically well–motivated in the ΛCDM model, in the present case there are imperative reasons for doing so, as is further discussed in §7.5.

In Fig. 3 we compare the difference in distance moduli of the FB model example and a ΛCDM model example with the distance modulus of the empty Milne universe, for the same value of the Hubble constant in all cases. We note that apparent acceleration starts later for the FB model, and is closer to the coasting Milne universe at late epochs.

The deceleration parameter as measured by a wall observer (64) is $q = -0.0428$, whereas a volume–average observer measures a deceleration parameter (61) given by

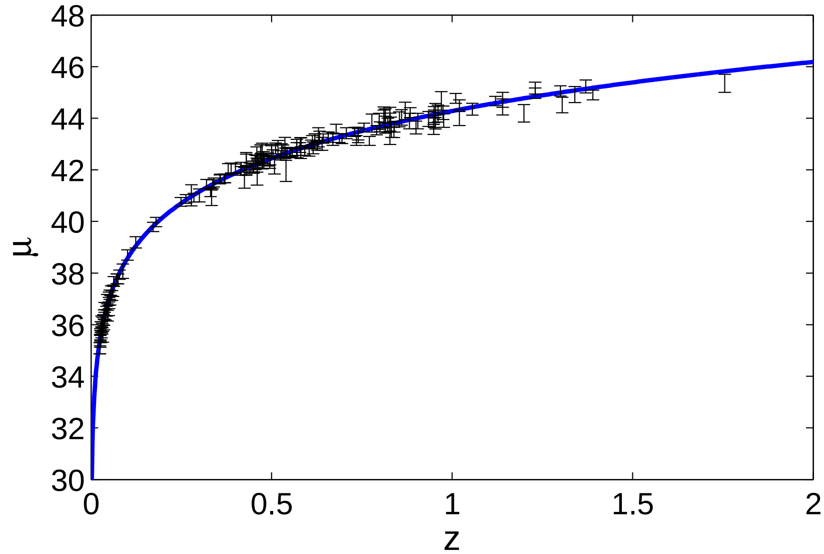

Figure 2. *Distance modulus, $\mu \equiv m - M = 5\log_{10}(d_L) + 25$, versus redshift, z, with d_L in units Mpc. The theoretical curve for a FB model with $H_0 = 62.0$ km sec^{-1} Mpc^{-1}, $\bar{\gamma}_0 = 1.38$, $f_{v0} = 0.759$ is compared to the 182 SneIa, excluding the "Hubble bubble" points at $z \le 0.023$ of the Riess* et al. *Gold06 data set [51]. For these parameter values $\chi^2 = 163.2$, or 0.9 per degree of freedom.*

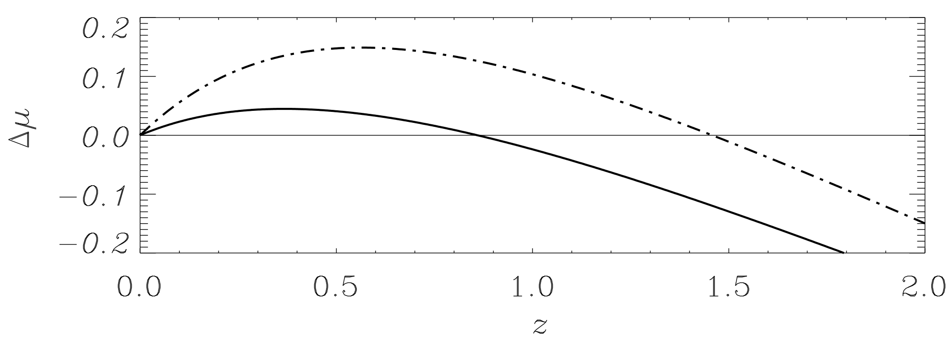

Figure 3. *The difference, $\Delta\mu = \mu_{\rm FB} - \mu_{\rm empty}$, of the distance modulus of the FB model with $H_0 = 62.0$ km sec^{-1} Mpc^{-1}, $\bar{\gamma}_0 = 1.38$, $f_{v0} = 0.759$ and the distance model that of an empty coasting (Milne) universe with the same Hubble constant, versus redshift (solid line). Positive values of $\Delta\mu$ correspond approximately to apparent acceleration. As a comparison we plot the corresponding difference of distance moduli, $\Delta\mu = \mu_{\Lambda\rm CDM} - \mu_{\rm empty}$ for a flat ΛCDM model with $\Omega_M = 0.268$, $\Omega_\Lambda = 0.732$, and the same value of the Hubble constant (dot–dashed line).*

$\bar{q} = 0.0153$. This gives one concrete example of our finding in §5.4 that cosmic "acceleration" is an apparent effect, depending crucially on the position of the observer and local clocks. Both observers register a deceleration parameter close to zero, a general feature of a universe which undergoes a void–dominance transition. According to a wall observer in a galaxy, apparent acceleration begins at an epoch $z = 0.909$ for the present parameters, when the universe is 7.07 Gyr old, a little under half its current age. The

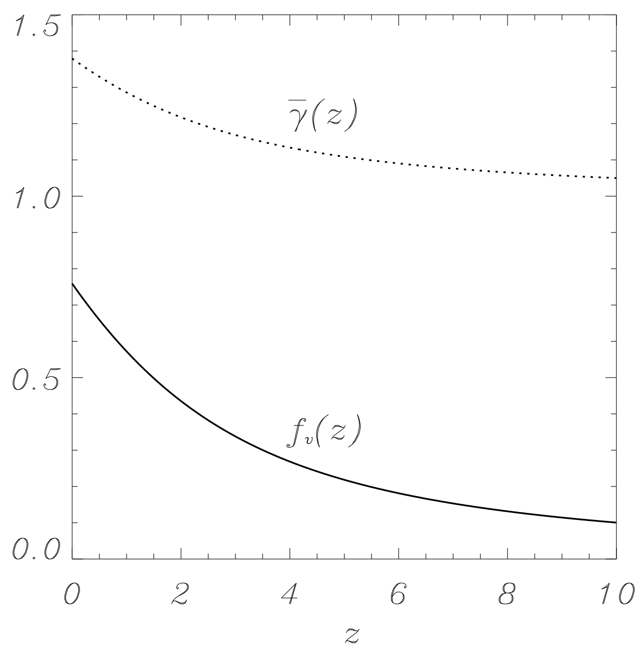

Figure 4. *The void fraction, $f_{\rm v}$, (solid line), and clock rate of volume–average observers with respect to wall observers in a galaxy, $\bar{\gamma}$, (dotted line) as a function of redshift for the FB model with $H_0 = 62.0\,km\ sec^{-1}\ Mpc^{-1}$, $\bar{\gamma}_0 = 1.38$, $f_{\rm v0} = 0.759$.*

void fraction at this epoch is $f_{\rm v} = 0.587$. The void fraction, $f_{\rm v}$, and the mean lapse function or "gravitational energy parameter", $\bar{\gamma}$, are shown in Fig. 4.

Since we are no longer dealing with a pure Gaussian curvature evolution, as discussed above the evolution of the luminosity distance cannot be characterised by one single parameter such as the deceleration parameter. In addition to the change of sign of the deceleration parameter as measured by observers in galaxies, another important cosmological milestone is the epoch at which the void fraction reaches $f_{\rm v} = 0.5$, since the coefficient of $\bar{H}^2$ in (58) changes sign at this epoch. For the present parameters, this event happens somewhat earlier than the apparent acceleration transition, at a redshift of $z = 1.49$ when the universe is 5.12 Gyr old as measured by a wall observer in a galaxy.

The expansion age history of the universe, as shown in Fig. 5, is also particularly interesting in comparison with the "concordance" ΛCDM model. As measured by wall observers in a galaxy, the universe is 14.7 Gyr old at present, 1 billion years older than the concordance model♯. The difference in ages is mainly due to the smaller Hubble constant in the FB model. The ΛCDM model would have an even greater age for the

♯ We are using best–fit parameters for the WMAP1 data with $h = 0.71$, which lie within the 1σ errors of the WMAP3 data, so as not to overstate our case. WMAP3 raises the best–fit Hubble parameter for ΛCDM to $h = 0.73$ and drops the matter density $\Omega_{M0} = 0.238$ for a spatially flat model, which combine to have little net effect on the expansion age. However, if WMAP3 is combined with other datasets a small amount of positive spatial curvature appears possible. Inclusion of a small positive spatial curvature, $\Omega_{k0} = -0.01$, would decrease the age of the universe slightly and increase the contrast with the FB model.

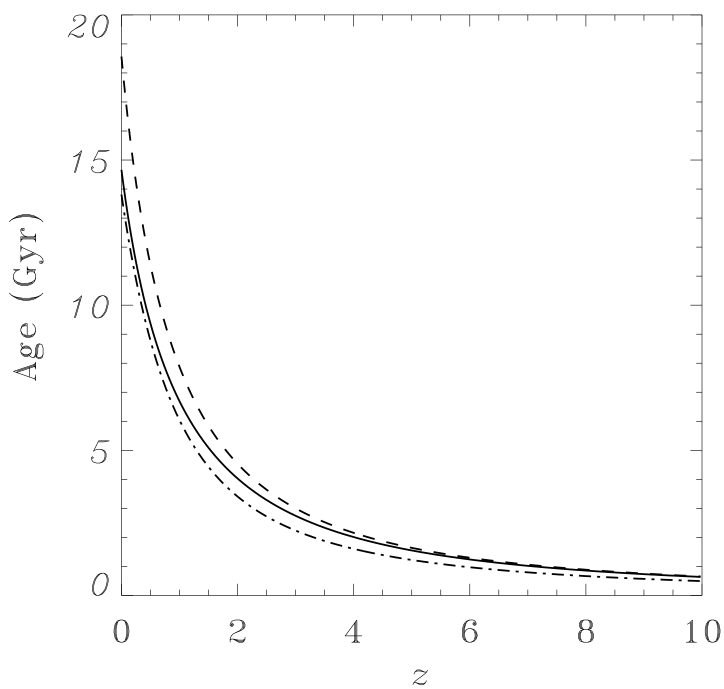

Figure 5. *The expansion age as seen by wall observers in galaxies for a FB model with* $H_0 = 62.0\,km\ sec^{-1}\,Mpc^{-1}$, $\bar{\gamma}_0 = 1.38$, $f_{v0} = 0.759$ *(solid line), is compared to the expansion age as seen by volume–average observers for the same parameters (dashed line). The expansion age for standard* Λ*CDM cosmology with the "concordance" parameters* $H_0 = 71.0\,km\ sec^{-1}\,Mpc^{-1}$, $\Omega_{M0} = 0.268$, $\Omega_{\Lambda 0} = 0.732$ *(dot–dashed line) is also shown for comparison.*

lower Hubble constant consistent with the measurement of Sandage *et al.* [52]; only that value is not the best fit to WMAP in the ΛCDM case. As we shall see in §7, by contrast, the lower Hubble constant provides a good fit very well to those features of the CMB that we have been able to determine for the FB model.

What is perhaps most significant about Fig. 5 is the earlier history since the fractional difference in the expansion age of the FB model example from the concordance ΛCDM model is much greater at higher redshifts, where complex structures are observed. At $z = 1$ with an age of 6.67 Gyr the FB model is 12% older than the ΛCDM model, at $z = 2$ with an age of 4.02 Gyr it is 20% older, and at $z = 6$ with an age of 1.24 Gyr it is 30% older. These values are for one particular choice of parameters. Other examples with an age of the universe greater than 15 Gyr are also in regions of parameter space which accord well with the supernovae data [17].

6.4. Averaging the optical equations

As noted above, local spatial curvature as related to volume average parameters via (40) is all that is required to determine quantities such as luminosity and angular diameter distances (68), (72) relevant to our own measurements. The determination of the volume average spatial curvature may therefore seem an academic exercise. However, it could be relevant for any calculations which may rely on the average differences in spatial curvature, including the integrated Sachs–Wolfe effect, cumulative gravitational lensing, and possibly even weak gravitational lensing measurements. The new interpretational framework presented in this paper requires a systematic re–examination of such effects from first principles. We leave such issues for future work. Here we will merely try to introduce a possible framework for calculating the volume average spatial curvature. As such, the results derived elsewhere in this paper are independent of the calculations in this section.

In the case of the standard FLRW models, given the comoving scale at which a null geodesic originated, it is trivial to compute the average focussing of the spatial geometry to that scale, as the geometry is completely homogeneous. Furthermore, we have found that since the wall geometry is spatially flat on average, we are still able to construct average luminosity and angular diameter distances (68), (72), because we still have an effective homogeneous reference system at the positions of both source and observer. In an arbitrary position in an inhomogeneous universe, however, the problem of determining the local spatial curvature is very nontrivial. In the present case the full volume average geometry (38) is not even conformally isometric to local wall geometry. Thus to determine the volume average spatial curvature we are left with the general problem of employing the equation of geodesic deviation for null geodesics in the form of the Sachs optical equation [54],

$$\frac{1}{\sqrt{\mathcal{A}}}\frac{\mathrm{d}^2\sqrt{\mathcal{A}}}{\mathrm{d}\lambda^2} = -\tfrac{1}{2}\left(\mathcal{R}_{\mu\nu}Y^\mu Y^\nu + w^{\mu\nu}w_{\mu\nu}\right) \tag{74}$$

where $\mathcal{A}$ is the cross–sectional area and $w_{\mu\nu}$ the shear tensor of a null congruence, with tangent $Y^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}$, λ being an affine parameter. The shear and expansion of the null congruence are, of course, distinct from those of the spacelike hypersurfaces used in Buchert's averaging scheme.

The aim here is to determine a quantity equivalent to the angular diameter distance in FLRW models at the volume average position. Thus rather than dealing with individual congruences we must deal with an average over the whole sky for congruences of some initial fiducial proper size, which originated at a given redshift of the average geometry. We will neglect the shear: such a term will have some small effect through cumulative gravitational lensing. However, based on the evidence of studies in FLRW models [55], it is expected to be much less than the dominant effect from the average geometry.

We will relate the expansion of the average null congruence to the volume expansion of the spacelike hypersurfaces in Buchert's scheme, by performing an average in the two–scale approximation of §5, whereby we assume that the geometry is approximately locally homogeneous and isotropic separately within the wall and void regions. Thus

$$\mathcal{R}_{\mu\nu}Y^\mu Y^\nu = -2\left(\frac{\mathrm{d}t}{\mathrm{d}\lambda}\right)^2\left(\frac{\ddot{a}_{\rm v}}{a_{\rm v}} - \frac{\dot{a}_{\rm v}{}^2}{a_{\rm v}{}^2} - \frac{k_{\rm v}}{a_{\rm v}{}^2}\right) \tag{75}$$

within a connected void region. Within such a region (74) then becomes

$$\left(\frac{\mathrm{d}t}{\mathrm{d}\lambda}\right)^2\left(\frac{\ddot{D}_A}{D_A}-\frac{1}{3}\theta_{\rm v}\frac{\dot{D}_A}{D_A}\right)=-\frac{2}{3}\left(\frac{\mathrm{d}t}{\mathrm{d}\lambda}\right)^2\left(\dot{\theta}_{\rm v}-\mathcal{R}_{\rm v}\right) \tag{76}$$

where $D_A=\sqrt{\mathcal{A}}$, and a similar expression applies to a wall region.

To perform a Buchert average of the optical equation requires care since the proper area, $\mathcal{A}$, of the null congruence of a source of some fiducial proper size is not a volume scalar defined on spacelike hypersurfaces of the sort appearing in the commutation rule (8). Since Buchert's scheme does not average the null geodesic equations themselves, we have assumed that the appropriate null geodesics are those of the average geometry (38) in the case of the two–scale model. Essentially, rather than taking an ensemble of null geodesics, we are looking for an idealised geodesic in the average spatial geometry that is the best fit to a homogeneous isotropic one at any epoch in the cosmological evolution.

Likewise, in the case of average null geodesic deviation, we will assume that the average area of a fiducial congruence of such geodesics responds dynamically to the spatial Buchert average of the r.h.s. of (74). Formally, for the purposes of the two–scale model we must assume two conditions in combining (76) with a similar expression for the wall regions:

(i) The affine parameter ratio $\left(\frac{\mathrm{d}t}{\mathrm{d}\lambda}\right)^2$ may be cancelled in numerator and denominator when performing the averages. In doing so we assume that just as the affine parameter, t, is related to timelike geodesics in the average geometry, λ is related to the null geodesics of the same geometry.

(ii) The average angular diameter scale is defined such that $\langle\theta\,\ln(D_A)\rangle=\langle\theta\rangle\langle\ln(D_A)\rangle$, or equivalently that $\langle\frac{\mathrm{d}}{\mathrm{d}t}\ln(D_A)\rangle=\frac{\mathrm{d}}{\mathrm{d}t}\langle\ln(D_A)\rangle$, where D_A is given in units $c/\bar{H}_0$. We note that $\dot{D}_A/D_A$ is the time derivative of an angular scale.

The two–scale average of (74) then becomes

$$\frac{\mathrm{d}^2\langle D_A\rangle}{\mathrm{d}t^2}-\frac{\dot{\bar{a}}}{\bar{a}}\frac{\mathrm{d}\langle D_A\rangle}{\mathrm{d}t}=\left(\frac{\ddot{\bar{a}}}{\bar{a}}-\frac{\dot{\bar{a}}^2}{\bar{a}^2}-\tfrac{1}{2}\mathcal{Q}-\tfrac{1}{6}\langle\mathcal{R}\rangle\right)\langle D_A\rangle\,, \tag{77}$$

where we emphasise that the commutation rule, $\langle\dot{\theta}\rangle=\frac{\mathrm{d}}{\mathrm{d}t}\langle\theta\rangle-\frac{3}{2}\mathcal{Q}$, has been applied only in the limit that the background shear averages to zero: $\langle\sigma\rangle=0$.

It is convenient to define an *effective volume–average comoving lensing scale* $\bar{r}(t)$ by

$$\langle D_A\rangle=\bar{a}(t)\,\bar{r}(t)\,\delta \tag{78}$$

for a fiducial source which subtends an angle, δ. We substitute (78) in (77) and use (32) and (33) to finally obtain

$$\ddot{\bar{r}}+\frac{\dot{\bar{a}}}{\bar{a}}\dot{\bar{r}}+\Big(\frac{\dot{f}_{\rm v}^{\,2}}{3f_{\rm v}(1-f_{\rm v})}-\frac{\alpha^2 f_{\rm v}^{\,1/3}}{\bar{a}^2}\Big)\bar{r}=0\,. \tag{79}$$

The important quantity to note in (79) is the kinematic back–reaction term, which reduces the effect of the spatial curvature term. In a smooth FLRW geometry obtained by integrating the Friedmann equation there is no need to go through an analysis similar to that presented above, since once one knows the comoving scale at which a null geodesic originated, one can read off the associated proper length scale from (78). Equivalently, when $\dot{f}_{\rm v} = 0$ and $f_{\rm v} = \text{const}$, (79) reduces to the FLRW null geodesic equation.

If we assume that $\bar{r} = 0$ is taken to denote the position of a volume average observer, then the relationship between the solution to (79) and the average spatial curvature is obtained by noting that the volume average area function, $A(\bar{\eta}, t)$, of (38) may be taken to satisfy $A \propto \bar{a}^2(t)\bar{r}^2(\bar{\eta})$, where $\bar{r}$ is the solution to (79) for radial null geodesics of (38) with volume average conformal time, $\bar{\eta}$, given by (41). We note that $\bar{r}(\bar{\eta})$ lies in the range $\bar{\eta} \le \bar{r}(\bar{\eta}) \le \sinh(\bar{\eta})$. The lower bound on $\bar{r}$ coincides with the purely spatially flat case, in which $f_{\rm v} = 0$, $\dot{f}_{\rm v} = 0$, so that we have an Einstein–de Sitter model. For a given value of $\bar{\Omega}_{M0}$, the upper bound on $\bar{r}$ coincides with the open FLRW model value for that choice of $\bar{\Omega}_{M0}$. For different void fractions and back–reaction, the actual value of $\bar{r}$ will lie in between these bounds, giving an average spatial curvature between both extremes, dependent on cosmological parameters.

We should be careful to note that if the solution to (79) were to be actually used to derive a luminosity distance or angular diameter distance, then the result could only apply to a situation in which both observer and source are at volume average locations in freely expanding space. Additional calculations would be needed if the source was in a bound system. However, since actual observers and sources are both in bound systems, the solution to (79) does not apply in this fashion.

7. The CMB and resolution of observational anomalies

Detailed data analysis of the new cosmological model, and a best–fit parameter determination will be presented elsewhere [17]. In this section we will illustrate how the recalibration of cosmological parameters, particularly those relating to the early universe as determined from the CMB, can resolve a number of observational anomalies.

An important consequence of the variation in clock rates between ideal comoving observers in galaxies and voids, is that the temperature of the CMB radiation will be lower when measured at the volume average in voids, taking a value

$$\bar{T} = \bar{\gamma}^{-1} T \tag{80}$$

at any epoch, where T is the temperature of the CMB as seen by wall observers. At the present epoch for the numerical example of §6.3, with a measured value of $T_0 = 2.725$ K locally, we have $\bar{T}_0 = 1.975$ K, for example, at the volume average location. Essentially, there is an extra factor of $\bar{\gamma}_0$ in the redshift of CMB photons as seen by void observers on account of gravitational energy differences. Since observed matter sources and absorbers of photons are located in galaxies and gas clouds within the filamentary walls at late epochs, this does not have any obvious direct implications for local physical processes.

Whereas the underlying physics at the epochs of primordial nucleosynthesis and recombination, when $\bar{\gamma} \simeq 1$, is no different than usual, the fact that volume average observers should measure a mean CMB temperature of $2.725\bar{\gamma}_0^{-1}$ at the present epoch will affect all the usual calibrations of radiation–dominated era parameters inferred relative to present epoch comoving observers. We are faced with the task of systematically rederiving all the standard textbook calculations [54, 56] associated with the hot big bang, and making recalibrations where necessary.

This section proceeds chronologically through the history of the universe. We explicitly resolve two anomalies in §7.1 and §7.2: those associated with the primordial lithium abundance, and with average spatial curvature. In §7.3 the anomalies associated with large angle multipoles in the CMB anisotropy spectrum are discussed. In §7.4 we outline qualitatively how the anomalies associated with the expansion age, early formation of galaxies and emptiness of voids in N–body simulations should be naturally resolved. In §7.5 we explain the Hubble bubble feature. Further observational questions concerning void and galaxy cluster scales will be dealt with in §8.

7.1. Primordial nucleosynthesis bounds and the baryon fraction

The first important parameter which is directly altered, given that a volume average observer would measure a lower CMB temperature to us, is the baryon–to–photon ratio inferred from primordial nucleosynthesis bounds, which will lead to a recalibration of the present epoch baryon fraction.

The number density of CMBR photons as measured by wall observers,

$$n_\gamma = \frac{2\,\zeta(3)}{\pi^2}\left(\frac{k_B T}{\hbar c}\right)^3 , \tag{81}$$

evaluated at the present epoch with $T_0 = 2.725$ K, yields the conventional value $n_{\gamma 0} = 4.105 \times 10^8\,\mathrm{m}^{-3}$. However, the corresponding number density measured at the volume average is

$$\bar{n}_\gamma = \frac{2\,\zeta(3)}{\pi^2}\left(\frac{k_B \bar{T}}{\hbar c}\right)^3 = \frac{n_\gamma}{\bar{\gamma}^3} , \tag{82}$$

yielding $\bar{n}_{\gamma 0} = 4.105\,\bar{\gamma}_0^{-3} \times 10^8\,\mathrm{m}^{-3}$ at the present epoch. The number density of baryons measured by a volume average observer is given by

$$\bar{n}_B = \frac{3\bar{H}_0^2\bar{\Omega}_{B0}}{8\pi G\, m_p} = \frac{11.2 f_B \bar{\Omega}_{M0} h^2}{(\bar{\gamma}_0 - \bar{\gamma}'_0)^2}\ \mathrm{m}^{-3}, \tag{83}$$

where prime denotes the dimensionless derivative, $\bar{\gamma}'_0 \equiv \frac{1}{\bar{H}_0}\frac{\mathrm{d}\bar{\gamma}}{\mathrm{d}t}\Big|_{t_0}$, m_p is the proton mass, $f_B \equiv \bar{\Omega}_{B0}/\bar{\Omega}_{M0}$ is the volume–average ratio of the baryonic matter density to the total clumped matter density (including any non–baryonic dark matter), and we have related the bare Hubble constant to our observed global average Hubble constant using (42). It follows that the volume–average baryon–to–photon ratio is

$$\eta_{B\gamma} = \frac{\bar{n}_B}{\bar{n}_\gamma} = \frac{2.736 \times 10^{-8} f_B \bar{\Omega}_{M0} \bar{\gamma}_0^3 h^2}{(\bar{\gamma}_0 - \bar{\gamma}'_0)^2} , \tag{84}$$

as compared to $\eta_{\rm FLRW} = 2.736 \times 10^{-8} f_B \Omega_M h^2$ in the standard FLRW model.

Early time calculations of nuclear processes are not affected in the new scenario, as they occur at epochs when spatial curvature is unimportant, and the effects of inhomogeneities are no more significant than in standard cosmological scenarios. The principal departure is firstly that (84) differs somewhat from the conventional value, and secondly, by (50), that a small "bare" value of $\bar\Omega_{B0}$ as determined by volume–average observers translates to a much larger equivalent value, $\Omega_{B\rm w} \simeq \bar\gamma_0^2 \bar\Omega_{B0}/(1-f_{\rm v0})$, as locally measured by observers within finite infinity regions, which have average critical density. This value of $\Omega_{B\rm w}$ can be translated to an equivalent number density of baryons within walls, n_B. One finds that the ratio n_B/n_γ, is a factor $\bar\gamma_0^{-1}(1-f_{\rm v0})^{-1}$ larger than the volume–average ratio (84). However, it is the volume–average ratio (84) which corresponds directly to that conventionally used in the standard FLRW models. Using conventional big bang nucleosynthesis bounds, we will find that the ratio of baryonic to non–baryonic dark matter will generally be found to be increased, and it is this difference which is the key to resolving the lithium abundance anomaly.

Prior to the the detailed measurements of the Doppler peaks in the CMBR, the values quoted for $\eta_{B\gamma}$ tended to be somewhat lower than the WMAP best–fit value. For example, Olive, Steigman and Walker [59] quoted two possible ranges at the 95% confidence level: $\eta_{B\gamma} = 1.2$–2.8×10^{-10} or $\eta_{B\gamma} = 4.2$–6.3×10^{-10}, depending on whether one accepted higher or lower values of the primordial D/H abundance. At a similar time, Tytler *et al.* [60], accepting the lower D/H abundances, quoted a range $\eta_{B\gamma} = 4.6$–5.6×10^{-10} at the 95% confidence level. The WMAP parameter estimates moved the best–fit range of the baryon to photon ratio to $\eta_{B\gamma} = 6.1^{+0.3}_{-0.2} \times 10^{-10}$, at the very edge [59], or beyond [60], the earlier 95% confidence limits. The underlying physical reason for this parameter shift is that by increasing the fraction of baryons relative to non–baryonic cold dark matter one increases baryon drag in the primordial plasma, which suppresses the height of the second Doppler peak relative to the first. Such a suppression was required in order to fit the WMAP data [6], even though it led to results at odds with primordial lithium abundance measurements, and pushed agreement with ^{4}He abundances to the previous 2σ confidence limit.

To demonstrate the resolution of the lithium abundance anomaly, let us assume a value of $\eta_{B\gamma} = 4.6$–5.6×10^{-10} in the range of Tytler *et al.* [60], the lower end of which accords with lithium abundance measurements, but which lies outside the best–fit range for WMAP using the standard ΛCDM model [61]. There is an intrinsic tension in the data between lithium and deuterium abundance measurements [62] at the opposite ends of the range we adopt. Using the parameters of the numerical example of §6.3, with $h = 0.62$, we find that the bare volume–average baryon density parameter is $\bar\Omega_{B0} \simeq 0.027$–$0.033$. Since the total clumped matter density is $\bar\Omega_{M0} = 0.127$, the relative ratio of non–baryonic dark matter to baryonic matter is roughly 3:1 in this particular example. Therefore it is likely that one can admit enough baryon drag to fit the ratio of the heights of the first two Doppler peaks, while simultaneously having a baryon–to–photon ratio favoured by big bang nucleosynthesis, with the bounds accepted prior to

WMAP [59, 60]. In particular, concordance with the lithium abundance observations can be obtained.

The example we have given is one simple illustration. Taking supernovae data alone, one finds similarly to the case of the zero back–reaction approximation [16], that any ratio of non–baryonic to baryonic matter from close to 0:1 up to the standard ΛCDM value of 5:1 is possible. Other data sets – in particular, the CMB data and baryon acoustic oscillation signatures in galaxy clustering statistics – place restrictions on this range. In general the range of non-baryonic to baryonic matter ratios is restricted from 2:1 to 5:1 by these other tests [17]. Thus while the present model leads to a recalibration of cosmological parameters, it appears that non–baryonic dark matter must be retained.

In a sense, the small baryon density parameters of order $\bar{\Omega}_{B0} \simeq 0.03$ favoured prior to WMAP are still correct for a volume–average void observer. It is simply the case that the mean CMB temperature and photon density as seen by the void observer is also lower, and thus a recalibration with respect to our measurements in galaxies is required, resulting in much larger possible ratios of baryons to non–baryonic dark matter, as locally measured in galaxies.

7.2. Average spatial curvature

Since the first measurement of the angular position of the first Doppler peak in the CMB anisotropy spectrum by the Boomerang experiment in 2000 [57], it has been assumed that the average spatial curvature of the universe is close to zero, and that models such as the present one, which has a sizeable negative spatial curvature in the observed portion of the universe at the present epoch, would be ruled out. In fact, the derivation of the average spatial curvature has assumed a FLRW geometry which evolves according to the Friedmann equation with the conventional identification of comoving clocks. In the present model the entire analysis of the CMB needs to be redone, and conclusions which pertain to FLRW evolution cannot be simply carried over.

The essential point, which has been overlooked, is that the angular positions of the Doppler peaks really are only a measure of *local* average spatial curvature, and an indirect one at that. It is only if one *assumes* that the locally measured spatial curvature is identical to the volume average that one can claim to "measure" the average spatial curvature of the universe. That average spatial curvature and local spatial curvature can be different was already borne out by the toy–model investigation of Einstein and Straus, who concluded that: "the field in the neighbourhood of an individual star is not affected by the expansion and curvature of space" [26]. In the present model, all bound systems lie within finite infinity regions, where the immediate local expansion is given by the geometry (36), which is spatially flat. There is a significant difference in spatial curvature between the voids and wall regions at the present epoch, which must be manifest in a relative focussing between the volume average position and galaxies within finite infinity regions.

If there is a sizeable negative average spatial curvature at the present epoch, then there must be ways of detecting it other than via the measurement of the angular scale of the Doppler peaks. Such effects do arise when one considers more subtle measurements associated with average geodesic deviation of null geodesics. Indeed one prediction is that there should be nontrivial ellipticity in the CMB anisotropies on account of greater geodesic mixing. This effect is in fact observed, and is an important anomaly for the standard ΛCDM paradigm, which has been overlooked by the majority of the community to date. The effect is most significant in the WMAP3 data [5], although it has been seen in all earlier data sets going back to the measurements by COBE [58]. Since this measurement of spectral ellipticity does not rely on a calibration of cosmological parameters which assume a standard FLRW evolution, it is arguably a much less model–dependent measurement of average spatial curvature than the angular position of the first Doppler peak is. This *better* determination of spatial curvature clearly supports the new paradigm, and conceivably can be further developed to constrain the model presented in this paper.

Since the present model implicitly solves the anomaly of refs. [5, 58], it remains for us to show that in the new paradigm, average negative spatial curvature at the present epoch is in fact *consistent* with the angular scale of the first Doppler peak, and the corresponding baryon acoustic oscillation scale in galaxy clustering statistics. We shall now proceed to do so.

The underlying physical effects which determine the angular size of the first Doppler peak in the angular power spectrum of CMB anisotropies boil down to chiefly two quantities:

(i) the proper length associated with the sound horizon, D_s, at the epoch of recombination;

(ii) the average integrated lensing of that scale by the average geometry in the intervening 15 Gyr; including the difference between our location and the volume average at the present epoch.

For a wall observer in a galaxy, the second quantity must be computed from the effective angular diameter distance (72), $d_A = D_s/\delta = a(t)\, r_{\rm w}(t)$, using a solution to (69). However, whereas in the FLRW model the average spatial curvature only affects the angular diameter distance, in the present model we must also recalibrate the sound horizon. We will now present the required steps.

7.2.1. Recalibration of the early universe: recombination and decoupling

In the standard FLRW models volume–average comoving observers are used to calibrate quantities associated with the CMB. Since the Buchert equations are also written with respect to these observers, in recalibrating the standard calculations associated with the radiation–dominated era, it is easiest to again use the volume average. However, the changed relationship between the volume–averaged quantities and our observations at the present epoch must be accounted for.

One direct route to the required recalibration comes from the observation that at early times $\bar{\gamma} \to 1$, an approximation which is valid not only in the radiation–dominated era but also for much of the early matter–dominated "dark ages" of the universe before galaxies form. From eq. (67) we see directly that our observed redshifts, z, are then related to the equivalent volume–averaged quantity, $\bar{z}$, by

$$1 + z \simeq \frac{1}{\bar{\gamma}_0} \frac{\bar{a}_0}{\bar{a}} = \frac{1}{\bar{\gamma}_0}(1 + \bar{z}) \tag{85}$$

for processes at early epochs.

The first quantities of interest are those associated with recombination and matter–photon decoupling. To discuss such effects we must add the energy density of radiation

$$\bar{\rho}_R = \frac{\bar{\rho}_{R0}\bar{a}_0^4}{\bar{a}^4} = \frac{\pi^2 g_*}{30} \frac{(k_B \bar{T})^4}{\hbar^3 c^5}, \tag{86}$$

to the r.h.s. of eq. (34). Here the degeneracy factor relevant for the standard model of particle physics, $g_* = 3.36$, is assumed. The bare cosmological parameters (43)–(45) are then supplemented by the bare radiation density parameter

$$\bar{\Omega}_R = \frac{8\pi G \bar{\rho}_{R0} \bar{a}_0^4}{3\bar{H}^2 \bar{a}^4}. \tag{87}$$

Similarly to (55) the bare parameter (87) is related to a conventional dressed parameter

$$\Omega_R = \bar{\gamma}^4(\tau)\, \bar{\Omega}_R, \tag{88}$$

relative to wall observers if referred to the frame (40) synchronous with our clocks at any epoch.

In addition to adding (86) to the energy density in the r.h.s. of eq. (34), we also note that with our chosen boundary conditions the universe is homogeneous and isotropic with close to zero spatial curvature initially, $f_{\rm v} \simeq 0$ and $\dot{f}_{\rm v} \simeq 0$, so that in the very early epochs near last scattering (34) simply becomes the Friedmann equation

$$\frac{\dot{\bar{a}}^2}{\bar{a}^2} = \frac{8\pi G}{3} \left(\frac{\bar{\rho}_{M0}\bar{a}_0^3}{\bar{a}^3} + \frac{\bar{\rho}_{R0}\bar{a}_0^4}{\bar{a}^4} \right), \tag{89}$$

pertinent to a spatially flat universe containing matter and radiation. Thus all the standard early time calculations apply, with the crucial proviso that *eq.* (85) *must be used to relate $\bar{a}_0/\bar{a}$ to the redshifts we determine as wall observers at the present epoch.* The epoch of matter–radiation equality occurs when

$$\frac{\bar{\Omega}_{M0}}{\bar{\Omega}_{R0}} = \bar{\gamma}_0 (1 + z_{\rm eq}) = \frac{\bar{\gamma}_0 \Omega_{M0}}{\Omega_{R0}} \tag{90}$$

so that $1 + z_{\rm eq} = \Omega_{M0}/\Omega_{R0}$ as is standard in terms of the "conventional" normalisation of parameters (55), (88) defined with respect to (40).

The conditions which define the epochs of decoupling and recombination do not differ at all from the standard case in so far as volume–averaged quantities are used. In particular, recombination may be defined as the epoch when the ionisation fraction decreases to $X_e = 0.1$, where X_e is given by the Saha equation for the equilibrium ionisation fraction,

$$\frac{1-X_e}{{X_e}^2} = \frac{4\sqrt{2}\,\zeta(3)\,\eta_{B\gamma}}{\sqrt{\pi}} \left(\frac{k_B\bar{T}}{m_e c^2}\right)^{3/2} \exp\left(\frac{B}{k_B\bar{T}}\right), \tag{91}$$

where m_e is the mass of the electron, B the binding energy of hydrogen and $\eta_{B\gamma}$ is given by (84). Decoupling occurs roughly at the epoch when

$$\bar{H} \simeq \frac{1}{\bar{n}_e \sigma_{\rm T}} = \frac{1}{\eta_{B\gamma}\bar{n}_\gamma X_e \sigma_{\rm T}}\,, \tag{92}$$

where $\bar{n}_\gamma$ is given by (82), $\bar{n}_e$ is the (bare) number density of electrons, X_e, is given by solving (91) with $\bar{T}/\bar{T}_0 = \bar{a}_0/\bar{a}$, and $\sigma_{\rm T}$ is the Thomson cross-section.

While the locally measured energy scales of recombination and decoupling are unchanged – modulo any small differences from small changes to $\eta_{B\gamma}$ – since these are physical scales determined by quantities such as the binding energy of hydrogen, what is important for cosmological calibrations is that decoupling occurs at a parameter value

$$\begin{aligned}\frac{\bar{a}_0}{\bar{a}_{\rm dec}} = \frac{\bar{T}_{\rm dec}}{\bar{T}_0} = \frac{\bar{\gamma}_0 T_{\rm dec}}{T_0} = \frac{\bar{\gamma}_0 a_0}{a_{\rm dec}} = \bar{\gamma}_0(1+z_{\rm dec}) \\ \simeq 1100\bar{\gamma}_0\,. \end{aligned} \tag{93}$$

In the numerical example of §6.3, we see that a volume–average void observer therefore estimates last–scattering to occur at a redshift, $\bar{z}_{\rm dec}+1 \simeq 1518$, as compared to the wall value $z_{\rm dec}+1 \simeq 1100$.

The sound horizon can be calculated in the standard fashion from (89), using the fact that the speed of sound in the plasma prior to decoupling is

$$c_s = \frac{c}{\sqrt{3(1+0.75\,\bar{\rho}_B/\bar{\rho}_\gamma)}}\,. \tag{94}$$

We therefore find that the proper distance to the comoving scale of the sound horizon at any epoch is

$$\bar{D}_s = \frac{\bar{a}(t)}{\bar{a}_0}\frac{c}{\sqrt{3}\,\bar{H}_0}\int_0^{\bar{x}_{\rm dec}} \frac{{\rm d}\bar{x}}{\sqrt{(1+0.75\,\bar{\Omega}_{B0}\bar{x}/\bar{\Omega}_{\gamma0})(\bar{\Omega}_{M0}\bar{x}+\bar{\Omega}_{R0})}}\,, \tag{95}$$

where $\bar{x} = \bar{a}/\bar{a}_0$, so that $\bar{x}_{\rm dec} = \bar{\gamma}_0^{-1}(1+z_{\rm dec})^{-1}$. There are small differences in each of the factors in the integrand as compared to the corresponding expression for the FLRW model, as well as the factor of ${\bar{\gamma}_0}^{-1}$ in the upper limit of integration. Furthermore, the overall Hubble distance is the bare one, referring to the volume average. Relative to the average Hubble constant measured by wall observers we have

$$\frac{c}{\bar{H}_0} = \frac{c(\bar{\gamma}_0 - \bar{\gamma}_0')}{H_0}\,. \tag{96}$$

For the numerical example of §6.3, $c/\bar{H}_0 = 6200$ Mpc, while the proper length scale of the sound horizon is $\bar{D}_s(t_{\rm dec}) \simeq 0.152 \pm 0.002$ Mpc at last scattering, or

$\bar{D}_s(t_0) \simeq 230 \pm 2$ Mpc at the present epoch. Since this second scale refers to the volume–average comoving rulers, in conversion to the conventional frame (40) synchronous with our clocks, we have $D_s(\tau_0) = \bar{\gamma}_0^{-1} \bar{D}_s(t_0) = 166.8 \pm 1.5$ Mpc. We note that the proper length scale at last–scattering is the same whether determined by the wall observer or the void observer, since $\bar{\gamma} \simeq 1$ at that epoch; i.e., it is a standard ruler set by local physics.

Since the wall measure $D_s(\tau_0) \simeq 167$ Mpc is the one that most directly compares to the corresponding scale in the FLRW model, it is the one which should be compared with the WMAP best–fit value of $D_{s(\Lambda\mathrm{CDM})}(t_0) \simeq 147$ Mpc [6]. Since these values are computed with values of the observed Hubble constant which differ by 14%, it actually makes more sense to compare them as $D_s(\tau_0) = 103.4 \pm 0.8\, h^{-1}$ Mpc for our case (with $h = 0.62$), and $D_{s(\Lambda\mathrm{CDM})}(t_0) = 104\, h^{-1}$ Mpc for the concordance ΛCDM model (with $h = 0.71$). Allowing for the different Hubble parameters, we see that the comoving scales in fact agree.

The comoving baryon oscillation scale of $\sim 100\, h^{-1}$ Mpc is now independently verified in the concordance ΛCDM model from its statistical signature in galaxy clustering statistics [63]. This determination is of course related to angular diameter distances in the standard FLRW geometry, and the details of the analysis should be thoroughly checked in the present model. Since the geometry (40) is by our arguments the closest thing to a FLRW geometry in terms of luminosity distance and angular diameter distance measurements, the fact that the scale determined by Eisenstein *et al.* and Cole *et al.* is reproduced in the present model should not be a surprise if it is correct. The numerical example of §6.3 was chosen as one case of parameters which fit both the supernovae data and the first Doppler peak while resolving the lithium abundance anomaly, rather than by its match to the baryon acoustic oscillation scale in galaxy clustering statistics. Other parameters exist which allow one to match the angular scale of the sound horizon but not the comoving baryon acoustic oscillation scale, but these also have a poor fit to the supernovae [17]. Since galaxy clustering statistics are an important model discriminator [64], the fact that our new concordance parameters appear to match the comoving baryon oscillation scale is encouraging. The details of galaxy clustering statistics in the new cosmological model still need to be thoroughly investigated from first principles.

7.2.2. Fitting the first Doppler peak in the CMB anisotropy spectrum

The angle subtended by the sound horizon provides the rough measure of the angular position of the first Doppler peak that we will adopt. Actual computation of the angular power spectrum would unfortunately require that we systematically rewrite the existing numerical codes which are calibrated to the Friedmann equation. Nonetheless, the proposed measure is close physically, and as a proof of principle we simply need to check whether the model gives the corresponding angular scale determined for the concordance ΛCDM model, $\delta_{s(\Lambda\mathrm{CDM})} = 0.01004 \pm 0.00004$ rad, where the error is a statistical one from parameter fits [6].

For the numerical example of §6.3 we find that the quantity $r_{\rm w}$ which solves (69) for the comoving distance to the last scattering surface at $\bar{a}_0/\bar{a}_{\rm dec} = 1518$ is $r_{\rm w\,dec} = 3.73c/(\bar{a}_0\bar{H}_0)$. Thus

$$d_{A\,\rm dec} = \frac{\bar{a}_0 r_{\rm w\,dec}}{\bar{\gamma}_0(1+z_{\rm dec})} = \bar{a}_{\rm dec} r_{\rm w\,dec} = 15.2\,{\rm Mpc}, \tag{97}$$

and since $\bar{D}_s(t_{\rm dec}) \simeq 0.152 \pm 0.002$ Mpc, the observed angular scale $\delta_s = \bar{D}_s/d_{A\,\rm dec} = 0.01$ rad, is reproduced, for values of the baryon–to–photon ratio which concord with lithium abundance measurements. We should note that the angular scale of the sound horizon does not coincide exactly with the angular position of the first peak in the ΛCDM model and cannot be expected to do so in the present model either. Although the entire detailed analysis of the WMAP data needs to be redone, since we are assuming entirely standard physics for the photon–matter plasma and since the effect of dark energy is negligible at the time of last scattering in the case of ΛCDM, we should expect that the small difference between the exact angular position of the first peak and the angular scale of the sound horizon should be similar in the present model. However, until the detailed calculations have been performed, in looking at the broader parameter space for the new cosmology we should only expect to match the angular scale of the sound horizon of the concordance ΛCDM model at the level of a few percent, rather than as precisely as the numerical example considered here.

We have therefore established that the model is viable in terms of the angular position of the first Doppler peak. In future, physicists and astronomers must take care in saying that this angular scale is a "measure" of average spatial curvature. It is not – it might perhaps be said to be a measure of local spatial curvature; in an inhomogeneous universe this need not coincide with the spatial curvature measured at the volume average. In changing the naïve assumption that our measurements coincide with the volume average, we have shown that a systematic analysis accounting for both gravitational energy and spatial curvature variations can agree quantitatively with observation. In the present model the angular position of the first Doppler peak might be said to be a "measure" of average negative spatial curvature at the present epoch, since zero average spatial curvature now gives a much smaller value of $r_{\rm w}$, and consequently an angle too large by a factor of two. However, I believe it would be safer to reserve statements about "measures" of spatial curvature to tests [5, 58] which do not make model–dependent assumptions about global averages.

7.3. Anomalies in large angle CMB multipoles

There are other anomalies which can be understood in the general framework of inhomogeneous cosmologies, whose resolution does necessarily rely on the particular recalibration of cosmological parameters presented here, but which may be naturally understood in the present framework. One particular anomaly in this class is the so–called "axis–of–evil" [4]. It has been known for some time that realistic large matter inhomogeneities nearby can generate a significant contribution to the CMBR dipole via the Rees–Sciama effect [65]. Foreground inhomogeneities cannot generate the entire dipole since generically the fractional temperature anisotropy in the quadrupole generated by the Rees–Sciama effect is of the same order as that due to the dipole [66]. However, it is very possible that realistic foreground inhomogeneities can produce fractional temperature anisotropies of order $|\Delta T|/T \sim 10^{-5}$ in both the dipole and quadrupole [67].

It is therefore very interesting to note that a statistical study of several possible systematic errors in the WMAP data by Freeman *et al.* [68] indicates that, of the several effects they studied, a 1–2% increase in the magnitude of the peculiar velocity attributed to the CMB dipole was the only one which may potentially resolve anomalies associated with large angle multipoles. This is precisely the order of magnitude of effect we would expect from a $|\Delta T|/T \sim 10^{-5}$ contribution from a Rees–Sciama dipole. Effectively, our current estimate of the magnitude of the peculiar velocity would include a 1–2% systematic error due to a small anomalous boost. Disentangling the small Rees–Sciama dipole from the dominant contribution of our own peculiar velocity with respect to the cosmic rest frame would require an enormous computational effort, and the sky maps would have to be redrawn. However, in the interests of our fundamental understanding of the universe, these steps should be taken.

7.4. Expansion age and structure formation

It is a general feature of the present model that the expansion age is larger at any given epoch than that of the currently favoured ΛCDM model, allowing more time for structure formation. Since old structures are observed at epochs which often seem somewhat too early in terms of the standard paradigm, this is particularly pleasing. Since the age of the universe is position–dependent in the new paradigm, there are in fact different aspects to the solution presented here.

First of all, in terms of wall time, which is relevant to all bound systems, the age of the universe is generally about 1 to 1.5 billion years older than the best–fit age to the WMAP data [6] using the standard ΛCDM model. Partly, this can be understood in terms of the value of the Hubble constant, which is about 14% lower in our case, in agreement with the recent value of Sandage *et al.* [52].

The second important issue surrounding the expansion age is that it is considerably larger for volume average observers than for observers in galaxies. Whereas wall time is the appropriate time parameter for actual measurements in bound systems, including the oldest globular clusters with nucleochronology dates of order 15 ± 1 Gyr, the question arises as to which parameter is the one that is closest to the parameter assumed in the N–body simulations in Newtonian gravity by which structure formation is modelled at present. Since such simulations refer to perturbations about a cosmic mean, it would appear logically that volume–average time is the relevant parameter. If this is the case, then the fact that too much structure is obtained within simulated voids, as compared to actual observation [9, 69], could have a very simple explanation: *the integration time assumed is too short*.

For the numerical example of §6.3 the relevant N–body simulations should be run for 18.6Gyr of volume–average time. Since the relative proportions of baryonic and non–baryonic dark matter are also recalibrated, the choice of boundary conditions in the simulations would also be changed.

One of the ultimate lessons of the present paper is that Newtonian concepts are inadequate to fully understand the large scale structure of the universe. We ultimately need a better post–Newtonian approximation scheme which accounts for the existence of a finite infinity scale, and gravitational energy gradients. Nonetheless, by simple recalibration of various parameters, accounting for the clock effects, it is quite possible that some of the puzzles of N–body simulations could be resolved. This issue should be fully investigated.

7.5. The Hubble bubble and scale of homogeneity

Recent analysis of SneIa data by Jha, Riess and Kirshner [70] confirms an effect which has been known about for some time, and has been interpreted as our living in a local void [11], the "Hubble bubble". By the latest estimates, if one excludes SneIa within the Hubble bubble at redshifts $z \lesssim 0.025$, then the value of the Hubble constant obtained is lower by $6.5 \pm 1.8\%$ [70]. It is for this reason that SneIa at $z \leq 0.023$ have been excluded from the Riess *et al.* Gold06 dataset [51], whereas they were included in the Gold04 dataset [50]. The Hubble bubble is particularly problematic for the standard ΛCDM paradigm, since there is no natural grounds for choosing a locally measured Hubble constant, or alternatively excluding the Hubble bubble. Jha, Riess and Kirshner show that the difference between these two choices using their MLCSk2 data–fitting techniques and a simulated sample for the ESSENCE survey, can lead to differences of up to order 20% in the parameter, w, deduced for the dark energy equation of state $P = w\rho$ [70]. Analysis of the first results of the ESSENCE survey [71] confirms that this effect could contribute as much as 0.065 in the systematic error budget for the w parameter in actual data obtained to date.

The existence of the Hubble bubble, if interpreted conventionally as a large local void expanding into a more slowly expanding surrounding space, gets to the heart of the Sandage–de Vaucouleurs paradox. If this is typical, rather than a statistical fluke, how can the overall expansion rate still be uniform? The answer in the present model is definitive. The expansion is actually uniform if differences in spatial curvature and clock rates are taken into account. If we sample standard candles below the scale of homogeneity, then the naïve assumption that all clock rates are equal will lead us to see a sizeable variance in the Hubble flow in spite of its statistical quietness.

If we make observations within a filamentary bubble wall, such as towards the Virgo cluster, we should infer a lower Hubble constant than the global average, H_0. Ideally this would approach the "bare" or "true" Hubble constant, $\bar{H}_0$, which for the example of §6.3 is 22% lower than H_0. The effect of minivoids will increase the inferred value. Similarly, if we make measurements to galaxies on the other side of the voids on the dominant scale, $\sim 30h^{-1}$ Mpc, then we should infer a Hubble constant which is greater than the global mean, H_0, by a commensurate amount. This occurs since the path integral of the photon geodesics through voids includes regions of high relative positive gravitational energy where clock rates are faster.

Since voids occupy a greater volume of space than the bubble walls, as long as we measure the Hubble constant by averaging isotropically over all directions on distance scales less than the scale of homogeneity we will infer a higher value for the Hubble constant than the global average, H_0. As we sample on larger and larger scales that approach the scale of homogeneity then the average Hubble constant will decrease until it levels out at the global average H_0.

The scale of homogeneity is only defined statistically, since structure formation is a non–linear process. However, there is a clear operational definition of this scale: *it is the baryon acoustic oscillation scale* [63], which effectively demarcates the regime of linear perturbations from non–linear ones, on account of the causal evolution of initial density fluctuations. For the numerical example of §6.3, this scale is at ~ 167 Mpc in terms of the fiducial geometry (40) synchronous with our clocks. This corresponds to a redshift $z = 0.033$, and thus the Hubble bubble feature presently observed is sampled at about 2/3 of the scale of homogeneity, consistent with our interpretation.

The Hubble bubble is not a statistical fluke which marks our location out as anything special. In the new paradigm it is a feature we can expect on average, consistent with the Copernican Principle.

8. Towards dynamical models on intermediate scales

8.1. Variance in the Hubble constant

The model proposed in §5 should not only be tested against all observational data in terms of standard tests of types that have already been considered, but also in terms of its unique predictions. One prediction is provided by eq. (42), which relates the local Hubble parameter, $\bar{H}$, at any location to the horizon volume average parameter, H.

It is difficult to empirically define a locally measured $\bar{H}$, as will be discussed below. However, I wish to stress that it is a feature of this model that it is not only the mean value of the Hubble parameter which is important but also its variance, as is given by the two extremes in (42). In particular, we see that resolving the Sandage–de Vaucouleurs paradox also provides new fundamental insights into another dispute associated historically with those two names, namely what is the value of the Hubble constant? Sandage and collaborators for many decades favoured lower values, and de Vaucouleurs higher values.

While it is beyond question that most of the history of the great Hubble debate has been associated with the resolution of tricky systematic issues to do with statistical biases and the astrophysics of the standard objects in the cosmic distance ladder [72], it is also evident that in the present model we should measure some underlying variance in the Hubble constant, depending on the scale over which it is determined, as discussed in §7.5. Once all astronomers agree on the scale of observations, as well as on the cosmic distance ladder, then values of the Hubble constant will converge to H_0. However, "local" measurements within an ideal bubble wall would give a lower value, $\bar{H}_0$. The determination of both of the parameters, H_0 and $\bar{H}_0$ is of crucial importance to the new paradigm, as it should ultimately be related to cosmic variance in the density perturbation spectrum, as will be discussed in §9.

8.2. Local Hubble flow

A crucial observational question, given the extent to which Newtonian gravity is employed by astronomers on large scales, and by those engaged in N–body structure formation simulations, is: on what scale may Newtonian gravity be applied in cosmology? There are three elements to the Newtonian limit: (i) the weak field limit of general relativity is assumed, $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, $|h_{\mu\nu}| \ll 1$; (ii) all characteristic velocities are smaller than that of light, $v \ll c$; and (iii) pressures are negligible in comparison to energy densities, $P \ll \rho c^2$. One point that is often overlooked in applying Newtonian gravity at cosmological scales is that even when the second and third approximations hold, since space is expanding assumption (i), the weak field limit, never strictly holds in cosmology and therefore application of Newtonian mechanics is completely *ad hoc* unless the validity of the weak field limit is considered within the context of an expanding space.

To identify the scales of averaging over which the weak field limit can be applied, one first needs to answer the question as to what is the largest scale on which the equivalence principle can be applied [44]? Having established such a scale, the Newtonian limit can be applied when in addition to matter satisfying properties (ii) and (iii) the expansion rate of space is small, or equivalently Newtonian dynamics may apply on scales over which quasilocal gravitational energy differences can be neglected. However, to determine this scale empirically is far from trivial. One might apply Newtonian gravity successfully to systems of tightly bound clusters of galaxies which have virialised, but it is clear that at some level Newtonian dynamics must be modified, and the determination of large–scale streaming motions is a point in issue.

I envisage that once the effects of quasilocal gravitational energy are better quantified then a new post–Newtonian approximation scheme relevant to expanding cosmological backgrounds might be developed. Within the context of such a scheme it may be possible to quantify the maximum quasilocal energy variations in minivoids of various sizes by application of the Traschen integral constraints [73] to the evolution of initial underdense perturbations.

The hope is that an effective post–Newtonian approximation for expanding space may yield much better fits to observation than is currently achieved. Techniques such as POTENT do not appear to be particularly successful, and detailed studies of the best observational environment within which a linear Hubble flow is observed, namely the 10 Mpc^3 "local volume", yield a picture which is full of puzzles [41, 42]. These puzzles may just be due to poor statistics; however, in the view of the work of the present paper, it is possible that model assumptions must be revised already at this local scale.

One key question is the extent of anisotropy in the Hubble law. There have been a number of attempts to determine anisotropy within the local volume, culminating in the work of ref. [42], where the principal eigenvalues of the local Hubble tensor are determined to be $\{86 \pm 15\ (31), 53 \pm 8\ (14), 40 \pm 20\ (34)\}\,\mathrm{km\ \ sec^{-1}\,Mpc^{-1}}$. Here Whiting's 1σ confidence limits are quoted with the 90% confidence limits in brackets. At the 90% level there is no evidence for anisotropy, and only tentative evidence at the 1σ level. Furthermore, the directions of the eigenvectors can vary by 90° within the large uncertainties, and so are effectively arbitrary. It is quite possible that the 10 Mpc^3 volume studied in refs. [41, 42] is too small, and that one needs to study the "local" Hubble flow galaxies on the far side of the closest void of a $30h^{-1}$ Mpc diameter, relative to filament directions, to see an effect.

On very large redshift scales, there is tentative evidence of anisotropy in SneIa data [74], which is limited by present statistics. It is vitally important to the present proposal that such studies are extended with new data, and any relevant scales determined. If any systematic anisotropies are observed it is possible and consistent with the present proposal that they could be due to large voids in the foreground, rather than the background.

At this stage it is too early to quantify what should be expected in the local volume. The greatest empirical difficulty lies in disentangling the peculiar velocities of bound systems within finite infinity regions from underlying expansion between such regions. A dense cluster such as Virgo should clearly be expected to be within a finite infinity region, and exhibit the increased peculiar velocity dispersion and tidal torques that one expects from Newtonian gravity. Those aspects of Newtonian intuition will remain empirically valid. It is only when significant regions of freely expanding space are incorporated in the scale under consideration that we should exercise caution: here a revised post–Newtonian approximation scheme is required.

Since regions of expanding space beyond finite infinity possess positive quasilocal gravitational energy, effectively the "energy required to escape" beyond finite infinity is larger than we would estimate from the dynamics of a bound system alone. In terms of the Newtonian Kepler problem finite infinity effectively sets the scale at which "$E_\infty = 0$". However, regions beyond finite infinity have $E > 0$ and the energy required to reach them is larger. A Newtonian physicist might think of this effect as "deepening" the gravitational wells of galaxy clusters. However, since space is not static and particles which escape beyond finite infinity do not return to the potential well from which they originated, this is a conceptually flawed interpretation which I do not wish to encourage. The actual circumstance is best compared with the case of a standard FLRW cosmology: one may solve the geodesic equations that follow from (6) to show that a particle with an initial peculiar velocity, v_i, as measured locally by an isotropic observer at an epoch when $\bar{a} = \bar{a}_i$ has a peculiar velocity (in units $c = 1$)

$$v(t) = \frac{v_i \bar{a}_i}{\sqrt{(1 - {v_i}^2)\bar{a}^2(t) + {v_i}^2 {\bar{a}_i}^2}}\,, \tag{98}$$

at later epochs as measured locally by isotropic observers. This well–known result may be interpreted as saying that peculiar velocities decay.

In the case of inhomogeneous cosmologies (98) must be replaced by a suitable average quantity. The question of averaging the geodesic equations becomes an interesting problem over and above the averaging in Buchert's scheme, which just considers the field equations. At the largest cosmological scales, we have assumed that light follows null geodesics of the average geometry (38), and the same could be expected to be true of timelike geodesics. However, below the scale of homogeneity – identified with the baryon acoustic oscillation scale – this average of geodesic motion would have to be refined.

For individual voids, where an observer will detect a differential rate of expansion from zero at the zero expansion surface to a maximum at the void centre, it would appear from quasilocal energy considerations that we should have sharper localised decays of peculiar velocities than for a homogeneous model. This needs to be confirmed by detailed modelling; the LTB models would be appropriate for studying individual voids. This question is left for future work; but has clear direct implications for the statistical quietness of the average Hubble flow, and the Sandage–de Vaucouleurs paradox.

8.3. Dynamics of galaxy clusters

Even though a revised post–Newtonian approximation scheme remains to be established, one clear physical difference that the present model will make is in any situation in which $3H^2/(8\pi G)$ is taken to be a measure of the closure density which demarcates bound from unbound systems. One situation in which this regularly occurs in observational cosmology is through the Navarro–Frenk–White (NFW) model [75] for galaxy clusters, which are estimated to have a density profile

$$\rho(r) = \rho_{\rm cr} \frac{\delta_C}{\left(r/r_S\right)\left(1 + r/r_S\right)^2}\,, \tag{99}$$

where $\delta_C = 200C^3/[3\ln(1 + C) - C/(1 + C)]$, r is a radial distance within the cluster, and r_S and $C = r_{200}/r$ are two empirical parameters, respectively the "scale radius", and "concentration parameter". When the NFW model is used in cosmological tests – for example, in order to determine the ratio of baryons to dark matter as a function of redshift [53], the critical density is calibrated with redshift as $\rho_{\rm cr}(z) = 3H^2(z)/(8\pi G)$.

Physically, as long as $\rho_{\rm cr}$ represents a closure density, then it must correspond to the true critical density, $3\bar{H}^2/(8\pi G)$, and as indicated earlier this may be typically 40–80% of the value of the critical density estimated from the global average Hubble constant at late epochs. Unfortunately, the NFW model is essentially an empirical fit to the results of N–body CDM simulations in Newtonian gravity. Thus one cannot make any simple qualitative statement about how its use might change if the closure density is to be recalibrated. If similar empirical fits apply then the only obvious deduction we can make is that if $\rho_{\rm cr}$ is effectively overestimated at late epochs, then the density contrast δ_C is effectively underestimated. Consequently, in the present model as long as eq. (99) remains empirically valid we would expect the density contrast in galaxy clusters to be higher than is usually assumed.

If there are any testable consequences that follow from such considerations, then they may possibly apply to highly dynamical non–equilibrium circumstances, such as that of the collision of the galaxy clusters observed in the "bullet cluster" 1E0657-56 [76]. The high velocity of the gas shock front trailing the smaller sub–cluster in 1E0657-56 appears anomalously high [77] as compared to expectations from the masses of the sub–clusters inferred by using weak gravitational lensing and with the NFW or King [78] profiles. Others have argued that such high velocities may not be all that rare statistically speaking [79]. Since it is quite possible that the dynamics of such systems may change in the present model, without invoking new forces of nature [77], this issue deserves further investigation.

It is clear that any model–dependent assumptions related to weak gravitational lensing need to be carefully re–examined. Furthermore, in the strong lensing regime time delays in gravitational lensing events, which exhibit some puzzles [80], are a clear circumstance in which the present model may give differences from the standard ΛCDM model. Much detailed modelling remains to be done. However, it is clear that the new paradigm is likely to give a number of predictions which differ from the standard ΛCDM model, which can be tested.

8.4. Corrections within finite infinity domains

If we demand that a revised post–Newtonian approximation scheme is required beyond finite infinity, a natural question to ask is: to what extent will such effects persist within the finite infinity domains, given the fact that there must be a nontrivial distance between the zero expansion surface and finite infinity? Could the zero expansion surface lie so close to galaxies that the (presumably small) differences of spatial expansion could influence their rotational dynamics?

I will not answer this question here, since if it extends to near galactic levels then we are dealing with scales where vorticity cannot be neglected, as we have in our averaging approximations. It is quite possible that given the non–linearities introduced by vorticity, the first step in the weak–field limit of assuming Minkowski space plus a small perturbation is inappropriate at the galactic scale, even with only a pressureless dust energy–momentum tensor. Arguments of this sort have been presented by Cooperstock and Tieu [81, 82]. While such arguments may have promise, unfortunately the analysis of refs. [81, 82] contains potential flaws which have been much debated††. The variant

††In particular, quite apart from any criticisms which have been raised concerning the physical nature of singularities in the $z = 0$ plane of the galactic matter distribution, Cooperstock and Tieu have given their results in terms of a radial coordinate which does not correspond to the physical proper radius.

of the Cooperstock–Tieu model proposed by Balasin and Grumiller [84] is particularly interesting, since these authors find a significant reduction in dark matter, but not its complete elimination – which accords well with our cosmological parameter estimates based on the baryon acoustic oscillation scale [17]. Whatever results from a careful analysis of the proposals of refs. [81, 84], it has only an indirect bearing on our discussion, since the present proposal relates to scales of averaging very much larger than galactic scales, and the problem of galactic dynamics could involve more ingredients than those considered by these authors.

The only direct consequence of the present proposal vis–à–vis dark matter, is that the recalibration of cosmological parameters will potentially result in substantial changes to the matter budget, both in the overall fraction of matter in our observed portion of the universe relative to the critical density, and in the ratio of non–baryonic to baryonic dark matter. The numerical example of §6.3, which fits observations well, has a 3:1 ratio of non–baryonic dark matter to baryonic matter. As indicated in §8.3, the exact value of this ratio cannot be fully resolved until gravitational lensing and the question of the estimate of masses of clusters of galaxies has been re–examined. This may involve some subtle recalibrations which are likely to change the overall dark matter budget in a fashion which is difficult to foresee.

If any further changes are to be made at the level of galactic dynamics, such changes should be made by solving geodesic equations within the context of general relativity, for appropriate non–asymptotically flat backgrounds. If any modification of Newtonian dynamics is encountered at the level of galactic bound orbits, then it may primarily be due to non–linear couplings between dust and spatial vorticity. Whether the present proposal has anything further to add at galactic scales can only be decided once we construct a dynamics for systems embedded in finite infinity regions. The scale at which the negative curvature of the voids becomes manifest may be an important boundary condition.

It is my view that we should think deeper and harder about general relativity, rather than resorting to ad hoc Newtonian forces or adding terms to the gravitational action which are not justified from any other observations or any fundamental physical principles. When non–baryonic dark matter candidates are postulated, then those which

Using the metric (1) of ref. [82] one sees that the invariant proper circumference of their axial Killing vector vanishes at $r = \mathrm{e}^{w(r,z)}N(r,z)$, and that the roles of $\partial/\partial\phi$ and $\partial/\partial t$ as spacelike and timelike vectors would be reversed for $r < \mathrm{e}^{w}N$. In fact, since the invariant norm of the Killing vector $\partial/\partial t$ diverges at $r = \mathrm{e}^{w}N$, there is a singularity there and the region $r < \mathrm{e}^{w}N$ is not physical. The true centre of the galaxy when expressed in terms of proper distances is at the coordinate surface $r = \mathrm{e}^{w}N$. Since the locally measured rotational velocity tends to the speed of light as one approaches this "galactic centre" the approximations used break down there. This is the explicit manifestation of the criticisms by others [83], which have been commented on in refs. [81, 82], that comoving coordinates cannot be chosen globally in the presence of vorticity. Cooperstock and Tieu point out that their choice of boundary conditions removes the potential problems at $r = \mathrm{e}^{w}N$. With different boundary conditions, Balasin and Grumiller [84] avoid any potentially singular source in the $z = 0$ plane, but have unfortunately made the error of failing to identify the true centre of the galaxy, so as to excise the unphysical region. It is possible that the main results of ref. [84] may survive the correction of this error.

naturally have almost no standard model interactions [85] would seem easiest to reconcile with the rest of cosmology, rather than those that result from fundamental modifications to gravity. While I do not see any obvious role for MOND [86], we should be open to the possibility of phenomenological changes to Newtonian gravity, and actively investigate models in which the assumptions of asymptotic flatness are changed.

Direct changes to Newtonian expectations could conceivably arise for particles in orbits unbound to galactic structures, as discussed at the end of §8.2. For example, if we were to send a spacecraft on a hyperbolic orbit with a large enough escape velocity to leave the galaxy, and ideally the local group, then we might well expect to register clock anomalies as compared to Newtonian predictions. It should be noted that the Pioneer spacecraft do *not* qualify as their escape velocities from the solar system should leave them still bound to the galaxy [87]. The present proposal relies on space in bound systems being non–expanding, and if it is correct attempts to explain the Pioneer anomaly [88] in terms of the expansion of the universe must fail. Indeed, careful attempts to consider the Pioneer anomaly in this fashion give an effective Newtonian acceleration in the *opposite* direction to the observed anomaly [89]. In my view, it is far more likely that the Pioneer anomaly – if it is anything other than instrumentational – is more likely to be explained in terms of a careful consideration of quasilocal gravitational binding energy differences between the solar system and the galactic frames.

A question of principle does remain. In the present context, gravitational binding energy differences may be larger than is assumed when dealing with ideal asymptotically flat geometries. We have assumed that differences in total gravitational energy between our location and finite infinity can be neglected for cosmological averaging. It is conceivable that the question of "where is infinity?" which is central to my proposal may also be related to issues associated to gravitational binding energy. The Pioneer and other flyby anomalies all appear to involve a potential misunderstanding of energy transfer processes [90] as spacecraft gain energy to move into orbits which are hyperbolic with respect to either the solar system frame, or planetary frame as relevant. Thus the understanding of gravitational energy in relation to the true notion of asymptotic infinity may reveal subtleties not yet considered. To date discussions of gravitational binding energy generally appeal to asymptotic flatness [25], and it is clear that at some level this must be corrected. The notion of finite infinity may be relevant to the framework within which such issues should ultimately be debated.

One experiment that would settle this issue would be a space probe mission to fly a cosmic microwave background imager on a trajectory similar to those of the Pioneer spacecraft, to measure the mean temperature of the CMB to the accuracy required if it is assumed that the observed anomaly is a clock effect related to gravitational energy. A very small deviation in the mean CMB temperature, showing a consistent decrease at very large distances, would be expected over the course of the mission. It is quite possible that measurements of the required accuracy are not yet technically feasible. Nonetheless, such a mission may become feasible over the course of the next century. A commensurate decrease in the angular scale of the CMB anisotropies would also be expected in principle, due to variations in spatial curvature, but at a level so tiny that it is likely to be beyond technical feasibility for even longer.

Collisions between galaxies could conceivably provide an astrophysical arena in which individual stars are ejected at high enough escape velocities for them to escape from galaxy clusters, and make the local clock effects I am concerned with readily measurable‖. In practice, such events are at such distances that it is unlikely that we could resolve individual stars ejected far from the galaxies. In the case of galaxy or galaxy cluster mergers, what we principally observe are the bound systems within finite infinity domains, for which there will be no direct clock effects. The only indirect tests might involve mass determinations, as in the case of the bullet cluster mentioned above. Since we observe such mergers at one epoch in the progress of the collision, and cannot play the movie forwards or backwards, the issue of statistics [79] is likely to make claims of a definitive test on these lines hard to justify, however.

8.5. Traversal of voids

The immediate question that is asked by many of those who are presented with this model is: surely we must observe clocks in voids that would rule out these effects? The answer, unfortunately, is that voids by definition have large negative density contrasts, and thus we do not actually observe matter within them. It is possible to find samples of galaxies in very thin filamentary structures within voids. However, a galaxy by definition is a bound system, and is within a finite infinity domain. The proper distance to finite infinity is much less for a void galaxy than a cluster galaxy; however, it is still within a region where clock differences can be expected to be negligible when measured at cosmological distances. Even the gas cloud precursors to galaxies, which are clumped with a sufficient density to be detectable via absorption signals, are contained within finite infinity domains.

It is only if the effects of quasilocal gravitational energy persist within finite infinity regions that we might expect to see differences, for example, in rotation curves of void galaxies as opposed to cluster galaxies. However, as I indicated in §8.4, without detailed modelling such a possibility remains a pure speculation, which does not follow in an obvious fashion from considerations about the quasilocal energy of the expansion of space or spatial curvature variations.

Apart from photons, the only particles that regularly traverse voids are high–energy protons and other cosmic rays. One might be concerned that the GZK bound [91] would be altered since the CMB temperature measured by isotropic observers is lower in voids in the present model, allowing higher charged particle energies to achieve the equivalent centre–of–momentum energy associated with the standard calculation of the GZK bound. However, since high–energy charged particles originate from sources in in galaxies within the filamentary walls, there will be a compensating decrease in their energies as measured by a void observer. This is equivalent to saying that the decay of

‖ I thank Alex Nielsen for this suggestion.

peculiar velocities of charged particles traversing voids is expected to be greater. Thus the GZK bound should not be grossly affected. There may be some changes due to the fact that the energy of massive and massless particles do not scale in the same way with redshift. However, detailed calculations, with gravitational energy changes expected from individual voids need to be considered. It is quite conceivable that there could be alterations to the statistics of charged particles travelling cosmological distances in relation to the GZK bound, due to something akin to an integrated Sachs-Wolfe effect for massive particles. To determine whether this is the case would require detailed calculations taking account of whatever relations replace (98).

9. Primordial inflation and the origin of inertia

I will now outline why the model described in the previous sections, with its two cosmic times, is the natural outcome of primordial inflation, via cosmic variance in the spectrum of primordial density perturbations.

We detect temperature fluctuations in the mean CMB temperature of order $|\Delta T|/T \sim 10^{-5}$, which themselves arise via the Sachs–Wolfe effect from primordial density perturbations believed to be of order $\delta\rho/\rho \sim 10^{-3}$ in non–baryonic dark matter at the time of decoupling. The precise value of $\delta\rho/\rho$ responsible for the observed temperature anisotropies via the Sachs–Wolfe effect may actually require recalibration since it turns out that the calibration makes use of present day values of the Hubble constant and matter density for a smooth FLRW model. However, it is unlikely that it should change much.

The new cosmological model requires systematic recalibration of all observed quantities, but in a manner which should ultimately be consistent with the inflationary spectrum and the early evolution of density perturbations according to established theory. Since the universe is homogeneous and isotropic at early epochs, there is no change to the underlying physics – it is merely the calibration which changes, wherever quantities have been determined with present epoch parameters of a smooth FLRW model. In practice, this occurs at very many steps.

It is a consequence of primordial inflation that the density fluctuations from which all structures ultimately grow are close to scale invariant. Thus provided our particle horizon volume is smaller than the scale of the largest perturbation at any epoch then at some scale of averaging we will always be sampling the non–linear structure that arose from perturbations nested inside some larger single perturbation which is still itself effectively in the linear regime, giving an effective almost–FLRW geometry on the largest averaging scale. We just have to be very careful about relating our observational parameters to that geometry.

One problem with the standard linearised analysis of cosmological perturbations, quite separate from the thorny issues associated with gauge choices, is that its methodology introduces language which becomes totally inappropriate for dealing with averaging at late epochs. Once large volumes of non–linear structures exist, it is more appropriate to think in terms of real space, rather than Fourier space. As an example, much of the debate that followed a suggestion of Kolb *et al.* [92], which was corrected in their later work [19], concerned the issue of "super–horizon sized" as opposed to "sub–horizon sized" modes.

Of course, only matter within our past lightcone can have a causal influence on the geometry of the universe we observe today. However, in reality all density perturbations are finite regions of space, nested within each other like Russian dolls. As with all other observers in bound systems we are sitting on overdense galaxy perturbations embedded in underdense regions embedded in overdense galaxy cluster perturbations etc. If we follow this hierarchy up to the largest observable scale, then we can imagine that we are sitting inside an underdense perturbation that is commensurate with the size of our present horizon volume. If this is the case, when did its effects become apparent to us? Those versed in linearised analysis at early epochs speak about perturbations starting to have an influence when they "cross the horizon". However, for perturbations which are relatively large in comparison to our Hubble volume, the perturbation "crossing time" is significant. Does a perturbation begin to have a noticeable effect when 50% of its spatial volume is within one's past light cone, or 75%, or what? Since general relativity is causal it is certainly not true that a density perturbation has an instantaneous impact on spacetime geometry the moment that the boundary of a perturbation pops through the past light cone at some event.

For the reasons above, I observe that while present geometry must result from Einstein's equations applied to sub–horizon–sized volumes, arguments about whether the perturbations responsible for back–reaction at late epochs are a few times smaller than our present horizon volume, or slightly larger than our present horizon volume, are not productive. Two scales are involved:

(i) The finite infinity region scales are very much smaller than the particle horizon volume. These surround bound systems, are truly in the "non–linear regime" and define the reference point of our clocks.

(ii) The large scales comparable to the horizon volume are characterised by perturbations still in the "linear regime" and define large–scale geometry. They define the clocks of observers at volume average positions in freely expanding space which differ from those in bound systems.

In dealing with the large scales we are talking not about smooth uniform perturbations with no substructure but about statistical correlations between disjoint regions that emerge over billions of years after galaxies have started to form. That is why an averaging scheme, which deals with a void volume fraction as in this paper, is preferred. Once a best–fit to cosmological parameters is found, one can determine the size of the primordial perturbations responsible.

The effects of individual perturbations will continue to have an influence on the geometry of spacetime over time scales commensurate with the size of each perturbation. As long as the volume of the observed universe is smaller than the scale of the largest perturbation, cosmic variance will ensure that back–reaction is important. If this notion is not familiar to the reader, I will now explain the idea from first principles.

9.1. Particle horizon volume selection bias and cosmic variance

Take an initial spatial hypersurface corresponding to the end of the inflationary epoch. Onto this hypersurface randomly scatter small blobs of all proper sizes, allowing blobs to fall inside other blobs, from some smallest scale, up to some largest scale $\mathcal{B}$. Larger blobs will naturally contain all smaller blobs, according to a distribution which could in principle be calculated. The fact that there exists a largest scale, $\mathcal{B}$, of perturbations is simply due to the fact that inflation ends, so there must be a cut–off to the spectrum of perturbations at some upper bound. Scales larger than $\mathcal{B}$ will be called the *bulk*, and the density averaged in the bulk will be extremely close to the true critical density.

The density of the blobs is assumed to be Gaussian–distributed about the true critical density, and the distribution is assumed to be scale–invariant in the following sense. Let us consider a volume, $\mathcal{V}$, of the spatial hypersurface very much larger than the cut–off $\mathcal{B}$. Then for density perturbations of any given proper volume scale the fraction of the proper volume of $\mathcal{V}$ contained in the blobs of any particular scale will be roughly equal, with the same mean density as the bulk. Tiny blobs, $\mathcal{T}$, each have a small volume but are much more numerous. Furthermore, when sampled on the bulk scale, a collection of perturbations on any scale will have the same mean density as the bulk, whether it is the tiniest scales, $\mathcal{T}$, or any intermediate scale, $\mathcal{S}$.

Now let us consider the intermediate scale, $\mathcal{S}$, which we will suppose is the scale of the perturbation which will be the dominant one determining the present epoch geometry of our observed portion of the universe, which is nearly FLRW. We will assume that $\mathcal{S}$ is commensurate with our present horizon volume, possibly somewhat smaller or larger. Since we are dealing with a single perturbation $\mathcal{S}$, rather than a statistical ensemble of them, its mean density can be different from the bulk critical density, and since our observed universe is void–dominated at present, we assume it to be underdense.

We also immediately run into a further sampling issue. While the tiny scales, $\mathcal{T}$, sampled within $\mathcal{S}$ will still have a mean density distributed about the bulk critical density, just as if they had been sampled on the bulk scale $\mathcal{V}$, as soon as we approach scales, $\mathcal{L}$, which are smaller than but close to the scale $\mathcal{S}$, then a discrepancy will arise. Each density perturbation, being a macroscopic region of space with a mean density has to be wholly contained in some other density perturbation. If two perturbations were to "overlap" then the overlap region is just a third perturbation with a mean density of the mean of the first two perturbations. As the scales of perturbations become comparable there are fewer ways to fit one perturbation inside another. For scales $\mathcal{L}$ close to that of $\mathcal{S}$, the number of perturbations in the sample will be very small indeed, and on average we expect the mean density of the limited sample to differ from the sample of the same scales within the bulk, through a $\sqrt{N}$ statistic. I will call these effects – combined with the fact that the overall single perturbation represented by the volume $\mathcal{S}$ has an average density different from the mean of the whole distribution – the *particle horizon volume selection bias*.

Density perturbations at last scattering will, of course, give rise to temperature fluctuations in the CMBR through the dominant Sachs–Wolfe effect, and what I have just defined as the particle horizon volume selection bias is the reason for *cosmic variance* in the two-point angular correlation function of CMBR temperature anisotropies [93] being much larger at low angular multipoles, as is well–understood. As the phrase cosmic variance is often used solely for the distribution of observed CMB temperature anisotropies, I have introduced an alternative terminology. However, the particle horizon volume selection bias is just one consequence of the variance in the underlying spectrum of nearly scale invariant density perturbations. It is that underlying variance in the perturbations, which I am referring to as cosmic variance in a broader sense.

9.2. Cosmic perturbations and cosmic evolution

Given the picture above of the macroscopic volume of space at last scattering from which our observable universe formed, the story of its evolution to the present day can be understood. If we consider the spatial extent of our present horizon volume at last scattering, then it will contain an initial underdense dust void volume fraction, $f_{\rm vi}$, consistent with (73) correlated on scales very large compared to the particle horizon at last scattering. Different values of $f_{\rm vi}$ consistent with the bound (73), and values of $(\delta\rho/\rho)_{\rm vi}$ consistent with the primordial spectrum, give different initial void volume fractions for numerical modelling.

An observer starting from an initial point which goes on to form a galaxy will be inside an overdense region, which itself will be a perturbation contained within the large fraction of $\mathcal{S}$ which averages to critical. Such perturbations can be treated as constrained perturbations in an Einstein–de Sitter background. An observer in such a perturbation will at first perceive that she lives in a closed universe while the perturbation remains within the linear regime, but after the perturbation turns around and collapses, it is other larger perturbations which remain in the linear regime and which define the dominating perturbation and back–reaction at any particular epoch. An observer starting at one of the initial parcels of fluid which develops to be a void at the present epoch will have quite a different interpretation of cosmic history.

The problem of explaining the existence of voids in structure formation simulations is now readily understood to be a consequence of the fact that the wrong mean density has been assumed. It is the true critical density which must be assumed in structure formation simulations: the average density we measure on the scale of $\mathcal{S}$ today has not evolved smoothly by the Friedmann equation from the true critical density. By assuming that it does we come to choose random perturbations about the wrong mean. Once a suitable initial void volume fraction, $f_{\rm v}$, is determined by fitting the overall cosmological evolution to Type Ia supernovae and gamma–ray burster luminosity distances, the CMB, and the acoustic oscillation scale in galaxy clustering statistics, then suitable boundary conditions for structure formation simulations can be specified.

9.3. Mach's principle

It is interesting to note that since primordial inflation ensures that the expansion rate of the universe is uniform at the time of last scattering, and since this gives a true universal critical density which defines the true surfaces of homogeneity, and therefore a reference point for inertial frames, primordial inflation effectively gives us a version of Mach's principle, even if not the one Mach would have originally envisaged. Because there is a true critical density there is a real sense in which it is all the matter in the Universe which defines the reference point of inertial frames. This just relies on the general properties of inflation, rather than any specific inflationary model. Specific versions of Mach's principle which relate to rotating frames can be understood in terms of the causal evolution of angular momentum perturbations [94, 95], with an initial distribution as expected from inflation. The arguments in the present paper amount to the statement that those aspects of the definition of inertial frames that relate to the normalisation of clocks can similarly be understood in terms of the causal evolution of density perturbations, with an initial distribution as expected from inflation. Further refinements of the formalism and predictions should be made by exploiting the Traschen integral constraints [73, 96].

10. Discussion

In this paper I have proposed a dramatically new interpretation of averaged cosmological quantities, in the context of standard general relativity. The Einstein equations are retained, as is the geodesic postulate (or theorem for those who regard it as such) that a test particle follows a timelike geodesic, and that it measures a proper time defined by the metric. What is abandoned is a simplifying assumption about averaging a cosmological spacetime geometry.

Given that space within bound systems is not expanding the pertinent question is not: "What effect does expanding space have on bound systems?" [26]; but "How does expanding space affect local determinations of cosmological parameters within expanding regions in a way which would distinguish them from local determinations of cosmological parameters made within bound systems?" The conventional assumptions that our clocks tick at a rate equal to that of a volume–averaged comoving observer, and that the average spatial curvature we measure in galaxy clusters also coincides with the volume average, are required neither by theory, principle nor observation.

The first key step in understanding this is that *total gravitational energy* – including the quasilocal energy of the expansion of space and spatial curvature variations, as well as gravitational potential differences in bound systems – can play a role in gravitational time dilation and this can be globally manifest in a universe whose spatial matter distribution is as inhomogeneous as the one we live in. Different classes of ideal isotropic observers can each measure a uniform CMB with no dipole anisotropy, while measuring a different mean temperature and different angular anisotropy scales. The second key understanding is that since the universe did initially expand at a uniform rate, there is a true critical density despite present day inhomogeneity, which does define a universal energy scale that demarcates bound systems from unbound ones. With the identification of finite infinity this allows us to identify a global mean expansion rate – we just need to define this expansion rate operationally, as it will generally differ from the average expansion when measured over our entire past horizon volume today according to a single clock. The third key understanding is that apart from the CMB all our cosmological observations are on bound systems, leading to an observer selection effect: even if our clock rates and spatial curvature measurements differ negligibly from those in distant galaxies, the same measurements can differ systematically at the volume average of our observable universe in freely expanding space. These understandings lead to a definition of average homogeneity by the obvious requirement that the expansion *rate* is homogeneous, once quasilocal expansion of space is operationally defined with respect to local measurements.

The new picture of the universe is both familiar and strange to twentieth century physicists. There is an average homogeneous isotropic geometry, but one which does not evolve according to the Friedmann equation. The fact that our clocks and measurements inferred from them do not coincide with those at a volume average position means that the universe has an age which is position–dependent, being as old as 21 billion years in the centres of voids. Yet all observers in bound systems will nonetheless agree on it being of order 15 billion years old. The ultimate fate of the universe – whether it will expand forever or not – is *undecidable*, at the present epoch at least. We must wait until the scale of the largest primordial perturbation is correlated within our past light cone before we can know whether the universe will ultimately collapse.

The new proposal renders the terminology of "open" and "closed" universes somewhat obsolete. I therefore propose to call this model the *Fractal Bubble Universe*. I use the word "fractal" here with some trepidation, as it has come to be popularly associated with those who wish to impose a mathematical structure on the universe as an extra principle of nature¶ [97]. In my view, that is taking one notion of mathematical beauty too far. Physics proceeds from principles to do with basic observable quantities such as energy and momentum; my proposal aims to refine the understanding of those quantities by better characterising quasilocal gravitational energy in cosmology. This relates to the operational understanding associated with averaging the left hand side of Einstein's equations, rather than imposing a fractal structure on the energy–momentum tensor [98]. While fractals are ubiquitous in nature, such structures arise from physical

¶ It should be noted that the Copernican Principle advanced in §2 is distinct from Mandelbrot's Conditional Cosmological Principle [97], which is further discussed by Mittal and Lohiya [98]. Mandelbrot assumes galaxies are fractally distributed and the universe appears the same at any galaxy, but not in voids. I argue that there is a notion of homogeneity whether viewed from galaxies or voids; one just needs to be careful about relating this to local measurements.

processes and are always limited to a range of scales. It is certainly true that a void–dominated universe at the present epoch would be consistent with a fractal distribution of galaxies at some scales, and on those scales the observed universe may appear hierarchical in the sense that de Vaucouleurs argued for [30]. I suggest the word "fractal" insofar as it aptly encompasses those observations. Whether the distribution of galaxies is strictly fractal or not in some precise mathematical sense is a point which has been much debated [31, 99], and will no doubt continue to be debated, but is not in any way directly relevant to the arguments I have presented.

This new proposal, if it is correct, means that present epoch "dark energy" is an historical accident resulting from a misidentification of gravitational energy, which is not local and cannot be fully described by the internal energy of a fluid–like quantity. Furthermore, this new understanding should provide a much richer and deeper framework for theoretical cosmology and observational modelling. The changes are not small, as every average cosmological parameter must be systematically recalibrated. However, once these changes are made many avenues of new research directions will open up. Qualitatively, cosmic variance becomes as important a feature as cosmic averages in the new paradigm. While the historical argument among astronomers over the value of the Hubble constant is for the most part due to systematic issues, there is an intrinsic variance related to the choice of averaging scale which may have contributed to these contentions. This variance, as quantified by eq. (42), should ultimately be related to cosmic variance in the primordial density perturbation spectrum.

The coincidence problem in the standard ΛCDM model, as to why the fraction of dark energy should be commensurate with the fraction of clumped matter at the present epoch, is eliminated since present epoch dark energy is eliminated. Furthermore, the other coincidence as to why cosmic "acceleration" should occur at the same epoch when the largest structures form is naturally solved. Voids are associated with negative spatial curvature, and negative spatial curvature is associated with the positive gravitational energy which is largely responsible for the gravitational energy gradient between bound systems and the volume average. Since gravitational energy directly affects relative clock rates, it is at the epoch when the gravitational energy gradient changes significantly that apparent cosmic acceleration is seen.

Some potential directions for future work and a number of potentially distinctive cosmological tests have been suggested in §7, §8 and §9.2. Clearly, variations in quasilocal energy should be modelled from a more suitable formalism, e.g., a generalisation of [25, 35, 36], accounting for the integral constraints [73, 94, 96]. Any better understanding of quasilocal gravitational energy in one dynamical situation can only help improve our understanding in other situations, such as the choice of average slicing for numerical studies of black hole encounters. The mean lapse or "gravitational energy parameter", $\bar{\gamma}$, which is in a sense related to the non–affinity of wall time, τ, on volume–average dust geodesics, appears to play a role analogous to the surface gravity in black hole models, although the latter is defined for null foliations. In the cosmological context, the first integral (50), which may also be written as $f_{\rm w} \simeq \bar{\gamma}^2 \bar{\Omega}_M$, is perhaps the single most important equation that follows from the definition of finite infinity.

As theoretical physicists we are altogether too much inclined to add all sorts of terms to the gravitational action, even if they potentially violate basic principles such as causality as in the case of "phantom energy", rather than thinking deeply about the basic operational issues of our subject. I believe we should guard the principles that have worked until they can be proved to fail. It is my own view that Einstein was correct about general relativity, and what I have presented here follows logically from his theory when combined with initial conditions given by primordial inflation. As a consequence even our understanding of Einstein's most famous equation for the rest energy of a test particle, $E = mc^2$, must be further refined in cosmology since in truly expanding regions two widely separated inertial observers cannot be at rest relative to each other. However, the refinement of the notion of "rest" energy merely clarifies an aspect of the theory, quasilocal gravitational energy, which Einstein knew to be incomplete.

Nature is of course the final arbiter of all ideas; so it is pleasing that in addition to achieving concordance with supernovae data and the baryon acoustic oscillation scale, the present model may resolve a number of observational anomalies. In addition to the fact that the expansion age is generally larger in the present model, allowing more time for structure formation, the specific observation of ellipticity in the CMB anisotropies [5, 58] is clearly consistent with the present paradigm. The problem of lithium abundances [7] can likewise be understood, since the baryon fraction obtained from standard nucleosynthesis bounds for a given volume–average baryon–to–photon ratio, $\eta_{B\gamma}$, can be higher. Although we have given generic arguments and explicit calculations to show that the generic features of the CMB anisotropy spectrum such as the angular scale of the first Doppler peak and the ratio of the heights of the first and second peaks will fit the recalibrated cosmological parameters, much work needs to be done to write detailed numerical codes to perform fits to the WMAP data in as much detail as has been done for the standard ΛCDM model [6].

Whatever the outcome of observational tests on the new cosmological model, the questions I have raised are, I believe, so fundamental to the foundations of cosmology that they must be seriously considered. If I am wrong, then the following questions remain: What is the energy content of an expanding space of varying curvature and what is its influence on particle clocks? Can average spatial curvature vary to the extent that we infer different angular scales in the CMB anisotropy spectrum to those at the volume average? What is the largest scale on which the equivalence principle can be applied? Where is the effective spatial infinity vis–à–vis a bound system? How do we resolve the Sandage–de Vaucouleurs paradox? How do we explain the existence of voids? Why does the observed universe appear to undergo "accelerated" expansion just at the epoch when structure forms? How do we define average surfaces of homogeneity in a very lumpy universe? A working framework which seeks to quantitatively answer all of these issues without changing the principles of our best theory of gravitation stands a better chance of success, I believe, than alternative hypotheses which ignore these foundational questions.

Acknowledgments

This work was supported by the Mardsen Fund of the Royal Society of New Zealand. I am particularly grateful to Jorma Louko and Jenni Adams for their detailed comments on initial drafts of this manuscript. I also wish to thank many people for discussions, correspondence, questions and criticisms which have helped me to clarify my ideas and presentation, including: Roy Kerr, Matt Visser, Paul Davies, Peter Szekeres, Thomas Buchert, Syksy Räsänen, Alan Whiting, Steve Maddox, Hui Yao, Benedict Carter, Alex Nielsen, Cindy Ng, Ben Leith, Tam Nguyen Phan, Marni Sheppeard, Ishwaree Neupane, Dharamvir Ahluwalia-Khalilova, Harvey Brown, Ewan Stewart, Brett Bochner, Tony Fairall, Richard Easther, and Don Page. I also thank Prof. Y.M. Cho for hospitality at Seoul National University.

References

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

[2] G.F.R. Ellis and W. Stoeger, Class. Quantum Grav. **4** (1987) 1697.

[3] K. Glazebrook *et al.*, Nature **430** (2004) 181; A. Cimatti *et al.*, Nature **430** (2004) 184.

[4] K. Land and J. Magueijo, Phys. Rev. Lett. **95** (2005) 071301; C.J. Copi D. Huterer, D.J. Schwarz and G.D. Starkman, Mon. Not. R. Astr. Soc. **367** (2006) 79.

[5] V.G. Gurzadyan, C.L. Bianco, A.L. Kashin, H. Kuloghlian and G. Yegorian, Phys. Lett. **A 363** (2007) 121;
V.G. Gurzadyan, P. de Bernardis, G. De Troia, C.L. Bianco, A.L. Kashin, H. Kuloghlian, S. Masi, F. Piacentini, G. Polenta and G. Yegorian, Mod. Phys. Lett. **A 20** (2005) 813.

[6] C.L. Bennett *et al.*, Astrophys. J. Suppl. **148** (2003) 1; D.N. Spergel *et al.*, Astrophys. J. Suppl. **170** (2007) 288.

[7] M. Asplund, D.L. Lambert, P.E. Nissen, F. Primas and V.V. Smith, Astrophys. J. **644** (2006) 229.

[8] A.J. Korn, F. Grundahl, O. Richard, P.S. Barklem, L. Mashonkina, R. Collet, N. Piskunov and B. Gustafsson, Nature **442** (2006) 657.

[9] P.J.E. Peebles, Astrophys. J. **557** (2001) 495.

[10] F. Hoyle and M.S. Vogeley, Astrophys. J. **566** (2002) 641; Astrophys. J. **607** (2004) 751.

[11] K. Tomita, Mon. Not. R. Astr. Soc. **326** (2001) 287; Prog. Theor. Phys. **106** (2001) 929.

[12] A.V. Tikhonov and I.D. Karachentsev, Astrophys. J. **653** (2006) 969.

[13] T. Buchert, Gen. Relativ. Grav. **32**, 105 (2000).

[14] R.M. Zalaletdinov, Gen. Relativ. Grav. **24**, 1015 (1992); A.A. Coley, N. Pelavas and R.M. Zalaletdinov, Phys. Rev. Lett. **95** (2005) 151102.

[15] D.L. Wiltshire, arXiv:gr-qc/0503099.

[16] B.M.N. Carter, B.M. Leith, S.C.C. Ng, A.B. Nielsen and D.L. Wiltshire, arXiv:astro-ph/0504192.

[17] B.M. Leith, S.C.C. Ng and D.L. Wiltshire, arXiv:0709.2535 [astro-ph].

[18] T. Buchert, Class. Quantum Grav. **22** (2005) L113; Class. Quantum Grav. **23** (2006) 817; M.N. Célérier, New Adv. Phys. **1** (2007) 29 [arXiv:astro-ph/0702416]; and references therein.

[19] E.W. Kolb, S. Matarrese and A. Riotto, New J. Phys. **8** (2006) 322.

[20] A. Ishibashi and R.M. Wald, Class. Quantum Grav. **23** (2006) 235.

[21] A. Krasiński, *"Inhomogeneous Cosmological Models"*, (Cambridge Univ. Press, 1997).

[22] G. Lemaître, Ann. Soc. Sci. Bruxelles **A 53** (1933) 51 [English translation in Gen. Relativ. Grav. **29** (1997) 641]; R.C. Tolman, Proc. Nat. Acad. Sci. **20** (1934) 169.

[23] H. Bondi, Mon. Not. R. Astr. Soc. **107** (1947) 410.

[24] P. Szekeres, Commun. Math. Phys. **41** (1975) 55.

[25] J. Katz, Class. Quantum Grav. **22** (2005) 5169; J. Katz, D. Lynden-Bell and J. Bičák, Class. Quantum Grav. **23** (2006) 7111; D. Lynden-Bell, J. Katz and J. Bičák, Phys. Rev. **D 75** (2007) 024040; Err. ibid. **D 75** (2007) 044901.
[26] A. Einstein and E.G. Straus, Rev. Mod. Phys. **17** (1945) 120; Err. ibid. **18** (1946) 148.
[27] E. Schücking, Z. Phys. **137** (1954) 595.
[28] M. Harwit, Astrophys. J. **447** (1995) 482.
[29] A. Sandage, G.A. Tammann and E. Hardy, Astrophys. J. **172** (1972) 253.
[30] G. de Vaucouleurs, Science **167** (1970) 1203.
[31] F. Sylos Labini, M. Montuori and L. Pietronero, Phys. Rep. **293** (1998) 61.
[32] Y. Baryshev, AIP Conf. Proc. **822** (2006) 23.
[33] A.D. Chernin, I.D. Karachentsev, M.J. Valtonen, V.P. Dolgachev, L.M. Domozhilova and D.I. Makarov, Astron. Astrophys. **415** (2004) 19.
[34] L.B. Szabados, Living Rev. Rel. **7**, 4 (2004).
[35] J.H. Yoon, Phys. Rev. **D 70** (2004) 084037.
[36] C.M. Chen, J.M. Nester and R.S. Tung, Phys. Rev. **D 72** (2005) 104020.
[37] T. Buchert and M. Carfora, Phys. Rev. Lett. **90** (2003) 031101; Class. Quantum Grav. **19** (2002) 6109.
[38] C. Hellaby, Gen. Relativ. Grav. **20** (1988) 1203.
[39] T. Buchert, Gen. Relativ. Grav. **33** (2001) 1381.
[40] A.R. King and G.F.R. Ellis, Commun. Math. Phys. **31** (1973) 209.
[41] A.B. Whiting, Astrophys. J. **622** (2005) 217.
[42] A.B. Whiting, Astrophys. J. **587** (2003) 186.
[43] R.H. Price, arXiv:gr-qc/0508052.
[44] D.L. Wiltshire, in preparation.
[45] T. Buchert, J. Larena and J.M. Alimi, Class. Quantum Grav. **23** (2006) 6379.
[46] W.H. Press and E.T. Vishniac, Astrophys. J. **239** (1980) 1.
[47] J.M. Bardeen, Phys. Rev. **D 22** (1980) 1882.
[48] S. Räsänen, JCAP **11** (2006) 003.
[49] D.L. Wiltshire, arXiv:0709.0732 [gr-qc].
[50] A.G. Riess *et al.*, Astrophys. J. **607** (2004) 665.
[51] A.G. Riess *et al.*, Astrophys. J. **659** (2007) 98.
[52] A. Sandage, G.A. Tammann, A. Saha, B. Reindl, F.D. Macchetto and N. Panagia, Astrophys. J. **653** (2006) 843.
[53] S.W. Allen, R.W. Schmidt, H. Ebeling, A.C. Fabian and L. van Speybroeck, Mon. Not. R. Astr. Soc. **353** (2004) 457.
[54] P.J.E. Peebles, *"Principles of Physical Cosmology"*, (Princeton Univ. Press, 1993).
[55] D.E. Holz and R.M. Wald, Phys. Rev. **D 58** (1998) 063501; C.M. Claudel, Proc. Roy. Soc. Lond. **A 456** (2000) 1455; H.G. Rose, Astrophys. J. **560** (2001) L15.
[56] E.W. Kolb and M.S. Turner, *"The Early Universe"*, (Addison–Wesley, Reading, Mass., 1990).
[57] P. de Bernardis *et al.* [Boomerang Collaboration], Nature **404** (2000) 955.
[58] V.G. Gurzadyan and S. Torres, Astron. Astrophys. **321** (1997) 19.
[59] K.A. Olive, G. Steigman and T.P. Walker, Phys. Rept. **333** (2000) 389.
[60] D. Tytler, J.M. O'Meara, N. Suzuki and D. Lubin, Phys. Scripta **T85** (2000) 12.
[61] R.H. Cyburt, B.D. Fields and K.A. Olive, Phys. Lett. **B 567** (2003) 227.
[62] G. Steigman, Int. J. Mod. Phys. **E 15** (2006) 1.
[63] D.J. Eisenstein *et al.*, Astrophys. J. **633** (2005) 560; S. Cole *et al.*, Mon. Not. R. Astr. Soc. **362** (2005) 505.
[64] A. Blanchard, M. Douspis, M. Rowan-Robinson and S. Sarkar, Astron. Astrophys. **449** (2006) 925.
[65] M.J. Rees and D.W. Sciama, Nature **217** (1968) 511.
[66] A. Meszaros and Z. Molnar, Astrophys. J. **470** (1996) 49.
[67] A. Rakic, S. Räsänen and D.J. Schwarz, Mon. Not. R. Astr. Soc. **369** (2006) L27.

[68] P.E. Freeman, C.R. Genovese, C.J. Miller, R.C. Nichol and L. Wasserman, Astrophys. J. **638** (2006) 1.
[69] P.J.E. Peebles, talk at *"Cosmic voids: Much ado about Nothing"*, Amsterdam, Dec, 2006. See: http://www.astro.rug.nl/~weygaert/tim1publication/knawvoid/voidknaw.peebles.pdf
[70] S. Jha, A.G. Riess and R.P. Kirshner, Astrophys. J. **659** (2007) 122.
[71] W.M. Wood-Vasey *et al.*, Astrophys. J. **666** (2007) 694.
[72] G.A. Tammann, Rev. Mod. Astron. **19** (2006) 1.
[73] J.H. Traschen, Phys. Rev. **D 29** (1984) 1563; Phys. Rev. **D 31** (1985) 283.
[74] B. Bochner, in Proc. 22nd Texas Symposium on Relativistic Astrophysics, Stanford University, eConf **C041213** (2004) 1301 [arXiv:astro-ph/0702730].
[75] J.F. Navarro, C.S. Frenk and S.D.M. White, Astrophys. J. **490** (1997) 493.
[76] D. Clowe, M. Bradac, A.H. Gonzalez, M. Markevitch, S.W. Randall, C. Jones and D. Zaritsky, Astrophys. J. **648** (2007) L109.
[77] G.R. Farrar and R.A. Rosen, Phys. Rev. Lett. **98** (2007) 171302.
[78] I.R. King, Astron. J. **67** (1962) 471.
[79] E. Hayashi and S.D.M. White, Mon. Not. Roy. Astron. Soc. Lett. **370** (2006) L38.
[80] C.S. Kochanek and P.L. Schechter, in W. Freeman (ed.), *Measuring and Modeling the Universe*, (Cambridge Univ. Press, 2004) pp. 117–137 [arXiv:astro-ph/0306040].
[81] F.I. Cooperstock and S. Tieu, Mod. Phys. Lett. **A 21** (2006) 2133.
[82] F.I. Cooperstock and S. Tieu, Int. J. Mod. Phys. **A 22** (2007) 2293.
[83] D.J. Cross, arXiv:astro-ph/0601191; T. Zingg, A. Aste and D. Trautmann, arXiv:astro-ph/0608299.
[84] H. Balasin and D. Grumiller, arXiv:astro-ph/0602519.
[85] D.V. Ahluwalia-Khalilova and D. Grumiller, Phys. Rev. **D 72** (2005) 067701; JCAP **07** (2005) 012.
[86] R.H. Sanders and S.S. McGaugh, Ann. Rev. Astron. Astrophys. **40**, 263 (2002).
[87] S.G. Turyshev, J.D. Anderson and M.M. Nieto, Am. J. Phys. **73** (2005) 1033.
[88] J.D. Anderson, P.A. Laing, E.L. Lau, A.S. Liu, M.M. Nieto and S.G. Turyshev, Phys. Rev. **D 65** (2002) 082004.
[89] M. Carrera and D. Giulini, Class. Quantum Grav. **23** (2006) 7483.
[90] J.D. Anderson, J.K. Campbell and M.M. Nieto, New Astron. **12** (2007) 383.
[91] K. Greisen, Phys. Rev. Lett. **16** (1966) 748; G.T. Zatsepin and V.A. Kuzmin, JETP Lett. **4** (1966) 78 [Pisma Zh. Eksp. Teor. Fiz. **4** (1966) 114].
[92] E.W. Kolb, S. Matarrese, A. Notari and A. Riotto, arXiv:hep-th/0503117.
[93] A.R. Liddle and D.H. Lyth, *"Cosmological Inflation and Large–Scale Structure"*, (Cambridge University Press, 2000); S. Dodelson, *"Modern Cosmology"*, (Academic Press, San Diego, 2003).
[94] D. Lynden-Bell, J. Katz and J. Bičák, Mon. Not. R. Astr. Soc. **272** (1995) 50; Err. ibid. **277** (1995) 1600.
[95] J. Bičák, D. Lynden-Bell and J. Katz, Phys. Rev. **D 69** (2004) 064011; C. Schmid, Phys. Rev. **D 74** (2006) 044031.
[96] G.F.R. Ellis and M.J. Jaklitsch, Astrophys. J. **346** (1989) 601; N. Deruelle, J. Katz and J.P. Uzan, Class. Quantum Grav. **14** (1997) 421; J. Katz, J. Bičák and D. Lynden-Bell, Phys. Rev. **D 55** (1997) 5957.
[97] B.B. Mandelbrot, "The Fractal Geometry of Nature", (W.H. Freeman, New York, 1983).
[98] A.K. Mittal and D. Lohiya, Fractals **11** (2003) 145.
[99] F. Sylos Labini, A. Gabrielli, M. Montuori and L. Pietronero, Physica **A226** (1996) 195; R. Scaramella *et al.*, Astron. Astrophys. **334** (1998) 404; K.K.S. Wu, O. Lahav and M.J. Rees, in *"49th Yamada Conference: Black Holes And High-Energy Astrophysics"*, eds. H. Sato and N. Sugiyama, (Tokyo Univ. Acad. Press, 1998), [arXiv:astro-ph/9804062]; M.N. Célérier and R. Thieberger, in Proc. 22nd Texas Symposium on Relativistic Astrophysics, Stanford University, eConf **C041213** (2004) 1404 [arXiv:astro-ph/0504442].