Dark energy and inhomogeneity

G F R Ellis
Mathematics Department, University of Cape Town, South Africa

Abstract. The standard understanding that distant supernova data implies existence of either
dark energy or a cosmological constant $\Lambda$ relies on the assumption of spatial homogeneity of the
universe. A large–scale smoothed–out model of the universe ignores small–scale inhomogeneities,
but the averaged effects of those inhomogeneities may alter both observational and dynamical
relations at the larger scale, mimicking the effect of a cosmological constant. Alternatively
large scale spatial inhomogeneity with zero $\Lambda$ might mimic the effect of a positive cosmological
constant. Whether this is the case or not is open to direct observational testing.

1. Inhomogeneity or Dark energy?
The explanation of dark energy is a central preoccupation of present day cosmology. Its presence
is indicated by the recent apparent speeding up of the expansion of the universe indicated by
supernova observations [51, 67, 2], which is usually taken to imply the existence of a cosmological
constant or quintessence causing acceleration of the universe at recent times [59], and is confirmed
by other observations such as those of cosmic background radiation anisotropies and Large Scale
Structure studies. Like dark matter, its existence was discovered, not predicted.

The astronomical observations are being refined in many sophisticated ways and used to
confirm the acceleration and test the equations of state of the hypothetical dark energy. Whether
it is constant or varying, its existence is a major problem for theoretical physics. It is therefore
crucial to pursue the possibility of other theoretical explanations.

The deduction of the existence of dark energy is based on the assumption that the universe
has a Robertson-Walker (RW) geometry - spatially homogeneous and isotropic on a large scale.
However the interpretation of these observations is ambiguous. The observations can at least in
principle be accounted for without the presence of any dark energy, if we consider the possibility
of inhomogeneity. This can happen in two ways: locally via backreaction and observational
effects, and globally via large scale inhomogeneity. I will consider these options in turn.

2. Different scale descriptions
Any mathematical description of a physical system depends on an averaging scale characterizing
the nature of the envisaged model. This averaging scale is usually hidden from view: it is
taken to be understood. Thus, when a fluid is described as a continuum, this assumes one is
using an averaging scale large enough that the size of individual molecules is negligible. If the
averaging scale is close to molecular scale, small changes in the position or size of the averaging
volume lead to large changes in the measured density and velocity of the matter, as individual

1 For full discussion, see the Journal of General Relativity and Gravitation special issue on dark energy: GERG
40: 219-660, 2008.
molecules are included or excluded from the reference volume. Then the fluid approximation is no longer applicable; rather one is using a detailed description of the fluid where individual molecules are represented. Usual work referring to the fluid density and velocity assumes a medium–size averaging scale: not so small that molecular effects matter, but not so large that spatial gradients in the properties of the fluid are significant ([3], p.5). The actual averaging scale, or rather the acceptable range of averaging scales, is not explicitly stated but is in fact a key–feature underlying the description used, and hence the effective macroscopic dynamical laws investigated. Indeed, different types of physics (particle physics, atomic physics, molecular physics, macroscopic physics, astrophysics) correspond to different assumed averaging scales. Thus, instead of referring to a density function \( \rho \), one should really refer to a function \( \rho_L \): the density averaged over volumes characterized by scale length \( L \). The key–point about the fluid approximation is that, provided this length scale is in the appropriate domain, then its actual value does not matter; i.e. when it is in this range, then changing \( L \) by a factor of 10, 100, or even much more makes no difference: the measured density and average velocity will not change. But if you change \( L \) by a very large amount until outside this range, this is no longer true. Hence, there is a range of validity \( L_1 < L < L_2 \) where the fluid approximation holds [3] and explicit mention of the associated averaging scale may be omitted.

In electromagnetic theory, polarization effects result from a large–scale field being applied to a medium with many microscopic charges. The macroscopic field \( E \) differs from the point–to–point microscopic field, which acts on the individual charges because of a fluctuating internal field \( E_I \), the total internal field at each point being \( D = E + E_I \) ([45], p.116). Spatially averaging, one regains the average field because the internal field cancels out: \( E = \langle D \rangle \), indeed this is how the macroscopic field is defined (implying invariance of the background field under averaging: \( E = \langle E \rangle \)). On a microscopic scale, however, the detailed field \( D \) is the effective physical quantity, and so is the field “measured” by electrons and protons at that scale. Thus, the way different test objects respond to the field crucially depends on their scale (a macroscopic device will measure the averaged field).

Now, exactly the same issue arises with regard to the gravitational field. Applications such as the solar system tests of general relativity theory are at solar system scales. We apply gravitational theory, however, at many other scales: to star clusters, galaxies, clusters of galaxies, and large–scale structures (walls and voids), as well as to black holes (occurring at solar system and star cluster scales, and possibly at much smaller scales).

2.1. Coarse–graining the gravitational field

A range of scales of description are relevant to cosmology. There are levels of approximation in modeling the universe, each with a hidden averaging scale. One can have a description in which every star is represented, or every galaxy (the stars averaged over), or only the largest scale cosmological structures (even galaxies averaged over, as in the fluid approximation).

The General Relativistic cosmological perturbation solutions used to study structure formation embody two interacting levels: the background (zero–order) model, almost always a Robertson–Walker metric, and the perturbed (first–order) model representing the growth of inhomogeneities, represented by a perturbed Robertson–Walker metric. This however may not represent accurately major inhomogeneities such as massive voids on scales bigger than clusters of galaxies. The question then is how do models on two or more different scales relate to each other in Einstein’s gravitational theory [31]. This is a difficult issue both because of the non–linearity of Einstein’s equations, and because of the lack of a fixed background spacetime – one of the core features of Einstein’s theory. This causes major problems in defining suitable averaging processes as needed in studying these processes. While there have been many analyses of this problem, there are still issues to be resolved in relation both to observations and dynamics, and in how this relates to gravitational entropy and the arrow of time. We will not consider the
latter here.

2.2. Non–commutativity of averaging and dynamics

The key–point in considering dynamical effects is that the two processes involved in relating the field equations at different scales do not commute [31]. These processes are:

E: calculating the Einstein tensor $G_{1ab} := R_{1ab} - \frac{1}{2} R_1 g_{1ab}$ from a metric tensor $g_{1ab}$, and, hence, determining the quantity $E_{1ab} := G_{1ab} - \kappa T_{1ab}$ for $g_{1ab}$, where $T_{1ab}$ is the matter tensor appropriate to the scale represented by $g_{1ab}$;

A: averaging the metric tensor $g_{1ab}$ to produce a smoothed metric tensor $g_{2ab}$:

$$g_{2ab} = \langle g_{1ab} \rangle$$

and the matter tensor $T_{1ab}$ to produce a corresponding smoothed matter tensor $T_{2ab}$:

$$T_{2ab} = \langle T_{1ab} \rangle.$$

Now in general the averaging process does not commute with taking derivatives: for a function $g$, usually $\partial_i \langle g \rangle \neq \langle \partial_i g \rangle$. Furthermore the inverse metric $g^{ab}$ (non–linearly dependent on the metric tensor components $g_{1ab}$) is not the smoothed version of $g^{ab}$. The resulting Christoffel terms $\Gamma^{ab}_{bc}$ are therefore not the smoothed version of $\Gamma^{ab}_{bc}$, hence the Ricci tensor components $R_{2ab}$, non–linearly dependent on $\Gamma^{ab}_{bc}$, are not the smoothed versions of $R_{1ab}$. Extra non–linearities occur in calculating the Einstein tensor $G_{2ab} = R_{2ab} - \frac{1}{2} R_2 g_{2ab}$ from the Ricci tensor $R_{2ab}$.

Thus, if you smooth first and then calculate the field equations, you get a different answer than if you calculate the field equations first and then smooth; symbolically $A(E(g_{1ab})) \neq E(A(g_{1ab}))$.

Suppose the field equations are true at the first scale:

$$E_{1ab} := G_{1ab} - \kappa T_{1ab} = 0,$$

then they will not be true at the second scale:

$$E_{2ab} := G_{2ab} - \kappa T_{2ab} \neq 0.$$  \hfill (1)

$$E_{2ab} := G_{2ab} - \kappa T_{2ab} \neq 0.$$  \hfill (2)

Thus, there will be an extra term in the equations at the smoother scale. We can either regard it as an extra term on the left–hand–side,

$$G_{2ab} - E_{2ab} = \kappa T_{2ab},$$

representing a modified curvature term, or as an extra term on the right–hand–side,

$$G_{2ab} = \kappa T_{2ab} + E_{2ab},$$

where it is regarded as an extra contribution to the matter tensor. Which is the more appropriate interpretation depends on the context.

Szekeres [77] developed a polarization formulation for a gravitational field acting in a medium, in analogy to electromagnetic polarization. He showed that the linearized Bianchi identities for an almost flat spacetime may be expressed in a form that is suggestive of Maxwell’s equations with magnetic monopoles. Assuming the medium to be molecular in structure, it is shown how, on performing an averaging process on the field quantities, the Bianchi identities must be modified by the inclusion of polarization terms resulting from the induction of quadrupole moments on the individual “molecules”. A model of a medium whose molecules are harmonic oscillators is discussed and constitutive equations are derived. This results in the form:

$$E^{2ab} = Q^{abcd} : cd,$$  \hfill (5)
that is $E_{2ab}$ is expressed as the double divergence of an effective quadrupole gravitational polarization tensor $Q^{abcd}$ with suitable symmetries:

$$Q^{abcd} = Q^{[ab][cd]} = Q^{cdab}.$$ (6)

Gravitational waves are demonstrated to slow down in such a medium.

The problem with such averaging procedures is that they are not covariant. They can be defined in terms of the background unperturbed space, usually either flat spacetime or a Robertson–Walker geometry, and so will be adequate for linearized calculations where the perturbed quantities can be averaged in the background spacetime (although even here the gauge problem arises, see below). But the procedure is inadequate for non–linear cases, where the integral needs to be done over a generic lumpy (non–linearly perturbed) spacetime that are not “perturbations” of a high–symmetry background. However, it is precisely in these cases that the most interesting effects will occur.

The only tensor integrals that are well–defined over a generic spacelike surface or spacetime region (and one interesting issue is which of these one should use in the averaging process) are for scalars [9, 10], unless one uses the bitensors associated with Synge’s world function [76], based on parallel propagation along geodesics, to compare tensors at different points in a normal neighbourhood. The problem is that they cannot be used for averaging the metric tensor, for it is the metric tensor itself that defines the parallel propagation used in this process, and so is left invariant by it (since $g_{ab, c} = 0$). So, one has to devise a procedure in which either the field equations are represented only in terms of scalars, possible for example if one takes components relative to a covariantly uniquely defined tetrad, or else bitensors are used to define averages of quantities other than the metric.

Zalaletdinov has taken this issue seriously, and provided the only sustained such attempt based on bitensors [88]. He proposes a macroscopic description of gravitation based on a covariant spacetime averaging procedure. The geometry of the macroscopic spacetime follows from averaging Cartan’s structure equations, leading to a definition of correlation tensors. Macroscopic field equations (averaged Einstein equations) can be derived in this framework. It is claimed that use of Einstein’s equations with a hydrodynamic stress–energy tensor means neglecting all gravitational field correlations, and a system of macroscopic gravity equations is given when the correlations are taken into consideration. This approach has not won many adherents, but is nevertheless a systematic and coherent attempt to set up the problem generically.

3. Small scale inhomogeneity: backreaction

The cosmological application is to understand the nature of the backreaction of perturbations in cosmology. The large-scale solution used is the Friedmann-Lemaître-Robertson-Walker (‘FLRW’) exactly spatially homogeneous and isotropic universe models. The real universe is not like this, it has vast inhomogeneities: walls of clusters of galaxies surrounding voids, forming a soap-bubble like structure. The FLRW models are only a smoothed approximation to this reality. All the issues discussed above arise.

Varied methods have been used to study this problem, including straightforward perturbation approaches. Isaacson’s method of averaging in the gravitational radiation case has been used in the cosmological context [37]. When Zalaletdinov’s approach to the averaging problem is applied to cosmology, the Einstein field equations on cosmological scales are modified by appropriate gravitational correlation terms [24]; for a spatially homogeneous and isotropic macroscopic spacetime, the correlation tensor is of the form of a spatial curvature term.

One sustained attack is that by Buchert [9, 10, 11]. Here one first thinks of averaging over scalar quantities like the density or the rate of expansion in an inhomogeneous universe model. As long as one works with exact equations for the evolution of those fields in a given foliation of
spacetime, such an averaging procedure is covariant. Cosmological parameters like the rate of expansion or the mass density are to be considered as volume–averaged quantities, so the relevant parameters are intrinsically scale–dependent unlike the situation in a Friedmann cosmology. Averaging scalar characteristics on a Riemannian spatial domain delivers the effective dynamical sources that an observer would measure, but although he measures within the lumpy spacetime, he is going to interpret his observations within a Friedmannian fitting model. This suggests a logical division of the averaging problem into 1) calculating averages in the real manifold, and 2) determining the mapping between averages in the real manifold and averages in the Friedmannian model. The first averaging is straightforward for scalars, and encounters non–commutativity of averaging and time–evolution: this is a purely kinematical property that can be expressed, for a scalar field $\psi$, through the rule

$$\partial_t \langle \psi \rangle - \langle \partial_t \psi \rangle = \langle \theta \psi \rangle - \langle \theta \rangle \langle \psi \rangle.$$  

(7)

The fluctuation part on the right–hand–side of this rule produces the kinematical backreaction. The result of this averaging is the Buchert modified Friedmann and Raychaudhuri equations [11]:

$$3(\dot{a}_D/a_D)^2 = \Lambda + 8\pi G \rho_D - (Q_D + \langle R \rangle_D)/2,$$

(8)

$$3(\dot{a}_D/a_D) = \Lambda - 4\pi G \langle \rho \rangle_D + Q_D,$$

(9)

where $a_D(t)$ is the volume averaged scale factor, $\langle R \rangle_D$ the averaged spatial curvature, and

$$Q_D := 2\langle II \rangle_D - \frac{2}{3} \langle I \rangle_D^2,$$

(10)

$$II := \theta^2/3 - \sigma^2, I := \Theta.$$  

(11)

This shows that averaging in principle allows acceleration terms to arise from the averaging process. The derivation of these equations, and their numerical implementation, has been studied by many others, see for example [4, 8].

The second “averaging” is more adequately thought of as a rescaling of the tensorial geometry. A (Lagrangian) smoothing as opposed to (Eulerian) rescaling of the metric on regional spatial domains has been proposed by Buchert and Carfora [12, 14], using a global Ricci deformation flow for the metric initially proposed by Carfora and Piotrkowska [17]. They introduced real–space renormalization group (RG) methods, based on properties of the Ricci-Hamilton flow. The smoothing of geometry implies a renormalization of averaged spatial variables, determining the effective cosmological parameters as they appear in the Friedmannian fitting model. Two effects that quantify the difference between background and real parameters were identified: curvature backreaction and volume effect [13]. Both are the result of an inherent non–commutativity of averaging and spatial rescaling. In this way we look at the averaging problem in two directions in function space: time–evolution (as a deformation in direction of the extrinsic curvature of the space sections encoding the kinematical variables), and scale–“evolution” (as a deformation in direction of the intrinsic 3–Ricci curvature).

Thus there are varied approaches to the averaging issue. However, it is not clear how they relate to the gauge problem.

3.1. The gauge issue

The gauge problem is the following: when you perturb a smooth background cosmological metric $g_{1ab}$ to obtain a perturbed metric $g_{ab} = \tilde{g}_{1ab} + h_{1ab}$, the inverse relation is not unique: there is no agreed averaging or fitting process that will give back a unique background metric $\tilde{g}_{ab}$ back from the “lumpy” metric $g_{ab}$ [35]. Some other smooth metric $\tilde{g}_{2ab}$ could have been
chosen as the background metric instead, leading to a different definition of the perturbations: $h_{ab} := g_{ab} - \bar{g}_{ab}$, instead of $h_{ab} := g_{ab} - \bar{g}_{ab}$. The choice of background metric $\bar{g}_{ab}$ for a specific “lumpy” metric $g_{ab}$ is called a ‘gauge choice’. The backreaction problem will look very different if described in terms of different gauges.

The best way to look at this is to think of a gauge choice as a mapping of a smooth background metric $\bar{g}_{ab}$ into the lumpy universe with metric $g_{ab}$ [34]. At each point in the real spacetime the density perturbation $\delta \rho$ is then defined by $\delta \rho := \rho - \bar{\rho}$, where $\rho$ is the actual density at that point, and $\bar{\rho}$ the background density at the same point. The key–issue is the choice of surfaces of constant time in the perturbed spacetime, conventionally taken to represent the image of surfaces of constant density of the background spacetime. It then becomes clear that one can for example set the density perturbation to zero by choosing the mapping so that the surfaces of constant background density $\bar{\rho}$ are the same as the surfaces of constant real density: for then at each point $\rho = \bar{\rho} \Rightarrow \delta \rho = 0$. However, with this choice, the fluid flow lines will not be orthogonal to the surfaces of constant density, so there will still be a non–zero density variation measured by comoving observers. Gauge issues arising in treating multi–component fluids raise extra issues because of the multiple possible choices of reference velocity field [26].

The remedy to this disconcerting behaviour is to choose gauge invariant variables, for example a set of covariantly defined variables that vanish in the background spacetime [34]. While many studies have been carried out for quantifying backreaction effects in cosmology, where the smoothed–out effect of the small–scale perturbations causes extra terms in the Friedmann equations for the background metric, none have been done that both fully and clearly take the gauge issue into account and go beyond linear order. This is a key–issue waiting to be resolved. One certainly wants to go at least to second order in understanding the effects of non–linear perturbations. Many of the crucial results at linear order no longer hold, for example scalar, vector and tensor perturbations are no longer independent of each other at second order [19], and then the backreaction in turn affects the perturbations themselves [61].

3.2. The fitting issue

Given the difficulty of handling averaging in a covariant and gauge invariant way, the conclusion may be that one should as far as possible use covariant variables, but choose a specific gauge when doing averaging. This means one needs to solve the Fitting Problem: one must decide which is the best way to fit a smooth background model to the lumpy real universe [35]. This decides what the magnitude of the backreaction effects is, because it determines the size of the "perturbation" away from the background at each point. One way of trying to do this is by averaging, for example one can fit the background model by averaging densities and pressures, or by fitting observable relations such as the $(m, z)$ and $(N, z)$ relations in order to determine $H_0$ and $q_0$ for the best-fit background model. These procedures will in general give different results, particularly because the former will be based on a spacelike averaging but the latter on a null cone averaging (because observations are made on the past null cone). Most observational studies (e.g. the classic study by Sandage [74]) implicitly use a null fitting procedure, without making this explicit.

A proper fitting procedure underlies detailed backreaction studies. If the wrong background is fitted then it may appear to create a back reaction which vanishes if a better fit is chosen. Thus a gauge choice is inherent in the backreaction issue; this will be apparent in any detailed studies, even if it is not clearly made manifest.

It is important to note the following: if you consider a FLRW model $G_1$ and perturb it to get an almost-FLRW model $H_1$, it is very tempting to assume that $H_1$ is the best-fit model to $G_1$: indeed this assumption is effectively hidden in most perturbation studies. But this is not necessarily the case. For example, if you add to $G_1$ a Gaussian distribution of density fluctuations suitably distributed over a spacelike surface, you will have raised the density on that
surface: so a process of fitting the background model by averaging densities will no longer result in determining $G_1$ as the best fit; it will rather yield a model $G_2$, with a higher average density than $G_1$. To preserve the best-fit model you started with, one needs to add a compensated set of perturbations (e.g. top-hat models) where an underdensity surrounds every overdensity, so that the average density is unchanged by the perturbation procedure. This is rarely done; but when it is not done, a change in the average properties will result in a spurious apparent back-reaction effect, which is in reality due to a failure to use the correct background model for the perturbed spacetime. A proper fitting procedure will lead to integral constraints related to the fitting procedure [78].

3.3. Importance of backreaction effects?
There is no doubt that interesting effects occur [36, 15, 68, 54]. The key issue is whether these backreaction effects are significant in cosmology. On the one hand, they might play a significant role in the inflationary era [63, 39]. On the other, it can possibly help explain the apparent dynamical existence either of dark energy and/or of dark matter as effective terms in the macroscopic dynamics at recent times. Various papers suggest the effect may indeed be significant, for example the observed acceleration of the universe could possibly be the result of the backreaction of cosmological perturbations rather than the effect of a negative–pressure dark energy fluid [82, 54, 62, 53] if astronomical parameters are restricted in specific ways [71]. However, other studies obtain different results [73, 15, 65, 66, 68, 69]. Gauge effects are problematic [38], and many doubt the effect is significant (e.g. [44]).

An interesting new development is the realisation that there are really large voids in the universe (see e.g. [72]) and these have important significance for averaging in cosmology, as emphasized by Wiltshire [85]. The geometry and dynamics in and out of a void are quite different: within voids where structures such as isolated galaxies are established, the geometry is locally quasi-static and vacuum dominated, completely different from the FLRW spatially homogeneous matter–dominated and evolving region outside the voids. They have to be joined with suitable junction conditions so that internal and external geometries agree.

One can model this situation by exact inhomogeneous models (next section), but often it is done by using standard perturbation methods based on a "Newtonianly perturbed FLRW metric", see Ishibashi and Wald [44]:

$$ds^2 = -(1 + 2\Psi)dt^2 + a^2(t)(1 - 2\Psi)\gamma_{ij}dx^idx^j$$

with $\gamma_{ij}$ the metric of a 3-space of constant curvature $k = +1, 0, \text{ or } -1$, and $\Psi$ satisfies

$$\left| \Psi \right| \ll 1, \quad |\partial\Psi/\partial t|^2 \ll (1/a^2)D^i\Psi D_i\Psi,$$

$$(D^i\Psi D_i\Psi)^2 \ll (D^iD^j\Psi)(D_iD_j\Psi),$$

where $D_i$ denotes the derivative operator associated with $\gamma_{ij}$. Note that the matter present is not assumed to be comoving in this frame. The assumption made in analyses such as that of Ishibashi and Wald is that one can use such coordinates globally. This is probably correct either in a perturbed FLRW domain, or in a locally isolated quasi-static domain, but it is not obvious such coordinates can be simultaneously used in both for any extended length of time. Such coordinates are probably only locally valid in a realistic description of the universe taking both expanding matter-filled domains and locally static vacuum domains into account; but it is the assumption that these coordinates can be used globally that leads to the conclusion that back-reaction effects are negligible (cf. [86, 52]). Its domain of validity in realistic situations needs to be investigated.

Thus a key issue is, How large in space and time is the domain where such quasi-Newtonian coordinates (12)-(14) can be used in a realistic model of an expanding universe with voids? One
may note here that the timelike reference congruence associated with such coordinates is shear-free, conformally mapping the 3-spaces onto each other. Assuming such a congruence exists is a major assumption, placing strong limits on the Weyl tensor, inter alia excluding occurrence of vector and tensor perturbations. It may be acceptable in the circumstances considered because matter is not comoving in this frame, and metric perturbations are small despite density perturbations being large; but this assumption certainly needs investigation.

An alternative to using these perturbation techniques is looking for suitable exact inhomogeneous solutions of the field equations. The first important set often used are the Lemaître-Tolman-Bondi (‘LTB’) spherically symmetric pressure-free models [7], discussed further below. The second ones are Swiss-cheese models, with spherically symmetric static vacuum domains (hence Schwarzschild solutions) imbedded in a FLRW universe $G_1$ (Einstein and Strauss [29], Harwit [40]). These can two sets can be blended by imbedding LTB models in a FLRW universe. One can then study averaging and observational issue in these models. If the fitting is done properly, matching metrics and masses at the junctions, no strange matter exists and the junctions and the existence of the lumps does not influence the background model FLRW $G_1$ first considered; however that may not then be the best fit model $G_2$, depending on the fitting procedure used (cf. section 3.2 above). However $G_2$ and $G_1$ are not likely to be very different; so the dynamical effect is not likely to be large. However the observational effect is a different matter, as we consider in the next section.

Before doing so, I remark on a rather different way of treating homogeneity: Lindquist and Wheeler [60] examined joining many Schwarzschild cells together, with boundaries subject to equations of motion that make the whole expand as an approximation to a $k = +1$ FLRW universe. They found that regular lattices of this type are possible only for $n = 5, 8, 16, 124$, or 600 vertices. The paper compares these lattice universe with closed Friedmann universes. This differs from the Swiss Cheese models, because in this case there are no FLRW domains used in addition to Schwarzschild cells: rather the composite of Schwarzschild cells IS the approximation to a FLRW universe model. locally static but globally expanding. The Schwarzschild cell method predicts the same relation between radius of the universe and proper time, except that the connection between maximum radius and mass is different. This is the way averaging effects lead to a different FLRW model than an exactly smooth distribution of the same matter. What would be interesting would be extension of such models to irregular lattices; but this might be technically difficult. Again the issue will arise of Best-Fitting of the background model derived in this way.

A tentative interim conclusion is that the dynamical effects are certainly there, but are small [4, 8]. In any case, a detailed investigation of backreaction effects helps to improve fitting models on regional scales for a better interpretation of observational data.

4. Small scale inhomogeneity: observations
The effect on observational relations of introducing local inhomogeneities into a given background spacetime is double: it alters redshifts, and it changes area distances.

4.1. Redshift effects
When a void is intervening between the source and the observer, photons drop into a potential well and then climb out, and they exit when the universe is larger than when they went in. If all is static inside the void the redshift changes in and out cancel, but when structure is forming the potential well is changing with time so one gets the Rees-Sciama effect [70]: a change in redshift due to change in the potential well as the photon traverses it. A non-zero cosmological constant will also lead to such an effect.

When the source is in the void, the photon has to climb out of the void, giving a contribution to observed redshifts. Wiltshire emphasizes [83] that time runs at different rate in and out
of voids in a void dominated universe, and the difference is cumulative. This can be thought of as the source of an extra redshift contribution, changing the observable relations between cosmological variables.

4.2. Area distance effects

The usual analysis of cosmological observations is based on the equations relating apparent magnitude and redshift in Robertson–Walker spacetimes. The Sachs optical scalar equations for hypersurface–orthogonal null geodesics in a general space-time are:

$$\frac{d\theta}{dv} = - R_{ab} K^a K^b - 2\sigma^2 - \frac{1}{2} \theta^2; \quad \frac{d\sigma_{mn}}{dv} = - E_{mn},$$

where $\theta$ is the rate of expansion of the null geodesics with tangent vector $K^a = dx^a/dv$ and affine parameter $v$, $\sigma_{mn}$ is their shear, $R_{ab}$ the Ricci tensor, and $E_{mn}$ a matrix of Weyl tensor components ([41], p.88; [75], pp.108–9). In the idealized Robertson–Walker case, the Weyl tensor $C_{abcd}$ vanishes, but the Ricci tensor is non–zero, being given via the Einstein field equations from the matter present. Thus, $C_{abcd} = 0 \Rightarrow E_{mn} = 0$, and the relevant solutions are shear–free:

$$\sigma^2 = 0 \Rightarrow \frac{d\theta}{dv} = - R_{ab} K^a K^b - \frac{1}{2} \theta^2.$$  

(16)

Integration gives the Mattig relations applicable to Friedmann universe models [74, 30, 75].

However, in the real universe, as pointed out by Bertotti [5], observations take place via null geodesics lying in the empty spacetime between galaxies. Thus, the real situation in a universe with no intergalactic medium (all the matter is concentrated in galaxies) is the opposite of that above: in the region of spacetime traversed by the geodesics, the Ricci tensor vanishes, so

$$\frac{d\theta}{dv} = -2\sigma^2 - \frac{1}{2} \theta^2, \quad \frac{d\sigma_{mn}}{dv} = - E_{mn},$$

(17)

The Weyl tensor (the tidal gravitational field caused by nearby matter) generates shear that then causes focussing. Thus, the microscopic description of the focussing caused by local inhomogeneities in the vacuum gravitational field ($\sigma \neq 0$, $R_{ab} = 0$, $E_{mn} \neq 0$) is radically different from the macroscopic one where focussing is caused by a smooth matter distribution only ($\sigma = 0$, $R_{ab} \neq 0$, $E_{mn} = 0$), and the area distance–redshift relation may be expected to be different on microscopic scales (i.e. the small solid angle bundles of null geodesics actually used in observations of individual objects), as compared with macroscopic scales (averaging over large solid angles).

Thus Ricci focusing on the large scale is replaced by Weyl focusing on small scales. This micro-focussing due to the Weyl tensor must give the FRW equations when averaged over whole sky, but how this happens is not obvious! It has been suggested that energy conservation will imply this FRW all-sky average [81], but this assumes that areas of a bundle of null geodesics are the same in the perturbed and background models, which will not be the case when one takes the effect of caustics into account [33]. Indeed, areas increase slower than in a Robertson–Walker model in the empty spaces between matter, where the Ricci term is zero, and faster in the high–density regions where matter is concentrated, so one might think these effects cancel out. However, the strongly lensed rays soon go through a caustic and emerge highly divergent, so that areas are rapidly increasing again. It is plausible that on average the overall effect is always an increase in area, that is a lesser area distance than in the smooth background model. But in any case supernova observations are preferentially in directions where there is no intervening matter: they will not correspond to an all-sky average.

Various proposals have been made to deal with this. The most popular is the Dyer–Roeder distance [27, 28], obtained by assuming only a fraction $f$ of the total mass density is encountered
by the light–rays but ignoring the shear. Thus, in (16) one replaces $R_{ab}K^aK^b$ by $fR_{ab}K^aK^b$ and works out the corresponding area distance ([75], pp.138–143; [25]). This may be a good approximation if galaxies are embedded in a fairly uniform intergalactic medium of dark matter, but clearly does not take shear effects and caustics properly into account. How good it is will depend on the nature of clustering in the universe and how the averaged distribution impacts along the line of sight [58]. If the dark matter is uniform, Dyer-Roeder is good; if dark matter is clustered, it is not so good.

One can approach the topic in other ways: for example by using stochastic methods [5], or detailed examination of geodesics in Swiss–Cheese universe models with FLRW regions joined to vacuum regions. In these exact inhomogeneous solutions, the null geodesic equations can be exactly integrated in each domain and matched; this includes shear effects.

Kantowski investigated this in a series of papers, and gave a nice geometrical confirmation of the effect on observational relations [46]. He obtained analytic expressions for distance–redshift relations that have been corrected for the effects of inhomogeneities in the density [48]. The values of the density parameter and cosmological constant inferred from a given set of observations depends on the fractional amount of matter in inhomogeneities and can significantly differ from those obtained by using the Mattig relations for the FLRW universes. As an example, a determination of $\Omega_0$ made by applying the homogeneous distance–redshift relation to SN 1997ap at $z = 0.83$ could be as much as 50% lower than its true value [48]. When $\Lambda = 0$ the equation is an associated Legendre equation and hence distance-redshift can be written using associated Legendre functions. When $\Lambda$ is not zero the equation is a Lame’ equation which is a special type of the Heun equation. The distance-redshift using Heun functions is in [48, 50], solutions depending on the fraction of the universe’s mass density that is transparent to the observing beams. The flat space case is special, even when $\Lambda$ is not zero; for this case distance-redshift can be written using associated Legendre functions [49].

These results make clear that the apparent acceleration term detected could be at least partly due to this optical effect: focussing of null geodesics is different in a lumpy universe than in a smooth one.

4.3. The combination of effects

Putting this together, averaging processes when there are many local inhomogeneities leads both to dynamical back reaction effects, and optical effects due both to altered redshifts and area distances. This combination of effects seems to have the potential to significantly influence interpretation of observational relationships such as the supernova data.

Is this influence of inhomogeneities sufficient to fully explain the apparent acceleration indicated by the supernova data? It is claimed by some that it is, see for example Leith, Ng and Wiltshire [55], utilizing Wiltshire’s solution of the Buchet equations [84]. However the modeling used is controversial, and not yet universally accepted. Investigations such as those by Kantowski [46, 47, 48] and Clifton et al [23] suggest it is certainly large enough to significantly affect the the Hubble diagram: "It is found that intervening voids, between the observer and source, have no noticeable effect, while sources inside voids can be effected considerably. By averaging observable quantities over many randomly generated distributions of voids we find that the presence of these structures has the effect of displacing the average magnitude from its background value, and introducing a dispersion around that average" [23]. A contrasting view, suggesting the effect is much smaller, is in [80]. Thus while it is very debatable if it is enough to fully account for the apparent acceleration, it may well be enough to significantly influence the concordance model parameter values, and hence the present understanding of inflationary effects in those models. Clearly, this topic needs further careful investigation.
5. Large scale inhomogeneity

Perhaps the Copernican assumption of large-scale spatial homogeneity is not true. Maybe there is a large scale inhomogeneity of the observable universe, such as that described by the Lemaître-Tolman-Bondi (LTB) pressure-free spherically symmetric models [7], and we are near the centre of a void. The redshift-distance relation for distant supernova may be measuring spatial inhomogeneity, rather than acceleration of a FLRW universe.

5.1. Can we fit the observations?

The LTB models have comoving coordinates

\[ ds^2 = -dt^2 + B^2(r,t) + A^2(r,t)(d\theta^2 + \sin^2\theta d\phi^2) \]  \hspace{1cm} (18)

where

\[ B^2(r,t) = A'(r,t)^2(1 - k(r))^{-1} \]  \hspace{1cm} (19)

and the evolution equation is

\[ (\dot{A}/A)^2 = F(r)/A^3 + 8\pi G\rho_M/3 - k(r)/A^2 \]  \hspace{1cm} (20)

with the energy density given by

\[ F'(A'A)^{-1} = 8\pi G\rho_M. \]  \hspace{1cm} (21)

There are two arbitrary functions of the spatial coordinate \( r \): namely \( k(r) \) (curvature) and \( F(r) \) (matter). This freedom enables us to fit the supernova observations with no dark energy or other exotic physics (this is a theorem, see Mustapha et al [64]). The idea that such models can explain the supernova observations without any dark energy is discussed by Celérié [18]. One can also fit the CBR observations because they refer to much larger values of \( r \) [1, 22]. One should note here that at least some of the observed CBR dipole can then arise because we are a bit off-centre in the void, so once one can re-evaluate the great attractor analysis in this context and the alignment of the dipole and quadrupole. Nucleosynthesis data can also be fitted; what is a bit more problematic is the Baryon Acoustic Oscillations (\'BAO\'). The key comment to make is that different scales are probed by different observations and can in principle all be fitted by adjusting the free spatial functions at different distances.

A typical observationally viable model is one in which we live roughly centrally (within 10% of the central position) in a large void: a compensated underdense region stretching to \( z \approx 0.08 \) with \( \delta\rho/\rho \approx -0.4 \) and size 160/\( h \) Mpc to 250/\( h \) Mpc, a jump in the Hubble constant of about 1.20 at that distance, and no dark energy or quintessence field present (Biswas et al [6], Ishak et al [43], Yoo et al [87]). The details are being tested by many further models, but the basic point is clear: it is indeed possible at first order to fit the observations by such inhomogeneous models with no dark energy present. In an era of precision cosmology, the present debate is about more accurate modeling, and obtaining a better model-dependent better fit to ever more detailed observations.

5.2. Dynamic evolution?

Given we can fit the observations by such a model, can we find dynamics (inflation followed by a HBB era) that can lead to such a model? It has the same basic dynamics as the standard model (evolution along individual world lines governed by the Friedmann equation) but with distant dependent parameters. Will inflation prevent it? This depends on the initial data, the amount of inflation, and the details of the unknown inflaton. If we are allowed the usual tricks of fiddling the inflationary potential and initial data, and adding in multiple fields as desired, then there is sufficient flexibility that it should certainly be possible.
5.3. Improbability
Many dismiss these models on probability grounds: It is improbable the universe is like this, and it is improbable we are near the centre of such a model. But there is always improbability in cosmology. We can can shift it around, but it is always there. It might be in the nature of a Robertson-Walker geometry (the old view), in the inflationary potential and initial conditions (the current mainstream position), which specific universe domain we are in within a multiverse, or the spatial position in an inhomogeneous universe (the present proposal). Note that we are competing with a probability of $10^{-120}$ for $\Lambda$ in a FLRW universe; we do not have to get very high probabilities to outdo that improbability, which is what multiverse proposal aim to handle.

Three comments are in order. First, a key feature of cosmology is that there is only one universe; and the very concept of probability does not apply to a single object, even though we can make many measurements of that single object to determine its detailed nature. Probability applies to the multiple measures we can make of the single universe, but not to issues doing with the existence of the universe itself (Ellis [32]). There is no physically realised ensemble to apply that probability to, unless a multiverse exists in physical reality which is not proven. In essence, there simply is no proof the universe is probable; that is a philosophical assumption, which may not be true. The universe may be improbable!! Secondly, there is no well-justified measure for any such probability proposal even if we ignore the first problem. This is still an issue of debate.

And thirdly, a study by Linde et al [57] shows that (given a particular choice of measure) this kind of inhomogeneity actually is a probable outcome of inflationary theory, with ourselves being located near the centre! One cannot dismiss such models out of hand for probability reasons.

6. Observational tests of spatial homogeneity
Given that we can find inhomogeneous models that reproduce the observations without any exotic energy, as well as homogeneous models that explain the same observations through some form of dark energy, how can we distinguish between the two? Ideally we need a model-independent test of the basic assumption of most present day cosmology: is a RW geometry the correct metric for the observed universe region? Is the Copernican principle correct? Four kinds of tests are possible, as I now discus.

6.1. CBR based tests
Some tests use scattered CBR photons to check spatial homogeneity (Goodman [42]; Caldwell and Stebbins [16]). If the CBR radiation is anisotropic around distant observers (as will be true in inhomogeneous models), Sunyaev-Zeldovich scattered photons have a distorted spectrum that reflects the spatial inhomogeneity. However this test is somewhat model dependent - it is good for void models but misses, e.g., conformally stationary spacetimes. It also has to take into account other possible causes of spectral distortion.

6.2. Direct observational tests: behaviour near origin
The universe must not have a geometric cusp at the origin, as this implies a singularity there. Thus there are centrality conditions that must be fulfilled in the inhomogeneous models (Vanderveld et al [79]). The distance modulus behaves as

$$\Delta dm(z) = -(5/2)q_0 z$$

in standard $\Lambda$CDM models for small $z$, but if this were true in a LTB void model without $\Lambda$ this implies a singularity (Clifton et al [21]). Observational tests of this requirement will be possible through intermediate redshift supernovae in the future.
6.3. Direct observational tests: constancy of curvature

There are two geometric effects on distance measurements: curvature $\Omega_k$ bends null geodesics, expansion $H(z)$ changes radial distances. These are coupled in RW models, as expressed in the relation

$$d_L(z) = \frac{(1 + z)}{H_0 \sqrt{-\Omega_k}} \sin \left( \sqrt{-\Omega_k} \int_0^z \frac{dz'}{H(z')} \right).$$

(23)

While these effects are strictly coupled in RW geometries, they are decoupled in LTB geometries.

In RW geometries, we can combine the Hubble rate and distance data to find the curvature today:

$$\Omega_k = \left[ \frac{H(z)D'(z)}{H_0 D(z)} \right]^2 - 1$$

(24)

This relation is independent of all other cosmological parameters, including dark energy model and theory of gravity. It can be used at single redshift to determine $\Omega_k$. The exciting result of Clarkson et al ([20]) is that since $\Omega_k$ is independent of $z$, we can differentiate to get the consistency relation

$$C(z) := 1 + H^2 (D''(z) - D'^2) + H H' D' D'' = 0,$$

(25)

which depends only on a RW geometry: it is independent of curvature, dark energy, nature of matter, and theory of gravity. Thus it gives the desired consistency test for spatial homogeneity. In realistic models we should expect $C(z) \approx 10^{-5}$, reflecting perturbations about the RW model related to structure formation. Errors may be estimated from a series expansion

$$C(z) = \left[ q_0^{(D)}(H) \right] z + O(z^2)$$

(26)

where $q_0^{(D)}$ is measured from distance data and $q_0^{(H)}$ from the Hubble parameter. It is simplest to measure $H(z)$ from BAO data. It is only as difficult carrying out this test as carrying out dark energy measurements of $w(z)$ from Hubble data, which requires $H'(z)$ from distance measurements or the second derivative $D''(z)$.

This is the simplest direct test of spatial homogeneity, and its implementation should be regarded as a high priority: for if it confirms spatial homogeneity, that reinforces the evidence for the standard view in a satisfying way; but if it does not, it has the possibility of undermining the entire project of searching for a physical form of dark energy.

6.4. Indirect Observational tests

If the standard inverse analysis of the supernova data to determine the required equation of state, as discussed at this meeting, shows there is any redshift range where $w := p/\rho < -1$, this may well be a strong indication that one of these geometric explanations is preferable to the Copernican (Robertson-Walker) assumption, for otherwise the matter model indicated by these observations is non-physical (it has a negative kinetic energy).

There is already data suggesting this may be the case, see e.g. Lima et al [56]. There are some attempts to generate matter models that will give this kind of behaviour without negative kinetic energies, for example through dark energy- dark matter interactions, but they are very speculative physically and with no tested physical basis, supposing multiple unknown forms of matter or energy with arbitrarily proposed interactions between them. It all seems rather reminiscent of the Ptolemaic epicycles for describing the solar system, with more and more complex mechanisms added in as the simpler ones failed to describe the phenomena. The physically most conservative approach is to assume no unusual dark energy or exotic interacting fields, but rather that an inhomogeneous geometry might be responsible for the observed apparent acceleration; this should be seriously considered as an alternative.
7. Conclusion
The issue of what is testable and what is not testable in cosmology is a key issue. The acceleration indicated by supernova data could possibly be due to small scale inhomogeneity that definitely exists, but may not be sufficiently significant to do the job. It could be due to large scale inhomogeneity that can probably do the job, but may not exist. Observational tests of the latter possibility are as important as pursuing the dark energy (exotic physics) option in a homogeneous universe. Theoretical prejudices as to the universe's geometry, and our place in it, must bow to such observational tests.

We should stand firm and insist that genuine science is based on observational testing of plausible hypotheses. There is nothing wrong with physically motivated philosophical explanation: but it must be labeled for what it is. Overall: theory must be subject to experimental and/or observational test; this is the central feature of science. Whatever position we may have on the issue of probability, in the end our philosophy on this question will have to give way to any such possible observational tests.

References
[1] Alexander S, Biswas T, Notari A and Vaid A 2007 Local void vs dark energy: confrontation with WMAP and Type IA supernovae Preprint arXiv:0712.0370
[2] Barausse E, Matarrese S and Riotto A 2005 Phys. Rev. D 71 063537
[3] Batchelor G K 1967 An Introduction to Fluid Dynamics (Cambridge, Cambridge University Press)
[4] Behrend J, Brown I A and Robbers G 2008 JCAP 08 013
[5] Bertotti B 1966 Proc. Roy. Soc. London A 294 195
[6] Biswas T, Monsouri R and Notari A 2007 JCAP 12 017
[7] Bondi H 1947 MNRAS 107 910
[8] Brown I A, Robbers G and Behrend J 2008 Averaging Robertson-Walker Cosmologies Preprint arXiv:0811.4495
[9] Buchert T 2000 Gen. Rel. Grav. 32 105
[10] Buchert T 2001 Gen. Rel. Grav. 33 1381
[11] Buchert T 2008 Gen. Rel. Grav. 40, 467
[12] Buchert T and Carfora M 2002 Class. Quant. Grav. 19 6109
[13] Buchert T and Carfora M 2003 Phys. Rev. Lett. 90 031101
[14] Buchert T and Carfora M 2008 Class. Quant. Grav. 25 195001
[15] Buchert T, Kerscher M and Sicka C 2000 Phys. Rev. D 62 043525
[16] Caldwell R and Stebbins A 2008 Phys Rev Lett. 100 191302
[17] Carfora M and Pietrokowska K 1995 Phys. Rev. D 52 4393
[18] Célérier M-N 2007 New Advances in Physics 1 29
[19] Clarkson C 2004 Phys. Rev. D 70 129902
[20] Clarkson C, Bassett B and Lu T H-C 2008 Phys. Rev. Lett. 101 011301
[21] Clifton T, Ferreira P G and Land K 2008 Phys. Rev. Lett. 101 131302
[22] Clifton T, Ferreira P G and Zuntz J 2009 What the small angle CMB really tells us about the curvature of the Universe Preprint arXiv:0902.1313
[23] Clifton T and Zuntz J 2009 Hubble Diagram Dispersion From Large-Scale Structure Preprint arXiv:0902.0726 (2009)
[24] Coley A A, Pelavas N and Zalaletdinov R M 2005 Phys. Rev. Lett. 95 151102
[25] Demianski M, de Ritis R, Marino A A and Piedipalumbo E 2003 Astron. Astrophys. 411 33
[26] Dunsby P K S, Bruni M and Ellis G F R 1992 Astrophys. J. 395 54
[27] Dyer C C and Roeder R 1974 Astrophys. J. 189 167
[28] Dyer C C and Roeder R 1975 Astrophys. J. 196 671
[29] Einstein A and Strauss E G 1945 Rev. Mod. Phys. 17 120
[30] Ellis G F R 1971 in General Relativity and Cosmology, Proc. Int. School of Physics “Enrico Fermi” (Varenna), Course XLVII, Ed R K Sachs (Academic Press) 104–179
[31] Ellis G F R 1984 in General Relativity and Gravitation, Ed B Bertotti et al (Reidel) 215
[32] Ellis G F R 2006 in Handbook in Philosophy of Physics Ed J Butterfield and J Earman (Elsevier) 1183
[33] Ellis G F R, Bassett B A C C and Dunsby P K S 1998 Class. Quant. Grav. 15 2345
[34] Ellis G F R and Bruni M 1989 Phys. Rev. D 40 1804
[35] Ellis G F R and Stoeger W R 1987 Class. Quant. Grav. 4 1697
[36] Futamase T 1991 Prog. Theor. Phys. 86 389
