(Mis)interpreting supernovae observations in a lumpy universe

Chris Clarkson,1⋆ George F. R. Ellis,1 Andreas Faltenbacher,2 Roy Maartens,2,3 Obinna Umeh1 and Jean-Philippe Uzan1,4,5

1Astrophysics, Cosmology & Gravity Centre, and Department of Mathematics & Applied Mathematics, University of Cape Town, Cape Town 7701, South Africa
2Physics Department, University of the Western Cape, Cape Town 7535, South Africa
3Institute of Cosmology & Gravitation, University of Portsmouth, Portsmouth PO1 3FX
4Institut d’Astrophysique de Paris, UMR-7095 du CNRS, Université Pierre et Marie Curie, 75014 Paris, France
5National Institute for Theoretical Physics, Stellenbosch 7600, South Africa

Accepted 2012 July 19. Received 2012 July 11; in original form 2012 May 22

ABSTRACT

Light from ‘point sources’ such as supernovae is observed with a beam width of the order of the sources’ size – typically less than 1 au. Such a beam probes matter and curvature distributions that are very different from coarse-grained representations in $N$-body simulations or perturbation theory, which are smoothed on scales much larger than 1 au. The beam typically travels through unclustered dark matter and hydrogen with a mean density much less than the cosmic mean, and through dark matter haloes and hydrogen clouds. Using $N$-body simulations, as well as a Press–Schechter approach, we quantify the density probability distribution as a function of beam width and show that, even for Gpc-length beams of 500 kpc diameter, most lines of sight are significantly underdense. From this we argue that modelling the probability distribution for au-diameter beams is absolutely critical. Standard analyses predict a huge variance for such tiny beam sizes, and non-linear corrections appear to be non-trivial. It is not even clear whether underdense regions lead to dimming or brightening of sources, owing to the uncertainty in modelling the expansion rate which we show is the dominant contribution. By considering different reasonable approximations which yield very different cosmologies, we argue that modelling ultra-narrow beams accurately remains a critical problem for precision cosmology. This could appear as a discordance between angular diameter and luminosity distances when comparing supernova observations to baryon acoustic oscillations or cosmic microwave background distances.

Key words: methods: numerical – supernovae: general – cosmology: theory – dark matter.

1 INTRODUCTION: ON NARROW BEAMS

Supernovae Ia (SNIa) observations play a critical role in evidence for a non-zero cosmological constant (see Amanullah et al. 2010 and references therein). SNIa are effective standard candles (we think their intrinsic luminosity can be calibrated from their light curve). In the standard cosmological model, their observed luminosity is used to infer their luminosity distance (or equivalently magnitude) by assuming that the geometry of the universe is well described by a Friedmann–Lemaître (FL) background geometry that describes the universe smoothed on large scales. Of course, photons do not propagate in the geometry of a smooth FL space–time but in the lumpy universe: a beam mostly propagates in underdense regions between clustered matter (overdense islands of matter). The problem of quantifying the effects of propagation in an inhomogeneous universe was first addressed, as far as we are aware, independently by Zel’dovich (1964) and R. P. Feynman (1964, unpublished talk). Zel’dovich introduced the empty-beam approximation to deal with light rays propagating in vacuum and this was extended to the case of a partially filled beam by Dashevskii & Slysh (1966). These results later came to be known as the Dyer–Roeder (DR) approximation – see below. Zel’dovich’s insight also led to other works on the problem in the 1960s (Bertotti 1966; Gunn 1967; Kantowski 1969; Refsdal 1970).

Light propagation in inhomogeneous space–times gives rise to both the distortion of images and the magnification of some images, because of gravitational lensing; in compensation, most images are demagnified. These effects induce, in particular, a dispersion of the observed SNIa luminosities and hence an extra scatter in the Hubble diagram (Kantowski, Vaughan & Branch 1995; Frieman 1996; Wang 2000a,b, 2005). 'Precision cosmology' within the standard
approach could be compromised by the effects of lensing on the interpretation of SNIa data — and thus it is crucial to characterize the magnitude of these effects precisely. A related key question is ‘what physical and angular sizes are relevant in estimating these effects on SNIa observations?’

A perturbative approach (i.e. with light propagating in a perturbed FL space–time) shows that the dispersion due to a large-scale structure becomes comparable to the intrinsic dispersion for redshifts \( z > 1 \) (Holz & Linder 2005). However, since the matter fluctuations responsible for the magnification of the SNIa also induce a shearing of the images of background galaxies, this dispersion can actually be corrected (Cooray, Holz & Huterer 2006a; Dodelson & Vallinotto 2006). A similar idea was pursued in the context of gravitational wave sirens (Shapiro et al. 2010). Nevertheless, a considerable fraction of the lensing dispersion arises from subarcminute scales, which are not probed by shear maps smoothed on arcminute scales (Dalal et al. 2003).

To estimate the dispersion induced by inhomogeneities, one first needs to determine the typical size of the geodesic bundle associated with SNIa. The typical observational aperture is of the order of 1 arcsec, whereas the beam is actually much thinner: \( \Theta(1) \) arc for a source at redshift \( z \approx 1 \), i.e. an aperture of \( \beta \sim 10^{-2} \) arcsec. This is typically smaller than the mean distance between any massive objects (galaxies, stars, H clouds, small dark matter haloes) — and on a scale where the fluid continuum model may not be suitable any more. Thus, the beam propagates in preferentially low-density regions with rare encounters of gravitationally collapsed, high-density patches (haloes) resulting in highly inhomogeneous geometry.

In contrast, the standard approach implicitly uses a perturbative analysis and a fluid continuum model by treating the beam as propagating in the background and perturbed FL geometry. From this viewpoint, it is surprising that the standard analysis of SNIa data leads to a consistent result, in particular with other cosmological probes. Is a smooth cosmological model a good description of our universe, and in particular for interpreting data such as SNIa? If so, can we understand clearly why this is the case?

Standard perturbation theory reveals that there are problems. If the angular diameter distance as a function of redshift in a perturbed FL model differs from the background by \( \kappa(\beta) \), where \( \beta \) is the angular scale of observation (see below for definitions), then the variance of \( \kappa \) scales as

\[
\langle \kappa^2(\beta) \rangle^{1/2} \sim 10^{-2} \sigma_8 \Omega_0^{0.75} z_s^{0.8} \left( \frac{\beta}{10} \right)^{-1-n/2},
\]

as derived in Bernardeau (1997) and Mellier (1999), using linear perturbation theory (\( n \) is the spectral index of the power spectrum of density fluctuations and \( z_s \) is the redshift of the source). This estimate was confirmed in Jain & Seljak (1997) considering the non-linear evolution of the power spectrum, but on scales below 10 arcmin, \( \langle \kappa^2(\beta) \rangle \) increases more steeply than the theoretical expectation of the linear theory and is two to three times higher. Note also that equation (1) implies that \( \langle \kappa^2(\beta) \rangle^{1/2} \) becomes of the order of \( 2 \times 10^{-2} \) on an angular scale of the order of 1 arcmin, so that the variance becomes much larger than unity on the typical angular size of the ray bundle for SNIa. Therefore, the large-scale structure induces a stochastic dispersion of the luminosity distance which is difficult to quantify with standard techniques for narrow beams. As soon as one goes down to much smaller scales, inhomogeneities also induce a systematic shift away from the background, since one cannot neglect the higher order terms. This is much more serious. The magnification of a source behaves as \( \mu \sim 1 + 2\kappa + 3\kappa^2 + |\gamma|^2 + \ldots \), where \( \gamma \) is the shear of the source (defined below), and hence the mean of the magnification is \( \langle \mu \rangle \sim 1 + 3\kappa^2 + |\gamma|^2 + \ldots \neq 1 \). Thus, if the variance is large on small angular scales, we expect the overall shift to be significant — potentially of order unity or larger — on these scales too. Modelling this shift accurately is critical for interpreting SNIa observations correctly.

To estimate the effect of the inhomogeneities on smaller scales, one needs to provide a better description of the matter distribution. Attempts to include a uniform component, high-density haloes and low-density zones (filaments and voids) have been proposed by Kaimailainen & Marra (2009, 2011), but none go down to the required scale.

Narrow light bundles travel large distances (\( > 100 \ h^{-1} \text{Mpc} \)) with a very low probability of encountering dark matter haloes of substantial mass, which we quantify in the following section. The cuspy density profiles of the haloes additionally reduce the probability of a bundle to cross the central, high-density regions. Thus, the bundles are subject mainly to Weyl focusing (i.e. induced by the gradient of the gravitational potential). These light rays are expected to be demagnified compared to light rays propagating in an FL space–time of mean matter density. When one averages such ray bundles over the whole sky, this dimming is compensated by a small number of ray bundles that encounter very large density inhomogeneities, which then results in high focusing for those directions and a magnification. The general averaging argument was put forward by Weinberg (1976).

A similar argument can be employed to determine the average density encountered by a light bundle from a supernova. The probability of a light bundle encountering massive haloes decreases with halo mass. Massive galaxies, groups and clusters (\( > 10^{12} \ h^{-1} \text{M}_\odot \)) are rarely encountered, yet, they comprise \( \sim 50 \) per cent of the total mass within the universe. Thus, if we measure SNIa in typical directions in the sky, we observe them in directions where the density of matter encountered by the relevant ray bundles may be expected to be less than the cosmological average (for almost all directions are of this nature). In these circumstances, one expects not only an extra dispersion in the Hubble diagram, induced by the spatial inhomogeneity of the intervening medium, but also a systematic shift, induced by an observational selection effect, which may well be significant. This argument does not apply to the much larger angular scales relevant to measurements of the baryon acoustic oscillations (BAO) and cosmic microwave background (CMB) peaks. These beams do encounter sufficient matter to on average correspond to the overall cosmological density, as argued by Weinberg (1976).

The objective of this paper is to investigate these effects. We first discuss modelling the matter distribution in the real universe. Then we consider the general relativistic problem of light rays in an arbitrary space–time, with a focus on how a light beam reacts to inhomogeneities compared to a smooth space–time. We then consider some different approximations used to model narrow beams, such as perturbed FL models, the DR approximation, as well as presenting two new approximations. Finally, we consider the problem of how Weyl focusing by many point sources is converted into Ricci focusing associated with a smooth matter distribution.

2 THE MATTER DISTRIBUTION

According to the most accepted variants of the \( \Lambda \) cold dark matter (ΛCDM) paradigm, gravitationally collapsed structures span a mass range from \( 10^{-8} \text{M}_\odot \) (determined by the free-streaming scale of the CDM particle) to \( 10^{15} \text{M}_\odot \). In addition to the matter bound in collapsed structures, a smooth component is expected which

© 2012 The Authors, MNRAS 426, 1121–1136
Monthly Notices of the Royal Astronomical Society © 2012 RAS
As a first attempt to answer this question, we employ an analytic approach based on a Press–Schechter (PS) model (Press & Schechter 1974). Press & Schechter derived an analytical expression for the cumulative mass function, \( F(M_{\text{halo}}) \), which gives the fraction of mass locked in haloes above a given mass. Integrating their formula over the whole mass spectrum yields one. Thus, the PS model predicts that all matter is bound in gravitationally collapsed haloes, with masses ranging from zero to infinity. With the increasing dynamical range of \( N \)-body simulations, quantitative differences with this model have become apparent (Jenkins et al. 2001; Reed et al. 2003, 2007; Warren et al. 2006; Tinker et al. 2008; Faltenbacher, Finoguenov & Drory 2010; Anderhalden & Diemand 2011; More et al. 2011). However, the difference between analytical and \( N \)-body predictions depends on the definition of a halo in the simulations. Recently, Prada et al. (2006) and Cuesta et al. (2008) showed that using ‘dynamical masses’ for \( N \)-body haloes yields good agreement at least for low redshifts.

The top panel of Fig. 1 shows our computation of \( F(M_{\text{halo}}) \) based on the 5-year WMAP data (Komatsu et al. 2009). The derivative of \( F \) determines the number density of haloes, \( N_{\text{halo}} \), as a function of mass (second panel). According to spherical collapse theory, the radius of a halo with mass \( M_{\text{halo}} \) is given by

\[
    r_{\text{halo}} = \left( \frac{3M_{\text{halo}}}{4\pi\Delta_1 \rho_{\text{crit}}} \right)^{1/3},
\]

where \( \rho_{\text{crit}} \) is the critical density of the universe and \( \Delta_1 = 95 \) (for cosmological parameters chosen above).

The average number of haloes, \( X_{\text{halo}} \), with mass \( M_{\text{halo}} \) encountered by a single infinitesimal ray per unit length is the product of \( N_{\text{halo}} \) and the surface area of the halo. Note that this quantity is computed at a constant time ignoring the presumably small contribution due to the global evolution of the halo population during the light travel time:

\[
    X_{\text{halo}} = N_{\text{halo}} \pi r_{\text{halo}}^2.
\]

\( X_{\text{halo}} \) is shown in the third panel from the top. The average number of haloes with masses above \( 10^{12} h^{-1} M_\odot \) encountered by a light ray is \( \sim 0.001 \). Thus, on average a single light ray of light hits one \( 10^{12} h^{-1} M_\odot \) halo every \( 1000 h^{-1} \) Mpc.

The bottom panel displays the probability that a ray intersects with at least one halo of a given mass while travelling a distance of \( 1 h^{-1} \) Mpc, which is equivalent to \( 1 - P_{\text{free}} \), where \( P_{\text{free}} \) is the probability of passing \( 1 h^{-1} \) Mpc freely, i.e. without encountering a single halo of that mass. To compute \( P_{\text{free}} \) we subdivide the volume into cubes with \( 1/N_{1D} = N_{\text{halo}}^2 \) on a side and assume that haloes within a given mass bin are distributed quasi-homogeneously, i.e. only one single halo is placed within each cube. The position of the halo within the cube is chosen randomly. With these assumptions, \( P_{\text{free}} \) can be approximated by

\[
    P_{\text{free}} = (1 - \pi r_{\text{halo}}^2 N_{1D})^{N_{1D}}.
\]

We conclude that a light ray travelling \( 1 h^{-1} \) Mpc through the present, non-linear, cosmic density field encounters with almost 100 per cent certainty several haloes with masses below \( 10^{13} h^{-1} M_\odot \). Since the PS model does not include a smooth component, the predictions for this mass range must be corrected accordingly. On the other hand, the probability to hit at least one \( 10^{12} h^{-1} M_\odot \) halo is of the order of 0.1 per cent. These results suggest that we must discuss the effects of small- and large-scale structures on the light propagation separately.

2.1 Effects of small-scale structure

Lensing can discriminate between a diffuse and smooth component (a gas of microscopic particles) and one of the macroscopic massive objects (gravitationally bound), and has been used by Metcalf & Silk (1999, 2007) to probe the nature of dark matter on galactic scales. The two components can be characterized by a mass scale, defined by the fact that the projected density be smooth on a scale of the order of the angular size of the source. This gives (Metcalf & Silk 1999)

\[
    M_s \sim 2 \times 10^{-23} M_\odot h^2 \left( \frac{\lambda_s}{1 \text{ au}} \right)^3.
\]

where \( \lambda_s \) is the physical size of the source.

Another important mass scale is set by the requirement that the angular size of the source, \( \beta = \lambda_s/D_A(z_s) \), is smaller than the Einstein angular radius \( \theta_E \) so that it can be considered as a true point source (Metcalf & Silk 1999):

\[
    M_{\text{point}} \sim 5 \times 10^{-7} M_\odot \left( \frac{\lambda_s}{1 \text{ au}} \right)^2 \left( \frac{10^3 \text{ Mpc}}{D_A(z_s)} \right).
\]
If $M < M_\epsilon$, the component can be considered as diffuse on the scale of the ray bundle. If $M_\epsilon < M < M_{\text{point}}$ there will be very few high magnification events with most of the lines of sight being demagnified, according to the standard lensing paradigm (see below). These two components affect the probability distribution function (PDF) for the magnification. In the extreme case where the matter is composed of macroscopic point-like objects, then most high-redshift SNIa would appear less bright than in a universe with the same density distributed smoothly, with some very rare events of magnified SNIa (Rauch 1991; Holz & Wald 1998; Metcalf & Silk 1999). It has been argued in Seljak & Holz (1999) that the magnification of high-redshift SNIa can be a powerful discriminator of the nature of dark matter. In particular, based on numerical simulations, a few hundred SNIa at $z \sim 1$ could allow a 20 per cent determination of the fraction of matter in compact objects.

The dispersion of SNIa data due to lensing has been estimated in various ways, but a complete analysis may require us to go down to scales where our knowledge of the distribution of matter is very poor. In particular, it is important to know the amount of diffuse matter compared to the amount of matter in small compact haloes. Knowledge of the spatial distribution of the haloes is required to determine the dispersion of the observations, keeping in mind that one also expects a bias since most lines of sight are demagnified.

The minimum mass of gravitationally bound structures is determined by the nature of the CDM particle itself. Its mass induces a free-streaming scale which in turn gives rise to a mass threshold below which no gravitationally bound structure can form (in contrast to the original PS approach where there is no such cut-off mass). Currently, neutralinos are the most promising CDM candidates. The lightest neutralino, with $m \sim 100$ GeV, is favoured, as it is both weakly interacting and stable (e.g. Bertone, Hooper & Silk 2005). Its free-streaming scale is $\sim 0.7$ pc, with a corresponding minimum halo mass $M_{\text{fs}} \sim 10^{-8} M_\odot$.

In principle, cosmological $N$-body simulations can be employed to determine the total amount of mass in the smooth and the halo component. Knowing the fundamental properties of the CDM particle, the initial conditions can be derived and propagated to the current epoch. However, the dynamical range required for this approach exceeds current computational resources.

One strategy to circumvent the computational limitations is to enclose a small region of very high resolution within a larger but lower resolution simulation. Using this technique, Diemand, Moore & Stadel (2005) found that at $z = 26$ the mass function is steep, $dn(M)/dM \propto M^{-2}$ down to $M_\epsilon$. At that time about 5 per cent of the mass in the high-resolution region has collapsed into gravitationally bound haloes (see fig. 3 in Diemand et al. 2005). Due to technical limitations, this simulation has not been run further than $z \approx 26$.

Another strategy to determine the total mass locked in haloes is via excursion set theory (e.g. Zentner 2007), which propagates density perturbations stochastically to generate the halo mass functions. For a $\Lambda$CDM model with 100-GeV neutralinos, 75–80 per cent of the matter is locked in haloes at $z \lesssim 1$ (Angulo & White 2010). The remaining 20–25 per cent is smoothly distributed without being associated with any collapsed structure.

### 2.2 Effects of large-scale structure

Current $N$-body simulations provide a very reliable picture of the cosmic large-scale structure. However, the small-scale structure can only be resolved down to the given mass resolution limit of these simulations (currently $\sim 10^9 M_\odot$). Analytical approaches, like those based on PS models, are not affected by mass resolution issues but are much more sketchy by nature. In the following we will investigate both approaches.

#### 2.2.1 Analytical approach

Based on the PS model discussed above, one can determine the PDF of densities averaged along infinitesimally thin light rays. For that purpose, we model the mass distribution of haloes by four bins with average masses from $10^4$ to $3 \times 10^{14} h^{-1} M_\odot$. The mass bins are chosen in such a way that the 1D number density, $N_{\text{1D}}$, increases by a factor of 10 for each subsequent (decreasing) mass bin. The remaining mass contained within haloes below the smallest mass bin is assumed to be homogeneously distributed. According to the PS model, 15 per cent of the total gravitational matter is found in haloes below $10^6 h^{-1} M_\odot$.

The probability of a ray encountering a given total number of haloes is computed as the product of the various binomial coefficients and probabilities for hitting the partial number of haloes per mass bin. As before the evolution of the halo population during the light travel time is ignored.

This approach allows us to compute the PDFs as a function of the path length as shown in Fig. 2. The distribution for a path length of $100 h^{-1}$ Mpc peaks at 0.4 times the mean density $\rho_{\text{mean}}$. For longer path lengths the peak shifts towards $\rho_{\text{mean}}$. However, even for a path length of $1000 h^{-1}$ Mpc (corresponding to a redshift of $z \approx 0.25$) the distribution peaks significantly below $\rho_{\text{mean}}$. It is worth noting that by construction the integral of $\rho$ from zero to infinity equals unity. A peak below $\rho_{\text{mean}}$ requires a high-density tail for counterbalance.

The model presented here does not include halo clustering (inherent to such kind of approaches) and assumes the mass contained in objects below $10^6 h^{-1} M_\odot$ to be distributed smoothly (to reduce computational cost). These shortcomings prevent us from making a quantitative interpretation at this point, but we can clearly see that the majority of light rays encounter averaged densities below the mean and that the shape of the PDFs is a function of path length.

![Figure 2](https://academic.oup.com/mnras/article-abstract/426/2/1121/973745)

**Figure 2.** Probability distribution of the averaged densities encountered by a single light ray of infinitesimal width for different path lengths based on the PS model.

© 2012 The Authors, MNRAS 426, 1121–1136

Monthly Notices of the Royal Astronomical Society © 2012 RAS
2.2.2 Numerical approach

Numerical simulations provide detailed insight into the non-linear matter evolution inaccessible to analytical descriptions. However, they are inevitably limited in their mass resolution. The resolution is proportional to the simulation volume and inversely proportional to the number of phase space elements (particles) used. For cosmological applications generally large volumes are desirable. The number of particles is limited by the available computational power. Particles of current state-of-the-art cosmological simulations have masses of a few times $10^8 \, M_\odot$. For comparison, the total mass, assuming homogeneous density distribution, contained within the volume of a light beam of 1-au diameter and 1000-Mpc length (corresponding to $z \approx 0.25$) is $\sim 10^{-9} \, M_\odot$.

In this section, we compute the PDFs of the mean densities at a constant time ($z = 0$) within long and narrow ‘beams’ based on a set of publicly available N-body simulations, namely the Bolshoi (Klypin, Trujillo-Gomez & Primack 2011), the Millennium (Springel et al. 2005) and the MultiDark R1 (Prada et al. 2012) simulations. (Descriptions of the data bases are given in Lemson & the Virgo Consortium 2006 and Riebe et al. 2011.) These simulations compute the CDM distribution within cubes of $250^3$, $500^3$ and $1000^3 \, h^{-1} \, \text{Mpc}^3$ with mass resolutions of $1.3 \times 10^3$, $8.6 \times 10^3$ and $8.7 \times 10^3 \, h^{-1} \, M_\odot$, respectively. These mass resolutions only allow the determination of the densities within beams wider than a few tens of kpc which is many orders of magnitude larger than the expected diameter of a light beam from a supernova. Nevertheless, the results derived here can give basic insight into the mean densities within the volume of the light beams from distant SNIa.

The left-hand panel of Fig. 3 shows the PDFs of the mean densities within beams of $500 \, h^{-1} \, \text{kpc}$ diameter as a function of their length. The dashed lines based on the MultiDark simulation are shown as a consistency check. They are expected to coincide with the Bolshoi results but are more affected by Poisson noise due to the lower mass resolution. The shape of the distributions is similar to those based on the PS model (Fig. 2). Independent of length all distributions peak below the mean density. The cumulative probability for a mean density below the cosmic mean for the 100, 250, 500 and $1000 \, h^{-1} \, \text{Mpc}$ beams is 75, 71, 68 and 65 per cent, respectively.

The middle panel of Fig. 3 shows the PDFs for different path lengths. There may be an indication that with decreasing diameters the location of the peak converges towards values close to $0.5 \rho_{\text{mean}}$, but we are unable to probe beam sizes below $\sim 50 \, h^{-1} \, \text{kpc}$. The shape of the distribution, and in particular that the location of the peak is below unity, is preserved independent of diameter and is expected to hold for much smaller diameters. The right-hand panel of Fig. 3 displays the PDFs based on cubic volumes, i.e. the densities are measured in cubes rather than beams (left-hand panel) while the volume remains the same. The overall shape of the PDFs is very similar to those shown on the left but the location of the peak is shifted towards smaller densities. Obviously, the geometry of the ‘test volumes’ has an impact on the PDFs. The PDFs cannot be accurately determined without incorporating their spatial information of the large-scale density distribution.

The picture arising from the N-body simulations is consistent with the PS model. The probability that a light beam from a supernova encounters an average density less than the cosmic mean is larger than 50 per cent. The exact value depends on the light path length. With shorter distances, the cumulative probability of sampling densities below the cosmic mean increases. This effect may be sufficient to induce biases in the luminosity distance relation.

3 LIGHT PROPAGATION

From a theoretical point of view, the effects of matter inhomogeneities can be described by the geodesic deviation equation, which describes the evolution of a bundle of geodesics $x^\nu (\nu, s)$, where $\nu$ is the affine parameter and $s$ labels the geodesics. The past light cones of the central observer are given by $w = \text{const}$, where $w$ is the phase. Then $k_\nu = \partial_\nu w$, so these curves are irrational null geodesics:

$$k^\mu k_{\mu} = 0, \quad k^\mu \nabla_\mu k_\nu = 0, \quad \nabla_\mu k_\nu = 0. \quad (7)$$

![Figure 3](http://example.com/fig3.png)

Figure 3. Left-hand panel: probability distribution of averaged densities within long narrow (0.5 $h^{-1}$ Mpc diameter) beams for different path lengths using different simulations (indicated). For short path lengths (e.g. $\sim 100 \, h^{-1}$ Mpc) most beams encounter low densities which are counterbalanced by comparatively few beams which encounter much higher densities. With increasing path length the peak of the PDF approaches the cosmic mean, but even for a beam length of 1 $h^{-1}$ Gpc the peak occurs significantly below the mean density. Independent of length the average density encountered by a sufficiently large number of beams is equal to the cosmic mean density, i.e. the density-weighted integrals for all curves shown above yield the cosmic mean density. Middle panel: PDF for beams of the same length (250$^{-1}$ Mpc) but different diameters (indicated). As the beam becomes narrower the PDF broadens and the location of the peak tends to shift to slightly smaller densities. Right-hand panel: the same as the left-hand panel except here the density is measured within cubes which have the same volume as the long and narrow beams. Note the striking difference between the PDFs: the peak position and widths change when comparing tubes to cubes, and the power-law tail which is prominent for beams is no longer present.
The connecting vector $\eta^\mu = dx^\mu/ds$ relates neighbouring geodesics with the tangent vector $k^\mu = dx^\mu/dt$ to an arbitrary reference geodesic of the bundle, $\tilde{x}^\mu(v) = x^\mu(v, 0)$, giving the distance between neighbouring geodesics and hence the physical size and shape of the bundle as one follows it down into the past. The connecting vector can always be chosen such that $k^\mu \eta_\mu = 0$ and it evolves according to the geodesic deviation equation:

$$k^\mu k^a \nabla_a \eta^\mu = R^{\mu}_{\alpha\beta\gamma} k^\alpha \eta^\beta. \tag{8}$$

This equation describes the change of shape of the bundle.

For fundamental observers with 4-velocity $u^\mu$ ($u^\mu u_\mu = -1$), the redshift is defined by

$$1 + z(v) = \frac{\langle k^\mu u^\mu \rangle_0}{\langle k^\mu u^\mu \rangle_v}, \tag{9}$$

where the past-directed photon 4-momentum is

$$k^\mu = (1 + z)(-u^\mu + e^\mu), \quad e^\mu u_\mu = 0, \quad e^\mu e_\mu = 1. \tag{10}$$

Here, $e^\mu$ is the spatial direction of observation, and the spatial direction of propagation is $n^\mu = -e^\mu$. The affine parameter increases monotonically along each ray and coincides in an infinitesimal neighbourhood of the observation point with the Euclidean distance in the rest frame of $u^\mu(0)$. Note that while it depends on the 4-velocity $u^\mu(0)$ of the observer, it does not depend on the 4-velocity $u^\mu(\tilde{x}^\mu(s))$ of the observed source.

The screen space at each point along a ray is in the observer’s rest space and orthogonal to the ray direction. It is spanned by unit vectors $n^\mu_n (a = 1, 2)$, with $g_{\mu n} n^\nu_n = \delta_{\mu b}$ and $n^\mu_n u^\mu_n = n^\mu_n k^\mu_n = 0$, that are parallel transported along the ray ($k^\mu \nabla_n n^\mu_n = 0$). We can choose the connecting vector to lie in the screen space, so that

$$\eta^\mu = n^\mu n^\nu_n + \eta^\nu_n n^\nu_n. \tag{11}$$

From equation (8)

$$\frac{d^2}{dv^2} \eta^\mu = R_{\alpha\beta\gamma\delta} \eta^\alpha \eta^\beta \eta^\gamma \eta^\delta, \tag{11}$$

where $R_{\alpha\beta\gamma\delta}$ is the Ricci tensor. We write

$$R_{\alpha\beta\gamma\delta} = \begin{pmatrix} \Phi_{00} & 0 \\ 0 & \Phi_{00} \end{pmatrix} + \left( -\text{Re} \Phi_0 \quad \text{Im} \Phi_0 \right), \tag{12}$$

with

$$\Phi_{00} = -\frac{1}{2} R_{\mu\nu\rho\sigma} k^\mu k^\nu k^\rho k^\sigma, \quad \Phi_0 = -\frac{1}{2} C_{\mu\nu\rho\sigma} m^\mu m^\nu m^\rho m^\sigma, \tag{13}$$

where $m^\mu \equiv n^\mu - v^\mu$. The Einstein equations give $R_{\mu\nu} k^\mu k^\nu = 8\pi G T_{\mu\nu} k^\mu k^\nu$, where $T_{\mu\nu}$ is the total energy-momentum tensor, $T_{\mu\nu} = (\rho + p) u^\mu u^\nu + p g_{\mu\nu} + \pi_{\mu\nu} + q_{\mu\nu} + u_{\mu} u_{\nu} + q_{\mu\nu}$. \(14\)

Here, $\pi_{\mu\nu}$ is the anisotropic stress and $q_{\mu\nu}$ is the momentum density. (For a perfect fluid $\pi_{\mu\nu} = 0 = q_{\mu\nu}$; for more general fluids, we can always choose $q_{\mu\nu} = 0$, corresponding to the frame where comoving observers see no momentum flux). Then we find

$$\Phi_{00}(v) = -4\pi G [1 + z(v)]^2 (\rho + p + 2q_{\mu\nu} + \pi_{\mu\nu} + q_{\mu\nu}) \left| \frac{d^2}{dv^2} \eta^\mu \right|^2. \tag{15}$$

Note that a cosmological constant $\Lambda$ makes no contribution to $\Phi_{00}$.

The linearity of (11) implies that

$$\eta^\mu(v) = D^\mu_{\alpha\beta} (v) \left. \frac{d\eta^\mu}{dv} \right|_{v=0} \tag{16},$$

where $D^\mu_{\alpha\beta}$ is the Jacobi map. From (11), we have the Jacobi matrix equation

$$\frac{d^2}{dv^2} D^\mu_{\alpha\beta} = \mathcal{R}^\mu_{\gamma\delta\beta} D^\gamma_{\alpha\beta} , \quad \eta^\mu(0) = 0, \quad D^\mu_{\alpha\beta}(0) = \delta^\mu_{\beta}. \tag{17}$$

This second-order linear equation can be rewritten as a first-order non-linear equation

$$\frac{d}{dv} S^\alpha_{\beta\gamma} + S^\beta_{\gamma\delta} S^\gamma_{\delta\alpha} = R^\mu_{\alpha\beta} D^\gamma_{\alpha\beta}, \tag{18}$$

by defining the deformation matrix

$$\frac{d}{dv} D^\mu_{\alpha\beta} = D^\gamma_{\alpha\beta} S^\gamma_{\delta\mu}. \tag{19}$$

The Jacobi map $D^\mu_{\alpha\beta}$ or equivalently the deformation matrix $S^\alpha_{\beta\gamma}$ are the central quantities to describe the distortion of the geodesic bundle. The deformation matrix is usually decomposed as

$$S^\alpha_{\beta\gamma} = \left( \begin{array}{ccc} \hat{\delta} & 0 \\ 0 & \hat{\beta} & \hat{\sigma} \end{array} \right), \tag{20}$$

which defines the optical scalars $\hat{\delta}$ (null expansion) and $\hat{\sigma} \equiv \hat{\delta}_1 + i \hat{\delta}_2$ (null shear). These satisfy the Sachs equations (Sachs 1961)

$$\frac{d\hat{\delta}}{dv} + \hat{\delta}^2 + |\hat{\sigma}|^2 = \Phi_{00}, \tag{21}$$

$$\frac{d\sigma}{dv} + 2\hat{\delta} \hat{\sigma} = \Psi_0, \tag{22}$$

$$\hat{\delta} \equiv \frac{1}{2} \nabla_v k^\mu, \quad |\hat{\sigma}|^2 \equiv \frac{1}{2} \nabla_v \nabla_v k^\mu - \hat{\delta}^2. \tag{23}$$

The evolution of a ray bundle can then be discussed in terms of Ricci focusing ($\Phi_{00}$) and Weyl focusing ($\Psi_0$). The first is generated by matter inside the beam (see equation 15) while the second derives from matter outside the beam, which can generate a non-vanishing Weyl tensor inside the beam. This distinction leads to the problem raised by Zel’dovich (1964) and R. P. Feynman (1964, unpublished talk), and posed in terms of the curvature tensor by Bertotti (1966): if the matter of the universe is clustered in massive galaxies, the bundle propagates almost exclusively in vacuum, or at least in underdense regions, and is thus mostly subject only to the Weyl focusing; in contrast, the cosmological effect is modelled using a homogeneous fluid which generates only Ricci focusing (the Weyl tensor vanishes in FL space–time). Dyer & Roeder (1972) (see also Dyer & Roeder 1973, 1974, 1981) effectively reproduced Zel’dovich’s idea and proposed an ansatz to model the propagation in regions with no intergalactic medium. Weinberg (1976) disputed this model, arguing that multiple Weyl deflections by individual masses average to mimic the Ricci effect of a fluid with equal average density. Weinberg’s argument, based on photon flux conservation, is effectively the basis for the standard perturbative approach – i.e. when averaged over distances and angles, the divergence from vacuum or underdense regions is compensated by the convergence due to clumping, so that the average luminosity distance is the same as the luminosity distance in the FL background (see e.g. Fang & Wu 1989).

Weinberg’s argument has been disputed in Ellis, Bassett & Dunsby (1998) and Rose (2001). Later work (e.g. Kibble & Lieu 2005; Kostov 2010) has not produced a definitive answer to the question, in particular for the case of the very narrow beams involved in SN1a observations.\(^1\)

\(^2\) Recall that the null rotation $\nabla_{[\mu} k_{\nu]}$ vanishes since $k_{\mu} = \partial_{\mu} w$.\(^3\)
In order to properly describe a thin geodesic bundle, we need to have a good description of the matter distribution on the scales of the extension of the bundle, and determine how the effect of the inhomogeneities average during the propagation of the bundle, with two main issues in mind: (1) determining the typical amplitude of the effect and (2) understanding why the description by a smooth universe seems to provide a good description and determine its validity.

3.1 Angular distance

The Jacobi matrix can be diagonalized by rotations:

\[ D''_{ab} = r(-\alpha) \begin{pmatrix} D_+ & 0 \\ 0 & D_- \end{pmatrix} r(\alpha), \]

where the shape parameters \( D_{\pm} \) are non-zero almost everywhere. Their absolute values give the semi-axes of the (elliptic) cross-section of the bundle. Once \( D_{\pm} \) are fixed, the angles \( \alpha \) are unique at all points where the bundle is non-circular.

The area distance or angular diameter distance is then defined as\(^3\)

\[ D_A(v) = \sqrt{\det[D(v)]} = \sqrt{|D_{ab}(v)D_{ab}(v)|}. \]

For a bundle converging at the observer, \( D_A \) relates the cross-sectional area \( A \) at the source to the opening solid angle at the observer. It depends on the 4-velocity of the observer, but not of the source. From (23), the null convergence is

\[ \dot{\theta} = \frac{1}{\sqrt{A}} \frac{d}{dv} \sqrt{A}, \]

and (21) becomes

\[ \dot{\theta}^2 \geq \frac{1}{A_{ab}} \frac{dA_{ab}}{dv}. \]

in any cosmological model, as long as the null energy condition holds, irrespective of the value of the cosmological constant.

In order to compare to observations, we need the relation between \( v \) and \( z \). We have

\[ \frac{dz}{dv} = d(v^\mu k_\mu)/dv = k^\mu \nabla_\mu (v^\mu k_\mu) = k^\mu k^\nu \nabla_\mu u_\nu. \]

Now

\[ \nabla_\mu u_\nu = \frac{1}{3} \Theta (g_{\mu\nu} + u_\mu u_\nu) + \sigma_{\mu\nu} + \omega_{\mu\nu} - u_\mu A_\nu, \]

where \( \Theta \) is the expansion, \( \sigma_{\mu\nu} \) is the shear, \( \omega_{\mu\nu} \) is the vorticity and \( A_\nu \) is the acceleration. In a universe containing CDM and baryons with 4-velocity \( \nu^\mu \), and \( \Lambda \) (where radiation is dynamically negligible), we have \( A_\nu = 0 \).

From equations (9) and (10), we obtain (Clarkson & Maartens 2010)

\[ \frac{dz}{dv} = (1 + z)^2 \frac{1}{3} \Theta + \sigma_{\mu\nu} e^\nu - A_\nu e^\mu. \]

For any quantity \( X \) evaluated along the ray bundle

\[ \frac{dX}{dv} = (1 + z)^2 H(z, e^\nu) \frac{dX}{dz}, \]

where \( H(z, e^\nu) \) is the observed expansion rate along the line of sight (Clarkson & Maartens 2010):

\[ H(z, e^\nu) = \frac{1}{3} \Theta + \sigma_{\mu\nu} e^\nu - A_\nu e^\mu. \]

The observed expansion rate is made up of an isotropic expansion monopole, an acceleration dipole and a shear quadrupole.

The set of equations (22), (27) and (30) is the basis for analysing the effect of inhomogeneities. There are four physical effects induced by inhomogeneities that need to be taken into account:

(i) Area distance modifications: due to the difference between Ricci focusing (when the rays move through a uniform medium) and Weyl focusing (due to the tidal effects of nearby matter).

(ii) Redshift adjustment: due to the differences between the true redshift of a source and its redshift in a smoothed-out model.

(iii) Affine parameter distortions: since inhomogeneities change the relation \( v(z) \) (this is actually where \( \Lambda \) affects observational relations).

(iv) Displacement of the light beam: since the ray path is shifted sideways by inhomogeneities and so experiences different Weyl and Ricci terms at the same \( v \) because it is at a different space–time point.

3.2 Links with weak lensing formalism

The weak lensing amplification matrix \( A \) relates the direction of observation,

\[ \theta^\nu \equiv \frac{dz}{dv} \bigg|_0, \]

to the direction of the source,

\[ \theta_0^\nu \equiv \frac{D_A(v)}{D_A(0)} = A^\nu_\mu \theta^\mu, \]

so that

\[ A^\nu_\mu(v) = \frac{D''_{ab}(v)}{D''_{ab}(0)}. \]

Here, \( D_A \) is the angular distance in an FL background and \( A^\nu_\mu = \delta_\nu^\mu \). We decompose \( A \) into a shear \( (\gamma_1, \gamma_2) \), a convergence \( \kappa \) and a rotation \( \omega \), so that

\[ A^\nu_\mu = \begin{pmatrix} 1 - \kappa - \gamma_1 & \gamma_2 - \omega \\ \gamma_2 + \omega & 1 - \kappa + \gamma_1 \end{pmatrix}. \]

The magnification is given by

\[ \mu \equiv \frac{S}{S_0} = \frac{1}{(1 - \kappa)^2 - |\gamma|^2 + \omega^2}. \]

where \( S, S_0 \) are the surface areas of the image and source \((S = S_0/\det A)\). Note that while \( R_{ab} \) is symmetric by construction (see equation 11) and \( S_{ab} \) is also symmetric for a bundle converging at the observer, it is not necessarily the case for \( D_{ab} \), which actually cannot be generically symmetric; see e.g. equation (19).

The amplification matrix (35) and the deformation matrix (19) are both related to the Jacobi matrix, and hence they are related by

\[ D_A \frac{d}{dv} A^\nu_\mu + A^\nu_\mu \frac{d}{dv} D_A = D_A A^\nu_\mu S^\mu_\nu. \]

This implies (away from caustics, where \( \det A = 0 \))

\[ (A^{-1})^\mu_\nu (A^\nu_\mu)' + \frac{D_A A^\nu_\mu}{D_A} = S^\mu_\nu. \]
with a prime denoting \( \partial / \partial v \) and where
\[
(A^{-1})^a_b = \mu \begin{pmatrix} 1 - \kappa + \gamma_1 & -\gamma_2 + \omega \\
-\gamma_2 - \omega & 1 - \kappa - \gamma_1 \end{pmatrix}.
\]
(40)

The Sachs optical scalars are then given by
\[
\tilde{\theta} = \left( \ln \frac{D_\Lambda}{\sqrt{\rho}} \right),
\]
(41)
\[
\tilde{\sigma}_1 = -\mu \left[ (1 - \kappa) \gamma'_1 + \gamma_1 k' + \gamma_2 \omega' - \omega \gamma'_2 \right],
\]
(42)
\[
\tilde{\sigma}_2 = \mu \left[ (1 - \kappa) \gamma'_2 + \gamma k' - \gamma_1 \omega' + \omega \gamma'_1 \right],
\]
(43)

with the constraint that \( \omega \gamma'_2 - \gamma_2 \omega' = (1 - \kappa) \gamma'_1 - \gamma_1 k' \) that arises from \( \nabla \mu k \). These relations are useful since the optical scalars are more general (they are defined for any space–time geometry), while the weak lensing scalars are widely used in cosmology (but they assume an FL background, see Pitrou, Uzan & Pereira 2012 for a general description). While generically \( \omega \neq 0 \), one can see that these equations imply that for an FL space–time only \( \tilde{\theta} = D_\Lambda / D_\Lambda \) is non-vanishing at the background level while the shear appears only at linear order in perturbation and one has \( (\tilde{\sigma}_1, \tilde{\sigma}_2) = (-\gamma_1, \gamma_2) \) and the rotation appears only at second order in perturbations.

### 3.3 From affine parameter to redshift dependence

The evolution equations (22) and (27) for the null shear and angular distance are in terms of the unobservable affine parameter \( v \). We need to convert these to the observed redshift, using equation (30). Using (31), we obtain
\[
\frac{d^2 z}{dz^2} = k^\mu k^\nu \nabla_\mu \nabla_\nu u_a
\]
\[
= -\frac{2}{3} (1 + z)^3 H_\theta(1) - \frac{1}{3} (1 + z)^3 k^\mu \nabla_\mu \Theta
+ (1 + z)^2 k^\mu \nabla_\mu A_\nu + (1 + z)^2 H k^\mu A_\nu
- k^\mu k^\nu \nabla_\mu \sigma_{\rho \nu}.
\]
(44)
The last term can be evaluated by expanding \( k^\mu \) with equation (10) and using \( u^\mu \nabla_\mu \sigma_{\rho \nu} = -\sigma_{\rho \nu} A^\mu \) and \( u^\mu \nabla_\mu \sigma_{\mu \nu} = -\sigma_{\mu \nu} e^\rho \nabla_\rho u^\nu \). It follows that for any quantity \( X \),
\[
\frac{d^2 X}{dz^2} = (1 + z)^3 H_\theta \left( \frac{d^2 X}{dz^2} + (1 + z) \frac{d X}{dz} \right),
\]
(45)
where
\[
Q = \frac{2}{3} \Theta H_\theta - \frac{1}{3} \Theta + A_\mu A^\mu
+ e^\nu \left( -\frac{1}{3} \nabla_\mu \Theta + H_\theta A_\mu - A_\mu - u^\rho \nabla_\rho A_\nu + 2 \sigma_{\rho \nu} A^\rho \right)
+ e^\nu e^\rho \left( \frac{2}{3} \Theta \sigma_{\mu \nu} - \sigma_{\mu \rho} - 2 \sigma_{\mu \rho} \sigma_{\nu \sigma} - 2 \omega_{\mu \rho} \omega_{\sigma \nu} \right)
+ \nabla_\mu A_\nu \right) - e^\nu e^\rho \nabla_\mu \sigma_{\rho \nu}.
\]
(46)

This form of \( Q \) is completely general, for any space–time geometry and energy-momentum tensor, and independent of the field equations. It is convenient to write \( H^2_\theta \) and \( Q \) in terms of covariant multipoles, using a covariant generalization of a spherical harmonic expansion (Clarkson & Maartens 2010). We expand in terms of the trace-free products \( e^\mu e^\rho, e^\nu e^\rho e^\mu \) and \( e^\mu e^\rho e^\sigma e^\delta \), and use the spatial covariant derivative \( \nabla_\mu \). Then, we obtain
\[
H^2_\theta = \frac{1}{9} \Theta^2 + \frac{1}{3} A_\mu A^\mu + \frac{2}{15} \sigma_{\mu \nu} \sigma^{\mu \nu}
- e^\mu \left( \frac{2}{3} \Theta A_\mu + \frac{4}{5} A \sigma_{\mu \nu} \right)
+ e^\mu e^\rho A_\rho A_\mu + \frac{4}{3} \Theta \sigma_{\mu \nu} + \frac{4}{3} \sigma_{\mu \nu} A_\rho A^\rho
- 2 e^\mu e^\nu e^\rho A_\mu \sigma_{\nu \rho} + e^\mu e^\nu e^\rho e^\sigma A_\mu \sigma_{\nu \rho} \sigma_{\sigma \rho}
\]
(47)
and
\[
Q = \frac{4\pi G}{3} (\rho + 3 p) - \frac{1}{3} \Lambda + \frac{1}{3} \Theta^2 + \sigma_{\mu \nu} \sigma^{\mu \nu}
- \frac{1}{3} \omega_{\mu \rho} \omega^{\mu \rho} + A_\mu A^\mu - \frac{2}{3} \nabla_\mu \Theta
+ e^\nu \left( \frac{1}{3} \nabla_\mu \Theta + \frac{2}{5} \nabla_\mu \sigma_{\mu \nu} \right)
+ A_\mu - \frac{4}{3} \Theta A_\mu - \frac{17}{5} \nabla_\mu A_\nu - A_\nu \omega_{\mu \rho}
+ e^\nu e^\rho A_\nu - 4 \pi G \sigma_{\mu \nu} + 2 \sigma_{\mu \nu} + 3 \sigma_{\nu \mu} \sigma_{\nu \mu}
+ \omega_{\mu \rho} \omega_{\nu \sigma} + 2 \omega_{\mu \rho} \omega_{\sigma \nu} - 2 \nabla_\mu A_\nu
+ e^\nu e^\rho e^\nu \left( \nabla_\mu \sigma_{\nu \rho} - A_\mu \sigma_{\nu \rho} \right),
\]
(48)
where we also used the covariant evolution and constraint equations of GR (see Tsagas, Challinor & Maartens 2008). Here, \( H_\theta = C_{\mu \nu \rho \sigma} e^\mu e^\nu e^\rho e^\sigma \) is the electric part of the Weyl tensor (generalizing the Newtonian tidal tensor). These expressions show clearly the covariant monopole and higher multipoles; for example, the octupole of \( Q \) is \( \nabla_{\mu \nu} \sigma_{\nu \rho} - A_\mu \sigma_{\nu \rho} \). Note that the monopole of \( H^2_\theta \) has contributions from the shear even though the monopole of \( H_\theta \) does not.

Finally, we can rewrite the evolution equation (27) for the angular distance in terms of redshift:
\[
(1 + z)^2 H^2_\theta \frac{d^2 D_\Lambda}{dz^2} + (1 + z) Q \frac{d D_\Lambda}{dz} = -4 \pi G (\rho + 3 p + \pi_{\mu \nu} e^\mu e^\nu e^\nu) + \frac{\left| \tilde{\sigma} \right|^2}{(1 + z)^2} D^\Lambda.
\]
(49)
This is a completely general and non-linear equation, valid in any space–time, with any matter content, where \( H^2_\theta \) and \( Q \) are given by equations (47) and (48). The null shear terms are given by the remaining Sachs equation (22); in terms of redshift, this is
\[
H_\theta \left( \frac{d \tilde{\sigma}}{dz} + \frac{1}{D^\Lambda} \frac{d D^\Lambda}{dz} \tilde{\sigma} \right) = -(E_{\mu \rho} - \epsilon_{\mu \rho \sigma} e^\sigma H^\rho_\theta) N^\nu_{\mu \rho},
\]
(50)
\[
N^\nu_{\mu \rho} \equiv \left( n_1^\nu n_1^\rho - n_1^\rho n_1^\nu, n_2^\nu n_2^\rho + n_2^\rho n_2^\nu \right),
\]
(51)
where \( H^\rho_\theta = \frac{1}{2} \epsilon_{\mu \rho \sigma} C^{\mu \rho \sigma} e^\nu e^\rho e^\sigma \) is the magnetic part of the Weyl tensor – which has no Newtonian analogue – and \( \epsilon_{\mu \nu} \) is the spatial alternating tensor.

Equations (49) and (50) form a closed system that determines \( D_\Lambda \) and \( \tilde{\sigma} \) in terms of \( z \) and \( e^\nu \). In particular, we see what is required to determine \( D_\Lambda \) in a lumpy universe: the total energy-momentum
tensor (i.e. $\rho, p, q_{\mu\nu}, \pi_{\mu\nu}$), the kinematics of the fundamental 4-velocity (i.e. $\Theta, \sigma_{\mu\nu}, \omega_{\mu\nu}, A_\nu$), the magnetic and electric part of the Weyl tensor, $H_{\mu\nu}$, and $E_{\mu\nu}$.

In a universe with dust matter (CDM and baryons, sharing the same 4-velocity), with dark energy in the form of $\Lambda$ and where we can neglect radiation (i.e. at late times), we have

$$ p = A_\mu = q_{\mu\nu} = \pi_{\mu\nu} = 0. $$

From now on we will make this assumption, together with $\omega_{\mu\nu} = 0$. Then, $H_{\mu\nu}$ and $Q$ simplify to

$$ H_{\mu\nu} = \frac{1}{9} \dot{\Theta}^2 + \frac{2}{15} \sigma_{\mu\nu} \sigma_{\mu\nu} + e^{\mu\nu} \left[ \frac{2}{3} \dot{\Theta} \sigma_{\mu\nu} + \frac{4}{7} \sigma_{\mu\alpha} \sigma_{\nu\alpha} \right] + e^{\mu\nu} e^{\rho\sigma} \sigma_{\mu\rho} \sigma_{\nu\sigma}, $$

$$\dot{Q} = \frac{4\pi G}{3} \rho - \frac{1}{3} \Lambda + \frac{2}{3} \dot{\Theta}^2 + \sigma_{\mu\nu} \sigma^{\mu\nu} + e^{\mu\nu} \left[ \frac{1}{3} \tilde{\nabla}_\mu \tilde{\Theta} + \frac{2}{3} \tilde{\nabla}^\rho \sigma_{\mu\rho} \right] + e^{\mu\nu} e^{\rho\sigma} \tilde{\nabla}_\mu \sigma_{\nu\rho}, $$

and the angular distance equation (49) becomes

$${(1 + z)^2} H_0^2 \frac{d^2 D_\Lambda}{dz^2} + (1 + z) Q \frac{dD_\Lambda}{dz} = - \left[ 4\pi G \rho + \frac{1}{(1 + z)^3} \right] D_\Lambda.$$  

The form of equation (50) is unchanged.

## 4 MODELS BASED ON DIFFERENT APPROXIMATIONS

We briefly review the standard FL approach and the DR approximation, and then we propose and investigate modifications of the DR model. (For other related reviews, see also Sasaki et al. 1999; Tomita et al. 1999; Räsänen 2009, 2010.) The set of equations (21), (22) and (30) – equivalently (49) and (50) – is completely general and does not depend on the choice of a particular space–time geometry. We show here how they lead to different expressions for the angular distance as a function of redshift, depending on the assumptions on the distribution of the matter.

### 4.1 Smooth FL model

If we assume the matter is smoothly distributed, then the universe can be described by an FL geometry,

$$ ds^2 = a^2(t) \left[ -dt^2 + d\chi^2 + f_K(\chi) d\Omega^2 \right], $$

$$ H(z) = H_0 \sqrt{\Omega_{m0}(1 + z)^3 + \Omega_{\Lambda0} + \Omega_{k0}(1 + z)^2}, $$

where $f_K(\chi) = \sin(\sqrt{K} \chi)/\sqrt{K}$ is the comoving angular distance. The Weyl tensor vanishes, so that $\Psi_0 = 0$, and $\sigma = 0$ from (23), consistent with (22). Moreover, $\tilde{\gamma}_a^b = \Phi_0 \delta_a^b$ and $H_1 = H$ from (31). Then, using (15), it follows that (49) reduces to

$$ \frac{d^2 \tilde{\gamma}_a^b}{dz^2} + \left( \frac{d \ln H}{dz} + \frac{2}{1 + z} \frac{dH}{dz} \right) \tilde{\gamma}_a^b + \frac{1}{H} \frac{dH}{dz} \tilde{\gamma}_a^b = - \frac{2}{2} \Omega_{m0} H_0^2 (1 + z) \tilde{D}_\Lambda.$$  

It is important to realize that $H(z)$ in this equation appears from the change of variable from $v$ to $z$.

This equation also follows directly from the Jacobi matrix equation (17), which is easily solved after a conformal transformation,

$$ \frac{d^2}{dv^2} D_\eta^b = -K D_\eta^b, $$

where $v$ is the affine parameter in the conformal space–time of the static metric $d\bar{s}^2 = -d\eta^2 + f_\bar{K}(\bar{\chi}) d\Omega^2$. The solution of equation (59) is $D_\eta^b = f_\bar{K}(\bar{\chi}) \delta_\eta^b$. One can choose $v$ either as $t$ or $\chi$ and the angular distance is then given by the standard formula (Schneider, Ehlers & Falco 1992; Perlick 2004; Peter & Uzan 2009)

$$ \tilde{D}_\Lambda(z) = \frac{a_0}{(1 + z)} f_\chi(\chi(z)]. $$

Along the past light cone $dv = -d\eta$ and $dz/d\eta = -a_0 H(z)$, so that $a_0 H_0 \tilde{D}_\Lambda(z) = \int_0^z \frac{dv}{H(z)}.$

Since $dz = dv/H$, equation (59) recovers equation (58), after using $H = -(1 + z) H_0 H/\frac{dz}{dz}$ and the Einstein equation for $H$. We can solve either equation (59) or equation (58) but the first is more direct. Moreover, the derivation of equation (58) required the relation $\nu(z)$, or equivalently $\tilde{v}(z)$.

For an FL universe $A_0^b = \delta_0^b$ (i.e. $\kappa = \gamma_1 = \gamma_2 = 0$) and then $S_\eta^b = f_\bar{K}(\bar{\chi}) \delta_\eta^b$, so that only $\tilde{\theta} = f_\bar{K}^{-1} d f_\bar{K} / d\bar{\chi}$ is non-vanishing, which is a consequence of the spatial homogeneity and isotropy. After integration of equation (22), $\sqrt{A} = f_\chi$ in the conformal space–time, so that again we recover the same expression for the angular distance.

### 4.2 Perturbed FL model

The simplest way to account for inhomogeneous matter is via perturbation theory. At first order for scalar perturbations,

$$ ds^2 = a^2(t) \left[-(1 + 2\Phi) d\eta^2 + (1 - 2\Psi) d\chi^2 \right], $$

in Newtonian gauge, where $\Phi$ and $\Psi$ are the Bardeen potentials. The angular distance is

$$ D_\Lambda = \tilde{D}_\Lambda(1 + \delta_\Lambda), $$

and the distance duality relation implies that the luminosity distance is

$$ D_L = (1 + z)^2 D_\Lambda(1 + \delta_\Lambda). $$

Thus,

$$ \delta_L(\zeta, n) = \delta_\Lambda(\zeta, n) + 2 \frac{\delta_\Lambda(\zeta, n)}{(1 + z)^2}, $$

where $n$ is the direction of observation. The second term encodes fluctuations of photon energy due to the local gravitational potentials as well as Doppler effects, and is similar to the Sachs–Wolfe effect on the CMB. It was investigated in Bonvin, Durrer & Gasparini (2006) and also estimated in Uzan, Bernardeau & Mellier (2008) in another context, but neglected in Cooray, Huterer & Holz (2006b), Sarkar et al. (2008), Cooray, Holz & Caldwell (2010), Dodelson & Vallinotto (2006) and Vallinotto, Dodelson & Zhang (2011), which assumed $\delta_L = \delta_\Lambda$.

As long as the bundle remains in the weak lensing regime, $\mu \simeq 1 + 2\kappa$,

$$ \delta_L(\zeta, n) = -\kappa(\zeta, n), $$

so that $\delta_L(\zeta, n) \simeq -\kappa(\zeta, n)$. This assumes that the inhomogeneities can be described by density fluctuations of a homogeneous field. Then $\delta_L$ is a stochastic field of zero mean, so that $\langle \mu \rangle = 1$ (in terms of ensemble average) and thus $\langle D_L(\zeta, n) \rangle = D_\Lambda(\zeta)$. In such a description, the FL angular distance is the mean distance that a
collection of observers will determine. There may indeed be a bias from this prediction arising from our actual position in the universe (see e.g. Marra, Pääkkönen & Valkenburg 2012; Valkenburg 2012). This is a cosmic variance problem.

4.2.1 Derivation from the Jacobi map equation

The Jacobi equation (59) reduces to

$$\frac{d^2}{dz^2}D_{ab}(v) + KT_{ab}(v) = f_k(v)R_{ab}(v),$$

(66)

where $D_{ab} = D_{ab}^{(0)} + D_{ab}^{(1)}$, and $D_{ab}^{(1)} = f_k(v)\delta_{ab}$ is the background Jacobi matrix derived above. This equation has the integral solution

$$D_{ab}^{(1)}(v) = \int_0^v f_k(v')f_k(v - v')R_{ab}^{(1)}(v') dv',$$

(67)

so that the amplification matrix is given by

$$A_{ab}^{(1)}(v) = \int_0^v f_k(v')f_k(v - v')R_{ab}^{(1)}(v') dv',$$

(68)

where $2R_{ab}^{(1)} = \delta_{ab}\delta_{\alpha\beta}k^\mu k^n n^\alpha n^\beta$. Since we are interested in modes smaller than the Hubble scale, we assume that the spatial curvature does not influence the perturbations and neglect it in the computation of $R_{ab}^{(1)}$ while we keep it in the geometrical factors. We find $R_{ab}^{(1)} = -\delta_{ab}\delta(\Phi + \Psi) = -2\delta_{ab}\Phi$, where $\Psi = \Phi$ since we can neglect anisotropic stress at late times and on sub-Hubble scales. Then the amplification matrix is $A_{ab} = \delta_{ab} - \delta_{ab}\psi(n, \chi)$ with

$$\psi = 2\int_0^v f_k(v')f_k(v - v')\Phi[f_k(v')n, v] dv'.$$

We conclude that $\delta_{ab} = -\kappa(n, \chi)$ with

$$\delta_{ab} = -\kappa(n, \chi) = -\frac{3}{2}H_0^2\Omega_m \times \int f_k(v')f_k(v - v')\delta[f_k(v')n, v'] dv'$$

(70)

where the Poisson equation has been used to replace $\nabla^2\Phi$ with $\delta$. This gives the fluctuation of the angular distance in the direction $n$, taking into account propagation in a perturbed space–time.

4.2.2 Derivation from the Sachs equations

The same result can, in principle, be derived from the Sachs equations (21), (22) and (30). The background and first-order equations are

$$D^{(0)}_{\Lambda} = \Phi^{(0)}D_{\Lambda},$$

(71)

$$z^{(0)} = \frac{1}{3} [1 + z^{(0)}]^2 \Theta^{(0)},$$

(72)

$$D^{(1)}_{\Lambda} = \Phi^{(0)}D_{\Lambda}^{(1)} + D_{\Lambda}^{(0)}\Phi^{(1)},$$

(73)

$$c^{(1)} = \frac{2}{3} \left[1 + z^{(0)} \right] \Theta^{(0)}\zeta^{(1)},$$

(74)

$$c^{(1)} = \frac{1}{3} \Theta^{(0)} + \Theta^{(0)} \zeta^{(1)},$$

(75)

$$\delta^{(1)} = 2\Theta^{(0)}\delta^{(1)} = \Psi^{(0)},$$

(76)

$$\delta_l = \delta_{\Lambda} + 2z^{(0)} \frac{z^{(1)}}{1 + z^{(0)}},$$

(77)

where primes are $\partial/\partial v$.

From equation (30) with $H_0 = H$, equation (72) reduces to the definition $H = \dot{a}/a$, and then (71) can be integrated to give (60). Then (73) is similar to (66), with the source term given by

$$\Phi^{(1)} = -4\pi G \rho^{(0)} \left[1 + z^{(0)}\right]^2 \left(\frac{\rho^{(1)}}{\rho^{(0)}} + 2\frac{z^{(1)}}{1 + z^{(0)}}\right).$$

(78)

Since $\hat{\delta}^{(0)}$ is known, equation (76) can be integrated once the source (which depends on gradients of the gravitational potentials) is known.

These two derivations have to give the same answer, which they actually do if the effect of perturbations is not neglected in any of the equations, and in particular in the equation for $v(z)$.

4.3 Dyer–Roeder approximation

It is important to stress that the perturbative description still assumes that the distribution of matter is continuous (i.e. it assumes that the fluid approximation holds on the scale of interest), so that, as long as one is in the weak lensing regime, the whole effect arises from Ricci focusing with the density of matter equal to the average density (because the effect of the shear appears only at second order). Zel’dovich (1964) pointed out that light is actually more likely to propagate in underdense regions so that an overall demagnification was expected. This idea was followed up by Dashevskii & Slysh (1966), Bertotti (1966), Gunn (1967), Kantowski (1969) andRefsdal (1970). The approach came to be named after the later work of Dyer & Roeder (1972, 1973, 1974).

The main assumptions of the DR approximation are (1) the Sachs equation (22) holds; (2) the relation $v(z)$ is the FL one, equation (30) with $H_0 = H$; (3) the null shear $\hat{\delta}$ vanishes, as in an FL universe; (4) $\Phi^{(0)}$ is replaced by $\sigma(z)\Phi^{(0)}$, where $\sigma(z)$ represents the fraction of the (mean) matter intercepted by the geodesic bundle. In summary, the DR model assumes that the bundle is propagating in an FL universe but that the Ricci focusing is reduced (to reproduce the fact that the beam propagates mostly in vacuum or underdense regions) and the Weyl focusing remains zero. This implies that the DR equation is

$$\frac{d^2\tilde{D}_\Lambda}{dz^2} + \left(\frac{d\ln H}{dz} + \frac{2}{1 + z}\right) \frac{d\tilde{D}_\Lambda}{dz} = -\frac{3}{2} \Omega_m H_0^2 \frac{H^2}{H^2} (1 + z)\sigma(z)\tilde{D}_\Lambda.$$  

(79)

This attempts to model the global effect of inhomogeneity in a ‘mean way’, while still assuming that the universe is isotropic and homogeneous.

The consistency of the DR approximations has been questioned by Ehlers & Schneider (1986) and others (e.g. Sasaki 1993; Tomita et al. 1999; Räsänen 2009), independently of Weinberg’s photon flux conservation argument (Weinberg 1976).

The smoothness parameter $\alpha$ was initially assumed to be constant (Dyer & Roeder 1972, 1973, 1974). Later it was refined to take into account its redshift dependence due to the growth of structure (Linder 1988; Tomita 1998; Mörtsell 2002) and then related to the statistical properties of the large-scale structure (Räsänen 2009; Bolejko 2011). A novel use of the DR approach was proposed in Wang (2000a,b) to correct SNIa observations for the matter distribution along the line of sight, which has important implications for parameter estimation (Wang, Chuang & Mukherjee 2012).

On large scales (Mpc) we expect a distribution of 3D compensated voids giving partially 1D compensated matter distributions along the line of sight (Bolejko 2011). The lensing effects will not be the...
same as a smooth distribution of matter. On smaller scales (2 kpc to 1 au) we expect to mainly move through voids, contaminated by remnant baryonic gas and non-baryonic dark matter, plus the nearly smoothly distributed photons and neutrinos. This suggests that the effect can be significant if matter is clustered on small scales with most of the light beams used in SNIa observations preferentially moving through voids.

4.4 Modifying the DR approximation

The previous discussions make it explicit that the DR approximation is not a satisfactory model for the effects of clumping on ray bundles. Consider the $D_A(z)$ equation, in the absence of pressure and null shear:

$$H_i(z) \frac{d}{dz} \left[ (1+z)^2 H_i(z) \frac{d}{dz} D_A(z) \right] = -4\pi G \rho(z) D_A(z),$$

where $H_i(z) = \Theta/3 + \sigma_{\mu\nu} e^\mu e^\nu$. The DR approximation assumes that we can model the encounters of photons with inhomogeneous matter, and not a homogeneous background space–time, by using $\rho(z)$ as the ‘true’ density along a ray – while leaving the rest of the equation as in the smooth background. Another way to think of this is that the relationship between the affine parameter and redshift is held smooth, and inhomogeneities are assumed not to affect the $v(z)$ relation significantly. However, photons only experience the local curvature, shear and expansion and the average FL behaviour must somehow emerge from integration along the line of sight.

As discussed above, even in the perturbed FL case the relation $v(z)$ fluctuates, and so the DR relation can not be relied upon as a useful approximation even in that situation. In particular, $\alpha$ must depend on the line of sight since each bundle experiences a different matter profile. In a more general sense, the DR approximation does not account for changes in the local expansion rate due to clumping (Rässänen 2009), and does not capture the essence of weak lensing unless $\alpha(z)$ is tuned to a specific form, with no apparent physical motivation (Bolejko 2011).

We do not aim to provide a detailed analysis of the DR model here. Rather we offer some possible alternatives in order to estimate how significant the effect of clumping could be, as well as to show how difficult it is to model in a simple but reliable way.

4.4.1 Modified DR

In a universe with irrotational dust and arbitrary inhomogeneity, the generalized Friedmann equation is (Tsagas et al. 2008)

$$\frac{1}{3} \Theta^2 = 8\pi G \rho + \Lambda + \frac{1}{2} \sigma_{\mu\nu} \sigma^{\mu\nu} - \frac{1}{2} R,$$

where $R$ is the Ricci curvature scalar of the 3-surfaces orthogonal to the matter 4-velocity. By holding $\Theta$ fixed to the FL background value $3H$, the DR approximation effectively assumes that the variations of $\rho$ on any null geodesic are compensated by corresponding fluctuations in the shear and curvature, which seems unphysical.

We expect a photon in the real universe to react to the non-local part of the gravitational field created by dark matter haloes through local curvature fluctuations in addition to the dynamics of the matter in the intervening space. A reasonable alternative to DR, then, is to first write out the $D_A(z)$ equation in a general FL model, using the Friedmann equation to evaluate $dH/dz$. Substituting for $H(z)$ using

(57) everywhere, we see that $\Omega_m$ appears in several places. Then, replacing $\rho_m \rightarrow \alpha \rho_m$ gives a plausible alternative to the usual DR approximation:

$$\frac{d^2 D_A}{dz^2} + \left\{ \frac{(1+z)H_0}{2H^2} \frac{\bar{\sigma}(z)\Omega_m(1+z) + \Omega_K}{(1+z)H_0^2} \right\} \frac{dD_A}{dz} = \frac{2}{1+z} \frac{\Omega_m H_0^2}{H^2} \left[ (1+z)\bar{\sigma}(z) D_A \right],$$

where

$$H(z)^2 = H_0^2 \left[ \sigma(z)\Omega_m(1+z)^3 + \Omega_{\Lambda,0} + \Omega_K(1+z)^2 \right].$$

This modified DR equation attempts to take into account some aspects of the change in expansion expected from an inhomogeneous matter distribution. There are clearly a variety of ways to do this (e.g. we have ignored $\alpha$ terms which could be important), but we have chosen just one. See Mattsson (2010) for an alternative approach.

4.4.2 Shell approximation

Consider a single line of sight, smoothed over some scale $\lambda$. The density profile along this line of sight takes some form $\rho(z)$. If we neglect the angular part of the shear $\sigma_{\mu\nu}$, then we can think of the beam as passing through shells of differing density. There exists a spherically symmetric Lemaître–Tolman–Bondi (LTB) model with non-zero $\Lambda$ that has the same density profile $\rho(z)$ along the past light cone from the centre. The $D_A(z)$ relation in the LTB model viewed from the centre will approximate the $D_A(z)$ relation along the line of sight we are trying to model. Each line of sight would have a different associated LTB model (a mosaic of cones around us, in the language of Wang 2000a). The utility of this approximation lies in the fact that we can specify a density profile on a surface of constant time, and use the exact LTB solution to evolve the density backwards on to the past light cone. We can then calculate $D_A(z)$ exactly for that line of sight. Most importantly, this will account for the variable expansion rate along the direction of propagation which also takes into account the radial component of the shear.

In the LTB model, the angular Hubble rate is given by an effective Friedmann equation (February et al. 2010)

$$\frac{H^2(z, r)}{H_{1,0}^2(r)} = \Omega_m(0) a_1^3 + \Omega_K(0) a_1^{-2} + \Omega_{\Lambda,0}(r),$$

where the angular scale factor $a_1(z, r)$ is normalized to unity today, the $\Omega$ values have an arbitrary radial degree of freedom in them, and $H_{1,0}(r)$ is calculated once the age is set (or the Hubble rate at the centre is chosen). If $\Omega_m(0)$ is chosen as a constant we have an FL model. To model a radial line of sight we can choose the density profile today as $\rho_0(z) = [1+\delta(r)]\rho_0(0)$, and then

$$\Omega_m(r) = \frac{H_{1,0}^2(r)}{H^2(z, r)} \int_0^r dr \frac{\rho_0(r')}{r'^2}.$$

Then equation (84) evolves the density back on to the past light cone, using

$$\frac{dr}{dz} = -\frac{1}{(1+z)H_i}, \quad \frac{dr}{dz} = \frac{\sqrt{1+\rho_0 H_{1,0}^2}}{(1+z)d_0(a_1 r)}.$$

where the radial Hubble rate is $H_i(t, r) = [\partial_i \partial_0(a_1 r)]/[\partial_0(a_1 r)]$. The area distance is then

$$D_A(z) = a_1(t(z), r(z)) r(z).$$
4.4.3 Numerical investigation

We can compare these different approximations numerically. First consider the case of a single density fluctuation. Fig. 4 shows the results of looking through a large void and large overdensity, 500 Mpc away, modelled with a Gaussian deviation from $\alpha = 1$ of width $\sim 100$ Mpc. An underdensity causes an increase in the distance modulus at redshifts beyond itself, in both the DR and modified DR cases; the opposite happens for an overdensity.

The DR and modified DR are qualitatively similar, while the shell approximation is very different. According to the (modified) DR approximations, we should expect SNIa to appear dimmer when located behind an underdensity as compared to an overdensity. In contrast, the shell approximation gives the opposite effect with a much larger amplitude: SNIa located behind a void appear brighter than in the fiducial cosmology; located behind an overdensity, they appear dimmer. The reason is as follows: although a void results in a negatively curved region (which would imply diverging light rays and larger distances), this is accompanied by an increase in the expansion rate in that region, which actually has a much stronger effect on distances (compare Räisänen 2009). In FL, increasing the expansion rate and decreasing the density while keeping the age fixed results in a model with smaller distances, and this is exactly what happens here. For an overdensity, the reverse applies. However, note that while the shell approximation captures the mean expansion rate down the line of sight nicely, it may not capture the radial expansion rate correctly, e.g. looking through a spherical void versus a shell with the same density profile have different shear along the line of sight, making our approximation somewhat exaggerated in such a case.

Now consider the case where the density along the line of sight is reduced by a fixed percentage below the background value. Fig. 3 shows that the main contribution of smoothing a simulation over smaller beam sizes is to reduce the mean value of $\alpha$, since particles of significant size are rarely encountered. In the DR and modified DR approximations, this amounts to fixing $\alpha = \text{const}$ and Monte Carlo-ing over the PDF. In the shell approximation we have $\delta = \alpha - 1$ (since $\rho$ is constant, this is really just an FL model with adjusted parameters). In Fig. 5 we show the distance modulus for two fiducial backgrounds: one a standard $\Lambda$CDM model and the other a curved CDM model with $\Lambda = 0$. We see that all the approximations give a systematic dimming effect to varying degrees. (There is actually an overall brightening in the shell case due to the change in $h$, but we have subtracted this off because SNIa observations are only sensitive to relative magnitudes, not absolute ones, which is marginalized over together with $h$.) While the DR and shell approximations are rather similar, giving changes to the distance modulus of at most $0.1$ at $z \sim 1$, we see that the modified DR approximation gives a much stronger effect, of several times that.

It is striking that in the case $\Lambda = 0$ for $\alpha \lesssim 0.3$ we see the modified DR distance modulus mimicking the behaviour of a dark energy component. More specifically, both the DR and modified DR approximations can mimic an $\Lambda$CDM model, though the DR
The average number of particle intersections before the redshift \( z \) is the optical depth,
\[
\tau(z) = \pi r_s^2 \int_0^z \frac{n(z')}{(1 + z')H(z')} \, dz'.
\] (89)
We assume particle number conservation, \( n(z) = n_0(1 + z)^3 \), with \( n_0 \sim 0.005 \, h^3 \, \text{Mpc}^{-3} \) (number density of haloes above \( 10^{12} \, h^{-1} \, M_\odot \) which comprise about 50 per cent of the mass in the universe) and \( r_s \sim 20 \, h^{-1} \, \text{kpc} \) for the central, high-density, region of the haloes where the galaxies are assumed to reside. This implies that at \( z = 1 \), \( \tau \sim 0.023 \) and \( \tau = 0.032 \), which means that only 2.3 and 3.2 per cent predictions for lensing observations. However, the dynamical range of scales of matter inhomogeneities that the simulations can reproduce is very limited. Ray tracing within \( N \)-body simulations is usually done by projecting all matter on to equally spaced ‘lens planes’ which are separated by about 100 Mpc. Such an approach is not able to clarify the effect that haloes below the simulations mass resolution have on light bundles from SNIa.

Exact solutions that are more general than the Swiss cheese models are usually less realistic, given the highly non-linear nature of GR. A class of Szekeres models, in which inhomogeneities may be modelled as non-linear perturbations on an FL background, has been used to investigate light propagation by Meures & Bruni (2009). Despite the idealized nature of clumping, the results show that for inhomogeneities of large spatial size, parameter estimation could be seriously affected (see also Krasiński & Bolejko 2011). A stronger conclusion follows from analysing light rays in a universe with regularly spaced point masses separated by vacuum (Clifton & Ferreira 2009). The average dynamics is close to \( \Lambda \)CDM, but the optics behaves very differently. Unlike \( N \)-body simulations, this model is self-consistent (i.e. the particles generate their own space–time geometry). However, the modelling of matter is necessarily oversimplified.

\section{5 Ricci and Weyl Focusing}

The DR approximation neglects the effect of point sources and the Weyl focusing they produce, in particular in the strong lensing regime where it is not negligible. This issue was addressed in Weinberg (1976), by considering the effect of the DR equation (for almost all directions, where Weyl effects can be neglected) and the effect of strong lensing (for the relatively few directions where light rays pass close to matter, with strong lensing occurring and leading to multiple images). These combined effects are shown to lead to the usual FL relations when averaged over the whole sky, the decreased flux in most directions being compensated by higher flux in a few directions.

Keeping in mind that strong lensing effects are negligible in most directions, and in particular for most SNIa (unless we observe a galaxy or a cluster on the line of sight), we can try to go a step further than the DR approximation by relating, at least heuristically, the Weyl focusing to an effective Ricci focusing. The approximation that the SNIa bundles remain in the weak-field regime can be supported as follows. If matter is modelled as a gas of particles of mass \( m \) and proper radius \( r_s \), with mean number density \( n \), then the mean energy density is \( \rho = mn \). The probability that a line of sight intersects such a matter particle within the redshift band \( z \rightarrow z + dz \) is proportional to the surface area of the particles, to their density, and to the distance light propagates in this redshift interval, i.e.
\[
dP = \pi r_s^2 n(z) \frac{dz}{(1 + z)H(z)}. \tag{88}
\]

Figure 6. The \( \alpha(z) \) required to give an \( \Lambda \)CDM \( D_A(z) \) curve, for a variety of \( \Omega_{m0} \) (with curvature making up the rest of the energy component). The DR approximation requires extremely negative values of \( \alpha \) to mimic dark energy, which might suggest that the effect of modelling narrow beams may not be important for dark energy, and certainly could not be the underlying cause of it. On the other hand, a simple modification to DR yields an approximation that much more readily produces dark energy-like effects. This suggests instead that modelling narrow beams properly, and getting the DR approximation right, could be vital for determining the nature of dark energy.

approximation requires a drastic change to \( \alpha \) at low redshift to do so. For the modified DR approximation this is not so – see Fig. 6 – and it is well known that an LTB distance modulus can mimic any FL one.

It is clear from Fig. 4 that the DR approximation and its plausible variations can give very different results – so that the basis of the DR approximation itself is suspect.

\subsection*{4.5 Other approaches}

Other approaches to the problem of light propagation in a clumpy universe are based on exact non-linear solutions or numerical methods or both.

In ‘Swiss cheese’ models, an FL universe contains one or more spherical inhomogeneous regions, with the same average density as the FL model, which could be Schwarzschild vacuoles, or more realistically, LTB balls. These have been used extensively to model the effect of inhomogeneities on light rays (e.g. Brouzakis, Tetradi & Tzavara 2007; Marra et al. 2007; Biswas & Notari 2008; Vanderveld, Flanagan & Wasserman 2008; Clifton & Zuntz 2009; Valkenburg 2009; Bolejko & Célerier 2010; Kantowski, Chen & Dai 2010; Szybka 2011). The results depend on the position and nature of the inhomogeneous regions, as well as on the method for randomization of light rays. A careful analysis that approximates observations in all directions and over a range of distances, Szybka (2011) concludes that there are only small corrections to the standard FL results. This is not surprising, since the average density along typical lines of sight is very close to the FL density. For the case of SNIa beams, which may preferentially sample underdense regions, the results could be different. However, an inherent limitation of this approach is that, by construction, and given the highly symmetric nature of the inhomogeneities, the clumps do not affect the expansion rate of the universe.

The method of ray tracing through \( N \)-body simulations (e.g. Takahashi et al. 2011) is very useful for statistical analysis and
of the lines of sight intersect a galaxy before \( z = 1 \), respectively, for an Einstein–de Sitter and a flat \( \Lambda \)CDM model with \( \Omega_m = 0.3 \). This is a rough estimate – \( \Lambda \) and inhomogeneous distribution of the luminous matter would tend to lower it.

In order to describe the transition from Weyl to Ricci focusing, we first recall the lensing effect of a single mass, whose gravitational potential is

\[
\Phi = -Gm(b^2 + u^2)^{-1/2},
\]

where \( b \) is the impact parameter and \( u \) is the distance along the line of sight. The deflection angle is thus

\[
\alpha = \frac{4Gm}{b^2}.
\]

The critical density and the Einstein angular radius are, respectively, defined by \( \Sigma_{\text{crit}} = D_{\text{OS}}/(4\pi G D_{\text{DS}} D_{\text{LS}}) \) and \( \theta_E^2 = 4GmD_{\text{DS}}/(D_{\text{OL}} D_{\text{OS}}) \), so that the lens equation takes the form

\[
\theta = \theta_i + \hat{\alpha}(\theta),
\]

where \( D_{\text{DS}}, D_{\text{OL}} \) and \( D_{\text{OS}} \) are the angular distances between the lens plane and the source, the observer and the lens plane, and the observer and the source, respectively. The angular position of the image on the lens plane is then given by \( \theta = b/D_{\text{OL}}. \) \( \hat{\alpha} \) is given by \( \hat{\alpha} = \theta_E^2/\theta^2 \), since \( \Sigma(\theta) = m\theta^3(2) = m\theta^3(2)/D_{\text{OL}}^2 \) for a point mass. The amplification matrix is then obtained as \( \hat{A}_E^\alpha = \theta_E^2/\theta^2 \), so that \( A_{\text{ab}} = \delta_{\text{ab}} - \delta_{\text{ab}} \hat{\alpha}_0 = \delta_{\text{ab}} - \delta_{\text{ab}} \hat{\alpha} \psi \) with \( \psi = \theta_E^2 \ln \theta \).

We conclude that the convergence and the shear are given by

\[
\kappa = \pi \theta_E^2 \delta^2(\theta), \quad (\gamma_1, \gamma_2) = \frac{\theta_E^2}{\theta^2} \left( \theta_1^2 - \theta_1^2, 2\theta_1 \theta_2 \right),
\]

which is more easily written in terms of the polar angle on the screen

\[
(\theta_1 = \theta \cos \varphi, \theta_2 = \theta \sin \varphi)
\]

as

\[
(\gamma_1, \gamma_2) = \frac{\theta_E^2}{\theta^2} (\cos 2\varphi, \sin 2\varphi).
\]

For a single mass and a line of sight that does not intersect it, \( \kappa = 0 \) and the magnification is \( \mu = [1 - |\varphi|^2]^{-1} \sim 1 + 2\theta_E^2/\theta^2 \), neglecting the image inside the Einstein radius. Now consider a shell of thickness \( \Delta z \) such that the density of particles is \( n(z) \) and \( dX \ll D_{\text{DS}}, D_{\text{OL}}, D_{\text{OS}} \). For a typical light ray, the total amplification matrix will have a shear given by

\[
\hat{\gamma}_1(z), \hat{\gamma}_2(z) = \theta_E^2 \sum \left[ \frac{\cos 2\varphi}{\theta^2}, \frac{\sin 2\varphi}{\theta^2} \right],
\]

where the sum is over all particles of the shell. If the particles are distributed homogeneously, this implies that they are isotropically distributed around the line of sight, so that we expect \( \hat{\gamma}_1(z), \hat{\gamma}_2(z) \sim 0 \).

Intuitively, this is understood by the fact that each point mass induces an ellipticity in a different direction and they should average to reduce the total ellipticity. The remaining effect of all the Weyl distortions is thus an effective convergence that can be determined from the magnification \( k(z) = \theta_E^2 \sum \theta^{-4} \). Estimating the sum by assuming we have a uniform distribution and that the typical smallest distance is \( \theta \sim n^{-1/3} D_{\text{OL}} \), we get that \( \sum \theta^{-4} \sim D_{\text{OL}}^2 n \), and thus

\[
k(z) \sim f_k(z) \left[ \chi(z) \right]^2 \left[ f_k(z) \left[ \chi(z) \right] \right]^{-2},
\]

where \( f_k(z) \) is the unfilled angular distance, i.e. using \( \hat{H} \) instead of \( H \) in (61). We can compare to the effect that a homogeneous distribution of density \( n_{\text{eff}}(z) = a(z)n(z) \) would generate through its Ricci focusing to get

\[
a(z) = 4Gm \frac{f_k(z) \left[ \chi(z) \right] - \chi(z)}{f_k(z) \left[ \chi(z) \right]} H(z).
\]

This is the effective DR parameter for the averaged Ricci convergence. It still depends explicitly on \( n \) because this is a second-order effect, hence scaling as \( \theta_E^2 \), while it is first order in a homogeneous medium. Such an estimate is indeed very crude but it confirms the statements of Dyer & Roeder (1981) and Weinberg (1976) and the results of the numerical simulations of Holz & Wald (1998).

6 DISCUSSION AND CONCLUSIONS

The effect of the inhomogeneity of the matter distribution induces a dispersion of the magnification of SNIa and thus of the luminosity distance. We have argued that this effect has not been properly modelled for SNIa since the beam is very narrow and far below the scales resolved in any numerical simulation.

For the first time, we have attempted to quantify the probability distribution for narrow beams, using a combination of \( N \)-body simulations and a PS approach. For a narrow beam of fixed length, the PDF is non-Gaussian, peaked at densities below the cosmic mean, with a power-law tail, whose power depends on the diameter of the beam, describing the relatively few lines of sight which have an overdense mean. These PDFs contrast sharply with distributions based on using cubes of the same volume. These estimates are based on current \( N \)-body simulations which do not have the resolution to probe beams with a diameter \( \lesssim 100 h^{-1} \) kpc. Nevertheless, the trend is clear: narrow beams typically experience a lower than average density, and do not sample the cosmic mean density until their length approaches the Hubble scale. Based on our results, we estimate that significantly more than 75 per cent of the beams experience less than the mean density.

From a theoretical point of view, the effect must be described by the set of equations (21), (22) and (30) that describe the distortion and magnification of any light bundle, whatever the space–time geometry. The explicit covariant form of these equations is given by equations (50)–(55). The main problem is that the solution of this set of equations requires a description of the distribution of matter on the scale of the beam size, i.e. on scales much smaller than those of our current understanding.

This dispersion has been modelled in some regimes but we argue that the distribution of matter on the scales relevant for the description of an SNIa bundle is not understood yet. On such small scales, the statistical dispersion comes with a bias that has two origins: (1) the fact that the non-linear terms in the expression for the magnification cannot be neglected a priori; (2) an observational selection effect due to the fact that most SNIa are observed in directions where they are not overshadowed by a galaxy. The bundles included in SNIa analysis are thus more likely to probe underdense regions.

Why is the Hubble diagram so compatible with that of an FL universe? In particular, the typical transverse size of the bundle is smaller than the typical mean distance between the smallest bound structures in the currently favoured CDM paradigm. Why does the fluid approximation used to interpret the data hold? This question is two-sided. It questions the robustness of the interpretation of the cosmological data but also offers a way to constrain the distribution of matter on small scales. We also gave a heuristic argument concerning the Ricci–Weyl focusing issue, leading to a prediction of an

\[\Sigma(\text{crit}) = \frac{D_{\text{OS}}}{4\pi G D_{\text{DS}} D_{\text{LS}}}; \theta_E^2 = \frac{4Gm D_{\text{DS}}}{D_{\text{OL}} D_{\text{OS}}}, \text{ and } \theta_i = \frac{4Gm b}{\theta^2}.\]
effective DR factor $\alpha(z)$ once the fractions of clustered and smooth matter are known.

Another description is provided by the DR equation. It has however some simplifying hypotheses that neglect the effects of changes in $v(z)$ and the local expansion due to clumping. We suggested two plausible modifications to the DR approach, and showed that the three models produce very different results – thus undermining confidence in the DR approximation. In particular, it is not clear whether underdensities lead to demagnification (due to negative curvature) or magnification (due to the increase in the expansion rate). Our shell approximation clearly points to the opposite effect calculated via normal lensing or the DR approximation: an SNIa located just behind an underdensity should appear brighter than it would in the fiducial cosmology. We estimate that an SNIa at the far side of a 100-Mpc void could be 0.1 mag brighter than it would be with no void present. Though likely an overestimation, this could have important implications for parameter estimation from SNIa. Quantifying this properly is an important open problem. In fact, it is striking to note from our investigations of $N$-body simulations that we should expect $\alpha$ to vary as a function of distance, from significantly below 1 locally and then approaching 1 on Hubble scales. Within our shell approximation, this would give exactly the kind of model used in Hubble scale ‘void models’, which require no dark energy, but without any kind of anti-Copernican fine tuning involved (Clarkson 2012). Every observer would observe such an effect.

While accurate modelling of such beams may be problematic for some time, we can still observationally test to see if there are problems and phenomenologically correct them.

(i) SNIa line of sight. Dividing up SNIa samples according to the estimated density along the line of sight may reveal a bias. If so, this may indicate that the effects we have discussed here must be taken into account.

(ii) Discordant distances. In any exact relativistic model the luminosity distance is $D_L = (1 + z)^2 D_A$, where $D_A$ is the area angular diameter distance. Measurements of the area distance on large scales will not be affected by the problems we have discussed here; however, we can expect a failure at some level of this relation when comparing measurements of $D_A$ from large-scale measurements such as the BAO and the CMB to measurements of $D_L$ from SNIa. On smaller scales, where the area distance is measured from radio or quasar sources, there could be an effective reciprocity breakdown because such sources still have much larger beam sizes than SNIa, and so smear the matter distribution to include many more overdense regions. Such a violation was found in Bassett & Kunz (2004), where a relative brightening of SNIa was found – as we would predict from the shell approximation.

(iii) Consistency conditions. A variety of consistency tests have recently been developed as a way of testing the standard model (see Clarkson 2012 for a review). It has recently been shown that these are strongly sensitive to changes of the DR form (Busti & Lima 2012). This implies that they can be used to probe the effects we have discussed here.

Generically, then, distances to the same object will depend on the scale over which the light from the source smears the intervening matter distribution.

We have found that the old problem of modelling narrow beams remains unsolved. As different interpretations of the problem give conflicting yet significant effects, we believe this problem needs considerably more attention. This is important not only from a theoretical perspective, but to ensure precision cosmology delivers correct answers as well as precise ones.

ACKNOWLEDGMENTS

We thank David Bacon, Krzysztof Bolejko, Ruth Durrer, Stefan Hilbert, Yannick Mellier, Bob Nichol, Brian Schmidt, Joe Silk, Zachary Slepian and Yun Wang for discussions and/or comments. RM is supported by a South African SKA Research Chair, and by the UK Science & Technology Facilities Council (grant no. ST/H002774/1). CC, GE, AF and OM are supported by a NRF (South Africa)/Royal Society (UK) exchange grant. JPU is partially supported by the ANR-Thales. The Millennium Simulation and the MultiDark data bases used in this paper and the web application providing online access to them were constructed as part of the activities of the German Astrophysical Virtual Observatory. Data for haloes and galaxies are publicly available at http://www.mpia-garching.mpg.de/millennium and http://www.multidark.org/MultiDark. The MultiDark data base is the result of a collaboration between the Leibniz-Institute for Astrophysics Potsdam (AIP) and the Spanish MultiDark Consolider Project CSD2009-00064. The Bolshoi and MultiDark simulations were run on the NASA’s Pleiades supercomputer at the NASA Ames Research Center.

REFERENCES

Amanullah R. et al., 2010, ApJ, 716, 712
Anderhalden D., Diemand J., 2011, MNRAS, 414, 3166
Angulo R. E., White S. D. M., 2010, MNRAS, 401, 1796
Bassett B. A., Kunz M., 2004, Phys. Rev. D, 69, 101305
Bernardeau F., 1997, A&A, 324, 15
Bertone G., Hooper D., Silk J., 2005, Phys. Rep., 405, 279
Bertotti B., 1966, Proc. R. Soc. Lond. A, 294, 195
Biswas T., Notari A., 2008, J. Cosmol. Astropart. Phys., 6, 21
Bolejko K., 2011, MNRAS, 412, 1937
Bolejko K., Céleriér M.-N., 2010, Phys. Rev. D, 82, 103510
Bonvin C., Durrer R., Gasparini M. A., 2006, Phys. Rev. D, 73, 023523
Brouzas N., Tetradis N., Tzavara E., 2007, J. Cosmol. Astropart. Phys., 2, 13
Busti V. C., Lima J. A. S., 2012, preprint (arXiv:1204.2083)
Clarkson C., 2012, preprint (arXiv:1204.5505)
Clarkson C., Maartens R., 2010, Classical Quantum Gravity, 27, 124008
Clifton T., Ferreira P. G., 2009, J. Cosmol. Astropart. Phys., 10, 26
Clifton T., Zuntz J., 2009, MNRAS, 400, 2185
Cooray A., Holz D. E., Huterer D., 2006a, ApJ, 637, L77
Cooray A., Huterer D., Holz D. E., 2006b, Phys. Rev. Lett., 96, 021301
Cooray A., Holz D. E., Caldwell R., 2010, J. Cosmol. Astropart. Phys., 11, 15
Cuesta A. J., Prada F., Klypin A., Moles M., 2008, MNRAS, 389, 385
Dalal N., Holz D. E., Chen X., Frieman J. A., 2003, ApJ, 585, L11
Dashevskii V. M., Slysh V. I., 1966, Sva, 9, 671
Dieudonné J., Moore B., Stadel J., 2005, Nat, 433, 389
Dodsonel A., Sillainnotno A., 2006, Phys. Rev. D, 74, 063515
Dyer C. C., Roeder R. C., 1972, ApJ, 174, L115
Dyer C. C., Roeder R. C., 1973, ApJ, 180, L31
Dyer C. C., Roeder R. C., 1974, ApJ, 189, 167
Dyer C. C., Roeder R. C., 1981, Gen. Relativ. Gravitation, 13, 1157
Ehlers J., Schneider P., 1986, A&A, 168, 57
Ellis J. R., 1994, Classical Quantum Gravity, 15, 2345
Faltenbacher A., Finoguenov A., Droy N., 2010, ApJ, 712, 484
Fang L., Wu X., 1989, Chin. Phys. Lett., 6, 233
February S., Larena J., Smith M., Clarkson C., 2010, MNRAS, 405, 2231
Frieman J. A., 1996, Comments Astrophys., 18, 323

© 2012 The Authors, MNRAS 426, 1121–1136
Monthly Notices of the Royal Astronomical Society © 2012 RAS

Supernovae observations in a lumpy universe 1135
