Approaches to understanding cosmic acceleration

Alessandra Silvestri\textsuperscript{1} and Mark Trodden\textsuperscript{2}

\textsuperscript{1} Department of Physics and Kavli Institute for Astrophysics and Space Research, MIT, 77 Massachusetts Ave., Cambridge, MA 02139, USA
\textsuperscript{2} Center for Particle Cosmology, Department of Physics and Astronomy, David Rittenhouse Laboratories, University of Pennsylvania, Philadelphia, PA 19104, USA

E-mail: asilvest@mit.edu and trodden@physics.upenn.edu

Received 2 April 2009, in final form 2 June 2009
Published 25 August 2009
Online at stacks.iop.org/RoPP/72/096901

Abstract

Theoretical approaches to explaining the observed acceleration of the universe are reviewed. We briefly discuss the evidence for cosmic acceleration, and the implications for standard general relativity coupled to conventional sources of energy–momentum. We then address three broad methods of addressing an accelerating universe: the introduction of a cosmological constant, its problems and origins; the possibility of dark energy and the associated challenges for fundamental physics and the option that an infrared modification of general relativity may be responsible for the large-scale behavior of the universe.

(Some figures in this article are in colour only in the electronic version)

This paper was invited by B Berger.

1. Introduction

The development of general relativity (GR) almost 90 years ago provided not only a new way to understand gravity, but heralded the dawn of an entirely new branch of science—cosmology. Of course, a discussion of the visible contents of the night sky was not a new development—these objects and phenomena have forever fascinated civilizations. However, a coherent theoretical framework within which to explore their distribution and evolution was lacking before Einstein’s crucial insight—that space and time are themselves players in the cosmic drama. Hubble’s 1921 observation that distant galaxies are receding from ours at speeds proportional to their distances, in agreement with Friedmann’s corresponding solution to general relativity, then provided the first experimental confirmation of this new science.

Almost a century has passed since the beginning of this era, and in the intervening years increasingly accurate predictions of this model of the cosmos, supplemented only by the presence of a dark matter component, have been
confronted, and spectacularly passed, a host of detailed tests—
the existence of the cosmic microwave background (CMB) 
radiation; the abundances of the light elements through big 
bang nucleosynthesis (BBN); the formation of structure under 
gravitational instability; the small temperature anisotropies in 
the CMB; the structure of gravitational lensing maps and many 
more. Many of these tests are highly non-trivial and provide 
remarkable support for the overall big bang model.

All this held true until approximately a decade ago, when 
the first evidence for cosmic acceleration was reported by 
two groups using the light curves of type Ia supernovae 
to construct an accurate Hubble diagram out to redshift 
greater than unity (figure 1). As we will describe in some 
detail, the natural expectation is that an expanding universe, 
evolving under the rules of GR, and populated by standard 
matter sources, will undergo deceleration—the expansion rate 
should slow down as cosmic time unfolds. Amazingly, those 
initial data, supplemented over the last decade by dozens of 
further, independent observations, showed that the universe is 
speeding up—what we observe is not deceleration, but cosmic 
acceleration!

The goal of this paper is to provide an overview of what 
is meant by cosmic acceleration, some proposed theoretical 
approaches to understanding the phenomenon, and how 
upcoming observational data may cast crucial light on these. 
We have not attempted to be completely comprehensive, but 
have tried to provide a flavor of the current research landscape, 
focusing on the better known approaches, and highlighting 
the challenges inherent in constructing a fundamental physics 
understanding of the accelerating universe. There are a number 
of excellent recent review papers on dark energy, which the 
reader might find as a useful complement to this paper [1–4].

The paper is organized as follows. In the next section 
we provide a brief introduction to the relevant cosmological 
solutions to general relativity, with perfect fluid sources. In 
section 3 we review the observational evidence, primarily 
from observations made during the last decade or so, for an 
accelerating universe. In section 4 we then discuss the minimal 
possibility that Einstein’s cosmological constant is responsible 
for late-time cosmic acceleration and discuss the challenges 
the required magnitude of this parameter raises for particle 
physics. In section 5 we then explore a class of dynamical 
dark energy models, in which the universe accelerates due 
to the evolution of a new component of the cosmic energy 
budget, under the assumption that the cosmological constant 
is negligibly small. In a similar vein, in section 6 we then 
consider the possibility that the universe contains only standard 
model and dark matter sources of energy and momentum, but 
accelerates because GR is modified at large distances, leading 
to new self-accelerating solutions. Finally, before concluding, 
in section 7 we then consider the challenge of distinguishing 
among these possible explanations for cosmic acceleration and 
discuss how upcoming observational missions may help us to 
better understand these phenomena.

A note on conventions. Throughout this paper we use 
units in which $h = c = 1$, adopt the (−, +, +, +) signature 
and define the reduced Planck mass by $M_p^{-2} \equiv \kappa^2 = 8\pi G$.

### 2. Essential elements of background cosmology

Our goal here is to lay out the bare minimum for understanding 
the background evolution of the universe. By background 
in this context, we mean the dynamics pertaining to the 
homogeneous and isotropic description of spacetime, valid 
on the very largest scales, without reference to spatial 
perturbations of either the metric or matter fields$^3$.

The most general homogeneous and isotropic metric 
ansatz is the Friedmann–Robertson–Walker (FRW) form

$$\text{d}s^2 = -\text{d}t^2 + a^2(t) \left[ \frac{\text{d}r^2}{1 - kr^2} + r^2 (\text{d}\theta^2 + \sin^2 \theta \text{d}\phi^2) \right],$$

where $k$ describes the curvature of the spatial sections (slices at constant cosmic time) and $a(t)$ is referred to as the scale factor 
or the universe. Without loss of generality, we may normalize $k$ so that $k = +1$ corresponds to positively curved spatial 
sections (locally isometric to 3-spheres); $k = 0$ corresponds 
to local flatness and $k = -1$ corresponds to negatively curved 
(locally hyperbolic) spatial sections. These local definitions 
say nothing about the global topology of the spatial sections, 
which may be that of the covering spaces—a 3-sphere, an 
infinite plane or a 3-hyperboloid—but it need not be, as 
topological identifications under freely acting subgroups of 
the isometry group of each manifold are allowed. As a specific 
example, the $k = 0$ spatial geometry could apply just as well 
to a 3-torus as to an infinite plane.

Note that we have not chosen a normalization such that 
a_0 = 1. We are not free to do this and to simultaneously 
normalize $|k| = 1$, without including explicit factors of the 
current scale factor in the metric. In the flat case, where $k = 0$, 
we can safely choose $a_0 = 1$.

$^3$ We follow closely the brief discussion in [5].
At cosmic time \( t \), the physical distance from the origin to an object at radial coordinate \( r \) is given by \( d(t) = a(t)r \). The recessional velocity of such an object due to the expansion of the universe is then given by

\[
v(t) = H(t) d(t), \quad (2)
\]

where \( H(t) \equiv \dot{a}/a \) (an overdot denotes a derivative with respect to \( t \)) is the Hubble parameter, the present day value of which we refer to as the Hubble constant \( H_0 \).

The Hubble parameter is a useful function through which to parametrize the expansion rate of the universe. As we shall encounter later, it is convenient to define a second function, describing the rate at which the expansion rate is slowing down or speeding up—the deceleration parameter, defined as

\[
q(t) = -\frac{\ddot{a}}{a} \quad (3)
\]

### 2.1. Dynamics: the Friedmann equations

The FRW metric is merely an ansatz, arrived at by requiring homogeneity and isotropy of spatial sections. The unknown function \( a(t) \) is obtained by solving the differential equations obtained by substituting the FRW ansatz into the Einstein equation

\[
R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G T_{\mu\nu}. \quad (4)
\]

The energy–momentum tensor \( T_{\mu\nu} \) describes the matter content of the universe. It is often appropriate to adopt the perfect fluid form for this

\[
T_{\mu\nu} = (\rho + p) U_\mu U_\nu + p g_{\mu\nu}, \quad (5)
\]

where \( U^\mu \) is the fluid four-velocity, \( \rho \) is the energy density in the rest frame of the fluid and \( p \) is the pressure in that frame.

Substituting (1) and (5) into (4), one obtains the Friedmann equation

\[
H^2 \equiv \left( \frac{\dot{a}}{a} \right)^2 = \frac{8\pi G \sum \rho_i - k}{3a^2}, \quad (6)
\]

and

\[
\frac{\ddot{a}}{a} + \frac{1}{2} \left( \frac{\dot{a}}{a} \right)^2 = -4\pi G \sum p_i - \frac{k}{2a^2}. \quad (7)
\]

Here \( i \) indexes all different possible types of energy in the universe. The Friedmann equation is a constraint equation, in the sense that we are not allowed to freely specify the time derivative \( \dot{a} \); it is determined in terms of the energy density and curvature. The second equation is an evolution equation.

We may combine (6) and (7) to obtain the acceleration equation

\[
\frac{\dddot{a}}{a} = -\frac{4\pi G}{3} \sum \rho_i + 3p_i, \quad (8)
\]

which will prove central to the subject of this paper.

In fact, if we know the magnitudes and evolutions of the different energy density components \( \rho_i \), the Friedmann equation (6) is sufficient to solve for the evolution uniquely.

The acceleration equation is conceptually useful, but rarely invoked in calculations.

The Friedmann equation relates the rate of increase of the scale factor, as encoded by the Hubble parameter, to the total energy density of all matter in the universe. We may use the Friedmann equation to define, at any given time, a critical energy density,

\[
\rho_c \equiv \frac{3H^2}{8\pi G}, \quad (9)
\]

for which the spatial sections must be precisely flat (\( k = 0 \)). We then define the density parameter

\[
\Omega_i \equiv \frac{\rho_i}{\rho_c}, \quad (10)
\]

which allows us to relate the total energy density in the universe to its local geometry via

\[
\Omega_{\text{total}} > 1 \Rightarrow k = +1, \quad \Omega_{\text{total}} = 1 \Rightarrow k = 0, \quad \Omega_{\text{total}} < 1 \Rightarrow k = -1. \quad (11)
\]

It is often convenient to define the fractions of the critical energy density in each different component by

\[
\Omega_i = \frac{\rho_i}{\rho_c}. \quad (12)
\]

Energy conservation is expressed in GR by the vanishing of the covariant divergence of the energy–momentum tensor,

\[
\nabla_\mu T^{\mu\nu} = 0. \quad (13)
\]

For the FRW metric (1) and a perfect fluid energy–momentum tensor (5) this yields a single energy-conservation equation,

\[
\dot{\rho} + 3H (\rho + p) = 0. \quad (14)
\]

This equation is actually not independent of the Friedmann and acceleration equations, but is required for consistency. It implies that the expansion of the universe (as specified by \( H \)) can lead to local changes in the energy density. Note that there is no notion of conservation of ‘total energy’, as energy can be interchanged between matter and the spacetime geometry.

One final piece of information is required before we can think about solving our cosmological equations: how the pressure and energy density are related to each other. Within the fluid approximation used here, we may assume that the pressure is a single-valued function of the energy density \( p = p(\rho) \). It is often convenient to define an equation-of-state parameter, \( w \), by

\[
p = w\rho. \quad (15)
\]

This should be thought of as the instantaneous definition of the parameter \( w \); it does not represent the full equation of state, which would be required to calculate the behavior of fluctuations. Nevertheless, many useful cosmological matter sources do obey this relation with a constant value of \( w \). For example, \( w = 0 \) corresponds to pressureless matter,
Thus, the energy density of radiation will scale as the expansion of the universe stretches the wavelengths of light. As for dust and by an extra factor of the scale factor as the physical volume of the redshift, the energy density in a fixed number of photons will drop with the physical volume, i.e. as $a(t)^{-3}$, in agreement with (16), with $w = 0$.

To make a similar argument for radiation, first note that the expansion of the universe (the increase of $a(t)$ with time) results in a shift to longer wavelength $\lambda$, or a redshift, of photons propagating in this background. A photon emitted with wavelength $\lambda_0$ at a time $t_0$, at which the scale factor is $a_0 = a(t_0)$ is observed today ($t = t_0$, with scale factor $a_0 = a(t_0)$) at wavelength $\lambda_a$, obeying

$$\frac{\lambda_a}{\lambda_0} = \frac{a_0}{a} \equiv 1 + z. \tag{17}$$

The redshift $z$ is often used in place of the scale factor. Because of the redshift, the energy density in a fixed number of photons in a fixed comoving volume drops with the physical volume (as for dust) and by an extra factor of the scale factor as the expansion of the universe stretches the wavelengths of light. Thus, the energy density of radiation will scale as $a(t)^{-4}$, once again in agreement with (16), with $w = 1/3$.

Thus far, we have not included a cosmological constant $\Lambda$ in the gravitational equations. This is because it is equivalent to treat any cosmological constant as a component of the energy density in the universe. In fact, adding a cosmological constant $\Lambda$ to Einstein’s equation is equivalent to including an energy–momentum tensor of the form

$$T_{\mu\nu} = -\frac{\Lambda}{8\pi G} g_{\mu\nu}. \tag{18}$$

This is simply a perfect fluid with energy–momentum tensor (5) with

$$\rho_{\Lambda} = \frac{\Lambda}{8\pi G}, \tag{19}$$

$$p_{\Lambda} = -\rho_{\Lambda},$$

so that the equation-of-state parameter is

$$w_{\Lambda} = -1. \tag{20}$$

This implies that the energy density is constant,

$$\rho_{\Lambda} = \text{constant}. \tag{21}$$

Thus, this energy is constant throughout spacetime; we say that the cosmological constant is equivalent to vacuum energy.

Similarly, it is sometimes useful to think of any non-zero spatial curvature as yet another component of the cosmological energy budget, obeying

$$\rho_{\text{curv}} = -\frac{3k}{8\pi Ga^2}. \tag{22}$$

$$p_{\text{curv}} = \frac{k}{8\pi Ga^2},$$

so that

$$w_{\text{curv}} = -1/3. \tag{23}$$

It is not an energy density, of course; $\rho_{\text{curv}}$ is simply a convenient way to keep track of how much energy density is lacking, in comparison with a flat universe.

### 2.2. Flat universes

Analytically, it is much easier to find exact solutions to cosmological equations of motion when $k = 0$. We are able to make full use of this convenience since, as we shall touch on in the next section, the combined results of modern cosmological observations show the present day universe to be extremely spatially flat. This can be seen, for example, from precision observations of the CMB radiation and independent measures of the Hubble expansion rate.

In the case of flat spatial sections and a constant equation-of-state parameter $w$, we may exactly solve the Friedmann equation (16) to obtain

$$a(t) = a_0 \left( \frac{t}{t_0} \right)^{2/(1+w)}, \tag{24}$$

where $a_0$ is the scale factor today, unless $w = -1$, in which case one obtains $a(t) \propto e^{Ht}$. Applying this result to some of our favorite energy density sources yields table 1.

Note that the matter- and radiation-dominated flat universes begin with $a = 0$; this is a singularity, known as the big bang. We can easily calculate the age of such a universe:

$$t_0 = \int_0^1 \frac{da}{aH(a)} = \frac{2}{3(1+w)H_0}. \tag{25}$$

Unless $w$ is close to $-1$, it is often useful to approximate this answer by

$$t_0 \sim \frac{1}{H_0^{-1}}. \tag{26}$$

It is for this reason that the quantity $H_0^{-1}$ is known as the Hubble time, and provides a useful estimate of the time scale for which the universe has been around.
3. Observational evidence for cosmic acceleration

An important breakthrough in cosmology occurred in the late 1990s with the measurements of type Ia supernovae (SNeIa) by the High-Z Supernova team [6] and the Supernova Cosmology Project [7]. As we review in this section, the data from these surveys, as well as from complementary probes, provide strong evidence that the universe has recently entered a phase of accelerated expansion.

3.1. Type Ia supernovae

Type Ia supernovae are stellar explosions that occur as a white dwarf, onto which mass is gradually accreting from a companion star, approaches the Chandrasekhar limit [8] (of about 1.4 solar masses) and the density and temperature in its core reach the ignition point for carbon and oxygen. This begins a nuclear flame that fuses much of the star up to iron. They are extremely bright events, with luminosities a significant fraction of that of their host galaxies during the peak of their explosions. Therefore they are relatively easy to detect at high redshift (z ∼ 1). This specific type of supernova is characterized by the absence of a hydrogen line (or, indeed, a helium one) in their spectrum (which is typical of type I SNe) and instead by the presence of a singly ionized silicon line at 615 nm, near peak light.

Despite a significant scatter, type Ia supernovae peak luminosities have been found to be very closely correlated with observed differences in the shapes of the light curves: dimmer SNe are found to decline more rapidly after maximum brightness, while brighter SNe decline more slowly [9–11]. This difference seems to be traceable to the amount of 56Ni produced in the explosion; more nickel implies both a higher peak luminosity and a higher temperature, and thus opacity, leading to a slower decline.

After one adjusts for the difference in the light curves, the scatter in peak luminosity can be reduced to 15%. In this sense, SNeIa are referred to as ‘standardizable candles’ and are very good candidates for distance indicators, since one would expect any remaining difference in their peak luminosity to be due to a difference in distance. Another important aspect of this uniformity is that it provides standard spectral and light curve templates that offer the possibility of singling out those SNe that deviate slightly from the norm [12].

After the necessary adjustments, all SNeIa have the same absolute magnitude M and any difference in their apparent magnitude m is attributable to a difference in their distance. The apparent and intrinsic magnitudes are related via the luminosity distance d_L,

\[ m = M + 5 \log \left( \frac{d_L}{10 \text{pc}} \right) + K, \]

where K is a correction for the shifting of the spectrum [6, 7, 9, 11, 13, 14] necessary because only a part of the spectrum emitted is actually observed. In an expanding universe, the expression of the luminosity distance as a function of redshift is

\[ d_L(z) = (1 + z) \cdot \int_0^z \frac{dz'}{H(z')} \]

By independently measuring d_L(z) and z, one can then constrain the expansion history H(z). Following pioneering work reported in [15], the High-Z Supernova team [6] and the Supernova Cosmology Project [7] measured the apparent magnitudes of many SNeIa (of redshift z ≲ 1) and directly determined their distances. They then compared these distances with those inferred from the redshifts of the host galaxies (measured from the spectra of the galaxies when possible, otherwise the spectra of the SNe themselves). The most distant SNe appeared dimmer than expected in a universe currently dominated by matter, and a careful comparison with low redshift supernovae allowed one to rule out the possibility that this dimming is due to intervening dust. Assuming a flat universe described by GR, and homogeneous and isotropic on large scales, the data are best fit by a universe which has recently entered a phase of accelerated expansion, i.e. a universe for which the deceleration parameter (3) is currently negative (q0 ≃ −1). The simplest model fitting the data is a universe in which matter accounts for only about a quarter of the critical density, while the remaining 70% of the energy density is in the cosmological constant Λ. This model is commonly referred to as the ΛCDM model. In fact, the SNe data by themselves allow a range of possible values for the matter and cosmological constant density parameters (Ω_M and Ω_Λ respectively). However, as we will discuss in the following subsections, we can use what we know from other observations, such as the CMB and the large scale structure (LSS), to further constrain these parameters.

In more recent years, distant SNe with redshifts up to z ∼ 1.7 have been observed by the Hubble space telescope (HST) [17, 18]. These observations show that the trend toward fainter SNe at moderate redshifts has reversed; therefore they play a key role in disregarding dust as a plausible explanation of the dimming of intermediate SNe, and in firmly establishing [19, 20] acceleration. More generally, SNe are important in the study of the expansion history. Nearby SNe (z ≲ 0.3) are useful for defining characteristics of type Ia SNe, understanding the explosion and exploring systematics; all combined, they are important for establishing distance indicators and are currently [21, 22] being measured, with more to come in the future [23, 24]. Intermediate redshift SNe (0.3 ≲ z ≲ 0.8) measure the strength of cosmic acceleration and are being measured by the ESSENCE project [25], the SuperNova Legacy Survey (SNLS) [26] and the Sloan Digital Sky Survey-II (SDSS-II) Supernova Survey [27]. Finally, distant SNe (z ≳ 0.8) are important to break the degeneracy among the cosmological parameters and they are currently being measured by the Higher-z Supernova Search Team (HZT) [28] and the GOODS team [29].

In the near future, the space-based Joint Dark Energy Mission (JDEM) [30], (a wide-field optical-infrared telescope), will offer precision measurements with the aim of determining the nature of cosmic acceleration. This mission may incorporate the Baryon acoustic oscillation (BAO), supernovae (SNe) and weak lensing (WL) techniques. Also, upcoming and future surveys such as the Earth-based Dark Energy Survey (DES) [31] and the Large Syncopic Survey...
Telescope (LSST) [32] will provide observations useful for constraining the expansion history of the universe.

3.2. Complementary probes

Supernovae are the first and most direct evidence for the late-time acceleration of the universe. However, independent evidence comes from complementary probes. Data from different observations are important not only to confirm the results from SNe, but also to help break some of the degeneracy in the cosmological parameters, such as the density parameters \( \Omega_M^0 \) and \( \Omega_{\Lambda}^0 \). As we have mentioned, the measurement of the expansion history via SNe is an example of the use of standard candles to infer the luminosity distance as a function of the redshift \( d_L(z) \). Alternatively, one may employ standard rulers, extracting the angular diameter distance \( d_A(z) \) (given by the ratio of the object’s actual size to its angular size) and then \( H(z) \). (Both methods are commonly referred to as geometrical probes). The expansion history can further be constrained with the growth of structure, as we will see shortly. Finally, some methods are based on inferring the age of the universe from stellar ages. In what follows we review several complementary probes of the expansion history of the universe.

The CMB is the radiation that reaches us from the surface of last scattering—that is, from the time at which photons decoupled from the thermal bath and started to free stream to us. It is in the form of an almost isotropic blackbody spectrum, with a temperature of approximately 2.7 K. The observed blackbody distribution, typical of sources that are in thermal equilibrium, is strong evidence for the theory of the big bang. At early times, when the universe was radiation dominated, photons were easily energetic enough to ionize hydrogen atoms. Therefore, the universe was filled with a photon–baryon plasma, and was opaque. As the universe expanded, the temperature decreased and the photons redshifted, losing energy until they were no longer able to ionize hydrogen. At \( T \sim O(1) \) eV, corresponding to a redshift of \( z \sim 1100 \), neutral hydrogen formed and the photons decoupled from the plasma. At early times, the number and energy densities of matter were high enough that the plasma constituents were in thermal equilibrium, ensuring a blackbody spectrum for the radiation distribution. This spectrum is maintained by photons after decoupling, just at lower and lower temperatures. Indeed, the effect of homogeneous and isotropic expansion on a blackbody distribution is that of preserving the blackbody structure with temperatures that redshift with the wavelength, \( T \propto 1/\lambda \).

As measured by the cosmic background explorer (COBE) [33], the temperature distribution of the photons is not completely isotropic, with fluctuations of one part in 10^5. These fluctuations contain a great deal of physical information about our universe, and the ongoing measurements from the Wilkinson microwave anisotropy probe (WMAP) [34–36] have reached unprecedented precision in using it to constrain cosmological parameters. It is worth exploring in a little more detail how this is achieved.

Matter in the universe today is highly inhomogeneous and clustered. Our understanding of how this LSS developed is that initially small density perturbations, in an otherwise homogeneous universe, grew through gravitational instability into the objects we observe today. The basic observable in CMB physics is the temperature fluctuation on the sky, for which we measure the 2-point correlation function. The Fourier counterpart of the correlation function, the power spectrum, is what is commonly shown in plots. As can be seen in figure 2, there are two major components of a typical CMB power spectrum: the Sachs–Wolfe plateau at large scales and a series of peaks and troughs at smaller scales. These both carry important information about the source of cosmic acceleration. The fluctuations we observe on the largest scales are essentially the primordial ones, since the largest scales have only recently entered the horizon and therefore little causal physics has affected them. On smaller scales we instead observe fluctuations which entered the horizon before decoupling and have undergone acoustic oscillations characteristic of the photon–baryon fluid.

Let us focus on the first peak of these oscillations. The acoustic oscillations are damped oscillations of the photon–baryon plasma under the competing effects of gravity and radiation pressure. The maximum amplitude of the oscillations is expected on fluctuations that entered the horizon just before decoupling. To see this, consider an overdense region of size \( R \), which therefore contracts under self-gravity over a timescale \( R/c \) (since we have set \( c = 1 \)). If \( R \gg H^{-1}_s \) then the region will not have had time to collapse over the lifetime of the universe at last scattering. If \( R \ll H^{-1}_s \) then collapse will be well underway at last scattering, matter will have had time to fall into the resulting potential well and result in a rise in temperature. This, in turn, gives rise to a restoring force from photons which acts to damp out the inhomogeneity. Clearly, therefore, the maximum anisotropy will be on a scale which has had just enough time to equilibrate, i.e. \( R \sim H^{-1}_s \). This means that we expect to see a peak in the CMB power spectrum at an angular size corresponding to the horizon size at last scattering. Since we know the physical size of the horizon at last scattering, this provides us with a ruler on the sky. The corresponding angular size will then depend on the spatial geometry of the universe. For a flat universe (\( k = 0 \), \( \Omega_M^0 = 1 \)), we expect a peak at \( l \simeq 220 \). As can be seen in figure 2, the first peak indeed appears at this angular scale, providing strong evidence for the spatial flatness of the universe. As represented in figure 3, the CMB strongly constrains the sum of the matter and dark energy densities, offering a probe which is complementary to the SNe data in the \( \Omega_M^0 - \Omega_{\Lambda}^0 \) space.

However, this is just one piece of the valuable information that can be extracted from the CMB. In general, the CMB power spectrum provides high precision constraints on essentially all of the cosmological parameters [36]. The overall CMB spectrum, as well as the matter power spectrum as measured from LSS, are best fit by models which undergo late-time accelerated expansion.

Information from the CMB concerning the total energy density of the universe can be combined with data about the distribution of matter, from which one extracts \( \Omega_M^0 \). In addition, the CMB power spectrum itself offers a way to constrain the matter density via the spacing of the peaks of
Figure 2. The cosmic microwave background angular power spectrum, representing the two-point correlation function of temperature fluctuations (in momentum space) on the sky as a function of angular separation, (courtesy of NASA/WMAP Science team). The solid line is the best-fit ΛCDM model. On larger scales, i.e. at lower multipoles $l$, we notice the integrated Sachs–Wolfe (ISW) plateau, at $l \sim 220$ we see the first acoustic peak and on smaller angular scales the subsequent acoustic peaks.

the acoustic oscillations. Complementary strong constraints on the matter density come from the observed distribution of matter in the universe. In particular, the turnover scale in the matter power spectrum is very sensitive to $\Omega_0^\Lambda$ and the abundance of clusters of galaxies constrain the $\sigma_8$ parameter, which encodes information about $\Omega_0^\Lambda$ and the fractional density perturbation.

The WMAP data together with LSS data from the Sloan Digital Sky Survey (SDSS) [40] yield the value $\Omega_0^\Lambda = 0.25 \pm 0.03$ (1σ, in the context of a flat ΛCDM model) [41]. Combining these with the supernovae data, we can pin down a narrow region in the $\Omega_0^M - \Omega_0^\Lambda$ space corresponding to $\Omega_0^M \sim 1 - \Omega_0^\Lambda \sim 0.27$, as shown in figure 3.

Gravitational lensing—the distortion of light paths by the inhomogeneous metric around matter in the universe—is also an important complementary probe as it offers a direct handle on the mass distribution, without relying on bias factors to convert from visible to total matter. Of particular importance in this context is the weak gravitational lensing of distant galaxies by intervening foreground mass overdensities, dubbed cosmic shear. This offers a direct measurement of mass fluctuations in the universe and their redshift evolution. The statistic associated with cosmic shear, the shear angular power spectrum, is sensitive to dark energy via the expansion history as well as via the growth of structure. From precise measurements of the cosmic shear and its cross-correlation with galaxies it is possible to extract geometrical probes of dark energy [42, 43] as well as an estimate of the matter density parameter [44]. In section 7 we will discuss in more detail the constraints on the growth of perturbations from WL tomographic surveys.

Further useful constraints on the matter density can be derived from measurements of the fraction $f_{\text{gas}}$ of x-ray emitting gas to total mass in galaxy clusters. This ratio is a good indicator of the overall baryon fraction in the universe, and depends on the distance to the cluster, offering a probe of the expansion history. The ratio of the gas (baryons) to total mass is found to be about 11%, which yields a cosmic baryon fraction of about 15% [45, 46], after estimated corrections. Combining these measurements with the baryon density inferred very precisely from BBN, one can thus derive an estimate of the matter density parameter.

The large scale plateau of the CMB power spectrum offers a useful probe of the expansion history. In a matter dominated universe, when the largest fluctuations enter the horizon, the gravitational potentials are constant, and therefore do not imprint any additional anisotropy on the photons that traverse them. Indeed, a photon entering a potential well gains energy, while a photon climbing out of a potential loses energy, as represented in figure 4; if the potential is constant in time, there is no overall change in the energy of the photon. However, if matter is not dominant and the late universe is undergoing a phase of accelerated expansion, then the gravitational potentials are expected to evolve. As a consequence, photons traveling through a shallowing (deepening) potential well will be blueshifted (redshifted), as they gain more energy entering the well than the energy they lose climbing out. This is known as the integrated Sachs–Wolfe (ISW) effect. However, this is a small contribution to the overall CMB power spectrum on the largest scales, approximately 10% of the total signal, and therefore is very hard to extract. One proceeds by cross-correlating the CMB power spectrum with the galaxy one, as the ISW signal is expected to be strongly correlated with the distribution of matter around us. At present, the ISW effect has been measured out to redshift $z \sim 1.5$ [47, 48] at a 4σ significance level, and constitutes a promising probe of dark energy that is likely to play an important role in distinguishing among candidate models of cosmic acceleration, as we will
In a redshift survey, this length scale can be measured both along and across the line of sight, allowing the observer to reconstruct the Hubble parameter and the angular diameter distance as a function of redshift, respectively, \( H(z) \) and \( d_A(z) \). The BAO have been detected in the SDSS data [38] at a redshift of approximately \( z \sim 0.35 \) and in a combination of the 2 degree Field Galaxy Redshift Survey (2dFGRS) and SDSS data [58].

From the recent WMAP 5-year results [36], the combined constraints from BAO, CMB and SNeIa on the density parameters are \( \Omega_M^0 = 0.0456 \pm 0.0015 \), \( \Omega_M^0 h^2 = 0.1358 \pm 0.0037 \) and \( \Omega_\Lambda^0 = 0.725 \pm 0.015 \) to 68% C.L.

4. The cosmological constant

The natural tendency of a universe filled with regular matter is to decelerate, as the mutual gravitational attraction fights against any initial expansion. This was clear even to Einstein, for whom this fact was an obstacle to obtaining a static universe, rather than an accelerating one. Indeed, as Einstein realized almost immediately, there is an extension of the field equations of general relativity, consistent with general covariance and second order field equations, that can battle contraction and ultimately lead to acceleration—the cosmological constant.

The cosmological constant \( \Lambda \) is the coefficient of the term in the field equations that is proportional to the metric. It appears in the action for gravity as

\[
S = \frac{M_p^2}{2} \int d^4x \sqrt{-g} (R - 2\Lambda),
\]

and in the resulting Einstein field equations as

\[
R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{1}{M_p^2} T_{\mu\nu},
\]

where \( T_{\mu\nu} \) is the energy–momentum tensor for mass–energy sources in the universe.

As we have mentioned earlier, for cosmological purposes, we may treat the cosmological constant as a component of the energy density in the universe, since adding a cosmological constant to the Einstein equation is equivalent to including an energy–momentum tensor of a perfect fluid with equation-of-state parameter \( w_\Lambda = -1 \).

As we have also seen, the cosmological constant is a prime candidate for the agent causing cosmic acceleration and, subject to ever more accurate data, provides a rather good fit to all known cosmological observations. Furthermore, in classical general relativity the cosmological constant is merely a constant of nature, with dimensions of \([\text{length}]^{-2}\), in the same sense that the Planck mass is nothing but a dimensionful constant of the theory. In this setting, it is meaningless to enquire about the value of the constant, rather it is just something that we determine through experiment. However, when we begin to think about the quantum nature of matter it becomes clear that the cosmological constant poses a deep puzzle for fundamental physics.

Particle physics, described by quantum field theory, is a remarkably accurate description of nature, and one of
Figure 4. Schematic representation of the ISW on CMB photons. On the left side of the figure, a photon travels through a constant potential, gaining and losing the same amount of energy while passing through the potential well. In this case, the outgoing wavelength corresponds to the incoming one, and there is no net gain (or loss) of energy. On the right side of the figure, a photon travels through a shallowing potential. As the potential well from which the photon exits is shallower than the one it entered, there is an overall gain in energy and the outgoing wavelength is shorter than the incoming one.

Figure 5. The large scale redshift-space correlation function of the SDSS LRG (luminous red galaxies) sample, from [38]. The bottom line shows a pure CDM model ($\Omega_m h^2 = 0.105$) which lacks the acoustic peak. The other lines represent models with a baryon fractional density of $\Omega_b h^2 = 0.024$ and a CDM density of, from top to bottom respectively, $\Omega_m h^2 = 0.12, 0.13, 0.14$. The bump at $100 h^{-1}$ Mpc corresponds to the acoustic peak. Details about the data set and error bars can be found in [38].

the best tested theories in all of science. Within this theory, we can attempt to calculate the expected contribution to the cosmological constant from quantum fluctuations in the vacuum. Taking the Fourier transform of a free (noninteracting) quantum field, each mode of fixed wavelength is a simple harmonic oscillator, which we know has zero-point (ground state) energy $E_0 = \frac{1}{2} \hbar \omega$, where $\omega$ is the frequency of the oscillator. The net energy of these modes is then given by a divergent integral. Thus, our most naive expectation of the vacuum energy calculated within field theory is that it should be infinite! Perhaps, however, we have been too hasty in performing this calculation. After all, we have included modes of arbitrarily small wavelength, and this presumes that we can trust our descriptions of matter and gravity down to such tiny scales. Without a well-understood and calculable quantum theory of gravity, this is surely not reasonable. A better estimate of the vacuum energy can therefore be obtained by introducing a cutoff energy, above which we ignore the contributions of any modes, assuming that this will be justified by a more complete theory. This approach corresponds to assuming that effective field theory correctly describes the appropriate limit of quantum gravity. In that sense, if a parameter has dimensions [mass]$^n$, then we expect the corresponding mass parameter to be driven up to the scale at which the effective description breaks down. Thus, if classical general relativity is valid up to the Planck scale, we expect the vacuum energy to be given by

$$\rho_{\text{vac}} \sim M_P^n. \quad (31)$$

Comparing this value with the value $\rho_{\text{vac}}^{\text{obs}}$ required to explain cosmic acceleration, we obtain

$$\rho_{\text{vac}}^{\text{obs}} \sim 10^{-120} \rho_{\text{vac}}^{\text{theory}}, \quad (32)$$

which is embarrassing, to say the least. It is perhaps more sensible to express the vacuum energy in terms of a mass scale as $\rho_{\text{vac}} = M_{\text{vac}}^4$, in terms of which the value required to explain our observations is $M_{\text{vac}}^{\text{obs}} \sim 10^{-3}$ eV, satisfying,

$$M_{\text{vac}}^{\text{obs}} \sim 10^{-30} M_{\text{vac}}^{\text{theory}}. \quad (33)$$
Nevertheless, this discrepancy of 30 orders of magnitude in energy is overwhelming, and is what is meant by the cosmological constant problem.

One may add to this problem the following puzzling observation. The ratio of the vacuum and matter densities changes as the universe expands according to

$$\frac{\Omega_\Lambda}{\Omega_M} = \frac{\rho_\Lambda}{\rho_M} \propto a^3. \quad (34)$$

Thus, only during a brief epoch of cosmic history is it possible for observers to witness the transition from matter domination to $\Lambda$ domination, during which $\Omega_\Lambda$ and $\Omega_M$ are of the same order of magnitude. This is known as the coincidence problem.

The issue of reliably calculating the cosmological constant, and finding a framework in which that calculation leads to a result dramatically different to the expected one has proven remarkably resistant to theoretical attack. It is fair to say that there are not currently any especially promising approaches. Nevertheless, there are two theoretical lines of research that are worth mentioning in this context.

The first is supersymmetry (SUSY). Supersymmetry is a spacetime symmetry that relates particles of different spins (there exists a matching fermionic degree of freedom for every bosonic one, and vice-versa). Since bosons and fermions make opposite sign contributions to the vacuum energy, in a truly supersymmetric theory the net vacuum energy should be zero, with cancellations occurring between, for example, the spin-1/2 electron and its spin-0 partner, the ‘selectron’ (with the same mass and charge as the electron).

It is, however, entirely obvious from experiment and observation, that the world is not supersymmetric—we would certainly know if there existed a scalar particle with the same mass and charge as the electron.

A second theoretical approach to the cosmological constant problem and cosmic acceleration problems is to consider the logical possibility that the vacuum energy is a feature of our local environment—the anthropic principle. This is an idea that may make logical sense if there exist many different possible values of physical parameters (and the cosmological constant in particular), and a mechanism in the early universe for sampling these possibilities, either over and over again in time, or in different regions of space.

Clearly, only a certain range of values of the cosmological constant is consistent with the development of life like us (and there are a number of assumptions here), since if $|\Lambda|$ is significantly larger than that required to yield the observed cosmic acceleration, then the universe would have recollapsed (if $\rho_{\text{vac}}$ were negative), or expanded so quickly that galaxies could not form (if $\rho_{\text{vac}}$ were positive). Given this, some authors have argued [59–64] that, under an assumption about the distribution of possible values of $\rho_{\text{vac}}$, we should expect a value of $\Lambda$ close to that we seem to have observed.

The recent resurgence of interest in this possibility stems from the realization that within string theory there seems to exist a large number (more than $10^{100}$) of possible vacuum states [65–70]. These vacua arise from different compactification possibilities, and the roles played in them by various brane and gauge field configurations. The low energy effective theories constructed around these metastable vacua should have different local values of the vacuum energy (and other physical parameters). An approach to addressing the cosmological constant problem then begins by asserting that it is highly likely that there must exist (and, eventually, hopefully identifying) one or more vacua with a cosmological constant (and perhaps other parameters) that take the values one might think it is reasonable given the number quoted above.

Thus, if there exist possible vacua with a cosmological constant of the right size to explain cosmic acceleration, parts of the universe may cease to inflate and drop out into different final vacuum states. In this model many different domains may be inflated separately, and then small parts of the universe may cease to inflate and drop out into different final vacuum states. Thus, if there exist possible vacua with a cosmological constant of the right size to explain cosmic acceleration, parts of the universe that populate those vacua will cease inflating, reheat the universe, undergo standard cosmic evolution, and ultimately begin accelerating again at late times.

This is a brief description of how the anthropic principle may be realized in the context of the string theory landscape. It is, however, too early to know whether it can be realistically
implemented, and even harder to say whether one will ever be able to falsify, and hence make science out of, such an approach.

Nevertheless, this is the best developed argument regarding the possible small value of \( \Lambda \), and perhaps for the value needed to explain cosmic acceleration.

5. Dark energy

As we discussed in section 3, while the data clearly point to an accelerating universe, they do not single out a cosmological constant as the clear source of this acceleration. Let us assume, therefore, that the cosmological constant is zero, or at least so small as to be unable to account for the current period of accelerated expansion, and explore what other sources of mass-energy in the universe may be responsible (see [72] for a discussion of logical possibilities, and [1] for a detailed review). Cosmologists have asked this question before, when confronted with the inflationary paradigm for solving the problems of early universe cosmology, and so we do not come to the issue completely cold.

The most faithful servant of the theoretical cosmologist is the trusty scalar field. It has served us dutifully not only through inflation, but also baryogenesis models, topological defect formation and some dark matter models, and so it is only natural that we turn to it now, in our epoch of need. Consider a real scalar field \( \phi \), with potential \( V(\phi) \), with Lagrangian density

\[
\mathcal{L} = \frac{1}{2} g^{\mu \nu} \partial_\mu \phi \partial_\nu \phi - V(\phi),
\]

for which the energy–momentum tensor is given by

\[
T_{\mu \nu} = \partial_\mu \phi \partial_\nu \phi - g_{\mu \nu} [\frac{1}{2} g^{ho \sigma} \partial_\rho \phi \partial_\sigma \phi + V(\phi)].
\]  

If the scalar field is homogeneously distributed in space, as we shall assume, then (39) takes the perfect fluid form, with energy density and pressure given by

\[
\rho = \frac{1}{2} \dot{\phi}^2 + V(\phi),
\]

\[
p = \frac{1}{2} \dot{\phi}^2 - V(\phi),
\]

and we may then associate with it an instantaneous equation-of-state parameter

\[
w_\phi = \frac{\dot{\phi}^2 - 2V(\phi)}{\dot{\phi}^2 + 2V(\phi)}.
\]

As we learned when studying inflation, in the slow-roll regime, in which the scalar field equation of motion

\[
\ddot{\phi} + 3H \dot{\phi} + V_\phi = 0,
\]

(where \( V_\phi \equiv \partial V/\partial \phi \)) becomes \( \dot{\phi} \simeq -\frac{1}{3M} V_\phi \), the kinetic energy contributions to the pressure and the energy density become subdominant, and the equation-of-state parameter can become sufficiently negative to drive cosmic acceleration.

Unlike inflation, however, we need not require this period of acceleration to end, or the universe to be reheated after it. Rather, our requirements are that the energy scale associated with the potential be such that acceleration is happening at the current epoch, and that the energy density in the scalar field comprise the measured value of \( \Omega_{\Lambda 1} \). Two of the most studied examples of this are inverse power law potentials

\[
V(\phi) = M^4 \left( \frac{\phi}{M_p} \right)^{-n},
\]

and exponential potentials

\[
V(\phi) = M^4 \exp(-\alpha \frac{\phi}{M_p}),
\]

where \( M \) is some mass scale and \( \alpha \) is a dimensionless parameter.

In the latter case, it is well known that in the limit that the scalar field becomes so dominant that one can neglect contributions from other sources (such as matter), the scalar field and Friedmann equations can be solved exactly, to yield

\[
\phi(t) = \frac{2}{\alpha} \ln \left( \frac{M^2 \alpha^2}{t} \right),
\]

\[
a(t) = a_0 t^{2/(\alpha^2)}.
\]

Therefore, acceleration \((\ddot{a} > 0)\) is achieved if \( \alpha^2 < 2 \).

Including matter changes this conclusion of course. One effect is that the actual functional form of the accelerating solution is altered from the simple power law one. However, a more interesting modification is the possible existence of tracker solutions (in which the energy density in the scalar field approaches a fixed fraction of that in the other components, rather insensitively to the initial conditions), which provides a way to ameliorate the coincidence problem of cosmic acceleration.

While these simple models provide a basic framework in which to obtain cosmic acceleration from a dynamical component, they are not without problems. Perhaps the largest issue is the extreme fine-tuning imposed for the same reasons as we described when discussing the cosmological constant. It is one thing to write down a potential, but quite another to provide a reason for the energy scales involved, and an argument why they should remain stable in any sensible particle physics implementation of the idea.

To understand the fine-tuning, consider the mass of the quintessence field by expanding the potential about a point as \( V(\phi) = (1/2) m_\phi^2 \phi^2 \). As we discussed in the previous section, to match observations, the energy density in whatever component is responsible for cosmic acceleration must be \(~ (10^{-12} \text{ GeV})^4\). Assuming slow-roll, and that \( \langle \phi \rangle \sim M_p \), we may then equate these two to obtain

\[
m_\phi \sim 10^{-33} \text{ eV}.
\]

Obtaining this kind of value from a particle physics model is an extreme challenge, consisting of two parts. First, it is not a simple matter to construct a model in which such low energy scales arise in the first place. Second, we face the issue confronting all scalar fields—how are we to protect an unnaturally small choice of mass scale against quadratic
divergences when we compute quantum corrections to the couplings?

As with inflation, one possibility is to choose a quintessence model in which a small mass is protected as a small parameter describing deviations from a point of enhanced symmetry in model space. This can be implemented [73] by making the quintessence field a pseudo-Nambu–Goldstone boson (PNGB) of a spontaneously broken shift symmetry. In the absence of soft breaking, the potential for such a field is a precisely flat function of an angular coordinate. Thus, if the potential is lifted by a small mass term the exact symmetry keeping the potential flat is only softly broken, and the potential takes the form

$$V(\phi) = M^4 \left[ 1 + \cos \left( \frac{\phi}{f} \right) \right],$$

(45)

where $M$ is the overall mass scale and $f \sim M_p$ measures the range of variation of $\phi$. Expanding this potential around $\phi = 0$ yields a field with a small mass (for appropriately small $M$), which is protected from quadratic divergences since the exact $U(1)$ axion symmetry is restored in the $f \to \infty$ limit.

We shall not dwell on the simplest quintessence models in this paper, since space is limited and our intention is to provide some idea as to more recent work in the field. The model we have presented above admits a number of generalizations which have their own interesting features and signatures.

5.1. Phantom models

As we have described earlier, observations allow for the possibility that the equation-of-state parameter for dark energy obeys $w < -1$. However, thus far the models we have constructed do not venture into this parameter region. There is a good reason for this: in general relativity, it is conventional to restrict the possible energy–momentum tensors by imposing energy conditions in order to prevent instabilities of the vacuum or the propagation of energy outside the light cone. Standard examples, where in parentheses we have evaluated the condition for a perfect fluid source, are

- The weak energy condition: $T_{\mu\nu}t^\mu t^\nu \geq 0 \forall$ timelike $t^\mu$ ($\rho \geq 0$ and $\rho + p \geq 0$).
- The null energy condition: $T_{\mu\nu}\ell^\mu \ell^\nu \geq 0 \forall$ null $\ell^\mu$ ($\rho \geq 0$).
- The dominant energy condition: $T_{\mu\nu}t^\mu t^\nu \geq 0 \forall$ timelike $t^\mu$ and $T_{\mu\nu}T^{\mu\nu}t^\nu < 0$ ($\rho \geq |p|$).
- The null dominant energy condition: $T_{\mu\nu}t^\mu t^\nu \geq 0 \forall$ null $t^\nu$ and $T^{\mu\nu}\ell_\mu$ is a non-spacelike vector (as for the dominant energy condition, except that negative densities are allowed if $p = -\rho$).
- The strong energy condition: $T_{\mu\nu}t^\mu t^\nu \geq \frac{1}{2}T_{\kappa\lambda}^\kappa t^\kappa t^\lambda \forall$ timelike $t^\mu$ ($\rho + p \geq 0$ and $\rho + 3p \geq 0$).

Since we are interested in an energy condition that implies stability, while allowing for the possibility of a cosmological constant, Garnavich [74] has advocated the use of the null dominant energy condition, boiling down to demanding that dynamical fields obey the dominant energy condition, while making an exception for the cosmological constant.

Nevertheless, given what is allowed by the data, we may ask what happens if we allow for dark energy that violates the null dominant energy condition; i.e. obeys $w < -1$.

The possibility of such phantom components has been suggested by many authors [75–85], and the simplest example [75] is provided by a scalar field $\chi$ with negative kinetic and gradient energy

$$\rho_\chi = -\frac{1}{2} \chi^2 - \frac{1}{2} (\nabla \chi)^2 + V(\chi).$$

(46)

The associated equation-of-state parameter is

$$w = \frac{p_\chi}{\rho_\chi} = -\frac{V(\chi) + \frac{1}{2} [\chi^2 + a^{-2} (\nabla \chi)^2]}{V(\chi) - \frac{1}{2} [\chi^2 + a^{-2} (\nabla \chi)^2]},$$

(47)

obeying $w \leq -1$.

While we will not describe the behavior of phantom fields in any detail, it is worth pointing out that, classically, the cosmology of such a system is relatively simple. For example, if the universe contains only dust and phantom matter. Then, if matter domination ceases at cosmological time $t_m$, then the solution for the scale factor is

$$a(t) = a(t_m) \left[ -w + (1 + w) \left( \frac{t}{t_m} \right)^{2/(3+w)} \right],$$

(48)

from which it is easy to see that phantom matter eventually comes to dominate the universe. Since the Ricci scalar is given by

$$R = \frac{4(1 - 3w)}{3(1+w)^2} \left[ 1 - \left( \frac{w}{1+w} \right)^{3/(1+w)} \right]^{-2},$$

(49)

there is a future curvature singularity at $t = w t_m/(1+w)$—the so-called big rip.

It is, however, simple to construct models of phantom energy in which a future singularity is avoided. With a suitable choice of potential, we may construct a simple scalar field model that exhibits a period of time in which the expansion proceeds with $w < -1$ and yet settles back to $w \geq -1$ at even later times, thus sidestepping the predictions of $w = \text{constant}$, and avoiding the big rip.

However, the background cosmology aside, it is relatively straightforward to see that fluctuations in a phantom field have a negative energy, and that it may therefore be possible for the vacuum to decay into a collection of positive-energy and negative-energy particles. Thus, if this instability is present on a timescale shorter than the age of the universe, then such a phantom component cannot be a viable candidate for dark energy. Such a question has been studied in some detail by several authors [86, 87]. For example, in [86], the authors considered a specific toy model in which the null dominant energy condition is violated, but in which the cosmology is well behaved and the theory is stable to small perturbations. In many ways, this model has been given the maximum opportunity to avoid a problem with instabilities. Nevertheless, what is relevant is a field theory calculation of the decay rate of phantom particles into gravitons. Clearly, this decay rate would be infinite if the phantom theory was fundamental, valid.
up to arbitrarily high momenta, and would render the theory useless as a dark energy candidate. Considering instead the phantom theory to be an effective theory valid below a scale \( \Lambda \), and including in the Lagrangian operators of all possible dimensions, suppressed by suitable powers of the cutoff scale, it was found that such higher order operators, even though they may be very high order, can lead to unacceptably short lifetimes for phantom particles unless the cutoff scale is less than 100 MeV. This is, of course, difficult to reconcile with current accelerator experiments.

For this, and other reasons, it has proven extremely difficult to construct sensible models for which an equation of state \( w < -1 \) is realized within general relativity with minimally coupled sources, although some progress has been made with ghost condensate models [88]. It is, however, important to emphasize that well-behaved energy–momentum sources can, nevertheless, yield an effective equation-of-state parameter that is less than \(-1\) if the dark energy is non-minimally coupled [82], or if the underlying gravitational theory is not GR [89–91].

5.2. K-essence

In addition to considering the behavior of general potentials for scalar fields, it is also possible to think about the effects of modifying the kinetic term, yielding a class of models called K-essence [92–96].

To simplify notation in doing this, let us define \( X = -\frac{1}{2}g^{\mu\nu}\partial_{\mu}\phi\partial_{\nu}\phi \). Thus, a general action for a scalar field takes the form

\[
S = \int d^4x \sqrt{-g} L(\phi, X),
\]

with \( L \) an unspecified function, and the canonical scalar field action corresponding to the special case \( L(\phi, X) = X - V(\phi) \). While not completely general, an interesting subset of these actions are those with Lagrangian of the form

\[
L(\phi, X) = K(X) - V(\phi),
\]

with \( K(X) \) a positive semi-definite kinetic function, to be specified.

If we are interested in background cosmological solutions, with the FRW ansatz for the metric, we must assume a homogeneous form for \( \phi \), yielding \( X = \frac{1}{2}\dot{\phi}^2 \). The energy–momentum tensor for this field is straightforward to calculate and yields the usual perfect fluid form with pressure \( p \) and energy density \( \rho \) given by

\[
p = L = K(X) - V(\phi),
\]

\[
\rho = [2XK_{,X}(X) - K(X)] + V(\phi),
\]

where the subscript \( , X \) denotes a partial derivative with respect to \( X \). Then defining \( w_\phi \equiv p/\rho \) one obtains

\[
w_\phi = \frac{K(X) - V(\phi)}{[2XK_{,X}(X) - K(X)] + V(\phi)}.
\]

Requiring that the energy density of the theory satisfy \( \rho > 0 \) yields

\[
g(X) = 2Xg_{,X}(X) < \frac{V(\phi)}{f(\phi)},
\]

and a further necessary condition that the theory be stable is that the speed of sound of \( \phi \) be positive [96]. This yields

\[
c_s^2 \equiv \frac{\partial p}{\partial \rho} = \frac{p_{,X}}{\rho_{,X}} = \frac{K_{,X}(X) + 2XK_{,XX}(X)}{K_{,X}(X) + 2XK_{,X}(X)} > 0.
\]

Solving the equations of motion in concert with the resulting Einstein equation, it has been shown [92–96] that tracking behavior can be obtained. This means that for a wide range of initial conditions, the energy density of the field \( \phi \) naturally evolves so as to track the energy density in matter, providing some insight into why dark energy domination began only recently in cosmic history. However, as in all rolling scalar models, some fine-tuning remains, since one must ensure the right amount of dark energy density today. Further, just as with minimal models, it is possible [97] to obtain phantom-like behavior, but once again, this comes at the risk of serious instabilities.

Most interestingly though, the varying speed of sound (compared with \( c_s^2 = 1 \) for canonical models) opens the window to possible observational signatures of these models (which can otherwise be difficult to distinguish from canonical ones [98]) in observations of LSS in the universe.

5.3. Coupled dark energy models

Our cosmological models seem to require both dark matter and dark energy. A logical possibility is that these dark sectors interact with each other or with the normal matter [99–109] and, indeed, a number of models have been proposed that exploit this possibility to address, for example, the coincidence problem.

There are, of course, many ways in which couplings may arise, and it is not possible to provide a comprehensive treatment here. Instead, let us mention an interesting subclass of models, defined by the action

\[
S[g_{\mu\nu}, \phi, \Psi_j] = \int d^4x \sqrt{-g} \left[ \frac{1}{2} M^2_p R - \frac{1}{2} \bar{\nabla}^2 \phi^2 - V(\phi) \right] + \sum_j S_j[e^{2\alpha_j(\phi)}g_{\mu\nu}, \Psi_j].
\]

where \( g_{\mu\nu} \) is the Einstein frame metric, \( \phi \) is a scalar field which acts as dark energy, and \( \Psi_j \) are the matter fields. The functions \( \alpha_j(\phi) \) are couplings to the \( j \)th matter sector.

This general action encapsulates many models studied in the literature [99, 110–116]. The field equations are

\[
M^2_p G_{\mu\nu} = \nabla_\mu \phi \nabla_\nu \phi - \frac{1}{2} g_{\mu\nu} (\bar{\nabla}^2 \phi^2) - V(\phi) g_{\mu\nu} + \sum_j e^{2\alpha_j(\phi)}[\bar{\rho}_j + \bar{p}_j] u_{j,\mu} u_{j,\nu} + \bar{p}_j g_{\mu\nu}],
\]

where we have treated the matter field(s) in the \( j \)th sector as a fluid with density \( \bar{\rho}_j \) and pressure \( \bar{p}_j \) as measured in the frame \( e^{2\alpha_i(\phi)}g_{\mu\nu} \), and with 4-velocity \( u_{j,\mu} \) normalized according to \( g^{\mu\nu} u_{j,\mu} u_{j,\nu} = -1 \).

An interesting subset of these models is that in which only the dark matter is coupled to the dark energy, with the baryonic
sector uncoupled. This corresponds to setting $\alpha_0(\phi) = 0$ and $\alpha_c(\phi) = \alpha(\phi)$. Defining $\rho_j = e^{\alpha_j} \dot{\rho}_j$ gives

$$M_j^2 G_{\mu\nu} = \nabla_\mu \phi \nabla_\nu \phi - \frac{1}{2} g_{\mu\nu} (\nabla \phi)^2 - V(\phi) g_{\mu\nu}$$

$$+ e^{\alpha(\phi)} \rho_c H_0 u_\mu u_\nu$$

(60)

and $\nabla_\mu V(\phi) = 0$, where we have defined an effective potential by $V_{\text{eff}}(\phi) = V(\phi) + e^{\alpha(\phi)} \rho_c$. The fluid obeys $\nabla_\mu (\rho_c u_\mu^C) = 0$ and $u_\mu^C \nabla_\nu u_\nu^C = -(\epsilon_{\mu\nu} + u_\mu^C u_\nu^C) \nabla_\mu \alpha$.

There are a number of interesting effects in coupled models, including a surprising problem—the adiabatic instability [117–121]. In the regime in which the scalar field adiabatically tracks the minimum of the effective potential, a careful study of linear perturbation theory implies that, within a certain regime of parameters, exponential growth can occur.

In general, one can constrain such coupled models using a combination of cosmological datasets. In [122] the datasets used included the measurements of the CMB temperature and polarization power spectrum from the WMAP 5-year data release [39, 123], the ‘union’ set of supernovae compiled by the Supernovae Cosmology Project [37], the matter power spectrum of Luminous Red Galaxies (LRG) as measured [41, 124] by the Sloan Digital Sky Survey (SDSS) survey, and a Gaussian prior on the Hubble constant today, $H_0 = 72 \pm 8$, using the HST measurements [125]. The resulting constraint plots are shown in figure 6, expressed in terms of the dimensionless coupling variable $\Omega_m$, defined by

$$C(\phi) = \frac{M_\phi \alpha_c \phi}{\beta}$$

(61)

with $\beta = \sqrt{2/3}$.

6. Modified gravity

As we have seen in the previous sections, within GR, a period of accelerated expansion can only be due to a dominant component of the energy density with a negative equation of state, the most common example being a static, or slowly rolling, scalar field (commonly referred to as dark energy or quintessence models). There is, of course, a natural alternative to be considered, namely the possibility that the dynamical rules determining how the geometry of spacetime responds to regular matter sources may differ from those given by GR. In other words, we may consider modifying GR in the low-curvature regime, such as to admit self-accelerating solutions in the presence of negligible matter [110–145].

While general relativity has been tested to high precision in laboratory, in the solar system and astrophysically, for example with binary pulsars [146], there are far fewer tests on cosmological scales. It is therefore natural to explore whether the laws of gravity could be modified on the larger scales and whether such modifications could be responsible for the late-time acceleration. Obviously though, just as with dark energy approaches, models of modified gravity need to explain why, or assume a reason that the vacuum energy would be zero. With respect to the cosmological constant approach, however, dark energy and modified gravity models have the advantage of adding dynamics to the theory, allowing for the possibility of alleviating the coincidence problem and increasing the experimental testability of the models.

When modifying gravity, one needs to ensure that GR is recovered at early times and on short scales, in regimes in which it has been well tested. Moreover, a viable model must agree with the precision measurements of the expansion history of the universe, of the CMB and of the growth of structure. As we shall see though, on theoretical grounds there are obstacles to constructing infrared modifications of GR that will escape detection on smaller scales. In a weak-field expansion around flat spacetime, GR describes the propagation of massless spin-2 gravitons coupled to the energy–momentum tensor $T_{\mu\nu}$, including that of the gravitons themselves. But such a theory is essentially unique; it has long been known that we can start with a field theory describing a transverse-traceless symmetric two-index tensor propagating in Minkowski space, couple it to the energy–momentum tensor of itself and other fields, and demonstrate iteratively that the background metric disappears, leaving instead a curved metric obeying Einstein’s equation.

It is therefore generally believed that infrared modifications of GR will necessarily involve the introduction of new degrees of freedom. This can be a good thing, in that the dynamics of these new degrees of freedom may be precisely
what one needs to drive the accelerated expansion of the universe. However, there are often problems associated with the extra dynamics and the extra forces mediated by these degrees of freedom. The problems may be of several different kinds. First, there is the possibility that, along with the desired deviations from GR on cosmological scales, one may also find deviations on smaller scales; e.g. in the solar system, where GR is well tested. Second is the possibility that the newly activated degrees of freedom may be badly behaved in one way or another; either having the wrong sign kinetic terms (ghosts), and hence being unstable, or leading to superluminal propagation. More generally, theories that contain higher than first order derivatives in the Lagrangian risk being plagued by the instability described by the Ostrogradski theorem [147, 148], i.e. the Hamiltonian of the system may be unbounded from below.

The previous arguments put surprisingly restrictive constraints on models of modified gravity, and this has motivated some authors to explore models that modify GR without introducing new degrees of freedom [129, 149, 150].

In this section we will give an overview of some of the main candidate models of modified gravity, focusing on the general challenges that such theories face and the peculiar signatures to which they can lead in cosmology.

There are several ways in which one can imagine of modifying the laws of gravity on the largest scales. A first approach could perhaps focus only on the phenomenology of cosmology, in which case one postulates specific modifications of the equations describing the dynamics of the universe such as to achieve the acceleration. An example is the Cardassian model [128], in which the Friedmann equation is modified such that the dependence of the Hubble parameter on matter sources is different than the standard one. Although models such as this can yield cosmic acceleration, it is desirable to begin with a theory described by a generally covariant action, from which we may derive not only a modified Friedmann equation, but also the solution describing any other physical setting we may wish to know about.

One broad class of modified gravity models that have been studied in detail contains those theories which involve covariant modifications of the 4-dimensional action by adding higher order curvature invariants to the Einstein–Hilbert action. The simplest examples in this class are the so-called $f(R)$ theories. It is instructive to go over some of the details of $f(R)$ models since they are representative of generalized modified theories of gravity. We will do so in the following subsection, while more details about these models can be found in the recent paper by Sotiriou and Faraoni [151].

Finally, modifications of gravity can originate from extra dimensional scenarios, in which all matter fields are confined to a four-dimensional brane embedded in a higher-dimensional bulk. The most prominent example is the Dvali–Gabadadze–Porrati (DGP) model [127], in which the 4D brane is embedded in a 5D Minkowski bulk of infinite volume, into which gravity may leak, weakening its strength. We dedicate the last subsection to this class of models and to some recent proposals that elaborate on them, e.g. degravitation [152] and cascading gravity [153, 154].

6.1. $f(R)$ theories

It is a well known fact that the Einstein–Hilbert action is non renormalizable, therefore standard GR cannot be properly quantized. However renormalizability can be cured by adding to the action higher order terms in curvature invariants [155–158]. Moreover, higher order terms are generally expected in low energy effective actions derived from string theory or grand unification theories (GUTs). In 1979 Starobinsky showed that one can achieve a de Sitter phase in the early universe by adding to the Ricci scalar $R$ in the Einstein action a function $f(R) \propto R^2$ [159, 160]. More recently, $f(R)$ models have been revisited in the context of cosmic acceleration, starting with [110, 130] and followed by many others. The underlying idea is to add functions of the Ricci scalar which become important at late times, for small values of the curvature.

Let us focus on the following action

$$S = \frac{1}{2\kappa^2} \int d^4x \sqrt{-g} \left[ R + f(R) \right] + \int d^4x \sqrt{-g} L_m[\chi_i, g_{\mu\nu}],$$

(62)

where $f(R)$ is a general function of the Ricci scalar, $R$. The matter Lagrangian, $L_m$, is minimally coupled and, therefore, the matter fields, $\chi_i$, fall along geodesics of the metric $g_{\mu\nu}$. The field equations obtained from varying the action (62) with respect to $g_{\mu\nu}$ are

$$\left(1 + f'_R\right) R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} (f(R) + f'_{\mu\nu} \square - \nabla_\mu \nabla_\nu) f_R = \kappa^2 T_{\mu\nu},$$

(63)

where we have defined $f_R \equiv \partial f/\partial R$. We take the energy–momentum tensor to be that of a perfect fluid (5). The metric $g_{\mu\nu}$ is minimally coupled to matter, hence the stress tensor and its conservation law will be the same as in standard GR.

The $f(R)$ term in the gravitational action leads to extra terms in the Einstein equations. The constraint equation of GR becomes a third order equation, and the evolution equations now become fourth order differential equations (in contrast to the second order equations of standard GR). In particular, when taking the metric to be of the flat FRW form (1), the Friedmann equation becomes

$$H^2 + \frac{f}{6} - \frac{\dot{a}}{a} f_R + H \ddot{f_R} = \frac{\kappa^2}{3} \rho,$$

(64)

and the acceleration equation is

$$\frac{\ddot{a}}{a} - f_R H^2 + \frac{f}{6} + \frac{\ddot{f_R}}{2} = -\frac{\kappa^2}{6} (\rho + 3P).$$

(65)

Cosmic acceleration in equations (64) and (65) can arise through the additional terms present for $f(R) \neq 0$. One can interpret these extra terms as the contribution from an effective fluid with an equation of state

$$w_{\text{eff}} = -\frac{1}{3} - \frac{2}{3} \left[ H^2 f_R - \frac{f}{6} - H \ddot{f_R} - \frac{\ddot{f_R}}{2} \right] / \left( -\frac{1}{2} \dot{f_R} R \right).$$

(66)

It is important to notice that equation (66) can be used as a differential equation for the function $f(R)$. Indeed, given an
expansion history, with a specified matter content, one can determine \( u_{\text{eff}} \) and then solve equation (66) for \( f(R) \). Since the resulting equation for \( f(R) \) is of second order, one has a family of \( f(R) \) models which reproduce the desired expansion history. In this sense, there is a degeneracy among models at the level of background cosmology, and, as we will discuss shortly, the growth of perturbations will play a key role in breaking this degeneracy.

There exists a complementary, and sometimes conceptually simpler, way in which to approach \( f(R) \) theories. It is possible to conformally transform the metric and bring the gravitational part of the action to the usual Einstein–Hilbert form of standard GR. The price one pays for this simplification is a non-minimal coupling between matter fields and the gravitational part of the action to the usual Einstein–Hilbert form (62). The transformation between the two frames is well defined, the classical results obtained using either description should be the same. The frame in which matter particles fall along geodesics of the metric, but in which the gravitational part of the action is modified, is referred to as the Jordan frame. The frame in which the gravitational part of the action is the same as in standard GR, while the matter may be non-minimally coupled to gravity, is called the Einstein frame.

The emergence of an additional scalar degree of freedom can also be seen directly in the Jordan frame. The role of the scalar field \( \varphi \) in which the gravitational part of the action is modified, is which the function \( f(R) \) (62). When the transformation between the two frames is well defined, the classical results obtained using either description should be the same. The frame in which matter particles fall along geodesics of the metric, but in which the gravitational part of the action is modified, is referred to as the Jordan frame. The frame in which the gravitational part of the action is the same as in standard GR, while the matter may be non-minimally coupled to gravity, is called the Einstein frame.

While \( \lambda_c \) is large at current cosmological densities, terrestrial, solar and galactic tests are not necessarily violated because the scalaron acquires a larger mass in regions of high matter density. This is essentially the Chameleon mechanism of [112, 113].

As we mentioned above, alternatively one may map the theory to the Einstein frame; following the approaches of [111, 161, 164], one can recast the gravitational action (62) into a dynamically equivalent form by applying the conformal transformation

\[
g_{\mu\nu} = e^{2\omega(x^\tau)} g_{\mu\nu},
\]

such that the function \( \omega(x^\tau) \) satisfies

\[
e^{-2\omega}(1 + f_R) = 1.
\]

With this choice of \( \omega(x) \) the action (62) transforms into an action with the usual Einstein–Hilbert form for gravity. If we now define the scalar field \( \phi = 2\omega/\beta x \), where \( \beta \equiv \sqrt{2/3} \), the resulting action becomes

\[
\begin{align*}
\hat{S} &= \frac{1}{2\kappa^2} \int d^4x \sqrt{-\hat{g}} \hat{R} \\
&\quad + \int d^4x \sqrt{-\hat{g}} \left[ -\frac{1}{2} \hat{g}^{\mu\nu} \hat{\nabla}_\mu \hat{\nabla}_\nu \phi - V(\phi) \right] \\
&\quad + \int d^4x \sqrt{-\hat{g}} \left[ e^{-\beta \phi} \mathcal{L}_m[\chi], e^{-\beta \phi} \hat{g}_{\mu\nu} \right].
\end{align*}
\]

where the potential \( V(\phi) \) is determined entirely by the original form (62) of the action and is given by

\[
V(\phi) = \frac{1}{2\kappa^2} \frac{R f_R - f}{(1 + f_R)^2},
\]

and where a tilde is used to indicate Einstein’s frame quantities.

The field equations obtained by varying the action with respect to the metric \( \tilde{g}_{\mu\nu} \) are

\[
\tilde{G}_{\mu\nu} = 8\pi G \tilde{T}_{\mu\nu} + \frac{1}{2} \tilde{\nabla}_\mu \tilde{\nabla}_\nu \phi + \frac{1}{4} (\tilde{\nabla}^\nu \tilde{\nabla}_\nu \phi \tilde{\nabla}_\mu \phi) \tilde{g}_{\mu\nu} - V(\phi) \tilde{g}_{\mu\nu}.
\]

It then follows that in order for the scalaron not to be tachyonic, (i.e. to have a positive squared mass), one must require \( f_{RR} > 0 \). Classically, \( f_{RR} > 0 \) is required to keep the expansion in the high-curvature regime stable against small perturbations [162, 163].

The scalaron mediates an attractive ‘fifth force’, which has a range determined by the Compton wavelength

\[
\lambda_c \equiv \frac{2\pi}{m_{\phi}}.
\]
Returning to the Jordan frame formulation, there is a series of conditions that must be imposed on the function \( f(R) \) in order to avoid instabilities and ensure agreement with local tests of gravity. These conditions can be summarized as follows:

(i) \( f_{RR} > 0 \) for \( R \gg f_{RR} \). Classically, this follows from requiring the existence of a stable high-curvature regime, such as the matter dominated universe [162].

(ii) \( 1 + f_R > 0 \) for all finite \( R \). The most direct interpretation of this condition is that the effective Newton constant, \( G_{\text{eff}} = G/(1 + f_R) \), is not allowed to change sign. This prevents the graviton from becoming ghost-like [165].

(iii) \( f_R < 0 \). Given tight constraints from BBN and the CMB, we must require GR to be recovered at early times, i.e. \( f(R)/R \to 0 \) and \( f_R \to 0 \) as \( R \to \infty \). Together with \( f_{RR} > 0 \), this implies that \( f_R \) must be a negative, monotonically increasing function of \( R \) that asymptotes to zero from below.

(iv) \( f_R \) must be small at recent epochs. This is not required if the only aim is to build a model of cosmic acceleration, without trying to satisfy the solar and galactic scale constraints. This condition ensures a small difference between the value of the scalaron field in the high density galactic center and the value in the outskirts of the galaxy [166]. As shown in [113], the difference between these values is effectively the potential difference that sources the attractive ‘fifth force’ acting on objects in the vicinity of the galaxy. The analysis of [166] suggests that the value of \( |f_R| \) today should not exceed \( 10^{-6} \). However, as stressed in [166], this bound assumes that in \( f(R) \) models galaxy formation proceeds similarly to that in GR. Hence, while it is certain that \( |f_R| \) must be small, any specific bound on its value today will be unreliable until galaxy formation in \( f(R) \) is studied in N-body numerical simulations—work that has only recently begun [167, 168].

Finally, it was recently shown [169, 170] that even models of \( f(R) \) that satisfy all of the above conditions might still face problems in the strong gravity regime. These problems are due to a curvature singularity \( R \to 0 \) appearing at the non-linear level, at a finite field value and energy, such that the scalar degree of freedom can easily access it in the presence of matter. As a consequence, \( f(R) \) theories appear to be incompatible with the existence of static dense compact objects like neutron stars. The problem seems to require an ultraviolet completion of \( f(R) \) theories, and indeed it was recently shown in [171, 172] that the addition of higher order curvature corrections might solve the curvature singularity problem, although an unpleasant degree of fine-tuning would be required.

### 6.2. Generalized theories of gravity

The \( f(R) \) theories studied in the previous subsection are the simplest example of generalized theories of gravity. Still, in their simplicity, they retain some of the features of more general higher order theories of gravity. Studying the cosmology of these models, we have learned how the introduction of higher order terms in curvature invariants can lead to instabilities and/or conflicts with local tests of gravity because of extra degrees of freedom that come into play. These problems are characteristic of generalized theories of gravity, and can be more severe in models that involve functions of curvature invariants other than the Ricci scalar.

There are, of course, any number of terms that one could consider, but for simplicity, we might consider those invariants of lowest mass dimension that are also parity-conserving

\[
P \equiv R_{\mu\nu} R^{\mu\nu},
\]

\[
Q \equiv R_{\alpha\beta\gamma\delta} R^{\alpha\beta\gamma\delta}.
\]

and actions of the form [142]

\[
S = \int d^4x \sqrt{-g} [f(R, P, Q)] + \int d^4x \sqrt{-g} \mathcal{L}_m,
\]

where \( f(R, P, Q) \) is a general function describing the deviations from general relativity.

It is convenient to define

\[
f_R \equiv \frac{\partial f}{\partial R}, \quad f_P \equiv \frac{\partial f}{\partial P}, \quad f_Q \equiv \frac{\partial f}{\partial Q},
\]

in terms of which the equations of motion are

\[
R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - \frac{1}{4} g_{\mu\nu} f_R R_{\alpha\beta}\nabla_\alpha \nabla_\beta f_R + 2 f_P R^{\mu\nu}_{\alpha\beta} R_{\alpha\beta} + 2 f_Q R_{\alpha\beta\gamma\delta} R^{\alpha\beta\gamma\delta} + 2 f_Q \Box f_R - \nabla_\mu \nabla_\nu f_R - 2 \nabla_\mu \nabla_\nu f_Q [f_P R^{\alpha\beta}_{\mu\alpha\beta}] + \Box (f_P R_{\mu\nu}) + g_{\mu\nu} \nabla_\alpha \nabla_\beta (f_P R^{\alpha\beta}) - 4 \nabla_\alpha \nabla_\beta [f_Q R^{\alpha\beta}_{\mu\nu}] = 8\pi G T_{\mu\nu}.
\]

It is straightforward to show that actions of the form (80) generically admit a maximally symmetric solution: \( R = a^2 \) a non-zero constant. However, an equally generic feature of such models is that this de Sitter solution is unstable (corresponding to one of the new degrees of freedom attaining a local potential maximum), and in many cases there exists another accelerating solution, analogous to those that we found for \( f(R) \) models.

What about constraints on these models? It has been shown [173] that solar system constraints, of the type we have described for \( f(R) \) models, can be evaded by large classes of these more general models. Roughly speaking, this is because the Schwarzschild solution, which governs the solar system, has vanishing \( R \) and \( P \), but non-vanishing \( Q \).

Far more serious though is the issue of ghosts and superluminal propagation. It has been shown [174, 175] that a necessary but not sufficient condition that the action be ghost-free is that \( b = -4c \), so that there are no fourth derivatives in the linearized field equations. What remained was the possibility that the second derivatives might have the wrong signs, and also might allow superluminal propagation at some time in a particular cosmological background. It has recently been shown that around a FRW background with matter, the theories are ghost-free, but contain superluminally propagating scalar or tensor modes over a wide range of parameter space [176, 177]. However, more generally it has
been argued [148], that only modifications of the action that are functions of the Ricci scalar and the Gauss–Bonnet term can avoid the instability of the Ostrogradski’s theorem. Thus, such general actions do not seem to be promising candidates for modified gravity theories explaining cosmic acceleration.

6.3. Modifying gravity without extra degrees of freedom

The challenges that higher order theories of gravity face are related to the dynamics of, and forces mediated by, the extra degree(s) of freedom introduced by the modifications. It is therefore interesting to investigate whether it is possible to achieve cosmic acceleration with models for which the constraint equations of gravity are modified while the graviton is still the only propagating degree of freedom [129, 149]. Since such models are a little outside the scope of this paper, we will not comment on them further, other than to point out that linear perturbation theory seems to show a possible conflict with observations of the formation of structure [129].

6.4. The DGP model and its derivatives

So far we have considered models of modified gravity based on a covariant four-dimensional action. In this subsection we shall review some of the proposals in the context of extra dimensions. We focus on brane-world models that modify gravity at low energies; i.e. models that can be used to achieve cosmic acceleration.

The most prominent proposal in this context is the well known DGP [127] model, the cosmology of which was described in [126]. In this model, the matter fields are constrained to lie on a 4D brane embedded in a flat 5D Minkowski bulk of infinite volume, into which the graviton can leak. In the DGP model gravity appears four-dimensional at short distances, while it is modified on distances large compared with a crossover scale $r_c$ through the evaporation of the graviton into the unseen fifth dimension.

The 5D action describing the model is

$$S_5 = -\frac{1}{16\pi} M^3 \int d^5x \sqrt{-g} \left( R - \frac{1}{16\pi} \frac{\rho}{M^3} L_m \right)$$

where $g$, $M$ and $R$ represent, respectively, the five dimensional metric, Planck mass and Ricci scalar, while $g^{(4)}$, $M_p$ and $R^{(4)}$ are their four-dimensional counterparts. Finally, $L_m$ is the 4D matter Lagrangian density. The crossover scale $r_c$ is given by the ratio of the observed 4D Planck mass to the 5D one

$$r_c = \frac{M^2}{2M^3}.$$  

This represents the distance over which metric fluctuations propagating on the brane dissipate into the bulk, weakening gravity. As a consequence, the 4D effective gravitational potential will behave as

$$\Phi \propto \begin{cases} r^{-1}, & r \ll r_c, \\ r^{-2}, & r \gg r_c. \end{cases}$$

Interestingly, significant departures from standard gravity persist down to scales smaller than $r_c$; for a mass source of Schwarzschild radius $r_s$, modifications will be relevant down to scales equal to the Vainshtein radius $r_v \approx (r_s M_p^2)^{1/3}$ [178, 179].

In order to address the phenomenon of cosmic acceleration it is necessary to choose $r_c \sim H_0^{-1}$. The Friedmann equation describing the expansion of a 4D brane with spatial curvature $k$ reads

$$H^2 + \frac{k}{a^2} - \frac{1}{r_c^2} \sqrt{H^2 + \frac{k}{a^2}} = \frac{8\pi G}{3} \rho.$$

where $\epsilon = \pm 1$ correspond to the two different branches of the model. The $\epsilon = +1$ choice corresponds to the self-accelerated branch, sDGP; this model naturally leads to the late-time acceleration and has been recently tested against SNeIa, BAO and CMB data [180]. It was found that the flat sDGP ($k = 0$) model, is a poor fit to the data and that the $\Lambda$CDM model gives a significantly better fit [89, 181–184]. Moreover, sDGP is plagued by the presence of a ghost in the gravitational background sector, which is a serious problem.

The choice $\epsilon = -1$ corresponds to the normal branch, nDGP, which is obtained via a different choice of embedding of the brane in the bulk. It requires the addition of a cosmological constant to achieve cosmic acceleration, but it has the advantage of being ghost-free. In this model, the cosmological constant is screened by 5D effects, and therefore it is the simplest example of degravitation, which we will mention shortly. However the screening effect is too small to solve the fine-tuning problem of the cosmological constant. Another interesting feature of nDGP is that it displays an effective phantom behavior [89, 181–183]. A comparison of the nDGP model with a combination of cosmological datasets [89, 181–184], found that closed models were a good fit to the data.

In [152] it was pointed out that massive-graviton theories, such as DGP, could lead to degravitation of the vacuum energy, offering a dynamical solution to the cosmological constant problem. Degravitation indeed allows for a large cosmological constant, but suppresses its backreaction by making gravity sufficiently weak on large scales. The degravitation of the vacuum energy is an intriguing solution to the fine-tuning problem of the cosmological constant, and can be achieved if Newton’s constant is promoted to a derivative operator [152]

$$G^{-1}(L^2 \Box) G_{\mu\nu} = 8\pi T_{\mu\nu}.$$  

The gravity described by the above Einstein equations will behave as a high-pass filter with characteristic scale $L$, such that $G \rightarrow G^{(0)}$ for $L^2(-\Box) \rightarrow \infty$ and $G \rightarrow 0$ for $L^2(-\Box) \rightarrow 0$. Equation (87) should be considered only as a phenomenological description of the degravitation model, the details of which can be found in [152].

Although DGP is effectively a massive-graviton theory, the degravitation phenomenon cannot be achieved in the context of standard DGP because the weakening of gravity is not sufficiently effective. A very promising way of realizing degravitation is through the so-called cascading gravity model [153, 154, 185, 186]. In this model, DGP is
extended to higher dimensions, in which our visible 3-brane is embedded within a succession of higher-dimensional branes, each with their own induced gravity terms, embedded in a flat $D$-dimensional bulk. The model is free of ghost-like instabilities (common to previous attempts at a higher-codimension DGP) and the 4D propagator is well behaved when calculated on the brane. Moreover, this model exhibits degravitation with a steep enough weakening of gravity—the gravitational force law cascades from a $1/r^2$ (4D) behavior, to $1/r^3$, (5D) to $1/r^4$, (6D) behavior, etc as one probes larger distances on the 3-brane.

These models are works in progress, but are particularly interesting examples of modified gravity models in which both cosmic acceleration and the cosmological constant problem may be addressed.

### 7. Distinguishing between models

The $\Lambda$CDM model is currently the best fit to the available cosmological data, and passes all local tests of gravity [146]. As we have already discussed, it is, however, important to explore the whole space of explanations that fit the data comparably well and have testable features. In the previous sections we saw that, in general, imposing cosmological and local viability of models of modified gravity, as well as generalized DE, requires a degree of fine-tuning. After this tuning, the predictions of these models for the expansion history of the universe are often identical to that of the $\Lambda$CDM model [152, 166, 187–190].

However, this degeneracy is typically broken at the level of cosmological structure formation; indeed, models of modified gravity that closely mimic the cosmological constant at the background level can nevertheless yield significantly different predictions for the growth of structure. In recent times it has become increasingly evident, that the LSS of the universe offers therefore a promising testing ground for GR. To this extent, it is important to explore to what degree one can use LSS to distinguish among all the candidate models of cosmic acceleration and, indeed, whether model-independent approaches are possible [191–194].

Scalar metric perturbations in the Newtonian gauge are described by two potentials, $\Psi(\vec{x}, t)$ and $\Phi(\vec{x}, t)$, defined via the perturbed FRW line element

\[ ds^2 = -(1 + 2\Psi(\vec{x}, t)) dt^2 + (1 - 2\Phi(\vec{x}, t)) d\vec{x}^2. \]

Thus, these potentials correspond, respectively, to perturbations in the time–time and space–space components of the metric tensor, representing the strength of gravity and the spatial curvature. In the $\Lambda$CDM model, the potentials are equal during the epoch of structure formation, and their time dependence is set by the same scale-independent linear growth function that describes the growth of matter density perturbations. This, it turns out, is a very peculiar feature of the $\Lambda$CDM model, and of models of dark energy with negligible shear and clustering. However, it no longer holds in theories of modified gravity or models of coupled dark energy, where one can have scale-dependent growth patterns. In these theories, the two Newtonian potentials generically differ, and their dependence on matter perturbations can be different. As a consequence, modifications of gravity generally introduce a time- and scale-dependent slip between the Newtonian potentials, as well as a time- and scale-dependent effective Newton’s constant describing the clustering of dark matter. These modifications are expected to leave distinct imprints on the LSS of the universe, which may help to break the degeneracy that characterizes models of cosmic acceleration at the background level. This explains the recent growing interest in studying the potential of current and upcoming surveys to detect and constrain any departure from GR in the growth of structure [195–202].

#### 7.1. Modified growth of structure

It is useful to review the evolution of linear scalar perturbations in general models of modified gravity or coupled dark energy, as it will allow us to describe in detail some of the recent efforts to constrain departures from GR using the growth of structure.

We focus on the evolution of linear matter and metric perturbations in a general metric theory of gravity. Assuming that the background evolution is correctly described by the flat FRW metric, we consider scalar perturbations in Newtonian gauge, as defined in (88). Representing all perturbed quantities in Fourier space, we use the standard notation for the energy–momentum tensor of the matter fields, which to first order in the perturbations, assumes the form

\[
T_{00}^0 = -\rho(1 + \delta),
\]

\[
T_{0i}^0 = -(\rho + P)v_i, \tag{89}
\]

\[
T_{ij}^0 = (P + \delta P)\delta_{ij} + \pi_{ij}, \tag{90}
\]

where $\delta \equiv \delta\rho/\rho$ is the density contrast, $v$ the velocity field, $\delta P$ the pressure perturbation and $\pi_{ij}$ denotes the traceless component of the energy–momentum tensor. Finally, we define the anisotropic stress $\sigma$ via $(\rho + P)\sigma \equiv -(k^2 k_i - \frac{1}{3}\delta)\pi_{ij}$.

In GR, the linearized Einstein equations provide two independent equations relating the metric potentials and matter perturbations, the Poisson and anisotropy equations, respectively:

\[
k^2\Phi = -\frac{\rho}{2M_p^2}\rho\Delta, \tag{91}
\]

\[
k^2(\Phi - \Psi) = \frac{3\rho}{2M_p^2}(\rho + P)\sigma, \tag{92}
\]

where $\rho\Delta \equiv \rho\delta + 3(aH/k)(\rho + P)v$ is the comoving density perturbation. In the $\Lambda$CDM and minimally coupled quintessence models, the anisotropic stress is negligible at times relevant for structure formation, and we have $\Psi = \Phi$.

In models of modified gravity, as well as in more exotic models of dark energy, the relation between the two Newtonian potentials, and between the potentials and matter perturbations, can be different [203–205]. The effect of these modifications can be encoded in two functions of time and space, a rescaling...
of the Newton’s constant $\mu(a, k)$ and a gravitational slip $\gamma(a, k)$, defined via
\begin{equation}
\kappa^2 \Psi \equiv -\frac{a^2}{2M_p^2} \mu(a, k) \rho \Delta, \quad (93)
\end{equation}
where $\mu(a, k)$ and $\gamma(a, k)$ are time- and scale-dependent functions. Note that $\mu$ is defined via the Poisson equation (93) written in terms of $\Phi$, the perturbation to the time–time component of the metric. This choice is natural, as it is $\Psi$ that enters the evolution equation for CDM density perturbations on sub-horizon scales:
\begin{equation}
\ddot{\delta} + 2H \dot{\delta} + \frac{k^2}{a^2} \Psi = 0. \quad (95)
\end{equation}
With the set of equations (93), (94) and (95), completed by an expression for $\mu$ and $\gamma$, one can evolve the linear perturbations forward in time and extract predictions for the growth of structure.

7.1.1. An example. It is instructive to look at the growth of structure in a specific model of modified gravity. For this purpose, we will briefly review some results for the evolution of linear perturbations in $f(R)$ theories. The growth of structure in these models has been studied in [188, 205, 206], where it was found that the perturbations grow with a characteristic scale-dependent pattern. The Compton wavelength of the scalaron (70) introduces a scale which separates two regimes of sub-horizon gravitational dynamics during which gravity behaves differently. On scales $\lambda \gg \lambda_c$, the scalaron is massive and the ‘fifth force’ it mediates is exponentially suppressed; thus deviations from GR are negligible. However, on scales inside the Compton radius, the scalaron is light and deviations are significant. The relations between $\Phi$ and $\Psi$, and the relation between them and the matter density contrast, will be different below the Compton scale and that affects the growth rate of structure. Indeed, on sub-horizon scales ($k \gg a H$), as an effect of the modifications to the Einstein–Hilbert action, the Poisson and anisotropy equations read
\begin{equation}
k^2 \frac{1}{a^2} (1 + f_R) \Phi = \frac{a^2}{2M_p^2} \rho \Delta \approx \frac{1}{a^2} R, \quad (96)
k^2 \frac{1}{a^2} (1 + f_R) \Phi \equiv \frac{1}{a^2} R,
\end{equation}
(1 + f_R)(\Phi - \Psi) = f_{RR} \delta R, \quad (97)
where we have neglected any shear from matter fields. From the sub-horizon limit of equations (96), (97), we can easily read off the resulting rescaling of the Newton’s constant and the gravitational slip as
\begin{align}
\mu(a, k) &= \frac{a^2}{a^2} (1 + f_R) + 4f_{RR} k^2 \equiv \frac{\lambda^2 + \frac{1}{2} \lambda_c^2}{\frac{3}{2} \lambda^2 + \frac{3}{2} \lambda_c^2}, \quad (98)
\gamma(a, k) &= \frac{a^2}{a^2} (1 + f_R) + 2f_{RR} k^2 \equiv \frac{\lambda^2 + 2\lambda_c^2}{3\lambda^2 + 4\lambda_c^2}, \quad (99)
\end{align}
where we have neglected a prefactor $(1 + f_R)^{-1}$ multiplying the expression for $\mu$, since it is very close to unity in models that satisfy local tests of gravity [166]. In particular, on scales below $\lambda_c$, the modifications introduce a slip between the Newtonian potentials, leading to the low-curvature regime $\Psi \sim 2\Phi$. The rate of growth of structure depends on the balance between the fifth force and the background acceleration. For modes that cross below $\lambda_c$ during matter domination, the effect of the modifications is maximized as the potentials grow in the absence of background acceleration. When, however, the background starts accelerating, the potentials begin to decay but at a lesser rate than in the $\Lambda$CDM model. Therefore, a characteristic signature of $f(R)$ theories would be a non-zero ISW effect during the matter era [188, 205]. The ISW constitutes just a small fraction of the overall CMB power on the larger scales, where cosmic variance dominates, and therefore it is very hard to extract. The way to proceed is to cross-correlate the CMB with galaxies, since one expects a correlation between the distribution of galaxies (i.e. DM potential wells) and the ISW signal [48, 207, 208].

In $f(R)$ theories such a correlation would be negative, since the potentials would be growing. In figure 7 we show such cross-correlation for an $f(R)$ model, while some constraints using existing ISW data can be found in [209]. One should note, however, that the ISW measurements are plagued by large statistical errors (because the ISW effect is only a part of the total CMB anisotropy) and cannot provide the percent level accuracy needed to test viable $f(R)$ models, (i.e. models that satisfy local tests of gravity and for which the ISW contribution is suppressed to the percent level). WL studies, on the other hand, are counted on to eventually provide highly accurate 3D maps of the sum of the gravitational potentials, $\Phi + \Psi$.

The features found in $f(R)$ models are common to the broader class of scalar–tensor theories, to which $f(R)$ belongs, as well as to models of coupled quintessence [105, 210]. These models can be described by a Chameleon-type action (57), representative of coupled dark energy models in the Jordan frame and of scalar–tensor theories in the Einstein frame. Focusing, as before, on dark matter, indicating its coupling simply with $\alpha$ (i.e. dropping the index $i$), and studying scalar linear perturbations around a flat FRW metric, on sub-horizon scales we can easily find [188, 205] that the effective Newton’s constant and the gravitational slip are
\begin{align}
\mu(a, k) &= \frac{1 + 1 + 2M_p^2 \alpha^2}{a^2} \frac{k^2}{m^2}, \quad (100)
\gamma(a, k) &= \frac{1 + 1 + 2M_p^2 \alpha^2}{a^2} \frac{k^2}{m^2}, \quad (101)
\end{align}
which are exactly of the form (98)–(99). (In (100) we have neglected e prefactor $e^{-2\alpha(\phi)} \equiv 1$). Here $m$ is the mass of the scalar field and $\alpha_{\phi} \equiv \partial \alpha / \partial \phi$.

To summarize, in models of modified gravity and coupled dark energy the growth pattern is in general scale-dependent
and the time evolution of the Newtonian potentials is different than that in \( \Lambda \)CDM. The introduction of an effective Newton’s constant and a slip between the potentials is a general feature of the modifications, independent of the background expansion. Therefore, even for models that at the background level closely mimic the cosmological constant, the predictions for the growth of structure can be significantly different. The degeneracy characterizing models of cosmic acceleration at the background level can henceforth be broken at the level of LSS. Characteristic imprints of such modifications will be different clustering rate for DM, a modified ISW signal as well as a different WL shear. Moreover, since the potentials \( \Phi \) and \( \Psi \) will in general differ, one expects that the function measured by ISW and WL, i.e. \( \Phi + \Psi \), will be different than the one to which peculiar velocities respond, i.e. \( \Psi \). All these features allow potentially powerful ways to distinguish between models of cosmic acceleration, which we review in the next subsection.

Before concluding, let us notice that, as we saw, models of generalized dark energy can induce a pattern of growth analogous to the one found in models of modified gravity. Therefore, the degeneracy between dynamical dark energy and modified gravity models appears to persist at the level of LSS. Indeed, it seems always possible to introduce a generic dark fluid with properties such as to mimic the modifications of gravity. This has been discussed recently by several groups [211–213] and the authors of [213] have noted that the breaking of such degeneracy would require the independent measurements of three observables, (e.g. \( \Psi, \Phi + \Psi \) and \( \delta \)).

7.2. Constraining departures from standard GR

From the preceding discussion it is clear that LSS offers a potentially powerful way of constraining models of cosmic acceleration. The task, however, is not trivial, and there are different ways in which one can proceed. A first approach could perhaps consist of assuming a functional form for the slip between the potentials and the effective Newton’s constant, motivated by a certain class of theories. Some representative parametrizations have been introduced in [197, 212, 214, 215]. From the previous examples we can easily infer that a quite generic parametrization would be

\[
\mu(a, k) = \frac{1 + \beta_1 \lambda_1^2 k^2 a^4}{1 + \lambda_1^2 k^2 a^4},
\]

(102)

\[
\gamma(a, k) = \frac{1 + \beta_2 \lambda_2^2 k^2 a^4}{1 + \lambda_2^2 k^2 a^4},
\]

(103)

where the parameters \( \lambda_i^2 \) have dimensions of length squared, while the \( \beta_i \) represent dimensionless couplings. Finally, from a scalar–tensor point of view, the parameter \( s \) encodes the time dependence of the scalar field mass. The expressions in (102) coincide with the scale-dependent parametrization introduced in [197]. In [201], the authors performed a Fisher matrix analysis to forecast the constraints on the parameters \( \{\lambda_i, \beta_i, s\} \) of (102) from a combination of power spectra from galaxies, WL and the ISW effect on the CMB. Such an analysis reveals the extent to which one can constrain these theories and also allows one to reconstruct the shape of the slip and effective Newton’s constant based on the chosen form. In [201], it was found that these data, even at the linear level, are quite powerful in constraining the modified growth parameters. The results, of course, depend on the choice of the parametrization. However, they are good indicators of the power of current and upcoming surveys to constrain departures from GR.

An alternative approach, which can work for certain estimators of the slip, is a direct reconstruction from data. In [199, 200], it was proposed to consider the ratio of the peculiar velocity–galaxy correlation with the WL–galaxy
correlation. Specifically, the authors propose the following estimator for the gravitational slip

\[ \dot{\gamma} = \frac{P_{\nabla^2(\Psi - \phi)g}}{P_{\nabla^2\Psi g}} - 1, \]

(104)

where \( P_{\nabla^2(\Psi - \phi)g} \) and \( P_{\nabla^2\Psi g} \) are the cross-power spectrum between the galaxy number overdensity and, respectively, WL and the potential \( \Psi \) (which is related to peculiar velocities via the continuity equation).

In such a ratio, the dependence on the galaxy bias cancels out. Such a ratio, if appropriately constructed, would directly probe any difference between \( \Phi \) and \( \Psi \). This is a more direct and model-independent way of testing GR with the growth of structure; however its power will depend on how well future experiments will be able to measure peculiar velocities. From the analysis in [199, 200], it appears however that the ratio measured by future surveys might have strong discriminating power for some dark universe scenarios.

Finally, another, non-parametric approach, consists of performing a principal component analysis (PCA) to determine the eigenmodes of the slip and the Newton constant, that can be constrained by data, in the same way as has been done for the dark energy equation of state [216–218]. This method allows one to compare different experiments and their combinations, according to the relative gain in information about the functions. PCA can also point to the ‘sweet spots’ in redshift and scale where data is most sensitive to variations in the slip and Newton’s constant, which can be a useful guide for designing future observing strategies. The PCA method does not allow one to reconstruct the shape of the functions from data. However, one can still reproduce the errors on parameters of any parametrization from the eigenvectors and eigenvalues found using PCA [217]. Thus, it is a promising method, but computationally challenging.

7.2.1. The observables. As we mentioned above, in the \( \Lambda \)CDM model the sub-horizon evolution of gravitational potentials and the matter density fluctuations are described by a single function of time—the scale-independent growth factor \( g(a) \). In models of modified gravity, on the other hand, the dynamics of perturbations can be richer and, generically, the evolution of \( \Phi \), \( \Psi \) and \( \delta \) (the matter density contrast), will be described by different functions of scale and time. Therefore, we expect that different observables will be described by different functions, and by combining different types of measurements, one can try to reconstruct these functions, or at least put a limit on how different they can be. In what follows, we shall give a brief review of the relation between the different types of observables and the gravitational potentials they probe. For a more thorough overview of the various ways of looking for modifications in the growth of perturbations we refer the reader to [213].

Galaxy counts (GC) probe the distribution and growth of matter inhomogeneities. However, to extract the matter power spectrum, one needs to account for the bias, which typically depends on the type of galaxies and can be both time- and scale-dependent. On large scales, where non-linear effects are unimportant, one can use a scale-independent bias factor to relate GCs to the total matter distribution. This relation becomes increasingly complicated and scale-dependent as one considers smaller and smaller scales. In principle, the bias parameters can be determined from higher order correlation functions [219–221]. On sub-horizon linear scales, the evolution of matter density contrast is determined by equation (95). Hence, measurements of GC over multiple redshifts can provide an estimate of \( \Psi \) as a function of space and time, up to a bias factor. A more direct probe of the potential \( \Psi \), would be a measurement of peculiar velocities, which follow the gradients of \( \Psi \). Such measurements would be independent of uncertainties associated with modeling the bias.

Peculiar velocity surveys typically use redshift-independent distance indicators to separate the Hubble flow from the local flow, and nearby SNeIa are therefore good candidates; a number of surveys, like the 6dFGS [222] and the 2MRS [223], use galaxies. An interesting alternative is offered by the kinetic Sunyaev–Zel’dovich effect in clusters [224], that arises from the inverse Compton scattering of CMB photons off high-energy electrons in the clusters. This effect provides a useful way of measuring the bulk motion of electrons in clusters, hence the peculiar velocity of clusters, but it is limited by low signal-to-noise ratio. Current measurements of peculiar velocities are limited in accuracy, and at this point it is not clear how to forecast the accuracy of future observations. However they are a potentially powerful probe [225].

In contrast to GCs and peculiar velocities, which respond to one of the metric potentials, namely \( \Psi \), WL of distant light sources by intervening structure is determined by spatial gradients of the sum \( (\Phi + \Psi) \). Hence, measurements of the WL shear distribution over multiple redshift bins can provide an estimate of the space and time variation of the sum of the two potentials. In the \( \Lambda \)CDM and minimally coupled models of dark energy, the two metric potentials coincide, and therefore WL probes essentially the same growth function that controls the evolution of galaxy clustering and peculiar velocities. In models of modified gravity, however, there could be a difference between the potentials, corresponding to an effective shear component. Measurements of the ISW effect in the CMB probe the time dependence of the sum of the potentials: \( d(\Phi + \Psi)/dt \).

By combining multiple redshift information on GC, WL and CMB, and their cross-correlations, one can constrain the differences between the metric potentials and the spacetime variation of the effective Newton constant defined in the previous section. Ideally, experimentalists would measure all possible cross-correlations, between all possible pairs of observables, in order to maximize the amount of information available to us. In practice, however, it can be difficult to obtain these cross-correlations, since their measurements require that each of the individual fields (CMB, GC, WL) be measured on the same patch of sky. This will be addressed with near and distant future tomographic LSS surveys (such as DES [31], LSST [32] and PAN-STARRS [226]). Even with conservative assumptions about the data, (i.e. considering only linear scales), it is hoped that DES will produce non-trivial constraints on modified growth, and that LSST will do even better (with \( \approx 10\% \) relative errors in the parameters).
From the discussions in the previous section, it is clear that the predictions for the growth of structure in models of modified gravity and coupled dark energy can significantly differ from those in $\Lambda$CDM or uncoupled quintessence models. Therefore, LSS and CMB data offer a valuable testing ground for gravity and cosmological tests are expected to play an important role in determining the physics of cosmic acceleration. In principle, a combination of galaxy number counts and WL measurements, along with CMB and SNe data, will allow a distinction between modified gravity/exotic dark energy models and the standard $\Lambda$CDM model of cosmology. However, as we have already mentioned, it will be harder to break the degeneracy between models of modified gravity and generalized models of dark energy, where the dark fluid is allowed to cluster and carry anisotropic stress. As discussed in [213], the latter task will require the independent measurements of at least three observables.

8. Conclusions

That the expansion rate of the universe is accelerating is now, just over a decade after the first evidence from observations of type Ia supernovae, a firmly established aspect of cosmology. The rapid progress in establishing this fact is a testament to the breathtaking convergence of techniques and technology that has emerged in observational cosmology. In turn, cosmic acceleration has introduced new wrinkles into almost every part of theoretical cosmology, ranging from the details of structure formation, through the CMB and gravitational lensing.

Beyond the fascinating new problems in theoretical cosmology, cosmic acceleration has presented an enormous and, as yet, unmet challenge to fundamental physics. A perfectly good fit to all known cosmological data is that cosmic acceleration is driven by a cosmological constant. However, accounting for the necessary magnitude of such an object seems at least as difficult as attempts to understand why it should be precisely zero. Anthropic arguments, fueled by results from the landscape of string theory and the idea of eternal inflation, may yield a way to understand such a small number. However, it is, at present, far too early to know if this is a sensible outcome of string theory, and there are no developed ideas of how such proposals might be tested. This is not to say, of course, that they are incorrect; merely that there are immense obstacles to establishing such ideas as a tested answer.

If cosmic acceleration is not due to a cosmological constant, then an exotic component of the energy budget of the universe may be the culprit. It is important to recognize up front that any such understanding of cosmic acceleration assumes that the cosmological constant itself is either precisely zero, or at least negligibly small—an unsolved problem in itself. If we make such an assumption, then proposals for a dynamical component of the cosmic energy budget, driving the accelerating expansion, come under the heading of dark energy. In their simplest forms, ideas of dark energy can be thought of as attempts to construct a model for late-time cosmological inflation, with the key differences being the extremely low energy scales involved, and the fact that this period of acceleration need not end in the way that inflation must. As with early universe inflation, a major challenge is finding a home for the abstract idea within a well-motivated and technically natural particle physics model. In fact, in a number of ways, this problem is more acute for dark energy models, since they must operate in a regime in which we already have a supremely well-tested low energy effective field theory of much of the matter content of the universe. As we have described, some tentative ideas exist to tackle this problem, but as yet no convincing particle physics implementation of dark energy exists, in contrast to the somewhat natural occurrence of dark matter candidates in theories of physics beyond the standard model.

But dark energy is not the only possible solution to the problem of cosmic acceleration. Instead we might imagine leaving the material contents of our theories untouched, comprised only of the standard model and dark matter, and instead revisiting the dynamical relationship between these components and the evolution of the background geometry of the universe. In other words, we might consider that cosmic acceleration is our first evidence for a modification of general relativity in the far infrared. One might imagine that this approach would hold out the possibility of a solution to the cosmological constant problem itself, as well as providing an understanding of cosmic acceleration. However, to date this possibility remains unrealized, and just as for dark energy, modified gravity approaches to cosmic acceleration must also merely assume that the cosmological constant is either zero, or negligibly small.

Making this assumption, a number of authors have considered how one might modify gravity. However, to date all attempts have run into problems, either with matching constraints on gravity at small (solar system) scales, or with the existence of ghost degrees of freedom, raising the possibility of problems similar to those we mentioned when discussing phantom fields. Nevertheless, if one views these models either as a long-wavelength approximation to a better-behaved theory of gravity, or as a very low energy effective theory, in need of an ultraviolet completion to cure its problems, one may extract a variety of predictions for signatures in upcoming cosmological observations.

Perhaps the biggest question to be answered is binary in nature—is cosmic acceleration due to a cosmological constant or not? In many ways, an affirmative answer is the most depressing, since a true cosmological constant varies in neither space nor time, and thus, if it is the driver of acceleration, we already know all that we will ever know about it. As such, we would not expect further observational insights into the required unnaturally small value. More generally, we would like to be able to distinguish among this minimal explanation, a possible dark energy component and the option of an infrared modification of gravity. It has become increasingly clear that our best hope for this is to compare geometrical measures, such as the distance to the surface of last scattering, with measures of the growth of LSS, which depend in detail, in a redshift-dependent manner on the gravitational force law underlying the collapse of overdensities.
In this paper we have endeavored to provide an overview of the major approaches to cosmic acceleration and to the techniques through which we hope to further our understanding. What should be clear is that the issues involved here are far from settled, and that the field is hungry for new ideas. It is unclear where such an idea will come from, and for this reason it seems important to pursue all possible avenues, no matter how unpromising they may seem. We hope to have provided a snapshot of some of these avenues of enquiry. It remains to be seen which, if any, of these will provide the key to the problem of cosmic acceleration.

Acknowledgments

The authors would like to thank Rachel Bean, Eanna Flanagan, Justin Khoury and Levon Pogosian for useful discussions during the writing of this paper, and all our collaborators for helping develop our understanding. The work of MT is supported in part by National Science Foundation grants PHY-0653563 and PHY-0930521, by Department of Energy grant DE-FG05-95ER40893-A020 and by NASA ATP grant NNX08AH27G. The work of AS is supported by the National Science Foundation under grant AST-0708501.

References

[1] Copeland E J, Sami M and Tsujikawa S 2006 Dynamics of dark energy Int. J. Mod. Phys. D 15 1753–936
[2] Eric V L 2008 Mapping the cosmological expansion Rep. Prog. Phys. 71 056901
[3] Frieman J, Turner M and Huterer D 2008 Dark energy and the accelerating universe Ann. Rev. Astron. Astrophys. 46 385–432
[4] Caldwell R R and Kamionkowski M 2009 The physics of cosmic acceleration arXiv:0903.0866
[5] Trodden M and Carroll S M 2004 TASI lectures: Introduction to cosmology arXiv:astro-ph/0401547
[6] Riess A G et al 1998 Observational evidence from supernovae for an accelerating universe and a cosmological constant Astron. J. 116 1009–38
[7] Perlmutter S et al 1999 Measurements of omega and lambda from 42 high-redshift supernovae Astrophys. J. 517 565–86
[8] Chandrasekhar S 1931 The maximum mass of ideal white dwarfs Astrophys. J. 74 81–2
[9] Phillips M M 1993 The absolute magnitudes of type Ia supernovae Astrophys. J. 413 L105–8
[10] Riess A G, Press W H and Kirshner R P 1996 A precise distance indicator: type Ia supernovae to z ~ 0.4 Astron. J. 112 1998
[11] Hamuy M et al 1996 The absolute luminosities of the Calan/Tololo type Ia supernovae Astron. J. 112 2391
[12] Perlmutter S 1999 Supernovae, dark energy, and the accelerating universe: the status of the cosmological parameters Talk given at 19th Int. Symp. on Lepton and Photon Interactions at High–Energies (LP 99) (Stanford, CA, 9–14 August 1999)
[13] Nugent P, Kim A and Perlmutter S 2002 K–corrections and extinction corrections for type Ia supernovae Publ. Astron. Soc. Pac. 114 803–819
[14] Perlmutter S and Schmidt B P 2003 Measuring cosmology with supernovae arXiv:astro-ph/0303428
[15] Jorgensen H E, Salamanca A A, Norgaard-Nielsen H H, Hansen L and Ellis R S
[16] Knop R A et al 2003 New constraints on \( \Omega_M, \Omega_l \), and \( \omega \) from an independent set of eleven high-redshift supernovae observed with HST Astrophys. J. 598 102
[17] Astier P et al 2006 The supernova legacy survey: measurements of \( \Omega_M, \Omega_l \), and \( \omega \) from the first year data Astron. Astrophys. 447 51–48
[18] Riess A G et al 2007 New Hubble Space Telescope discoveries of type Ia supernovae at \( z \geq 1 \): narrowing constraints on the early behavior of dark energy Astrophys. J. 659 98–121
[19] Starkman G, Trodden M and Vachaspati T 1999 Observation of cosmic acceleration and determining the fate of the universe Phys. Rev. Lett. 83 1510–13
[20] Huterer D, Starkman G D and Trodden M 2002 Is the universe inflating? Dark energy and the future of the universe Phys. Rev. D 66 043511
[21] Hicken M et al 2009 Improved dark energy constraints from 100 new CfA supernova type Ia light curves arXiv:0901.4804
[22] Hicken M et al 2009 CfA3: 185 Type Ia supernova light curves from the CfA arXiv:0901.4787
[23] Wood-Vasey W M et al 2004 The nearby supernova factory New Astron. Rev. 48 637–40
[24] Freedman W L 2004 The Carnegie Supernova Project arXiv:astro-ph/0411176
[25] Wood-Vasey W M et al 2007 Observational constraints on the nature of the dark energy: first cosmological results from the ESSENCE supernova survey Astrophys. J. 666 694–715
[26] Palanque-Delabrouille N 2005 Supernova Legacy Survey (SNLS): real time operations and photometric analysis arXiv:astro-ph/0509425
[27] Frieman J A et al 2008 The Sloan digital sky survey-II supernova survey: technical summary Astron. J. 135 338–47
[28] Louis-Gregory S et al 2004 The hubble higher-Z supernova search: supernovae to \( z = 1.6 \) and constraints on type Ia progenitor models Astrophys. J. 613 200–23
[29] Giavalisco M et al 2004 The great observatories origins deep survey: initial results from optical and near-infrared imaging Astrophys. J. 600 L93
[30] http://pds.gsfc.nasa.gov
[31] http://www.darkenergysurvey.org
[32] http://www.lsst.org
[33] Smoot G F et al 1992 Structure in the COBE differential microwave radiometer first year maps Astrophys. J. 396 L1–5
[34] Spitzer D N et al 2003 First year Wilkinson Microwave Anisotropy probe (WMAP) observations: determination of cosmological parameters Astrophys. J. Suppl. 148 175
[35] Spiegel D N et al 2007 Wilkinson Microwave Anisotropy probe (WMAP) three year results: implications for cosmology Astrophys. J. Suppl. 170 377
[36] Komatsu E et al 2008 Five-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological interpretation
[37] Kowalski M et al 2008 Improved cosmological constraints from new, old and combined supernova datasets Astrophys. J. 686 749–78
[38] Eisenstein D J et al 2005 Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies Astrophys. J. 633 560–74
[39] Dunkley J et al 2009 Five-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: likelihoods and parameters from the WMAP data Astrophys. J. Suppl. 180 306–29
[40] http://www.sdss.org
[41] tegmark M et al 2006 Cosmological constraints from the SDSS luminous red galaxies Phys. Rev. D 74 123507
[198] Daniel S F, Caldwell R R, Cooray A and Melchiorri A 2008 Large scale structure as a probe of gravitational slip Phys. Rev. D 77 103513

[199] Zhang P, Liguori M, Bean R and Dodelson S 2007 Probing gravity at cosmological scales by measurements which test the relationship between gravitational lensing and matter overdensity Phys. Rev. Lett. 99 141302

[200] Zhang P, Bean R, Liguori M and Dodelson S 2008 Weighing the spatial and temporal fluctuations of the dark universe arXiv:0809.2856

[201] Zhao G-B, Pogosian L, Silvestri A and Zylberberg J 2008 Searching for modified growth patterns with tomographic surveys Phys. Rev. D 77 083513

[202] Song Y-S and Dore O 2008 Testing gravity using weak gravitational lensing and redshift surveys arXiv:0812.0002

[203] Zhang P 2006 Testing $f(R)$ gravity against the large scale structure of the universe Phys. Rev. D 73 123504

[204] Bertschinger E 2006 On the growth of perturbations as a test of dark energy Astrophys. J. 648 797–806

[205] Song Y-S, Hu W and Sawicki I 2007 The large scale structure of $f(R)$ gravity Phys. Rev. D 75 044004

[206] Bean R, Bernat D, Pogosian L, Silvestri A and Trodden M 2007 Dynamics of linear perturbations in $f(R)$ gravity Phys. Rev. D 75 064020

[207] Crittenden R G and Turok N 1996 Looking for $\Lambda\chi$ with therees-sciana effect Phys. Rev. Lett. 76 575

[208] Boughn S and Crittenden R 2004 A correlation of the cosmic microwave sky with large scale structure Nature 427 45–7

[209] Song Y-S, Peiris H and Hu W 2007 Cosmological constraints on $f(R)$ acceleration models Phys. Rev. D 76 063517

[210] Amendola L 2004 Linear and non-linear perturbations in dark energy models Phys. Rev. D 69 103524

[211] Kunz M and Sapone D 2007 Dark energy versus modified gravity Phys. Rev. Lett. 98 121301

[212] Hu W and Sawicki I 2007 A parameterized post-Friedmann framework for modified gravity Phys. Rev. D 76 104043

[213] Jain B and Zhang P 2008 Observational tests of modified gravity Phys. Rev. D 78 063503

[214] Hu W 2008 Parameterized post-Friedmann signatures of acceleration in the CMB Phys. Rev. D 77 103524

[215] Amin M A, Wagoner R V and Blandford R D 2007 A sub-horizon framework for probing the relationship between the cosmological matter distribution and metric perturbations arXiv:0708.1793

[216] Huterer D and Starkman G 2003 Parameterization of dark-energy properties: a principal-component approach Phys. Rev. Lett. 90 031301

[217] Crittenden R G and Pogosian L 2005 Investigating dark energy experiments with principal components

[218] de Putter R and Linder E V 2008 To bin or not to bin: decorrelating the cosmic equation of state Astropart. Phys. 29 424

[219] Fry J N 1994 Gravity, bias, and the galaxy three-point correlation function Phys. Rev. Lett. 73 215–19

[220] Matarrese S, Verde L and Heavens A F 1997 Large-scale bias in the universe: bispectrum method Mon. Not. R. Astron. Soc. 290 651–62

[221] Sefusatti E and Komatsu E 2007 The bispectrum of galaxies from high-redshift galaxy surveys: primordial non-Gaussian and non-linear galaxy bias Phys. Rev. D 76 083004

[222] Jones D H et al 2004 The 6dF galaxy survey: samples, observational techniques and the first data release arXiv:astro-ph/0403501

[223] Ergoddu P et al 2006 Reconstructed density and velocity fields from the 2MASS redshift survey Mon. Not. R. Astron. Soc. 373 45–64

[224] Sunyaev R A and Zeldovich Ya B 1980 The velocity of clusters of galaxies relative to the microwave background. The possibility of its measurement Mon. Not. R. Astron. Soc. 190 413–20

[225] Song Y-S and Percival W J 2008 Reconstructing the history of structure formation using peculiar velocities arXiv:0807.0810

[226] http://pan-starrs.ifa.hawaii.edu/public/home.html