Dark energy: The cosmological challenge of the millennium

T. Padmanabhan

Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune 411 007, India

Recent cosmological observations suggest that nearly seventy per cent of the energy density in the universe is unclustered and has negative pressure. Several conceptual issues related to the modelling of this component (‘dark energy’), which is driving an accelerated expansion of the universe, are reviewed with special emphasis on the cosmological constant as the possible choice for the dark energy.

I. The cosmological paradigm

The last couple of decades have been the golden age for cosmology, in much the same way as the mid-1900s were a golden age for particle physics. Data of exquisite quality confirmed the broad paradigm of standard cosmology and helped us to determine the composition of the universe. As a direct consequence, the cosmological observations have thrusted upon us a rather preposterous composition for the universe which defies any simple explanation, thereby posing the greatest challenge theoretical physics has ever faced.

To understand these exciting developments, it is best to begin by reminding ourselves of the standard paradigm for cosmology. Observations show that the universe is fairly homogeneous and isotropic at scales larger than about 150 h⁻¹ Mpc, where 1 Mpc = 3 × 10⁸ cm is a convenient unit for extragalactic astronomy and h = 0.7 characterizes the current rate of expansion of the universe in dimensionless form. (The mean distance between galaxies is about 1 Mpc while the size of the visible universe is about 3000 h⁻¹ Mpc.)

The conventional – and highly successful – approach to cosmology separates the study of large scale (l ≥ 150 h⁻¹ Mpc) dynamics of the universe from the issue of structure formation at smaller scales. The former is modelled by a homogeneous and isotropic distribution of energy density; the latter is addressed in terms of gravitational instability which will amplify the small perturbations in the energy density, leading to the formation of structures like galaxies.

In such an approach, the expansion of the background universe is described by a single function of time a(t) which is governed by the equations (with c = 1):

\[ \frac{a^2 + k}{a^2} = \frac{8\pi G\rho}{3}; \quad d(p a^3) = -p da^3. \]

The first one relates expansion rate to the energy density ρ and k = 0, ± 1 is a parameter which characterizes the spatial curvature of the universe. The second equation, when coupled with the equation of state ρ = p(a) in terms of the expansion factor of the universe. In particular if p = wρ with (at least, approximately) constant w then, ρ ∝ a⁻³(1+w) and (if we assume k = 0), a ∝ t³[1+(1+w)]

It is convenient to measure the energy densities of different components in terms of a critical energy density (ρc) required to make k = 0 at the present epoch. (Of course, since k is a constant, it will remain zero at all epochs if it is zero at any given moment of time.) From eq. (1), it is clear that ρc = 3H₀²/8πG where H₀ = (d/a₀) is the rate of expansion of the universe at present. The variables Ω = ρ/ρc will give the fractional contribution of different components of the universe (i denoting baryons, dark matter, radiation, etc.) to the critical density. Observations then lead to the following results:

- Our universe has 0.98 ≤ Ωtot ≤ 1.08. The value of Ωtot can be determined from the angular anisotropy spectrum of the cosmic microwave background radiation (CMBR) (with the reasonable assumption that h > 0.5) and these observations now show that we live in a universe with critical density².³

- Observations of primordial deuterium produced in big bang nucleosynthesis (which took place when the universe was about 1 min in age) as well as the CMBR observations show that the total amount of baryons in the universe contributes about Ω_h = (0.024 ± 0.0012)h⁻². Given the independent observations on the Hubble constant¹ which fix h = 0.72 ± 0.07, we conclude that Ω_h ≡ 0.04–0.06. These observations take into account all baryons which exist in the universe today irrespective of whether they are luminous or not. Combined with previous item we conclude that most of the universe is non-baryonic.

- Host of observations related to large scale structure and dynamics (rotation curves of galaxies, estimate of cluster masses, gravitational lensing, galaxy surveys, etc.) all suggest⁵ that the universe is populated by a non-luminous component of matter (dark matter, DM) made of

---

e-mail: nabhan@iucaa.ernet.in
weakly interacting massive particles which does cluster at galactic scales. This component contributes about $\Omega_{DM} \approx 0.15-0.30$.

- Combining the last observation with the first we conclude that there must be (at least) one more component to the energy density of the universe contributing about 70% of critical density. Early analysis of several observations indicated that this component is unclustered and has negative pressure. This is confirmed dramatically by the supernova observations (see ref. 7; for a critical look at the data, see refs 8, 9). The observations suggest that the missing component has $w = p/\rho \leq -0.78$ and contributes $\Omega_{DE} \approx 0.7-0.85$.

- The universe also contains radiation contributing and energy density $\Omega_k h^2 = 2.56 \times 10^{-5}$ today most of which is due to photons in the CMBR. This is dynamically irrelevant today but would have been the dominant component in the universe at redshifts larger that $z_{eq} \approx \Omega_{DM}/\Omega_k \approx 4 \times 10^5 \Omega_{DM} h^2$.

- Together we conclude that our universe has $\Omega_{DE} \approx 0.7$, $\Omega_{DM} \approx 0.26$, $\Omega_k \approx 0.04$, $\Omega_R \approx 5 \times 10^{-5}$. All known observations are consistent with such an admittedly weird composition for the universe.

Before discussing the composition of the universe in greater detail, let us briefly consider the issue of structure formation. The key idea is that if there existed small quantum fluctuations of the field driving the inflationary phase then the universe would be characterized by a power spectrum of the form $P(k) = Ak^n$ with $n = 1$. The models cannot predict the value of the amplitude $A$ in an unambiguous manner but it can be determined from CMBR observations. The CMBR observations are consistent with the inflationary model for the generation of perturbations and gives $A \approx (28.3 \ h^{-4} \text{Mpc})^4$ and $n = 0.97 \pm 0.023$ (The first results were from COBE$^{14}$ and WMAP$^{16}$ has confirmed them with far greater accuracy).

So, to the zeroth order, the universe is characterized by just seven numbers: $\Omega_{DE} \approx 0.7$, $\Omega_{DM} \approx 0.26$, $\Omega_k \approx 0.04$, $\Omega_R \approx 5 \times 10^{-5}$ giving the composition of the universe; the amplitude $A \approx (28.3 \ h^{-4} \text{Mpc})^4$ and the index $n = 1$ of the initial perturbations. The challenge is to make some sense out of these numbers from a more fundamental point of view.

II. The dark energy

It is rather frustrating that we have no direct laboratory evidence for nearly 96% of matter in the universe. (Actually, since we do not quite understand the process of baryogenesis, we do not understand $\Omega_B$ either; all we can theoretically understand now is a universe filled entirely with radiation!) Assuming that particle physics models will eventually come of age and (i) explain $\Omega_B$ and $\Omega_{DM}$ (probably as the lightest supersymmetric partner) as well as (ii) provide a viable model for inflation predicting correct value for $A$, one is left with the problem of understanding $\Omega_{DE}$. While the issues (i) and (ii) are by no means trivial or satisfactorily addressed, it is probably correct to say that the issue of dark energy is lot more perplexing, thereby justifying the attention it has received recently.

The key observational feature of dark energy is that treated as a fluid with a stress tensor $T^\mu_\nu = \text{diag}(\rho - p, -p, -p, -p)$ it has an equation state $p = w \rho$ with $w \leq -0.8$ at the present epoch. The spatial part $g$ of the geodesic acceleration (which measures the relative acceleration of two geodesics in the spacetime) satisfies an exact equation in general relativity (see e.g., page 332 of ref. 15) given by:

$$\nabla \cdot \mathbf{g} = -4\pi G(\rho + 3p). \quad (2)$$

This shows that the source of geodesic acceleration is $(\rho + 3p)$ and not $\rho$. As long as $(\rho + 3p) > 0$, gravity remains attractive while $(\rho + 3p) < 0$ can lead to repulsive gravitational effects. In other words, dark energy with sufficiently negative pressure will accelerate the expansion of the universe, once it starts dominating over the normal matter. This is precisely what is established from the study of high redshift supernova, which can be used to determine the expansion rate of the universe in the past$^{9}$. Figure 1 presents the supernova data as a phase portrait of the universe (plotting the ‘velocity’ $\dot{a}$ against ‘position’ $a$). It is clear that the universe was decelerating at high redshifts and started accelerating when it was about two-thirds of the present size.

The simplest model for a fluid with negative pressure is the cosmological constant (for a review, see ref. 16) with $w = -1$, $\rho = -p = \text{constant}$ (which is the model used in Figure 1). If the dark energy is indeed a cosmological constant, then it introduces a fundamental length scale in the theory $L_k = H_0^{-1}$, related to the constant dark energy density $\rho_{DE}$ by $H_0^2 = 8\pi G \rho_{DE}$/3. In classical general relativity, based on the constants $G$, $c$ and $L_k$, it is not possible to construct any dimensionless combination from these constants. But when one introduces the Planck constant, $\hbar$, it is possible to form the dimensionless combination $H_0^2 = (Gc/H)^2 = (L_k^2 / L_\hbar^2)$. Observations then require $(L_k^2 / L_\hbar^2) \approx 10^{-123}$. As has been mentioned in the literature, this would require enormous fine tuning. What is more, in the past, the energy density of normal matter and radiation would have been higher while the energy density contributed by the cosmological constant does not change. Hence we need to adjust the energy densities of normal matter and cosmological constant in the early epoch very carefully so that $\rho_A \approx \rho_{NR}$ around the current epoch. This raises the
second of the two cosmological constant problems: Why is it that \((\rho_v/\rho_{br}) = \mathcal{O}(1)\) at the current phase of the universe?

Because of these conceptual problems associated with the cosmological constant, people have explored a large variety of alternative possibilities. The most popular among them uses a scalar field \(\phi\) with a suitably chosen potential \(V(\phi)\) so as to make the vacuum energy vary with time. The hope then is that, one can find a model in which the current value can be explained naturally without any fine tuning. A simple form of the source with variable \(w\) are scalar fields with Lagrangians of different forms, of which we will discuss two possibilities:

\[
L_{\text{quin}} = \frac{1}{2} \partial_i \phi \partial^i \phi - V(\phi); \quad L_{\text{tach}} = -V(\phi) \left[1 - \partial_i \phi \partial^i \phi\right]^{1/2}. \tag{3}
\]

Both these Lagrangians involve one arbitrary function \(V(\phi)\). The first one, \(L_{\text{quin}}\), which is a natural generalization of the Lagrangian for a non-relativistic particle, \(L = (1/2) q^2 - V(q)\), is usually called quintessence (for a sample of models, see ref. 17). When it acts as a source in Friedman universe, it is characterized by a time dependent \(w(t)\) with

\[
\rho_q(t) = \frac{1}{2} \dot{\phi}^2 + V; \quad p_q(t) = \frac{1}{2} \dot{\phi}^2 - V; \quad w_q = \frac{1 - (2V/\dot{\phi}^2)}{1 + (2V/\dot{\phi}^2)}. \tag{4}
\]

The structure of the second Lagrangian in eq. (3) can be understood by a simple analogy from special relativity (see the first reference in ref. 18). A relativistic particle with (one dimensional) position \(q(t)\) and mass \(m\) is described by the Lagrangian \(L = -m(1 - \dot{q}^2)^{1/2}\). It has the energy \(E = m(1 - \dot{q}^2)^{1/2}\) and momentum \(k = m \dot{q}(1 - \dot{q}^2)^{1/2}\) which are related by \(E^2 = k^2\). As is well known, this allows the possibility of having massless particles with finite energy for which \(E^2 = k^2\). This is achieved by taking the limit of \(m \to 0\) and \(q \to 1\), while keeping the ratio in \(E = m(1 - \dot{q}^2)^{1/2}\) finite. The momentum acquires a life of its own, unconnected with the velocity \(\dot{q}\), and the energy is expressed in terms of the momentum (rather than in terms of \(q\)) in the Hamiltonian formulation. We can now construct a field theory by upgrading \(q(t)\) to a field \(\phi\). Relativistic invariance now requires \(\phi\) to depend on both space and time \([\phi = \phi(t, x)]\) and \(\dot{q}\) to be replaced by \(\partial_i \phi \partial^i \phi\). It is also possible now to treat the mass parameter \(m\) as a function of \(\phi\), say, \(V(\phi)\) thereby obtaining a field theoretic Lagrangian \(L = -V(\phi) \left[1 - \partial_i \phi \partial^i \phi\right]^{1/2}\). The Hamiltonian structure of this theory is algebraically very similar to the special relativistic example we started with. In particular, the theory allows solutions in which \(V \to 0\), \(\partial_i \phi \partial^i \phi \to 1\) simultaneously, keeping the energy (density) finite. Such solutions will have finite momentum density (analogous to a massless particle with finite momentum \(k\)) and energy density. Since the solutions can now depend on both space and time (unlike the special relativistic example in which \(q\) depended only on time), the momentum density can be an arbitrary function of the spatial coordinate. This provides a rich gamut of possibilities in the context of cosmology\(^{18-21}\). This form of scalar field arises in string theories\(^{22}\) and – for technical reasons – is called a tachyonic scalar field. (The structure of this Lagrangian is similar to that analysed in a wide class of models called \(K\)-essence; see for example, ref. 23.)

Figure 1. The ‘velocity’ \(\dot{a}\) of the universe is plotted against the ‘position’ \(a\) in the form of a phase portrait. The different curves are for models parameterized by the value of \(\Omega_{\text{dm}}(= \Omega_{\Lambda})\) keeping \(\Omega_{br} = 1\). The topmost curve has \(\Omega_{\Lambda} = 1\) with no dark energy and the universe is decelerating at all epochs. The bottommost curve has \(\Omega_{\Lambda} = 0\) and \(\Omega_{br} = 1\) and the universe is accelerating at all epochs. The in-between curves show universes which were decelerating in the past and began to accelerate when the dark energy started dominating. The supernova data clearly favours such a model. (For a more detailed discussion of the figure, see refs 8, 9.)

The stress tensor for the tachyonic scalar field can be written as the sum of a pressure-less dust component and a cosmological constant (see the first reference in ref. 19). To show this explicitly, we break up the density \(\rho\) and the pressure \(p\) and write them in a more suggestive form as

\[
\rho = \rho_\Lambda + \rho_{\text{DM}}; \quad p = p_V + p_{\text{DM}}, \quad \text{where}
\]

\[
\rho_{\text{DM}} = \frac{V(\phi) \dot{\phi} \dot{\phi}}{\sqrt{1 - \dot{\phi} \dot{\phi}}}; \quad p_{\text{DM}} = 0;
\]

\[
\rho_\Lambda = V(\phi) \sqrt{1 - \dot{\phi} \dot{\phi}}; \quad p_V = -\rho_\Lambda. \tag{5}
\]

This means that the stress tensor can be thought of as made up of two components – one behaving like a pressure-less fluid, while the other having a negative pressure.

When \(\phi\) is small (compared to \(V\) in the case of quintessence or compared to unity in the case of tachyonic field), both these sources have \(w \to -1\) and mimic a cosmological constant. When \(\dot{\phi} \gg V\), the quintessence has \(w \approx 1\) leading
to $\rho_\phi \propto (1 + z)^2$; the tachyonic field, on the other hand, has $w = 0$ for $\phi \to 1$ and behaves like non-relativistic matter. In both the cases, $-1 < w < 1$, though it is possible to construct more complicated scalar field Lagrangians with even $w < -1$. (For some alternatives to scalar field models see, for example, ref. 25.)

Since the quintessence field (or the tachyonic field) has an undetermined free function $V(\phi)$, it is possible to choose this function in order to produce a given $H(a)$. To see this explicitly, let us assume that the universe has two forms of energy density with $\rho(\alpha) = \rho_{\text{known}}(\alpha) + \rho_{\phi}(\alpha)$, where $\rho_{\text{known}}(\alpha)$ arises from any known forms of source (matter, radiation, etc.) and $\rho_{\phi}(\alpha)$ is due to a scalar field. Let us first consider quintessence. Here, the potential is given implicitly by the form

$$V(\alpha) = \frac{1}{16\pi G} H(1 - Q) \left[ 6H + 2ah' - \frac{aHQ'}{1 - Q} \right], \quad (6)$$

$$\phi(\alpha) = \left[ \frac{1}{8\pi G} \right]^{1/2} \left[ \frac{d\alpha}{a} \left( aQ'(1 - Q) \frac{d\ln H}{d\ln a} \right)^{1/2} \right], \quad (7)$$

where $Q(\alpha) \equiv [8\pi G\rho_{\text{known}}(\alpha)/3H^2(\alpha)]$ and prime denotes differentiation with respect to $\alpha$. Given any $H(\alpha)$, $Q(\alpha)$, these equations determine $V(\alpha)$ and $\phi(\alpha)$ and thus the potential $V(\phi)$.

Every quintessence model studied in the literature can be obtained from these equations. We shall now briefly mention some examples:

- **Power law expansion of the universe can be generated by a quintessence model with $V(\phi) = \phi^{-\alpha}$**. In this case, the energy density of the scalar field varies as $\rho_{\phi} \propto t^{-2(1+2\alpha)}$; if the background density $\rho_{bg}$ varies as $\rho_{bg} \propto t^{-2}$, the ratio of the two energy densities changes as $(\rho_{\phi}/\rho_{bg}) \propto t^{2(1+2\alpha)}$. Obviously, the scalar field density can dominate over the background at late times for $\alpha > 0$.

- **A different class of models arise if the potential is taken to be exponential with, say, $V(\phi) \propto \exp(-\lambda\phi/M_p)$. When $k = 0$, both $\rho_\phi$ and $\rho_{bg}$ scale in the same manner leading to

$$\frac{\rho_\phi}{\rho_{bg}} + \rho_\phi \propto \frac{3(1 + w_{bg})}{\lambda^2}, \quad (8)$$

where $w_{bg}$ refers to the background parameter value. In this case, the dark energy density is said to ‘track’ the background energy density. While this could be a model for dark matter, there are strong constraints on the total energy density of the universe at the epoch of nucleosynthesis. This requires $\Omega_\phi \leq 0.2$ requiring dark energy to be subdominant at all epochs.

- **Many other forms of $H(\alpha)$ can be reproduced by a combination of non-relativistic matter and a suitable form of scalar field with a potential $V(\phi)$. In fact, one can make the dark energy to vary with $a$ in an unspecified manner as $a^\gamma$. In this case we need $H^2(a) = H^2_0 [\Omega_{\text{NR}}a^{-3} + (1 - \Omega_{\text{NR}})a^{-\gamma}]$ which can arise if the universe is populated with non-relativistic matter with density parameter $\Omega_{\text{NR}}$ and a scalar field with the potential, determined using eqs (6), (7). We get

$$V(\phi) = V_0 \sinh^{2\alpha(1-n)}\left[\alpha(\phi - \psi)\right], \quad (9)$$

where

$$V_0 = \frac{(6-n)}{16\pi G} \left[ \frac{\Omega_{\text{NR}}}{(1 - \Omega_{\text{NR}})} \right]^{1/2}; \quad \alpha = (3-n)(2\pi G/n)^{1/2} \quad (10)$$

and $\psi$ is a constant.

Similar results exist for the tachyonic scalar field as well. For example, given any $H(\alpha)$, one can construct a tachyonic potential $V(\phi)$ so that the scalar field is the source for the cosmology. The equations determining $V(\phi)$ are now given by:

$$\phi(\alpha) = \frac{d\alpha}{a} \left( aQ'(1 - Q) - \frac{2aH'}{3} \right)^{1/2}, \quad (11)$$

$$V = \frac{3H^2}{8\pi G} (1 - Q) \left[ 1 + \frac{2aH'}{3} - \frac{aQ'}{(1 - Q)} \right]^{1/2}. \quad (12)$$

Equations (11) and (12) completely solve the problem. Given any $H(\alpha)$, these equations determine $V(\phi)$ and $\phi(\alpha)$ and thus the potential $V(\phi)$. As an example, consider a universe with power law expansion $a = t^\gamma$. If it is populated only by a tachyonic scalar field, then $Q = 0$; further, $(aH'/H)$ in eq. (11) is a constant making $\phi$ a constant. The complete solution is then given by

$$\phi(t) = \left( \frac{2}{5n} \right)^{1/2} t + \phi_0; \quad V(t) = \frac{3n^2}{8\pi G} \left( 1 - \frac{2}{3n} \right)^{1/2} \frac{1}{t^2}. \quad (13)$$

where $n > (2/3)$. Combining the two, we find the potential to be

$$V(\phi) = \frac{n}{4\pi G} \left( 1 - \frac{2}{3n} \right)^{1/2} (\phi - \phi_0)^{-2}. \quad (14)$$

For such a potential, it is possible to have arbitrarily rapid expansion with large $n$. (For the cosmological model, based on this potential, see ref. 20.) A wide variety of phenomenological models with time-dependent cosmological...
constant have been considered in the literature, all of which can be mapped to a scalar field model with a suitable $V(\phi)$.

While the scalar field models enjoy considerable popularity (one reason being they are easy to construct!) it is very doubtful whether they have helped us to understand the nature of the dark energy at any deeper level. These models, viewed objectively, suffer from several shortcomings:

- They completely lack predictive power. As explicitly demonstrated above, virtually every form of $a(t)$ can be modelled by a suitable ‘designer’ $V(\phi)$.
- These models are non-predictive and degenerate in another sense. The previous discussion illustrates that even when $w(a)$ is known/specified, it is not possible to proceed further and determine the nature of the scalar field Lagrangian. The explicit examples given above show that there are at least two different forms of scalar field Lagrangians (corresponding to the quintessence or the tachyonic field) which could lead to the same $w(a)$. (See ref. 8 for an explicit example of such a construction.)
- All the scalar field potentials require fine tuning of the parameters in order to be viable. This is obvious in the quintessence models in which adding a constant to the potential is the same as invoking a cosmological constant. So to make the quintessence models work, we first need to assume the cosmological constant is zero. These models, therefore, merely push the cosmological constant problem to another level, making it somebody else’s problem!
- By and large, the potentials used in the literature have no natural field theoretical justification. All of them are non-renormalizable in the conventional sense and have to be interpreted as a low energy effective potential in an adhoc manner.
- One key difference between cosmological constant and scalar field models is that the latter lead to a $w(a)$ which varies with time. If observations have demanded this, or even if observations have ruled out $w = -1$ at the present epoch, then one would have been forced to take alternative models seriously. However, all available observations are consistent with cosmological constant ($w = -1$) and – in fact – the possible variation of $w$ is strongly constrained\textsuperscript{29} as shown in Figure 2. (Also see ref. 29).

Given this situation, we shall now take a more serious look at the cosmological constant as the source of dark energy in the universe.

III. …For the Snark was the Boojum, you see

If we assume that the dark energy in the universe is due to a cosmological constant then we are introducing a second length scale, $L_\Lambda = H_\Lambda^2$, into the theory (in addition to the Planck length $L_P$) such that $(L_P/L_\Lambda) \approx 10^{-60}$. Such a universe will be asymptotically deSitter with $a(t) \propto \exp(t = L_\Lambda)$ at late times. We will now explore several peculiar features of such a universe.

Figure 3 summarizes these features\textsuperscript{30,31}. Using the characteristic length scale of expansion, the Hubble radius $d_H \equiv (a/a_0)^{-1}$, we can distinguish between three different phases of such a universe. The first phase is when the universe went through an inflationary expansion with $d_H = \text{constant}$; the second phase is the radiation/matter dominated phase in which most of the standard cosmology operates and $d_H$ increases monotonically; the third phase is that of re-inflation (or accelerated expansion) governed by the cosmological constant in which $d_H$ is again a constant. The first and last phases are time translation invariant (that is $t \to t + \text{constant}$ is an (approximate) invariance for the universe in these two phases); the universe satisfies the perfect cosmological principle and is in steady state during these phases!

In fact, one can easily imagine a scenario in which the two deSitter phases (first and last) are of arbitrarily long duration\textsuperscript{32}. If $\Omega_\Lambda = 0.7, \Omega_{DM} = 0.3$ the final deSitter phase does last forever; as regards the inflationary phase, nothing prevents it from lasting for arbitrarily long duration. Viewed from this perspective, the in between phase – in which most of the ‘interesting’ cosmological phenomena occur – is of negligible measure in the span of time. It merely connects two steady state phases of the universe.

Given the two length scales $L_p$ and $L_\Lambda$, one can construct two energy scales $\rho_p = 1/L_p^2$ and $\rho_\Lambda = 1/L_\Lambda^2$ in natural units.
The geometrical structure of a universe with two length scales $L_P$ and $L_\Lambda$ corresponding to the Planck length and the cosmological constant. Such a universe spends most of its time in a deSitter phase which is (approximately) time translation invariant. The first deSitter phase corresponds to the inflation and the second corresponds to the accelerated expansion arising from the cosmological constant. Most of the perturbations generated during the inflation will leave the Hubble radius (at A) and re-enter (at B). However, perturbations which exit the Hubble radius earlier than C will never re-enter the Hubble radius, thereby introducing a specific dynamic range CE during the inflationary phase. The epoch F is characterized by the redshifted CMB temperature becoming equal to the deSitter temperature ($H_\Lambda/2\pi^2$) which introduces another dynamic range DF in the accelerated expansion after which the universe is dominated by vacuum noise of the deSitter spacetime.

Let us first discuss the case of re-inflation in the late universe. As the universe grows exponentially in the phase 3, the wavelength of CMBR photons is being red shifted rapidly. When the temperature of the CMBR radiation drops below the deSitter temperature (which happens when the wavelength of the typical CMBR photon is stretched to the $L_\Lambda$), the universe will be essentially dominated by the vacuum thermal noise of the deSitter phase. This happens at the point marked F. From the equation $T_0(a_F/a_0) = (1/2\pi L_\Lambda)$ we find that the dynamic range of DF is

$$\frac{a_F}{a_0} = 2\pi T_0 L_\Lambda \left( \frac{\Omega_\Lambda}{\Omega_{DM}} \right)^{1/3} = 3 \times 10^{30}. \quad (15)$$

Interestingly enough, one can also impose a similar bound on the physically relevant duration of inflation. We know that the quantum fluctuations generated during this inflationary phase could act as seeds of structure formation in the universe$^{12}$. Consider a perturbation at some given wave-
length scale which is stretched with the expansion of the universe as $\lambda \propto a(t)$. (See the line marked AB, in Figure 3.) During the inflationary phase, the Hubble radius remains constant while the wavelength increases, so that the perturbation will ‘exit’ the Hubble radius at some time (point A in Figure 3). In the radiation dominated phase, the Hubble radius $d_H \propto t \propto a^2$ grows faster than the wavelength $\lambda \propto a(t)$. Hence, normally, the perturbation will ‘re-enter’ the Hubble radius at some time (point B in Figure 3). If there was no re-inflation, this would make all wavelengths re-enter the Hubble radius sooner or later. But if the universe undergoes re-inflation, then the Hubble radius ‘flattens out’ at late times and some of the perturbations will never re-enter the Hubble radius! The limiting perturbation which just ‘grazes’ the Hubble radius as the universe enters the re-inflationary phase is shown by the line marked CD in Figure 3. If we use the criterion that we need the perturbation to re-enter the Hubble radius, we get a natural bound on the duration of inflation which is of practical relevance. This portion of the inflationary regime is marked by CE and can be calculated as follows: Consider a perturbation which leaves the Hubble radius ($H_{in}^0$) during the inflationary epoch at $a = a_i$. It will grow to the size $H_{in}^0(a_f)$ at a later epoch. We want to determine $a_f$ such that this length scale grows to $L_{\Lambda}$ just when the dark energy starts dominating over matter; that is at the epoch $a = a_\Lambda = a_\Lambda(\Omega_{DM}/\Omega_{\Lambda})^{1/3}$. This gives $H_{in}^0(a_\Lambda) = L_{\Lambda}$ so that $a_\Lambda = (H_{in}^0/L_{\Lambda})(\Omega_{DM}/\Omega_{\Lambda})^{1/3}a_0$. On the other hand, the inflation ends at $a = a_{end}$ where $a_{end}/a_0 = T_{reh}/T_{reh}$ where $T_{reh}$ is the temperature to which the universe has been reheated at the end of inflation. Using these two results we can determine the dynamic range of CE to be

$$\frac{a_{end}}{a_i} = \left( \frac{T_0 L_{\Lambda}}{T_{reh} H_{in}^{-1}} \right) \left( \frac{\Omega_\Lambda}{\Omega_{DM}} \right)^{1/3} = \frac{(a_f/a_\Lambda)}{2\pi T_{reh} H_{in}^{-1}} \approx 10^{25}.$$

(16)

where we have used the fact that, for a GUTs scale inflation with $E_{GUT} = 10^{14}$ GeV, $T_{reh} = E_{GUT}$, $\rho_{in} = E_{GUT}$ we have $2\pi H_{in}^{-1} T_{reh} = (3\pi/2)^{1/2}(E_f/E_{GUT}) = 10^5$. For a Planck scale inflation with $2\pi H_{in}^{-1} T_{reh} = O(1)$, the phases CE and DF are approximately equal. The region in the quadrilateral CEDF is the most relevant part of standard cosmology, though the evolution of the universe can extend to arbitrarily large stretches in both directions in time. It is possible that this figure is telling us something regarding the time translation invariance of the universe (‘the perfect cosmological principle’) and – more importantly – about the breaking of this symmetry, but it is not easy to translate this concept into a workable theory.

Let us now turn our attention to few of the many attempts to understand the cosmological constant. This is of course a non-representative sample and a host of other approaches exist in the literature.

Dark energy from a nonlinear correction term

One of the least esoteric ideas regarding the dark energy is the realization that the cosmological constant term in the FRW equations arises because we have not calculated the energy density driving the expansion of the universe correctly. The motivation for such a suggestion arises from the following fact: The energy momentum tensor of the real universe, $T_{ab}(t, \mathbf{x})$ is inhomogeneous and anisotropic and will lead to a very complex metric $g_{ab}$ if only we could solve the exact Einstein’s equations $G_{ab}[\rho] = \kappa T_{ab}$. The metric describing the large scale structure of the universe should be obtained by averaging this exact solution over a large enough scale, to get $\langle g_{ab} \rangle$. But what we actually do is to average the stress tensor first to get $\langle T_{ab} \rangle$ and then solve Einstein’s equations. But since $G_{ab}[\rho]$ is nonlinear function of the metric, $\langle G_{ab}[\rho] \rangle \neq G_{ab}[\langle \rho \rangle]$ and there is a discrepancy. This is most easily seen by writing

$$G_{ab}[\langle \rho \rangle] = \kappa [\langle T_{ab} \rangle] + \kappa^{-1} \langle G_{ab}[\rho] \rangle - \langle G_{ab}[\langle \rho \rangle] \rangle.$$

(17)

If – based on observations – we take the $\langle g_{ab} \rangle$ to be the standard Friedman metric, this equation shows that it has, as its source, two terms: The first is the standard average stress tensor and the second is a purely geometrical correction term $\langle T_{ab} \rangle = \kappa^{-1} \langle G_{ab}[\rho] \rangle - \langle G_{ab}[\langle \rho \rangle] \rangle$ which arises because of nonlinearities in the Einstein’s theory that leads to $\langle G_{ab}[\rho] \rangle \neq G_{ab}[\langle \rho \rangle]$. If this term can mimic the cosmological constant at large scales, there will be no need for dark energy! Unfortunately, it is not easy to settle this question to complete satisfaction. One possibility is to use some analytic approximations to nonlinear perturbations (usually called nonlinear scaling relations, see e.g. ref. 35) to estimate this term. This does not lead to a stress tensor that mimics dark energy (Padmanabhan, unpublished) but this is not a conclusive proof either way. We mention this mainly because this problem deserves more attention than it has got.

Unimodular gravity

Another possible way of addressing this issue is to simply eliminate from the gravitational theory those modes which couple to cosmological constant. If, for example, we have a theory in which the source of gravity is $(p + p)$ rather than $(p + 3p)$ in eq. (2), then cosmological constant will not couple to gravity at all. (The nonlinear coupling of matter with gravity has several subtleties; see e.g. ref. 36.) Unfortunately it is not possible to develop a covariant theory of gravity using $(p + p)$ as the source. But we can achieve the same objective in different manner. Any metric $g_{ab}$ can be expressed in the form $g_{ab} = f^2(\lambda)q_{ab}$ such that $\det q = 1$ so that $\det g = f^4$. From the action functional for gravity
It is obvious that the cosmological constant couples only to the conformal factor $f$. So if we consider a theory of gravity in which $f^4 = (-g)^{12}$ is kept constant and only $q_{ab}$ is varied, then such a model will be oblivious of direct coupling to cosmological constant. If the action (without the $\Lambda$ term) is varied, keeping $\det g = -1$, say, then one is lead to a unimodular theory of gravity with the equations of motion $R_{ab} - (1/4)g_{ab}R = \kappa (T_{ab} - (1/4)g_{ab}T)$ with zero trace on both sides. Using the Bianchi identity, it is now easy to show that this is equivalent to a theory with an arbitrary cosmological constant. That is cosmological constant arises as an undetermined integration constant in this model. While this is interesting, we need an extra physical principle to fix its value.

One possible way of doing this is to interpret the $\Lambda$ term in the action as a Lagrange multiplier for the proper volume of the spacetime. Then it is reasonable to choose the cosmological constant such that the total proper volume of the universe is equal to a specified number. While this will lead to a cosmological constant which has the correct order of magnitude, it has several obvious problems. First, the proper four volume of the universe is infinite unless we make the spatial sections compact and restrict the range of time integration. Second, this will lead to a dark energy density which varies as $f^{-2}$ (corresponding to $w = -1/3$) which is ruled out by observations.

Scale dependence of the vacuum energy

The conventional discussion of the relation between cosmological constant and vacuum energy density is based on evaluating the zero point energy of quantum fields with an ultraviolet cutoff and using the result as a source of gravity. Any reasonable cutoff will lead to a vacuum energy density $\rho_{\text{vac}}$ which is unacceptably high. This argument, however, is too simplistic since the zero point energy -- obtained by summing over the $(1/2)\hbar \omega_k$ -- has no observable consequence in any other phenomena and can be subtracted out by redefining the Hamiltonian. The observed non trivial features of the vacuum state of QED, for example, arise from the fluctuations (or modifications) of this vacuum energy rather than the vacuum energy itself. This was, in fact, known fairly early in the history of cosmological constant problem and, in fact, is stressed by Zel’dovich who explicitly calculated one possible contribution to fluctuations after subtracting away the mean value. This suggests that we should consider the fluctuations in the vacuum energy density in addressing the cosmological constant problem.

If the vacuum probed by the gravity can readjust to take away the bulk energy density $\rho_B \approx L_p^3$, quantum fluctuations can generate the observed value $\rho_{\text{DE}}$. One of the simplest models which achieves this uses the fact that, in the semiclassical limit, the wave function describing the universe of proper four-volume $V$ will vary as $\psi \propto \exp[-i(\lambda_{DE}/V\Delta\delta)]$. If we treat $(\lambda_{DE}/V, V)$ as conjugate variables then uncertainty principle suggests $\Delta\lambda \approx \Delta V \approx 1/\sqrt{V} = H_0$. (This idea can be a more quantitative; see ref. 39.)

Similar viewpoint arises, more formally, when we study the question of detecting the energy density using gravitational field as a probe. Recall that an Unruh–DeWitt detector with a local coupling $L = M(\tau)\phi(x,\tau)$ to the field $\phi$ actually responds to the two-point function $\langle 0|T_{ab}(x)T_{cd}(y)|0 \rangle$ rather than to the field itself. Similarly, one can use the gravitational field as a natural ‘detector’ of energy momentum tensor $T_{ab}$ with the standard coupling $L = \kappa h_{ab} T^{ab}$. Such a model was analysed in detail in ref. 41 and it was shown that the gravitational field responds to the two-point function $\langle 0|T_{ab}(x)T_{cd}(y)|0 \rangle$. In fact, it is essentially this fluctuations in the energy density which is computed in the inflationary models as the seed source for gravitational field, as stressed in ref. 13. All these suggest treating the energy fluctuations as the physical quantity ‘detected’ by gravity, when one needs to incorporate quantum effects. If the cosmological constant arises due to the energy density of the vacuum, then one needs to understand the structure of the quantum vacuum at cosmological scales. Quantum theory, especially the paradigm of renormalization group, has taught us that the energy density -- and even the concept of the vacuum state -- depends on the scale at which it is probed. The vacuum state which we use to study the lattice vibrations in a solid, say, is not the same as vacuum state of the QED. Using this feature, it is possible to construct systems in condensed matter physics wherein the quantity analogous to vacuum energy density has to vanish on the average because of dynamical reasons.

In fact, it seems inevitable that in a universe with two length scale $L_\Lambda, L_p$, the vacuum fluctuations will contribute an energy density of the correct order of magnitude $\rho_{\text{DE}} = (\rho_A \rho_B)^{1/2}$. The hierarchy of energy scales in such a universe has the pattern

$$\rho_{\text{vac}} = \frac{1}{L_p^4} + \frac{1}{L_p^4} \left( \frac{L_p}{L_\Lambda} \right)^2 + \frac{1}{L_p^4} \left( \frac{L_p}{L_\Lambda} \right)^4 + \ldots$$

The first term is the bulk energy density which needs to be renormalized away (by a process which we do not understand at present); the third term is just the thermal energy density of the deSitter vacuum state; what is interesting is
enough, a rigorous calculation which is provided by the cosmic horizon. Remarkably of exhibit fluctuations cosmological horizon which blocks information, the natural developed more generally in ref. 44.) If the spacetime has a without specifying this scale. (These ideas have been de-
violet cutoff is imposed. Similar calculations have been done
that quantum fluctuations in the matter fields inevitably generate the second term.

The key new ingredient arises from the fact that the properties of the vacuum state depends on the scale at which it is probed and it is not appropriate to ask questions without specifying this scale. (These ideas have been developed more generally in ref. 44.) If the spacetime has a cosmological horizon which blocks information, the natural developments in ref. 44.) If the spacetime has a without specifying this scale. (These ideas have been de-

The operator \( H(< L_\Lambda) \), corresponding to the total energy inside a region bounded by a cosmological horizon, will exhibit fluctuations \( \Delta \) since vacuum state is not an eigenstate of this operator. The corresponding fluctuations in the energy density, \( \Delta \rho \propto (\Delta E)/L_\Lambda = f(L_\rho, L_\Lambda) \) will depend on both the ultraviolet cutoff \( L_\rho \) as well as \( L_\Lambda \). To obtain \( \Delta \rho_{\text{vac}} \propto \Delta E/L_\Lambda^2 \) which scales as \((L_\rho L_\Lambda)^{-2}\) we need to have \((\Delta E)^2 \propto L_\rho^2 L_\Lambda^2 \); that is, the square of the energy fluctuations should scale as the surface area of the bounding surface which is provided by the cosmic horizon. Remarkably enough, a rigorous calculation\(^5\) of the dispersion in the energy shows that for \( L_\Lambda \gg L_\rho \), the final result indeed has the scaling

\[
(\Delta E)^2 = c_1 \frac{L_\Lambda^2}{L_\rho},
\]

(20)

where the constant \( c_1 \) depends on the manner in which ultraviolet cutoff is imposed. Similar calculations have been done (with a completely different motivation, in the context of entanglement entropy) by several people and it is known that the area scaling found in eq. (20), proportional to \( L_\Lambda^2 \), is a generic feature\(^5\). For a simple exponential UV-cutoff, \( c_1 = (1/30\pi^2) \) but cannot be computed reliably without knowing the full theory. We thus find that the fluctuations in the energy density of the vacuum in a sphere of radius \( L_\Lambda \) are given by

\[
\Delta \rho_{\text{vac}} = \frac{(\Delta E)}{L_\Lambda} \propto L_\rho^{-2} L_\Lambda^{-2} \propto \frac{H_\Lambda^3}{G}.
\]

(21)

The numerical coefficient will depend on \( c_1 \) as well as the precise nature of infrared cutoff radius (like whether it is \( L_\rho \) or \( L_\rho/2\pi \), etc.). It would be pretentious to cook up the factors to obtain the observed value for dark energy density. It is a fact of life that a fluctuation of magnitude \( \Delta \rho_{\text{vac}} \approx H_\Lambda^3/G \) will exist in the energy density inside a sphere of radius \( H_\Lambda t \) if Planck length is the UV cut off. One cannot get away from it. On the other hand, observations suggest that there is a \( \rho_{\text{vac}} \) of similar magnitude in the universe. It seems natural to identify the two, after subtracting out the mean value by hand. Our goal was more towards explaining why there is a surviving cosmological constant which satisfies \( \rho_{\text{DE}} = (\rho_0 \rho_{\text{vac}})^{1/2} \) which—in our opinion—is the problem.

There is a completely different way of interpreting this result based on some imaginative ideas suggested by Bjorken\(^3\) recently. The key idea is to parametrize the universes by the value of \( L_\Lambda \) which they have. It is a fixed, pure number for each universe in an ensemble of universes but all the other parameters of the physics are assumed to be correlated with \( L_\Lambda \). This is motivated by a series of arguments in ref. 31 and, in this approach, \( \rho_{\text{vac}} \approx L_\Lambda^{-2} \) almost by definition; the hard work was in determining how other parameters scale with \( L_\Lambda \). In the approach suggested here, a dynamical interpretation of the scaling \( \rho_{\text{vac}} \approx L_\Lambda^{-2} \) is given as due to vacuum fluctuations of fields. We now reinterpret each member of the ensemble of universes as having zero energy density for vacuum (as any decent vacuum should have) but the effective \( \rho_{\text{vac}} \) arises from the quantum fluctuations with the correct scaling. One can then invoke standard anthropic-like arguments (but with very significant differences as stressed in ref. 31) to choose a range for the size of our universe. This appears to be much more attractive way of interpreting the result.

Finally, to be fair, this attempt should be judged in the backdrop of other suggested solutions almost all of which require introducing extra degrees of freedom in the form of scalar fields, modifying gravity or introducing higher dimensions etc. and fine tuning the potentials. At a fundamental level such approaches are unlikely to provide the final solution.

3. See Subramanian, K., this volume for a discussion of theoretical aspects of CMBR.
5. For a critical discussion of the evidence, see Peebles, P. J. E., astro-ph/0410284.
SPECIAL SECTION: COSMOLOGY


36. Padmanabhan, T. [gr-qc/0409089].


43. Padmanabhan, T. [hep-th/0406060].


Acknowledgements. I thank K. Subramanian for two decades of discussion about various aspects of cosmological constant and for sharing the view that any quick-fix solution to this problem will be futile. I also thank Apoorva Patel for useful discussions.