Preprint typeset in JHEP style - HYPER VERSION

PRELIMINARY DRAFT

Dark Energy and the Curious Symbiosis Between Micro-physics and Cosmology (Naturally)*

C.P. Burgess

Department of Physics & Astronomy, McMaster University,
1280 Main Street West, Hamilton, Ontario, Canada, L8S 4M1.
Perimeter Institute for Theoretical Physics,
31 Caroline Street North, Waterloo, Ontario, Canada, N2L 2Y5.
School of Theoretical Physics, Dublin Institute for Advanced Studies,
10 Burlington Rd., Dublin, Co. Dublin, Ireland.

ABSTRACT: These notes aim to highlight the remarkable symbiosis that currently exists between the physics of the very small and the physics of the very large, using the unsolved puzzle of the nature of Dark Energy as a vehicle for so doing. The notes first summarize what we know observationally about the properties of Dark Energy (and the Dark sector more broadly) and then discuss several approaches to explain them. Along the way this involves determining the types of interactions that would on general grounds be expected to be present in the low-energy limit of fundamental theories involving the many hierarchy of scales we see around us. This includes (but is not limited to) a discussion of technical naturalness (and ‘t Hooft naturalness) as well as the arguments for their use as a criterion for distinguishing amongst candidate theories. Some recent approaches I find promising are briefly summarized at the end.

*Lectures given to the Les Houches Summer School “Dark Universe,” 7 July - 1 August 2025

Contents

1. The Facts in the Sky

These notes use a discussion of Dark Energy as a vehicle for illustrating the peculiarly effective symbiosis that currently exists between our understandings of physics at the smallest and largest scales. This symbiosis is peculiar because it seems to fly in the face of an important fact of Nature: decoupling.

Decoupling states that details of small-distance physics tend not to be important for understanding long-distance physics. This is indeed partially why science makes progress at all – although Nature comes to us with many scales we are not required to understand them all at once. This is why it was possible to figure out how atoms work before also understanding the nature of the atomic nucleus. It turns out that atomic physics mostly depends only on a few nuclear properties – its charge and mass and spin, for example – but not on the rest of the nuclear nitty gritty. This is also why it is not that surprising that the Standard Model of particle physics gets right all of the details of (say) condensed matter physics or of quantum optics. *Any* theory of micro-physics that properly predicts Quantum Electrodynamics at low energies automatically gets all condensed matter and optical phenomena right for free. This is a good thing because it means our understanding of the properties of matter in bulk or of light in matter is robust to changes to our understanding of currently unknown microphysics.

The situation is different in cosmology, where different micro-physical theories can differ radically in their cosmological implications and the observational success or failure of these implications are often used to constrain what might be possible at the shortest distances to which we have access. Perhaps even more interesting: many popular models in cosmology seem not to be obtainable from sensible micro-physics – if true this might be a useful clue that allows us to choose amongst the very many models on the market.

1.1 Vanilla cosmology

Let's start with a very brief recap of cosmology basics (a classic textbook for this is [1]). The vast majority of cosmological models start with the premise that the geometry of the Universe around us can be described by the classical solutions to Einstein equations of General Relativity (GR):[1]

$$\mathcal{R}_{\mu\nu} - \tfrac{1}{2}\,\mathcal{R}\,g_{\mu\nu} + \kappa^2 T_{\mu\nu} = 0\,. \tag{1.1}$$

Here $\mathcal{R}_{\mu\nu}$ is the Ricci tensor built from the spacetime metric, $g_{\mu\nu}$, while $\mathcal{R} = g^{\mu\nu}\mathcal{R}_{\mu\nu}$ is its Ricci scalar and $T_{\mu\nu} = T_{\nu\mu}$ is the stress energy tensor of all of the forms of matter that are currently present (or were present in the past). The parameter $\kappa^2 = 8\pi G_N$ denotes the gravitational coupling, where G_N is Newton's constant of universal graviation. In fundamental units (for which $\hbar = c = 1$) its value defines the (reduced) Planck mass: $M_p = \kappa^{-1}$.

The combination $\mathcal{G}_{\mu\nu} = \mathcal{R}_{\mu\nu} - \tfrac{1}{2}\,\mathcal{R}\,g_{\mu\nu}$ satisfies a well-known *Bianchi identity*: $\nabla^\mu \mathcal{G}_{\mu\nu} = 0$, where ∇_μ is the covariant derivative built from the metric. Consistency requires that whatever the matter is that is present, its total stress energy must be covariantly conserved: $\nabla^\mu T_{\mu\nu} = 0$.

[1]We denote spacetime coordinates by $x^\mu = \{x^0, x^i\} = \{x^0 = t, x^1 = x, x^2 = y, x^3 = z\}$ and choose the metric signature $(-+++)$ together with Weinberg's curvature conventions [2] (which differ from those of Misner, Thorne and Wheeler [3] – more commonly used in the relativity community – only in the overall sign of the Riemann tensor.

In cosmology it often happens that the matter of interest is a homogeneous and isotropic fluid whose elements move through spacetime with 4-velocity $u^\mu(x)$. As is true for any 4-velocity, $u^\mu(x)$ must satisfy $g_{\mu\nu}u^\mu u^\nu = -1$ and so the fluid rest frame is defined as the frame where the spatial components satisfy $u^i = 0$ (and so in this frame $u^0 = |g_{00}|^{-1/2}$). Denoting the fluid's rest-framepressure and energy density by p and ρ respectively, the fluid's stress-energy to be used in (1.1) is

$$T_{\mu\nu} = p\, g_{\mu\nu} + (p+\rho)\, u_\mu u_\nu \qquad \text{(fluid)}\,. \tag{1.2}$$

In the special case where the universe is homogeneous and isotropic the spacetime metric can always be written in the Friedmann, LeMaitre, Robertson, Walker (FLRW) form

$$\begin{aligned} \mathrm{d}s^2 &= -\mathrm{d}t^2 + a^2(t)\,\left[\frac{\mathrm{d}r^2}{1-\mathfrak{K}r^2/R_0^2} + r^2\,\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2\right] \\ &= -\mathrm{d}t^2 + a^2(t)\,\left[\mathrm{d}\ell^2 + r^2(\ell)\,\mathrm{d}\theta^2 + r^2(\ell)\sin^2\theta\,\mathrm{d}\phi^2\right]\,, \end{aligned} \tag{1.3}$$

where R_0 is a constant and $\mathfrak{K}$ can take one of the following three values: $\mathfrak{K} = 1, 0, -1$. The coordinate ℓ is related to r by $\mathrm{d}\ell = \mathrm{d}r/(1-\mathfrak{K}r^2/R_0^2)^{1/2}$, and so

$$r(\ell) = \begin{cases} R_0\,\sin(\ell/R_0) & \text{if} \quad \mathfrak{K} = +1 \\ \ell & \text{if} \quad \mathfrak{K} = 0 \\ R_0\,\sinh(\ell/R_0) & \text{if} \quad \mathfrak{K} = -1\,. \end{cases} \tag{1.4}$$

The geometry at fixed t is in this case a 3-sphere when $\mathfrak{K} = 1$, flat space (when $\mathfrak{K} = 0$) or a hyperbolic space ($\mathfrak{K} = -1$). It is usually convenient to rescale $\ell \to R_0\ell$ when $\mathfrak{K} = \pm 1$. For $\mathfrak{K} = 0$ it is convenient instead to rescale ℓ to ensure that the scale factor is unity at a particular time, $a(t_0) = 1$ (with the particular time chosen to be now). These rescalings amount to choosing convenient units of length

With these choices the fundamental evolution equation (1.1) boils down to two independent differential equations relating $a(t)$ to $\rho(t)$ and $p(t)$. These may be chosen to be the *Friedmann equation*,

$$H^2 + \frac{\mathfrak{K}}{a^2} = \frac{8\pi G}{3}\,\rho\,, \tag{1.5}$$

as well as the equation describing the *Conservation of Stress-Energy* ($\nabla^\mu T_{\mu\nu} = 0$),

$$\dot\rho + 3H(\rho + p) = 0\,. \tag{1.6}$$

In these expressions over-dots denote differentiation with respect to t and the Hubble function is defined by $H(t) = \dot a/a$. Equation (1.6) has an intuitive interpretation if it is rewritten

$\mathrm{d}(\rho\, a^3) + p\, \mathrm{d}(a^3) = 0$, which relates the rate of change of the total energy, $\rho\, a^3$, to the work done by the pressure as the universe expands. For a thermodynamic fluid this is consistent with the First Law of Thermodynamics when the evolution is at constant entropy.

Eqs. (1.5) and (1.6) provide two differential equations for the three unknown functions $\rho(t)$, $p(t)$ and $a(t)$ and so can only be fully integrated after more information is provided. Typically this information comes from identifying the types of matter making up the fluid. Any specific type of fluid – a gas of photons, for example, or nonrelativistic electrons – has an equation of state: a relation relating ρ to p. Once an equation of state is specified there is enough information to integrate eqs. (1.5) and (1.6) to obtain the histories $a(t)$, $p(t)$ and $\rho(t)$.

For instance, if it happens that the equation of state has the commonly occuring form

$$p = w\, \rho\,, \tag{1.7}$$

where w is a t-independent constant then eq. (1.6) integrates to give

$$\rho = \rho_0 \left(\frac{a_0}{a} \right)^{\sigma} \qquad \text{with} \quad \sigma = 3(1+w)\,. \tag{1.8}$$

In the special case that $\mathfrak{K} = 0$ this allows eq. (1.5) to be integrated to give

$$a(t) = a_0 \left(\frac{t}{t_0} \right)^{\alpha} \qquad \text{with} \quad \alpha = \frac{2}{\sigma} = \frac{2}{3(1+w)}\,. \tag{1.9}$$

1.1.1 ΛCDM

The core theory of Hot Big Bang cosmology postulates that all ordinary matter starts off in the remote past as a hot dense fluid, and then asks what evidence for this exists in the later universe. It turns out it does: as the universe expands it cools and bound states form as the temperature falls below the relevant binding energy. The formation of nuclei leads to the successful Big Bang nucleosynthesis prediction for the abundances of light elements; atom formation leads to the universe becoming transparent and the associated Cosmic Microwave Background (CMB) relic radiation, and so on.

All told, a minimal successful description of cosmological observation requires four main types of components to the cosmic fluid, each of which (in the later universe at least) does not exchange energy with the others – so their stress-energies are individually conserved and satisfy (1.6) separately.

- **Radiation:** Photons (and neutrinos) are relativistic through (most of) the universe's history and so turn out to have a pressure-to-energy-density ratio of $w_{\rm rad} \simeq \frac{1}{3}$ (we collectively call such species 'radiation'). Eq. (1.8) then implies $\rho_{\rm rad}(a)/\rho_{\rm rad\,0} = (a_0/a)^4$.

- **Ordinary Matter (baryons):** Ordinary matter is nonrelativistic for much of the epoch to which we have observational access (electrons and neutrinos are the exception for earlier parts of the universal history). Since p/ρ involves the ratio of some measure of particle kinetic energy (like temperature) over rest mass we have $w_b \simeq 0$ for nonrelativistic species and (1.8) implies $\rho_b(a)/\rho_{b0} = (a_0/a)^3$. Because of electromagnetic interactions this fluid is only uncoupled from the radiation fluid in the relatively late universe (after neutral atoms are able to form).

- **Cold Dark Matter:** There is considerable evidence for the existence of another fluid that behaves gravitationally much like baryons do (*i.e.* it clumps together in galaxies and clusters of galaxies due to gravitational attraction) but which does not otherwise interact with ordinary matter. If this fluid describes the bulk behaviour of a new type of matter then this matter must be moving slowly – *i.e.* be 'cold' – in order to clump sufficiently efficiently, and so is also well-described as a fluid with a nonrelativistic equation of state parameter $w_c \simeq 0$. As a result its density also falls in an expanding universe like $\rho_c(a)/\rho_{c0} = (a_0/a)^3$.

- **Vacuum Energy:** There is good evidence the vacuum is Lorentz invariant to high accuracy and so its stress energy tensor must be proportional to the metric:[2]

$$T^{\mu\nu}_{\rm vac} = -\rho_{\rm vac}\, g^{\mu\nu} \,. \tag{1.10}$$

  Conservation of stress energy ($\nabla_\mu T^{\mu\nu}_{\rm vac} = 0$) then implies $\rho_{\rm vac}$ must be a constant. Because it is a constant it contributes to Einstein's equations (1.1) in the same way as would Einstein's cosmological constant term $\Lambda g_{\mu\nu}$. Comparing (1.10) with the general fluid stress-energy (1.2) then shows that $p_{\rm vac} = -\rho_{\rm vac}$ and so the equation of state parameter is $w_{\rm vac} = -1$. In this case we have $\rho(a) = \rho_{\rm vac}$ is independent of a. Because either $\rho_{\rm vac}$ or $p_{\rm vac}$ must negative (observations say $\rho_{\rm vac} > 0$ and so it is $p_{\rm vac}$ that is negative) this is distinct from Dark Matter, and so is given its own name: Dark Energy.

The above fluids are part of the definition of the ΛCDM model of cosmology, and taken together they imply the relative abundance of the different fluid components changes as the universe expands (see Fig. 1). In particular the total energy density and pressure have the

[2]The vacuum is usually meant as the lowest-energy state but there is no reason it has to have zero energy density, particularly for $\langle T_{\mu\nu}\rangle$ evaluated in a quantum vacuum state. More about quantum effects below.

form

$$\begin{aligned} \rho(a) &= \rho_{\rm vac} + \rho_{\rm m0}\left(\frac{a_0}{a}\right)^3 + \rho_{\rm rad\,0}\left(\frac{a_0}{a}\right)^4 \\ p(a) &= -\rho_{\rm vac} + \tfrac{1}{3}\,\rho_{\rm rad\,0}\left(\frac{a_0}{a}\right)^4 , \end{aligned} \tag{1.11}$$

if the energy exchange between fluids is neglible. Here $\rho_{\rm m0} := \rho_{b0} + \rho_{c0}$ sums the contributions of the two types of nonrelativistic fluids (which is only appropriate after the baryons have decoupled from the radiation). Using the above expression for $\rho(a)$ in the Friedmann equation (1.5) gives $H = \dot{a}/a$ as a function of a, which can be integrated to get $a(t)$. Comparing to (1.9) shows this implies in particular that $a \propto t^{1/2}$ when radiation dominates the energy density ('radiation domination') and $a \propto t^{2/3}$ when nonrelativistic matter dominates ('matter domination').

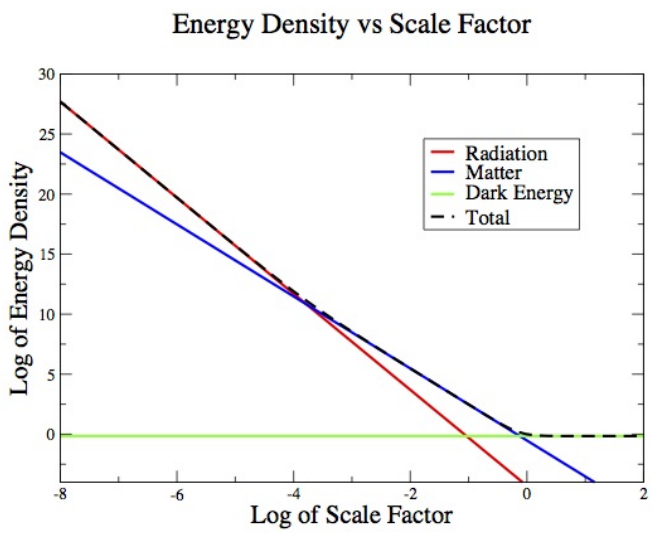

Figure 1: $\log \rho(a)$ as given in (1.11) vs a with realistic choices for the present-day energy densities (with present-day defined by $a(t_0) = 1$).

For a given Hubble parameter, H, it is conventional to define the critical density by $\rho_{\rm crit}(a) := 3H^2/(8\pi G_N)$. Given the current measurement $H_0 \simeq 70$ km/sec/Mpc, the critical density's numerical value today becomes $\rho_{\rm crit0} \simeq 9 \times 10^{-30}$ g/cm^3. $\rho_{\rm crit}$ is defined this way because the Friedmann equation becomes[3]

$$H^2 + \frac{\mathfrak{K}}{a^2} = \frac{8\pi G_N}{3}\,\rho \qquad \text{or} \qquad 1 + \frac{\mathfrak{K}}{(aH)^2} = \frac{\rho}{\rho_{\rm crit}} =: \Omega(a) \,, \tag{1.12}$$

[3] It is sometimes convenient to put the $\mathfrak{K}$ term on the right-hand side of the Friedmann equation and define $\Omega_\kappa := -\mathfrak{K}/(aH)^2$ so that the Friemann equation becomes $\Omega_\Lambda + \Omega_{\rm m} + \Omega_{\rm rad} + \Omega_\kappa = 1$.

and so if there should be a time t_0 when $\rho(t_0) = \rho_{\rm crit}(t_0)$ then $\mathfrak{K} = 0$ and so $\rho = \rho_{\rm crit}$ for all times. Similarly if $\mathfrak{K} = +1$ then we must have $\rho > \rho_{\rm crit}$ and if $\mathfrak{K} = -1$ then $\rho < \rho_{\rm crit}$. The last equality of (1.12) defines $\Omega(a) := \rho(a)/\rho_{\rm crit}(a)$: the total energy density in units of this critical density.

Normalizing densities in terms of $\rho_{\rm crit}$ proves to be useful because the best evidence currently is consistent with $\mathfrak{K} = 0$ and so at present $\rho_0 \simeq \rho_{\rm crit0}$ and therefore $\Omega_0 \simeq 1$. When this is true the densities of the cosmic fluid components when normalized to $\rho_{\rm crit}$ – *i.e.* $\Omega_{\rm rad} := \rho_{\rm rad}/\rho_{\rm crit}$, $\Omega_\Lambda := \rho_{\rm vac}/\rho_{\rm crit}$ and $\Omega_{\rm m} := \Omega_b + \Omega_c$ (with $\Omega_b := \rho_b/\rho_{\rm crit}$ and $\Omega_c := \rho_c/\rho_{\rm crit}$) – give their fraction of the total energy density and so should all sum to unity: $\Omega_{\rm vac} + \Omega_{\rm m} + \Omega_{\rm rad} \simeq 1$.

Although we do not pursue this further here, there is also good theoretical reasons why $\Omega = 1$ should be true to a very good approximation: it is what would be expected if the much earlier universe were to have undergone a significant period of accelerated expansion, such as proposed by inflationary models [4] for which it is hypothesized that the scale factor evolves like $a(t) = a_0\, e^{H_I(t-t_0)}$ for roughly 50 e-foldings or more. Such an expansion would quickly drive $\mathfrak{K}/(aH)$ to be extremely small even if $\mathfrak{K}$ were nonzero. Inflationary models are attractive inasmuch as they provide a dynamical explanation for some of the initial conditions that are required for successful description of our later universe, including providing a mechanism for the origins and properties of the primordial density fluctuations that ultimately source the distribution of matter we now find around us [5].

1.1.2 Observations

The above picture proves to be a spectacularly successful description of the universe we see around us. This agreement is even better than the above discussion might suggest because the assumption that the universe is exactly homogeneous and isotropic can be relaxed to follow how perturbations around homogeneity and isotropy evolve in time. Although a full description goes beyond the scope of this survey the result provides a beautifully accurate picture of how the much clumpier distribution of galaxies we see around us now arises from the gravitational attraction of initially very small deviations from homogeneity and isotropy (for textbook discussions see *e.g.* [6]). Agreement with observations determines the various parameters of the cosmology (such as Ω_{c0}, Ω_{b0} and Ω_Λ) to the percent level or better. In particular fits to observations confirm that $\Omega_{c0} \simeq 0.28$ and $\Omega_\Lambda \simeq 0.67$ are nonzero and Ω_κ is consistent with zero (see *e.g.* Fig. 2).

Because Ω_κ is consistent with zero cosmologists define a 6-parameter model – called ΛCDM – for which $\kappa = 0$ is set by hand. The results for the six cosmological parameters is

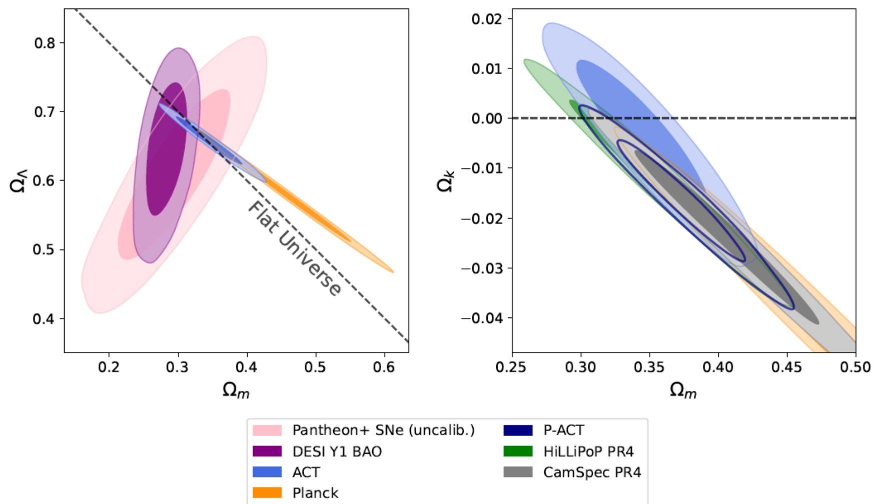

Figure 2: Left panel: Best-fit values for $\Omega_\Lambda := \Omega_{\rm vac}$ vs $\Omega_{\rm m}$ evaluated in the present day where the diagonal line corresponds to $\mathfrak{K} = 0$ (a spatially flat universe). Right panel: Best fits for the present-day values of $\Omega_\kappa = -\mathfrak{K}/(aH)^2$ vs $\Omega_{\rm m}$. Different colours correspond to fits to different data sets. Both figures taken from [7]

then obtained by fitting to observations and given these parameters many other observables can be computed. The results of such a process are listed in Fig. 3, which shows the precision of agreement is currently better than the percent level.

This successful description in particular tells us $\Omega_{b0} \simeq 0.05$ and $\Omega_{\rm rad\,0} \sim 10^{-4} \ll 1$, which allows at most only about 5% of the current energy density to be matter we understand in detail. The remarkable fact that we can describe the universe so accurately while being almost completely ignorant about the fundamental nature of 95% of what is in it is one of the central scientific puzzles of our times.[4] We have direct observational evidence that we are missing something important but with not (yet) enough information to pin down decisively what is going on. Indeed an unusual opportunity.

The rest of these notes explore this opportunity further, but before doing so it is worth reassessing the validity of the big picture. On one hand it is claimed that we have a detailed description of cosmology that is very accurate. On the other hand this detailed description requires the universe to have many unknown ingredients. Perhaps this is really tellng us our overall conceptual framework is flawed. Questions like these have stimulated much study of the foundations on which cosmology is based (some of which is summarized below).

[4]One attitude is that the vacuum having an energy density is not unexpected and so the evidence for Dark Energy is not really that mysterious. Even if so Dark Matter – 28% of what is out there – is still a puzzle. The rest of these notes make the case that interpreting Dark Energy as a vacuum energy contains many puzzles.

Parameter	ACT	Planck	W-ACT	P-ACT	P-ACT-LB
Sampled					
$10^4\theta_{\rm MC}$	104.056 ± 0.031	104.088 ± 0.031	104.066 ± 0.029	104.073 ± 0.025	104.086 ± 0.025
$10^2\Omega_{\rm b}h^2$	2.259 ± 0.017	2.237 ± 0.015	2.263 ± 0.012	2.250 ± 0.011	2.256 ± 0.011
$10^2\Omega_{\rm c}h^2$	12.38 ± 0.21	12.00 ± 0.14	12.20 ± 0.18	11.93 ± 0.12	11.79 ± 0.09
$\log(10^{10}A_{\rm s})$	3.053 ± 0.013	$3.054^{+0.012}_{-0.013}$	$3.057^{+0.010}_{-0.012}$	3.056 ± 0.013	$3.060^{+0.011}_{-0.012}$
$n_{\rm s}$	0.9666 ± 0.0077	0.9651 ± 0.0044	0.9660 ± 0.0046	0.9709 ± 0.0038	0.9743 ± 0.0034
τ [%]	$5.62^{+0.53}_{-0.63}$	$5.90^{+0.55}_{-0.65}$	$5.71^{+0.54}_{-0.64}$	$6.03^{+0.55}_{-0.65}$	$6.32^{+0.55}_{-0.66}$
Derived					
H_0 [km/s/Mpc]	66.11 ± 0.79	67.31 ± 0.61	66.78 ± 0.68	67.62 ± 0.50	68.22 ± 0.36
$\Omega_{\rm m}$ [%]	33.7 ± 1.3	31.58 ± 0.85	32.6 ± 1.1	31.16 ± 0.71	30.32 ± 0.48
Ω_b [%]	5.17 ± 0.12	4.937 ± 0.070	5.075 ± 0.098	4.920 ± 0.063	4.847 ± 0.044
Ω_c [%]	28.3 ± 1.2	26.50 ± 0.78	27.37 ± 0.96	26.10 ± 0.65	25.34 ± 0.44
Ω_Λ [%]	66.3 ± 1.3	68.41 ± 0.85	67.4 ± 1.1	68.83 ± 0.71	69.67 ± 0.48
$10^2\Omega_m h^2$	14.70 ± 0.21	14.31 ± 0.13	14.53 ± 0.18	14.25 ± 0.12	14.11 ± 0.08
$n_{\rm s} - 1$ [%]	-3.34 ± 0.77	-3.49 ± 0.44	-3.40 ± 0.46	-2.91 ± 0.38	-2.57 ± 0.34
σ_8	0.8263 ± 0.0074	0.8151 ± 0.0066	0.8221 ± 0.0070	0.8149 ± 0.0063	0.8126 ± 0.0046
S_8	0.875 ± 0.023	0.836 ± 0.016	0.857 ± 0.020	0.830 ± 0.014	0.8169 ± 0.0087
Age [Gyr]	13.801 ± 0.023	13.800 ± 0.024	13.788 ± 0.019	13.789 ± 0.018	13.772 ± 0.015
$10^4\theta_\star$	104.075 ± 0.031	104.109 ± 0.031	104.085 ± 0.029	104.094 ± 0.025	104.107 ± 0.025
$10^4 Y_{\rm He}$	2459.50 ± 0.71	2458.55 ± 0.64	2459.66 ± 0.51	2459.10 ± 0.48	2459.37 ± 0.46
$10^{10}\eta_b$	6.185 ± 0.046	6.124 ± 0.041	6.196 ± 0.033	6.159 ± 0.030	6.177 ± 0.029
$z_{\rm reio}$	$7.88^{+0.54}_{-0.61}$	$8.15^{+0.55}_{-0.62}$	$7.93^{+0.54}_{-0.61}$	8.23 ± 0.59	$8.47^{+0.54}_{-0.61}$
$\tau_{\rm rec}$ [Mpc]	593.6 ± 3.1	599.5 ± 2.0	596.2 ± 2.6	600.4 ± 1.8	602.4 ± 1.3
$z_\star$	1089.96 ± 0.30	1089.92 ± 0.29	1089.75 ± 0.24	1089.68 ± 0.21	1089.47 ± 0.18
$r_{s,\star}$ [Mpc]	143.32 ± 0.54	144.43 ± 0.31	143.74 ± 0.45	144.53 ± 0.29	144.85 ± 0.22
z_d	1060.72 ± 0.39	1059.94 ± 0.29	1060.67 ± 0.28	1060.17 ± 0.23	1060.21 ± 0.23
r_d [Mpc]	145.88 ± 0.56	147.09 ± 0.30	146.30 ± 0.46	147.14 ± 0.29	147.45 ± 0.23

Figure 3: The current (mid 2025) state of the art for cosmological parameters obtained by fitting observations to ΛCDM cosmology. The different columns fit to different datasets – see [7] for details.

Ultimately our confidence in the existence of things like Dark Matter relies on the redundancy of the evidence in its favour. Redundancy is convincing in two ways. First it provides protection from some of the observations simply being wrong (due to unknown mistakes). Redundant evidence survives even when some individual experiments are thrown away. Second, redundancy provides confidence in an overall picture. The situation is much like it was for the discussion of atoms at the turn of the 20th century: they could not be directly detected but the properties of bulk matter provided multiple independent lines of evidence for their existence. After all, if atoms did not exist there is no reason why independent inferences of their mass, size and abundance from the properties of bulk matter should all give the same answer. But they did and it is the agreement of multiple independent lines of evidence that is compelling. Although not explored in detail in these notes, the evidence for Dark Matter is similarly redundant, coming from several different sources within cosmology, but also from myriad observations of galaxies, galaxy clusters, and the large-scale distribution of matter.

1.2 Playing the field

The purpose of these notes is to explore what we know about Dark Energy in two separate ways. We start here by summarizing the extent to which evidence is building that the Dark Energy density might not be constant in time which, if true, would mean the Dark Energy could not just be a vacuum energy. This evidence is one of a small set of 'tensions' within ΛCDM cosmology. Tensions arise as degradation of the quality of the fit to observations if the agreement between observations and predictions starts to deteriorate as either or both become more accurate. We later explore the extent to which there are useful clues in demanding consistency of viable cosmological proposals with more microscopic physics.

1.2.1 Anomalies?

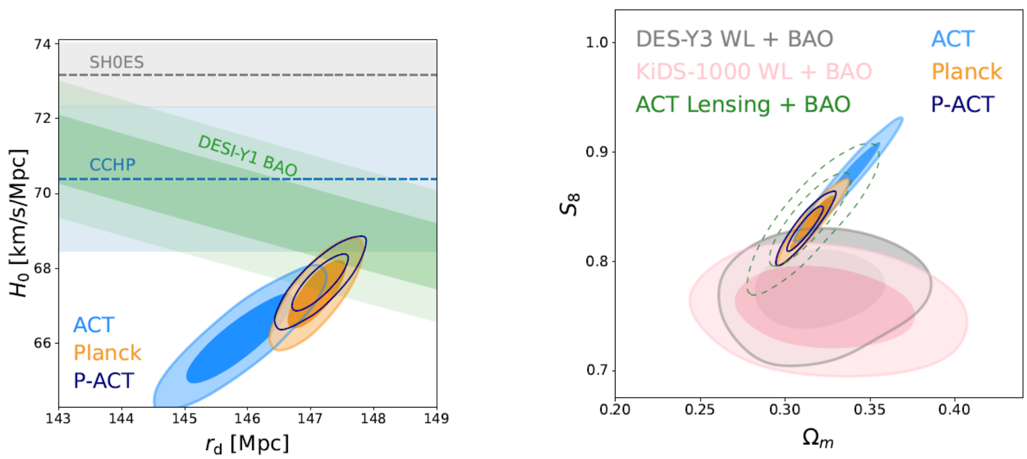

Figure 4: Left panel: Summary of the constraints on H_0 and the sound horizon r_d coming from different kinds of observations, illustrating the Hubble tension. Right panel: Summary of the constraints on S_8 and the density of nonrelativistic matter $\Omega_{\rm m}$ coming from different kinds of observations, illustrating the S_8 tension. Both figures taken from [7].

Three types of tensions are normally discussed when asking about the ability of ΛCDM model to fit cosmological data.

- **Hubble tension:** The value of H_0 inferred from measurements of the Cosmic Microwave Background (CMB) seems to disagree with the value of H_0 obtained by measuring the luminosity of and distance to relatively nearby objects. Inferences based on the CMB (such as those in Fig. 3) tend to prefer $H_0 \simeq 67$ km/sec/Mpc with an error of just under 1 km/sec/Mpc. Observations using relatively nearby supernovae instead give a value of $H_0 \simeq 73$ km/sec/Mpc with an error of around 1 km/sec/Mpc [8] (see

Fig. 4). It is not yet clear whether this disagreement is telling us about non-ΛCDM physics or about the difficulty of performing the relevant measurements [9, 10].

- S_8 **tension:** A similar tension has arisen for inferences of the size of clumping at particular scales – parameterized by a quantity S_8 – as measured in the distant and closer-by universe. CMB-based inferences (such as those in Fig. 3) give $S_8 \simeq 0.85$ to within a few percent but those using more recent observables instead find $S_8 \simeq 0.77$ with similar errors [11] (see Fig. 4). Again there is uncertainty as to how much of this discrepancy is associated with systematic errors [10].

- **Time-dependent Dark Energy:** More recent than the previous two are tentative indications that the energy density of Dark Energy is time-independent. This is illustrated in Fig. 5 which plots the equation of state parameter $w_{\rm vac}$ as a function of redshift (which is a proxy for scale factor). This result, if it survives scrutiny, would directly contradict the interpretation of Dark Energy as a constant vacuum energy.

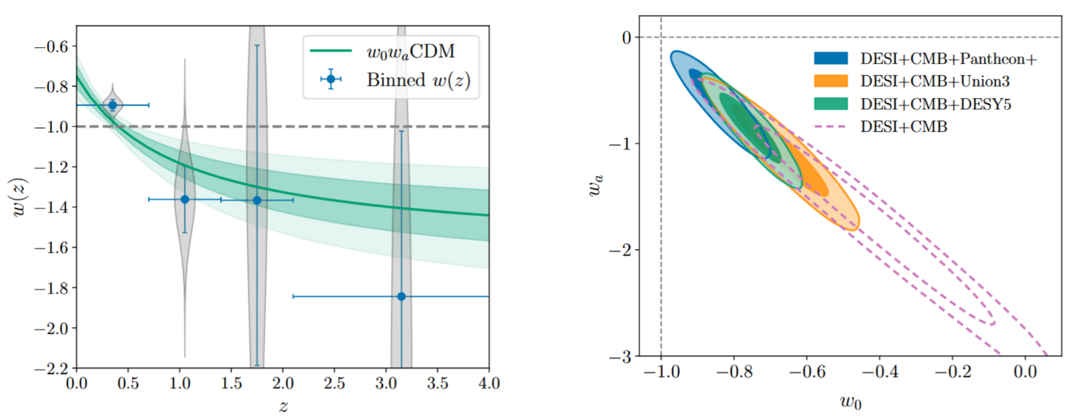

Figure 5: Left panel: constraints on the evolution of the Dark Energy equation of state parameter w as a function of redshift (and so also of universal scale factor). The green swathe denotes the expected shape given a phenomenological parameterization $w(a) = w_0 + w_a\, a$. Right panel: best-fit values for the parameters w_0 and w_a. Both figures taken from [14].

Many cosmological models have been designed to explain these tensions in terms of new fundamental physics (for the Hubble tension see [12, 13] for recent reviews), or as solutions to other problems that happen also to have cosmological significance. The second direction these notes explore is the extent to which one can differentiate amongst the many cosmological models by asking whether they can plausibly emerge at low-energies given what we know about the physics of higher energies elsewhere in physics. This turns out to be fairly restrictive and it is argued that the criteria required to interface with higher-energy physics is an important clue when figuring out what is going on. We use the third anomaly – evolution of Dark Energy density with time – as a vehicle for having this discussion.

1.2.2 Simple Scalar Model

To get things going it is useful to have a straw man: a concrete example of a model that could give a time-evolving Dark Energy density. This can be used to illustrate the kinds of difficulties that can arise. A vanilla starting point of this type is to postulate the existence of a scalar field ϕ that evolves homogeneously over cosmological time scales. This will look approximately like Dark Energy if its kinetic energy, K, is much less than its kinetic energy, V – becoming exactly like a vacuum energy in the limit $K \to 0$.

Consider for instance supplementing the cosmological model with a scalar field whose action has the form

$$S = -\int \mathrm{d}^4x \sqrt{-g} \left[\tfrac{1}{2}\, g^{\mu\nu}\partial_\mu\phi\, \partial_\nu\phi + V(\phi) \right], \tag{1.13}$$

where $V(\phi)$ is a scalar potential to be specified below. Such a field satisfies the classical field equations

$$\Big[-\Box + V'(\phi) \Big] = 0\,, \tag{1.14}$$

where $\Box = g^{\mu\nu}\nabla_\mu\nabla_\nu$, and contributes to Einstein's equations (1.1) by adding a new term to the stress energy. Applying the definition $\frac{1}{2}\sqrt{-g}\, T^{\mu\nu} = \delta S/\delta g_{\mu\nu}$ to (1.13) one finds

$$T^{(\phi)}_{\mu\nu} = \partial_\mu\phi\, \partial_\nu\phi - g_{\mu\nu}\Big[\tfrac{1}{2}\, g^{\lambda\rho}\partial_\lambda\phi\, \partial_\rho\phi + V(\phi) \Big]\,. \tag{1.15}$$

Working within an FLRW geometry (1.3) and assuming ϕ depends only on t in the rest-frame of the cosmological fluid then reduces (1.14) to an ordinary differential equation:

$$\ddot\phi + 3H\,\dot\phi + V'(\phi) = 0\,, \tag{1.16}$$

where $H = \dot a/a$ as usual and V' is the derivative of $V(\phi)$ with respect to ϕ. In the special case $V = \frac{1}{2}m^2\phi^2$ (1.16) is a linear equation that describes damped oscillations with a frequency set by m and a damping rate set by H. When $m \gg H$ these oscillations are rapid on cosmological timescales and the damping ensures the energy density in these oscillations drops with universal expansion like $\rho_\phi \propto 1/a^3$. (Exercise: prove this.) As a consequence nontrivial scalar evolution within a potential is normally only important over long times in cosmology[5] if the scalar mass is not large compared with H. For the present-day Hubble constant this is an extremely small scale, $H_0 \sim 10^{-32}$ eV, relative to microphysical scales.

Homogeneity and isotropy also imply $T^{(\phi)}_{00} = \rho_\phi$, $T^{(\phi)}_{0i} = 0$ and $T^{(\phi)}_{ij} = p_\phi \, g_{ij}$ where

$$p_\phi = \tfrac{1}{2}\dot{\phi}^2 - V(\phi) \qquad \text{and} \qquad \rho_\phi = \tfrac{1}{2}\dot{\phi}^2 + V(\phi)\,, \tag{1.17}$$

so the Einstein equations governing homogeneous evolution remain (1.5) and (1.6), but with the pressure and energy density of (1.11) supplemented by adding (1.17), evaluated at the solution to (1.16).

The ratio

$$w_\phi := \frac{p_\phi}{\rho_\phi} = \frac{\frac{1}{2}\dot{\phi}^2 - V(\phi)}{\frac{1}{2}\dot{\phi}^2 + V(\phi)} \tag{1.18}$$

is in general time-dependent and so this kind of theory gives a cosmic fluid with a time-dependent equation-of-state parameter. It in general does not conform to the special case discussed in eqs. (1.7) through (1.9) apart from in a few special limits. For instance when the motion is rapid enough that $\frac{1}{2}\dot{\phi}^2 \gg V(\phi)$ – the so-called *kination* regime – then $p_\phi \simeq \rho_\phi$ and so $w_\phi \simeq +1$. In this case (1.8) and (1.9) apply and give $\rho_\phi \propto a^{-6}$ and (if the scalar energy dominates) $a(t) \propto t^{1/3}$.

Another limit for which eqs. (1.7) through (1.9) apply is the *slow-roll* regime, for which $\frac{1}{2}\dot{\phi}^2 \ll V(\phi)$. In this case (1.17) implies $p_\phi \simeq -\rho_\phi$ and so $w_\phi \simeq -1$, mimicking a vacuum energy along which $\rho_\phi \simeq V(\phi)$ is approximately constant. This suggests that choosing $V(\phi)$ to be sufficiently shallow provides a candidate for Dark Energy where both ρ_{DE} and w_{DE} can vary with time. A possible difficulty with such a candidate is the observation that reproducing $\Omega_\Lambda \simeq 0.67$ requires $\rho_\phi \simeq V > 0$ during the slow roll. But for $V > 0$ eq. (1.18) implies $-1 \leq w_\phi \leq +1$ and so in particular one should never enter the regime $w_{DE} < -1$ seen in the left-hand panel of Fig. 5. We return to the question of how discouraging we should find this in later sections.

2. Prior Knowledge

A great many models for Dark Energy can be (and have been) built in this way (see for example the reviews [15]), and at first sight there seems to be little chance of being able to distinguish amongst so many models using only the limited data available to us from cosmology (wonderful though this data surely is). The next few sections step back and ask

[5] An important exception to this is if the rapid oscillations themselves are the Dark Matter, whose energy also falls like $1/a^3$.

whether what we know about the rest of physics (outside of cosmology) can usefully constrain the search for phenomenologically successful models.

It turns out that it can, and this might come as something of a surprise since experience in other areas of physics tells us that the details of short-distance physics usually are not important when computing long-distance properties – a phenomenon called 'decoupling' – and cosmology deals with the longest distances of all. As we shall see, the theories that describe cosmology well rely heavily on the few things that do depend on what happens at higher energies. One of the things we learn along the way by asking this question is what controls the corrections to the basic semiclassical limit that lies behind the logic of solving equations like (1.1) and (1.14) classically in the first place.

2.1 Semiclassical methods in gravity (and why they work)

Since our goal is to ask what we can learn by thinking about cosmological models as the low-energy limit of some more fundamental theory, the first step is to systematize what kinds of things emerge in general for the low-energy limit of physical systems. The answer to this question is best answered using the tools of Effective Field Theories (EFTs) [16] and so we start with a brief digression to summarize these (leaning heavily on the reviews [17, 18, 19]). Although quantum effects are often small in practical applications to gravity we nonetheless explore low-energy EFTs for quantum systems. For applications to gravity this will bring the later payoff of showing us what controls the semiclassical approximation in the first place.

2.1.1 EFT methods

Suppose we have a physical system with a characteristic scale M, such as a collection of 'heavy' degrees of freedom with masses of order M represented by fields collectively denoted $h(x)$. Suppose the theory also have very 'light' degrees of freedom whose masses are much smaller than M, represented by fields collectively denoted $\ell(x)$. Our interest is in observables, $\mathcal{A}(E, M)$, involving the light fields that involve energies much smaller than M, such as scattering of ℓ particles with centre-of-mass energy $E \ll M$. These observables inevitably simplify once Taylor expanded in powers of E/M. The hard way to find the simple $E \ll M$ limit is to compute $\mathcal{A}(E, M)$ in all of its glory and then Taylor expand. EFT methods seek instead to do the Taylor expansion as early as possible in a calculation in order to exploit the simplicity as effectively as possible.

A conceptually simple way to do so is to write out the path-integral expression for $\mathcal{A}$ and then 'integrate out' the heavy degrees of freedom once and for all early in the calculation.

For instance, suppose $\mathcal{A}$ has the path-integral representation

$$\mathcal{A} = \int \mathcal{D}\ell\,\mathcal{D}h\;\mathcal{O}(\ell)\;\exp\left[i\int \mathrm{d}^4x\,\mathcal{L}(\ell,h)\right]\,, \tag{2.1}$$

where $\mathcal{O}$ is some operator built only from the light degrees of freedom. (This would be true in particular for correlation functions from which a great many observables can be derived.) There is a great deal of freedom in choosing precisely how to separate the path-integral into integrals $\mathcal{D}\ell$ and $\mathcal{D}h$ over light and heavy degrees of freedom – for instance on flat space one might imagine dividing up all field modes in momentum space as heavy or light depending on whether or not[6] $p^2 + m^2 > \Lambda^2$ or $p^2 + m^2 < \Lambda^2$, where p is the corresponding particle momentum and m is its mass. Λ here is an arbitrary cutoff chosen to be much smaller than the heavy scale but much larger than the energies of interest in $\mathcal{A}$: that is $E \ll \Lambda \ll M$.

Although the details of the heavy-light split can affect intermediate steps in the calculation, these choices must drop out of the final physical predictions because they are just an artefact of how we organize the calculation. They do not appear at all in the original integral (2.1) before trying to separate the fields into high and low energy parts. So one is free to use calculational convenience as a guide when making this split.

Now comes the main point: the observables (by assumption) do not depend on the heavy degrees of freedom and so the integration over h does not depend on the choice for $\mathcal{O}$ and can be evaluated once and for all right at the very beginning, with the efficiency of performing an expansion in powers of $1/M$ reaped very early on. When this is done the influence of heavy fields on any dynamics at low energies is completely encoded in the following *effective action*:

$$e^{iS_{\rm eff}[\ell,\Lambda]} := \int_\Lambda \mathcal{D}h\;\exp\left[\int d^4x\,\mathcal{L}(\ell,h)\right]. \tag{2.2}$$

The dependence of $S_{\rm eff}$ on Λ is a shorthand for a dependence on all of the details of precisely how one makes the low-energy/high-energy split.

Physical observables at low energies are now computed by performing the remaining path integral over the light degrees of freedom only:

$$\mathcal{A} = \int^\Lambda \mathcal{D}\ell\;\mathcal{O}(\ell)\;\exp\Big[iS_{\rm eff}(\ell,\Lambda)\Big]\,, \tag{2.3}$$

showing that the integration over light fields is weighted by $S_{\rm eff}(\ell)$ in precisely the same way as the classical action $S = \int \mathrm{d}^4x\,\mathcal{L}(\ell,h)$ does for the original integral over both heavy and light degrees of freedom. This derivation also makes clear that any dependence on Λ (*i.e.* on

[6] It is usually most convenient to do this in Euclidean signature.

the details of the high/low-energy split) must cancel between the explicit Λ-dependence of $S_{\rm eff}(\Lambda)$ and the implicit dependence in the definition of the low-energy integration $\int^\Lambda \mathcal{D}\ell$.

Although $S_{\rm eff}$ obtained in this way is in general a hot mess, it simplifies dramatically once it is expanded to a finite order in $1/M$, in which case it become local in spacetime,[7] so

$$S_{\rm eff} = \int {\rm d}^4 x \; \mathcal{L}_{\rm eff}(\ell, \Lambda) \,, \tag{2.4}$$

where $\mathcal{L}_{\rm eff}$ is (to any finite order in $1/M$) a simple function – usually a polynomial – of ℓ and its derivatives all evaluated at the same spacetime point. This happens in detail because the low-energy expansion of massive particle propagators is local

$$\Big(-\Box + M^2 \Big)^{-1} = \frac{1}{M^2} \sum_{n=0}^\infty \left(\frac{\Box}{M^2} \right)^n \,, \tag{2.5}$$

when truncated to any finite order. Physically this has its roots in the uncertainty principle: high-energy states can only get into low-energy predictions by violating energy conservation, which the uncertainty principle allows[8] provided they are only done over times $\Delta t \ll 1/M$, making them effectively local for low-energy observers who cannot resolve such small intervals.

The upshot of all of this is that *all* low-energy contributions of the heavy degrees of freedom are encoded in an effective lagrangian (or Wilson action) that is a product of powers of the light field ℓ and derivatives. Because the integral in (2.2) defining $S_{\rm eff}$ involves only high-energy states any dimensionful parameters in this lagrangian will involve the heavy scale M (such as is true for each additional power of $\Box$ in (2.5)). On dimensional grounds any additional powers of derivatives and fields generically cost additional powers of $1/M$ and so become negligible once one restricts to a finite order.

Only a very small number of terms can involve absolutely no suppression by powers of $1/M$ and the lagrangian obtained by keeping all such unsuppressed terms is called *renormalizable*. We expect renormalizable theories to describe the dominant physics at low energies, and this is indeed what we find in successful theories like Quantum Electrodynamics, Quantum Chromodynamics and the Weinberg-Salam model of electroweak unification. This is the modern understanding of why the renormalizable theories like the Standard Model of particle physics (which includes the other three mentioned) work so well: what we are looking at in practice in Nature is consistent with being the low-energy tip of an iceberg: some more fundamental theory describing physics at much higher energies.

[7]In nonrelativistic systems the low-energy expansion makes $S_{\rm eff}$ local in time but locality in space depends on whether or not short distance degrees of freedom can also be low-energy degrees of freedom.

[8]More precisely, energy conservation can be violated in old-fashion Rayleigh-Schrödinger perturbation theory, but is regidly conserved in Schwinger-Feynman perturbation theory (in which case the same conclusions follows from off-shell contributions as described in (2.5).

2.1.2 GREFT

This is all very nice, but although cosmology involves the longest distances to which we have access (and so the lowest energies of all) the theory most relevant to it is General Relativity, which is *not* renormalizable. Why is the above EFT discussion relevant?

In the EFT picture renormalizable interactions usually dominate nonrenormalizable ($1/M$-suppressed) interactions *when renormalizable interactions exist*. But sometimes no renormalizable interactions are possible and in such cases nonrenormalizable interactions can dominate. An example of this is when a renormalizable theory contains an 'accidental' symmetry (like baryon number B or lepton-number L conservation in the Standard Model for instance). Accidental symmetries are symmetries not built in as assumptions; they instead emerge as accidental consequences of renormalizability for a given field content. If the Standard Model emerges as the low-energy limit of a more fundamental theory in which baryon number is not conserved (such as a Grand Unified Theory, or GUT) then the leading rates for B- or L-violating processes at low energies would be described by nonrenormalizable interactions because the renormalizable ones of the Standard Model preserve baryon number.

An even more informative example – for which both the low-energy and high-energy theories are well understood – is the low-energy effective Fermi theory of the weak interactions that are responsible for many radioactive decays. In this case the fundamental high-energy theory is the Standard Model itself and the low-energy theory is obtained once the W boson (with mass $M_W \simeq 80$ GeV) is integrated out (together with other, heavier, particles). In this effective theory the renormalizable interactions preserve particle flavours (like charm, strangeness, up-ness or down-ness *etc*) and so radioactive decays mediated by the weak interactions that violate the conservation of these quantities (like $\pi^+ \to e^+\nu$ or nuclear β-decays) are well described by the nonrenormalizable Fermi theory. In this theory the effective Fermi coupling constant, G_F, has dimension M^{-2} for a scale M much larger than the energies in the decays – the characteristic suppression by inverse powers of a heavy scale typical of nonrenormalizable interactions. But because we also understand the more fundamental high-energy theory we can in this case explicitly relate the size of G_F to the scale of the heavy physics that was integrated out (in this case the W-boson mass, M_W), with $G_F \simeq g^2/M_W^2$ (where $g \sim 10^{-1}$ is a measure of the coupling strength of the W boson).

A similar story applies to gravity: it turns out there are no renormalizable couplings possible for the graviton and so its interactions must be nonrenormalizable. There is even a candidate for a more fundamental theory – string theory – that produces General Relativity in its low-energy limit with Newton's constant $G_N \simeq g_s^2/M_s^2$ calculable in terms of the couplings

g_s and masses M_s of very high-energy states.[9] As we shall see, even without such a UV completion the EFT framework also seems to be required if the theoretical error associated with quantum effects in gravity are to be reliably estimated.

The main practical consequence of regarding GR as part of a low-energy EFT within a more fundamental theory (much as we do for the Standard Model) is that the action can no longer be limited to the vanilla Einstein-Hilbert action. The Einstein-Hilbert action should instead just be regarded as the leading term in an expansion in powers of derivatives of the metric divided by some heavy scale $1/M$.

Recall for these purposes that the field relevant for GR is the metric, $g_{\mu\nu}$, of spacetime itself, and that its action is required to be invariant under general covariance and local Lorentz invariance. Invariance under these symmetries dictate the metric can appear in the action only through curvature invariants built from the Riemann tensor,

$$R^\mu{}_{\nu\rho\lambda} = \partial_\lambda \Gamma^\mu_{\nu\rho} + \Gamma^\mu_{\lambda\alpha}\Gamma^\alpha_{\nu\rho} - (\lambda \leftrightarrow \rho) \quad \text{with} \quad \Gamma^\mu_{\nu\lambda} = \tfrac{1}{2}\, g^{\mu\beta}\Big(\partial_\nu g_{\beta\lambda} + \partial_\lambda g_{\beta\nu} - \partial_\beta g_{\nu\lambda}\Big)\,, \tag{2.6}$$

and its contractions – such as the Ricci curvature $R_{\mu\nu} = R^\alpha{}_{\mu\alpha\nu}$ and Ricci scalar $R = g^{\mu\nu}R_{\mu\nu}$ – and their covariant derivatives. What is important in what follows about these definitions is that the curvature tensors involve precisely two derivatives of the metric.

The low-energy EFT for the metric (called GREFT) is defined as the local action involving all possible powers of derivatives of the metric, which general covariance then requires must be built from powers of the curvature tensors and their derivatives,

$$\begin{aligned} - \frac{\mathcal{L}_{GREFT}}{\sqrt{-g}} &= \lambda + \tfrac{1}{2} M_p^2\, R + c_{41}\, R_{\mu\nu}\, R^{\mu\nu} + c_{42}\, R^2 + c_{43}\, R_{\mu\nu\lambda\rho} R^{\mu\nu\lambda\rho} + c_{44}\, \Box R \\ &\quad + \frac{c_{61}}{M^2}\, R^3 + \frac{c_{62}}{M^2} \partial_\mu R\, \partial^\mu R + \cdots \,. \end{aligned} \tag{2.7}$$

The first two terms here are the only ones possible involving just the metric and two or fewer derivatives and these agree with the Einstein-Hilbert action of General Relativity with cosmological constant λ. The rest of the first line includes all possible terms involving precisely four derivatives, and (for brevity) the third line includes only two representative examples of the many possible terms involving six or more derivatives.

The constants c_{dn} appearing in (2.7) are labelled using the convention that d counts the number of derivatives of the corresponding effective operator and $n = 1, \cdots, N_d$ runs over the number of such couplings. These couplings are dimensionless because the appropriate power

[9] Although we do not yet know whether string theory correctly describes nature, much of the attention it receives hinges on it being a rare example of a consistent fundamental UV completion for General Relativity, including quantum effects.

of a high-energy mass scale M has been extracted to ensure this is so (assuming 4 spacetime dimensions). Although it is tempting to use $M \simeq M_p$ everywhere for this high-energy mass scale (given that this is what appears in front of the Einstein-Hilbert term) this would in general be a mistake.

To see why, imagine generating a contribution to these effective couplings by integrating out a heavy particle of mass M. All of the terms listed are generically generated when doing so, with M appearing in each coefficient as required on dimensional grounds (leading to terms like $M^2 R$ and R^2 and R^3/M^2 and so on). The complete coefficient of any one term in $\mathcal{L}_{GREFT}$ would then be obtained by summing over all of the possibly many particles appearing in the fundamental theory, making the coefficient of R in this lagrangian a sum of the schematic form $\sum_n k_n M_n^2$ while the coefficient of R^3 would instead be something like $\sum_n \tilde{k}_n M_n^{-2}$.

Here comes the point: although it is the largest mass that dominates in any sum over positive powers of M_n, it is the *smallest* mass that dominates a sum over negative powers of M_n. Consequently we are not surprised at all to find a large coefficient like $M_p \sim 10^{18}$ GeV appearing in front of the Einstein-Hilbert term, but this does *not* provide evidence for the scale M appearing in the curvature-cubed and higher terms in (2.7) also being this large. Instead one should expect M in any given application to be of order the lightest of the heavy particles whose integrating out generates $\mathcal{L}_{GREFT}$. For instance, for applications to the solar system M might be the electron mass; for applications to post-nucleosynthesis Big-Bang cosmology M might be of order the QCD scale, and so on.) Of course, contributions like $M^2 R$ or R^3/M_p^2 could also exist, but when $M \ll M_p$ these are completely negligible compared to the terms displayed in eq. (2.7).

The first, cosmological constant, term in eq. (2.7) is the only one with no derivatives and the alert reader will notice that we did not write its coefficient as $\lambda = c_{01} M_p^4$. This was not done because (as discussed in §1.1.1) it contributes to observables in the same way as does the vacuum energy and so plays the role of Λ in the ΛCDM model. As a consequence its value has already been measured, with observations implying $\lambda \simeq \rho_{\rm vac} \simeq 0.67\rho_c \sim (3 \times 10^{-3} \text{ eV})^4$ and so is roughly 122 orders of magnitude smaller than M_p^4. Since $c_{01} \sim 10^{-122}$ it is an extremely good approximation for most applications to neglect it completely when asking for the implications of GREFT in noncosmological settings. Much of the rest of this review will be devoted to how puzzling we should find it that λ should be so small, which is called the cosmological constant problem [20, 22, 23]. (We return to this issue below when discussing implications for cosmology.)

Redundant interactions

The attentive reader might notice that not all of the interactions listed in (2.7) are equally important, even when only comparing interactions having the same number of derivatives. For instance, the freedom to drop total derivatives[10] from the lagrangian allows us to ignore the coupling c_{44}, because $\sqrt{-g}\,\Box R = \partial_\mu(\sqrt{-g}\,\partial^\mu R)$ is a total derivative. A similar argument applies as well (in 4 dimensions) as c_{43} since the quantity

$$\sqrt{-g}\; X = \sqrt{-g}\Big(R_{\mu\nu\lambda\rho}R^{\mu\nu\lambda\rho} - 4R_{\mu\nu}R^{\mu\nu} + R^2\Big)\,, \tag{2.8}$$

is locally also a total derivative (it integrates to give a topological invariant in 4 dimensions). Dropping total derivatives allows us to replace, for example, $R_{\mu\nu\lambda\rho}R^{\mu\nu\lambda\rho}$ with the linear combination $4\,R_{\mu\nu}R^{\mu\nu} - R^2$, with no consequences for any observables (provided these observables are insensitive to the overall topology of spacetime, as are the classical equations or perturbative particle interactions).

It is also possible to ignore any effective interactions in (2.7) that involve the Ricci tensor $R_{\mu\nu}$ (and so also its trace $R = g^{\mu\nu}R_{\mu\nu}$), provided we work only perturbatively in powers of $1/M$. This is because the variation of the leading Einstein-Hilbert action under a field redefinition $\delta g_{\mu\nu}(x)$ is (dropping total derivatives)

$$\delta S_{EH} = \int \mathrm{d}^4x \, \left(\frac{\delta S_{EH}}{\delta g_{\mu\nu}}\right)\delta g_{\mu\nu} = \tfrac{1}{2}M_p^2 \int \mathrm{d}^4x\, \sqrt{-g}\,\Big(R^{\mu\nu} - \tfrac{1}{2}Rg^{\mu\nu}\Big)\delta g_{\mu\nu} \tag{2.9}$$

This means that any term in the GREFT action that vanishes for a Ricci-flat geometry, like

$$S_{GREFT} \ni -\int \mathrm{d}^4x\sqrt{-g}\,\Delta^{\mu\nu}R_{\mu\nu}\,, \tag{2.10}$$

can be removed at leading order by choosing

$$\delta g_{\mu\nu} = 2\kappa^2\Big(\Delta_{\mu\nu} - \tfrac{1}{2}g^{\lambda\rho}\Delta_{\lambda\rho}\,g_{\mu\nu}\Big)\,. \tag{2.11}$$

This argument is a special case of a more general statement that also applies when matter is present: any effective interaction that vanishes when the lowest-order equations of motion are used can be similarly removed by performing an appropriate field redefinition.

Any interaction that is a surface term or can be removed using a field redefinition in this way is called a *redundant* interaction because most observables (except perhaps those sensitive to boundary terms) cannot depend on their coefficients. It is useful to remove all

[10]These cannot be dropped if one cares about boundary information or topology, so we when making these arguments we have in mind the vast majority of other local effects for which surface terms are irrelevant.

such interactions from the effective theory because carrying them around is not wrong but is needlessly time-consuming since they have no effects.

In practice this means that for pure gravity (no other fields, like matter) *all* of the effective interactions beyond the Einstein-Hilbert term that are written explicitly in (2.7) are redundant (in 4 spacetime dimensions) because they are either total derivatives or they vanish when $R_{\mu\nu} = 0$ (or both). The first nontrivial non-redundant effective interaction involves cubic or higher powers of the Riemann tensor.

2.1.3 Power counting (gravity only)

We see there can be a large number of interactions in an EFT – potentially arbitrarily large if one works to arbitrary fixed order in $1/M$. How can a theory with so many effective couplings ever be predictive? This is a central question whose general answer is given by power-counting [16] (as we describe for gravity in this section).

In any EFT we imagine expanding all observables in powers of q/M where q is a typical energy scale of interest in the low-energy sector (perhaps a centre-of-mass scattering energy or the Hubble expansion rate) and so a very important question asks which interactions are relevant when computing observables at a specific order in powers of q/M (and q/M_p in the case of the lagrangian (2.7)). We here briefly recap the result without repeating the details (see however [17]).

To see how various interactions contribute to physical processes consider using the lagrangian (2.7) to calculate a correlation function or a scattering amplitude involving a path integral like in (2.3). For simplicity we ignore here the cosmological constant term λ, but return to it when we consider cosmology in the next sections. The integral is evaluated semi-classically by expanding around some classical background spacetime $\overline{g}_{\mu\nu}$ that we assume to be a stationary point of the action built from (2.7). We then write the full metric as[11] $g_{\mu\nu} = \overline{g}_{\mu\nu} + h_{\mu\nu}/M_p$ and do a double expansion of the action S_{GREFT} in powers of both $h_{\mu\nu}$ and of derivatives, keeping in mind that the curvature involve all possible powers of $h_{\mu\nu}$, but precisely two derivatives.

One finds in this way the expansion

$$S_{GREFT}[\overline{g} + h] = S_{GREFT}[\overline{g}] + S_{EH(2)}[\overline{g}, h] + S_{\rm int}[\overline{g}, h]\,, \tag{2.12}$$

where

$$S_{EH}[g] = -\tfrac{1}{2} M_p^2 \int {\rm d}^4x \sqrt{-g}\; R\,, \tag{2.13}$$

[11]Strictly speaking a factor of 2 would pre-multiply $h_{\mu\nu}$ if fluctuations were to be canonically normalized, but our focus here is on how the scales M and M_p appear.

is the Einstein-Hilbert lagrangian – or equivalently the terms in (2.7) involving precisely two derivatives – and $S_{EH(2)}$ contains those terms in the expansion of the Einstein-Hilbert action arising at quadratic order[12] in $h_{\mu\nu}$. The 'interaction' term contains everything else:

$$S_{\rm int}[\overline{g}, h] = S_{EH\,{\rm int}}[\overline{g}, h] + S_{\rm eff\,int}[\overline{g} + h]\,, \tag{2.14}$$

where $S_{EH\,{\rm int}}$ contains two-derivative terms coming from the expansion of $S_{EH}[\overline{g}+h]$ that are cubic or higher in $h_{\mu\nu}$ and $S_{\rm eff\,int}$ contains terms involving any number of powers of $h_{\mu\nu}$ but with no fewer than 4 derivatives – *i.e.* the higher-derivative terms in (2.7).

The integrand of the path integral is then written perturbatively in $S_{\rm int}$

$$e^{iS_{GREFT}[\overline{g}+h]} = e^{S_{EH}[\overline{g}]+S_{EH(2)}[\overline{g},h]} \sum_{r=0}^{\infty} \frac{1}{r!} \Big[S_{\rm int}[\overline{g}, h] \Big]^r\,, \tag{2.15}$$

so that the path integration becomes gaussian and can be evaluated using the standard Feynman procedure (including covariant gauge fixing and ghosts in the usual way, the details of which do not change the arguments to be made below). Evaluating the gaussian integrals can still be hard in practice because we so far make no assumptions about the nature of the background metric $\overline{g}_{\mu\nu}$, but it can be done explicitly for simple spacetimes like Minkowski space or anti-de Sitter space, say.

Our goal here is less ambitious than full evaluation, however. We wish only to perform a power-counting exercise to identify what must be small in order for this expansion to be a good approximation. This involves identifying how an arbitrary Feynman graph depends on the scales M_p and M appearing in the lagrangian (2.7), which can be done in great generality in some circumstances. In particular, it can be done in situations when there is only one scale of interest in the low-energy theory[13] – call it q say – since in this case it boils down to a dimensional argument.[14]

Consider an arbitrary graph that contributes at L loops to the amputated[15] E-point $h_{\mu\nu}$ correlation function, $\mathcal{A}_E(q)$, performed with all external background curvatures and mode

[12]Any linear term in the expansion of S_{EH} simply contributes to cancellation of the 'tadpole' graphs (those with one external leg) that determine how the background metric changes from the solution to Einstein's equations once higher-derivative terms are included.

[13]A nontrivial example of this might be if we follow fluctuations about de Sitter space and focus only on fluctuations whose physical momenta k/a are roughly the same size as the background curvature scale H. In this case we could choose $q \sim H$.

[14]Dimensional arguments become more complicated if UV divergences are regularized using a cutoff but go through as expected naively if one instead uses dimesional regularization, for instance.

[15]Amputation means that the graphs have no external lines, such as might be encountered when computing the size of coefficients in a low-energy effective action itself.

numbers characterized by a single low-energy scale $q \ll M \ll M_p$. Suppose the graph contains V_{id} vertices involving d derivatives and i factors of the fluctuation field $h_{\mu\nu}$. The dependence of $\mathcal{A}_E(q)$ on the scales M and M_p can be read off from the Feynman rules that determine the propagators and vertices of the graph in question and then all of the remaining dimensions are taken to be captured by the appropriate power of the low-energy scale q. This leads [17] to the following prediction for the q, M and M_p dependence of $\mathcal{A}_E(q)$:

$$\mathcal{A}_E(q) \sim q^2 M_p^2 \left(\frac{1}{M_p}\right)^E \left(\frac{q}{4\pi M_p}\right)^{2L} \prod_i \prod_{d>2} \left[\frac{q^2}{M_p^2}\left(\frac{q}{M}\right)^{(d-4)}\right]^{V_{id}} . \tag{2.16}$$

Notice that since d is even for all of the interactions, the condition $d > 2$ in the product implies there are no negative powers of q in this expression. The argument leading to (2.16) is sketched out in a bit more detail in the next section.

Eq. (2.16) is this section's main result, and it contains lots of information.

- First, the appearance of only positive powers of q verifies that it is indeed self-consistent to organize calculations using a derivative expansion when computing using (2.7). The weakness of gravitational self-couplings comes purely from the low-energy approximations $q \ll M_p$ and $q \ll M$.

- For a fixed process (*i.e.* for a fixed number, E, of external lines) each additional loop costs a factor of $q^2/(4\pi M_p)^2$. But it is the number of loops that also counts the factors of $\hbar$ that premultiply the action in non-fundamental units $e^{iS/\hbar}$, making the loop expansion also the semiclassical expansion. Why is the classical approximation good in GR? We see it is ultimately the hierarchy $q \ll 4\pi M_p$ that justifies the use of semiclassical methods: the semiclassical approximation *is* the low-energy approximation.

- Notice that there is no low-energy penalty for using as many 2-derivative interactions as we like. This shows that there is nothing in the low-energy limit that allows us to neglect the full nonlinearity of GR.

- Even though the ratio q/M could be much larger than q/M_p, it only arises in $\mathcal{A}_E$ together with a factor of q^2/M_p^2, making it hard in practice to exploit the hierarchy $M \ll M_p$ to obtain surprisingly large effects.

Eq. (2.16) can be used to identify the dominant contributions to any low-energy process (graviton scattering amplitude or correlation function) that is characterized by a single scale $q \ll M \ll M_p$. Eq. (2.16) shows that the least suppressed contributions come from graphs with $L = 0$ and $V_{id} = 0$ for all $d > 2$. That is to say, using only tree graphs ($L = 0$) constructed purely from the Einstein-Hilbert ($d = 2$) action. As might have been expected, it is classical General Relativity that dominantly governs the low-energy dynamics of gravitational fluctuations.

For instance, the above estimate applies in particular to graviton-graviton scattering, in which case we take $E = 4$. Specializing eq. (2.16) to this case (with $L = 0$ and $V_{id} = 0$ for all $d > 2$) then says that at low energies we have $\mathcal{A}_4 \simeq (q/M_p)^2$, which agrees well with the result obtained by explicit calculation [25], which for 2-body graviton scattering on flat space gives (at tree level)

$$\mathcal{A}_4 \simeq 8\pi i G_N \left(\frac{s^3}{tu}\right) , \tag{2.17}$$

where s, t and u are the Mandelstam invariants for 2-body scattering, defined in terms of the initial 4-momenta p_i^μ and final 4-momenta $p_i'^\mu$ by $s = -(p_1 + p_2)^2$, $t = -(p_1 - p_1')^2$ and $u = -(p_1 - p_2')^2$. What is important for comparing with the estimate $(q/M_p)^2$ is that $8\pi G_N = M_p^{-2}$ and s, t and u when evaluated in the centre-of-mass frame are $s = 4E_{cm}^2$, $t = -2E_{cm}^2(1 - \cos\vartheta)$ and $u = -2E_{cm}^2(1 + \cos\vartheta)$, where E_{cm} is the center-of-mass energy and ϑ is the angle between the incoming momentum $\mathbf{p}_1$ and the outgoing momentum $\mathbf{p}_1'$ in this frame. These imply $\mathcal{A}_4 \propto (E_{cm}/M_p)^2$ in agreement with (2.16) with E_{cm} playing the role of the low-energy scale q.

But (2.16) also identifies which graphs give the next-to-leading contributions. These come in one of the following two ways:

- $L = 1$ and $V_{id} = 0$ for any $d \neq 2$ but V_{i2} is arbitrary, or
- $L = 0$, $\sum_i V_{i4} = 1$, V_{i2} is arbitrary, and all other V_{id} vanish.

That is to say: the next to leading contribution comes from one-loop graphs constructed using only the interactions of General Relativity, or by working to tree level and including precisely one insertion of a curvature-squared interaction in addition to any number of interactions from GR. Both of these are suppressed compared to the leading term by a factor of $(q/M_p)^2$. The next-to-leading tree graphs provide precisely the counter-terms required to absorb the UV divergences in the one-loop graphs. And so on to any desired order in the expansion. Despite being nonrenormalizable the theory is predictive *provided* one works only to a fixed order in q/M and q/M_p.

2.2 Power-counting in cosmology

We are now ready for the main event: asking more systematically whether the observation that any successful theory of cosmology must emerge as the low-energy EFT for some more fundamental theory carries any practical consequences.

To this end – and partly with scalar models of Dark Energy and/or Dark Matter in mind – we repeat the power-counting estimates made above for GREFT but this time do so for gravity coupled to a collection of N dimensionless scalar fields, θ^i. A generic EFT containing these low-energy fields can be expanded in a derivative expansion, leading to a lagrangian that extends (2.7) to include new scalar interactions:

$$\begin{aligned} -\frac{\mathcal{L}_{\rm eff}}{\sqrt{-g}} &= v^4 U(\theta) + \tfrac{1}{2}\, g^{\mu\nu}\Big[M_p^2\, W(\theta)\, R_{\mu\nu} + f^2\, G_{ij}(\theta)\, \partial_\mu\theta^i\partial_\nu\theta^j\Big] \\ &\quad + A(\theta)(\partial\theta)^4 + B(\theta)\, R^2 + C(\theta)\, R\,(\partial\theta)^2 + \frac{E(\theta)}{M^2}\,(\partial\theta)^6 + \frac{F(\theta)}{M^2}\, R^3 + \cdots\,, \end{aligned} \tag{2.18}$$

where all terms involving up to two derivatives are written explicitly in the first line, with the rest written schematically on the second line.[16]

The explicit mass scales M_p and M are explicitly written, as before, so that the functions $W(\theta)$, $G_{ij}(\theta)$, $A(\theta)$, $B(\theta)$ *etc*, are dimensionless. The functions $W(\theta)$ and $G_{ij}(\theta)$ are positive definite and there can be positivity conditions on some of the other functions as well [24]. Here M is the *lowest* scale integrated out to obtain $\mathcal{L}_{\rm eff}$ (since this is what dominates in the denominator). We allow the scalar kinetic term to be normalized differently, with some new scale f appearing there instead of M_p, but our interest here is scalars that interact with gravitational strength for which $f = M_p$. We assume $M \ll M_p$ and we take $f < M_p$ in situations where $f \neq M_p$.

A new scale, v, is also added so that the scalar potential $V(\theta)$ is order v^4 when the dimensionless function $U(\theta)$ is order unity. With cosmological applications in mind we imagine v to be a low-energy scale and take $v \ll M \ll M_p$. Notice that if we were to use a more canonical normalization where $\phi^i \sim f\,\theta^i$ then if $U(\theta) = \lambda_0 + \lambda_i\theta^i + \lambda_{ij}\theta^i\theta^j + \cdots$ is order unity for θ^i order unity then this choice for the scalar potential implies

$$V(\phi) = v^4\left[\lambda_0 + \lambda_i\frac{\phi^i}{f} + \lambda_{ij}\frac{\phi^i\phi^j}{f^2} + \lambda_{ijk}\frac{\phi^i\phi^j\phi^k}{f^3} + \cdots\right], \tag{2.19}$$

which shows that V changes by order v^4 as ϕ^i ranges through values of order f. This form captures qualitative features of many explicit cosmological models when $f \sim M_p$ and $\lambda_{ijk\cdots} \sim \mathcal{O}(1)$, but for future purposes it is worth keeping in mind that these choices make V remarkably shallow compared to most scalar potentials considered in particle physics (more

[16]These higher derivative terms are schematic inasmuch as R^3 collectively represents all possible independent curvature invariants involving six derivatives, and so on.

about which in later sections). In this language a typical particle physics potential would change by order v^4 when ϕ goes through scales much smaller than M_p (which amounts to taking $f \ll M_p$ in (2.19)).[17]

As usual, there is considerable freedom to simplify the action (2.18) by performing field redefinitions, and we use this freedom to Weyl rescale the metric to set $W(\theta) = 1$ (*i.e.* go to Einstein frame). The function $G_{ij}(\theta)$ can often similarly be simplified through redefinitions of the form $\theta^i \to f^i(\theta)$, but it cannot be made θ-independent if the Riemann tensor $\mathcal{R}^i{}_{jkl}$ built in the usual way from the 'target-space metric' G_{ij} is flat.[18]

2.2.1 Power counting

From here the argument proceeds much as in the earlier section discussing pure gravity (for details see [26, 27, 28]). To this end, as above, we expand $S_{\rm eff} = \int {\rm d}^4x\, \mathcal{L}_{\rm eff}$ about a classical solution using fields that have canonical dimension, $\theta^i(x) = \bar\theta^i(x) + \phi^i(x)/f$ and $g_{\mu\nu}(x) = \bar g_{\mu\nu}(x) + h_{\mu\nu}(x)/M_p$. As above we keep track of the scales appearing in the action (2.18) by reading them off from the Feynman rules for each vertex and propagator, and we assign the dependence on any low energy scale purely on dimensional grounds.

The dependence of a Feynman graph on the scales M, M_p, f and v are found by expanding the lagrangian (2.18) in powers of the fluctuation fields ϕ^i and $h_{\mu\nu}$, leading to a sum of interactions of the form

$$\mathcal{L}_{\rm eff}(\theta, g_{\mu\nu}) = \mathcal{L}_{\rm eff}(\bar\theta, \bar g_{\mu\nu}) + M^2 M_p^2 \sum_n \frac{c_n}{M^{d_n}}\, \mathcal{O}_n\left(\frac{\phi^i}{f}, \frac{h_{\mu\nu}}{M_p}\right) \tag{2.20}$$

where the functions, $\mathcal{O}_n$, are monomials involving $N_n = N_n^{(\phi)} + N_n^{(h)} \geq 2$ powers[19] of the fields ϕ^i and $h_{\mu\nu}$ and their derivatives. The parameter d_n counts the number of derivatives appearing in $\mathcal{O}_n$, and the coefficients c_n are dimensionless and calculable in terms of the functions $U(\theta)$, $G_{ij}(\theta)$ and so on (and their derivatives) evaluated at the background fields.

The prefactor, $M^2 M_p^2$, ensures the kinetic terms (*i.e.* the terms with $N_n^{(h)} = d_n = 2$) for $h_{\mu\nu}$ are independent of M and M_p, and so the same is also true for its propagator. The

[17]For instance the Higgs potential in the Standard Model would correspond to choosing $v \sim f \sim 100$ GeV while the QCD axion potential has $v \sim \Lambda_{QCD} \sim 0.2$ GeV and $f \sim 10^{10}$ GeV.

[18]$G_{ij}(\theta)$ transforms like a covariant tensor under redefinitions of the θ^i fields, and since it also is positive and symmetric it can be regarded as a metric on the 'target space' (*i.e.* the range of the function $\theta^i(x)$).

[19]Terms linear in the fluctuations only arise at subdominant order, where they cancel the 'tadpole' graphs (those with exactly one external line), since this is the condition that defines the background fields. At leading order this makes the background fields solve the classical equations of motion, with corrections order by order in perturbation theory.

same is true for the kinetic term for ϕ^i if $f = M_p$. For more general f the expansion of the 'sigma-model' term $f^2\, G_{ij}\, \partial\phi^i \partial\phi^j$ term of (2.18) gives operators of the form (2.20) that are proproportional to f^2/M_p^2, so

$$c_n^{\sigma\,\mathrm{mod}} = \hat{c}_n \left(\frac{f^2}{M_p^2} \right) , \tag{2.21}$$

where $\hat{c}_n$ is order unity. This ensures that the ϕ^i propagators are also scale independent.

Similarly, the lagrangians of (2.20) and (2.18) only make equivalent predictions for the M and M_p dependence in Feynman graphs if the coefficients c_n for all of the higher-derivative terms in (2.18) are proportional to M^2/M_p^2, so

$$c_n = \left(\frac{M^2}{M_p^2} \right) g_n \qquad (\text{if } d_n > 2) \,, \tag{2.22}$$

where g_n is at most order unity. For terms with no derivatives — *i.e.* those coming from the scalar potential, $V(\theta)$ — one instead finds that agreement requires

$$c_n = \left(\frac{v^4}{M^2 M_p^2} \right) \lambda_n \qquad (\text{if } d_n = 0) \,, \tag{2.23}$$

where the dimensionless couplings λ_n are also independent of M_p and M (up to logarithms).

Consider first the case where $f = M_p$ and ask how the various scales enter into an amputated E-point correlator of $h_{\mu\nu}$ and ϕ^i fields at L loops. As before we keep track of the coupling for each vertex to see how the scales M, M_p and v appear in the graph. Using dimensional analysis – with dimensional regularization to remove the need for a confounding cutoff scale – then gives the dependence on the (assumed) single low-energy scale (call it H this time since it is often the Hubble scale in cosmology):

$$\mathcal{A}_{\scriptscriptstyle E}(H) \simeq M_p^2 H^2 \left(\frac{1}{M_p} \right)^E \left(\frac{H}{4\pi\, M_p} \right)^{2L} \mathfrak{F}_{d=0}\, \mathfrak{F}_{d=2}\, \mathfrak{F}_{d>2} \,, \tag{2.24}$$

where vertices coming from the $d_n = 0$ terms of the lagrangian (the scalar potential) contribute the factor

$$\mathfrak{F}_{d=0} = \prod_n \left[\lambda_n \left(\frac{v^4}{H^2 M_p^2} \right) \right]^{V_n} \qquad (\text{if } f = M_p) \,, \tag{2.25}$$

while the 2-derivative terms and higher-derivative terms contribute

$$\mathfrak{F}_{d=2} = \prod_n c_n^{V_n} \qquad (\text{if } f = M_p) \,, \tag{2.26}$$

and

$$\mathfrak{F}_{d>2} = \prod_n \left[g_n \left(\frac{H}{M_p} \right)^2 \left(\frac{H}{M} \right)^{d_n - 4} \right]^{V_n} \qquad (\text{if } f = M_p) \,. \tag{2.27}$$

In all three of these expressions the product only runs over those values of n that correspond to vertices that have the number of derivatives as indicated (so $d_n = 0$ for (2.25), $d_n = 2$ for (2.26) and $d_n > 2$ for (2.27)).

Eqs. (2.24) through (2.27) are key results because they quantify very explicitly the size of corrections to semiclassical methods.

- The appearance of only positive powers of H (except for within $\mathfrak{F}_{d=0}$ – more about which below) again verifies that the derivative expansion controls perturbative calculations made using (2.18), with perturbation theory relying on there being a hierarchy $H \ll M_p$ and $H \ll M$.

- For a fixed number of external lines E each loop costs a factor of $H^2/(4\pi M_p)^2$, so this is again what controls the corrections to the semiclassical expansion.

- Once again there are no low-energy penalties for using as many 2-derivative interactions as we like, either from the Einstein-Hilbert lagrangian and from the sigma-model term $M_p^2 \, G_{ij} \, \partial\phi^i \partial\phi^j$. At low energies the full nonlinearity of GR remains important and the two-derivative interactions in the sigma-model term like to compete with GR.

- The scalar-potential terms appearing in $\mathfrak{F}_{d=0}$ are generically dangerous because in them H appears in the *denominator* rather than the numerator. This acts to undermine the validity of the semiclassical approximation because including zero-derivative interactions in a graph can amplify its size and make it no longer subdominant to other graphs that were naively bigger. This can make semiclassical methods suspect in surprising ways. This issue does *not* pose a problem for cosmological models if the low-energy scale H is the Hubble scale and if the potential is what generates the Hubble curvature, since in this case the Friedmann equation implies $H \simeq v^2/M_p$, and so connects the size of H to the scale v in the potential. When this is true the potentially dangerous factor $\mathfrak{F}_{d=0}$ becomes

$$\mathfrak{F}_{d=0} = \prod_n \left[\lambda_n \left(\frac{v^4}{H^2 M_p^2} \right) \right]^{V_n} \simeq \prod_{d_n=0} \lambda_n^{V_n} \,, \tag{2.28}$$

As before, for any low-energy process that is characterized by a single scale $H \ll M \ll M_p$ the dominant contributions to observables come from graphs with $L = 0$ and $V_{id} = 0$ for all $d > 2$; *i.e.* tree graphs constructed using just from the $d \leq 2$ terms in the action: the Einstein-Hilbert term with the sigma model and scalar-potential terms. The leading corrections are again generated by loops involving these interactions plus tree graphs that include precisely one 4-derivative interaction, and so on.

Notice that there is no particular low-energy penalty working with large fields $\phi \gtrsim M_p$ *provided* the functions $U(\theta)$, $G_{ij}(\theta)$ and the like remain order unity when θ is order unity. That is, if large fields do not also imply large energy then they need not cause difficulties with the low-energy limit.

2.2.2 More general f

Before continuing we pause here briefly to record how power-counting formulae like (2.24) change if we relax the condition $f \simeq M_p$, typically with scales $f \ll M_p$ in mind. Rather than setting things up from scratch again it is easier to just flag the changes relative to the estimate made above when $f = M_p$.

Inspection of (2.20) and (2.21) shows that there are two sources of change. One of these is the explicit factor of f^2/M_p^2 that now appears in the Feynman rule for any vertex coming from expanding the sigma-model interaction $G_{ij}\,\partial\phi^i\,\partial\phi^j$ about the background. This has the effect of multiplying the factor $\mathfrak{F}_{d=2}$ given in (2.26) by a factor $\prod_n (f^2/M_p^2)^{V_{\sigma n}}$ where $V_{\sigma n}$ counts the number of vertices of type 'n' that come specifically from the sigma-model interaction. For 2-derivative interactions coming from the Einstein-Hilbert term there is no change and $\hat{c}_n = c_n$ (rather than their being related by (2.21)).

The other change comes because the scalar fields appear as ϕ/f in (2.20) rather than as ϕ/M_p as was used earlier when studying the limit $f = M_p$. This is corrected by taking every appearance of ϕ in any interaction and rescaling $\phi \to (M_p/f)\,\phi$. This introduces a new factor into the amplitude $\mathcal{A}_{E}(H)$ of the form $\prod_n (f/M_p)^{-\mathfrak{s}_n V_n}$, where the product is over all vertices (not just those coming from the sigma-model interaction) and $\mathfrak{s}_n$ counts the number of scalar lines that converge on vertex n (and $\mathfrak{h}_n$ similarly counts the number of $h_{\mu\nu}$ lines that converge on this vertex).

Combining both sources of f-dependence again leads to expression (2.24) – repeated again here:

$$\mathcal{A}_{E}(H) \simeq M_p^2 H^2 \left(\frac{1}{M_p}\right)^{E} \left(\frac{H}{4\pi\, M_p}\right)^{2L} \mathfrak{F}_{d=0}\, \mathfrak{F}_{d=2}\, \mathfrak{F}_{d>2}\,, \tag{2.29}$$

but with (2.25) through (2.27) replaced by

$$\mathfrak{F}_{d=0} = \prod_n \left[\lambda_n \left(\frac{v^4}{H^2 M_p^2}\right)\right]^{V_n} \left(\frac{f}{M_p}\right)^{-\mathfrak{s}_n V_n} \qquad (\text{general } f)\,, \tag{2.30}$$

while the 2-derivative terms and higher-derivative terms contribute

$$\mathfrak{F}_{d=2} = \prod_n \hat{c}_n^{V_n} \left(\frac{f}{M_p}\right)^{2V_{\sigma n} - \mathfrak{s}_n V_n} \qquad (\text{general } f)\,, \tag{2.31}$$

and

$$\mathfrak{F}_{d>2} = \prod_n \left[g_n \left(\frac{H}{M_p}\right)^2 \left(\frac{H}{M}\right)^{d_n-4}\right]^{V_n} \left(\frac{f}{M_p}\right)^{-\mathfrak{s}_n V_n} \qquad (\text{general } f)\,. \tag{2.32}$$

As before the products run over the vertices that have the number of derivatives given by d.

The power of f/M_p can be rewritten using three very useful identities, which hold for any graph consisting of I_s scalar internal lines and I_h tensor internal lines. The first of these is simply the definition of the number of loops in a graph:[20]

$$1 = L - I + \sum_n V_n \qquad (\text{definition of } L) \tag{2.33}$$

where $I = I_s + I_h$ is the total number of internal lines. The other two identities express 'conservation of ends' (which says the number of ends of external and internal lines must equal the number of ends appearing in all vertices, separately for both scalar and tensor lines):

$$E_s + 2I_s = \sum_n \mathfrak{s}_n V_n \qquad \text{and} \qquad E_h + 2I_h = \sum_n \mathfrak{h}_n V_n\,. \tag{2.34}$$

Recall here that $\mathfrak{s}_n$ and $\mathfrak{h}_n$ respectively count the number of scalar and tensor lines that converge at vertex 'n'. These last two identities can be used to eliminate I_s and I_h from any expression, and after this is done (2.33) implies the number of vertices and external lines are related to the number of loops by the following expression:

$$E + 2(L-1) = \sum_n \Big[(\mathfrak{s}_n + \mathfrak{h}_n) - 2\Big] V_n\,, \tag{2.35}$$

where $E = E_s + E_h$.

This is useful because it allows the power of the product of the factors $(f/M_p)^{-\mathfrak{s}_n V_n}$ appearing in (2.30) through (2.32) to be written as

$$-\sum_n \mathfrak{s}_n V_n = 2(1-L) - E + \sum_n \Big(\mathfrak{h}_n - 2\Big) V_n\,. \tag{2.36}$$

This in turn allows (2.29) to be cast in a way where the factors of f/M_p mostly have positive powers:

$$\mathcal{A}_E(H) \propto f^2 H^2 \left(\frac{1}{f}\right)^E \left(\frac{H}{4\pi f}\right)^{2L} \widetilde{\mathfrak{F}}_{d=0}\, \widetilde{\mathfrak{F}}_{d=2}\, \widetilde{\mathfrak{F}}_{d>2}\,, \tag{2.37}$$

where

$$\widetilde{\mathfrak{F}}_{d=0} = \prod_n \left[\lambda_n \left(\frac{v^4}{H^2 f^2}\right)\right]^{V_n} \left(\frac{f}{M_p}\right)^{\mathfrak{h}_n V_n} \qquad (\text{general } f)\,, \tag{2.38}$$

[20] This definition gives the intuitive number of loops for any graph that is topologically a disc – *i.e.* can be drawn on a page (draw some graphs and check) – but is the definition of L regardless of the graph's topology.

while the 2-derivative terms and higher-derivative terms contribute

$$\widetilde{\mathfrak{F}}_{d=2} = \prod_n \hat{c}_n^{V_n} \left(\frac{f}{M_p}\right)^{(\mathfrak{h}_n - 2)V_{EHn} + \mathfrak{h}_n V_{\sigma n}} \qquad \text{(general } f\text{)}, \tag{2.39}$$

and

$$\widetilde{\mathfrak{F}}_{d>2} = \prod_n \left[g_n \left(\frac{H}{f}\right)^2 \left(\frac{H}{M}\right)^{d_n-4}\right]^{V_n} \left(\frac{f}{M_p}\right)^{\mathfrak{h}_n V_n} \qquad \text{(general } f\text{)}. \tag{2.40}$$

Eq. (2.39) uses that 2-derivative interactions must either come from the sigma-model term or the Einstein-Hilbert term – *c.f.* the lagrangian (2.18) – so for them $V_n = V_{\sigma n} + V_{EHn}$.

As a check, consider evaluating a graph with L loops and E_s external scalar lines and E_h external tensor lines, for which *all* of the vertices come from the sigma-model interaction $G_{ij}\,\partial\phi^i\,\partial\phi^j$). In this case the only vertices present have $d_n = 2$ derivatives so $\mathfrak{F}_{d=0} = \mathfrak{F}_{d>2} = 1$. Since every vertex comes from the sigma-model interaction we also have $V_n = V_{\sigma n}$ for all nonzero V_n's. Under these assumptions the dependence on M, f, v, and M_p comes from using (2.31) in (2.29). Combining everything implies either (2.29) or (2.37) can be written

$$\mathcal{A}_E(H) \propto f^2 H^2 \left(\frac{1}{f}\right)^E \left(\frac{H}{4\pi f}\right)^{2L} \left(\frac{f}{M_p}\right)^{\sum_n \mathfrak{h}_n V_n}. \tag{2.41}$$

In the special case where $h_{\mu\nu}$ does not appear in the graph we have $\mathfrak{h}_n = 0$ and this reproduces the standard sigma-model power-counting expression [16] – which is basically (2.24) with $\mathfrak{F}_{d=0} = \mathfrak{F}_{d>2} = 1$ and $M_p \to f$. More generally $\sum_n \mathfrak{h}_n V_n \geq E_h$ since any external metric line must end somewhere, so an amplitude with E_s external scalars and E_h external metric fluctuations always has a proportionality constant of at least $(1/M_p)^{E_h}(1/f)^{E_s}$.

A second useful special case is the result for graphs with no metric external or internal lines: $E_h = I_h = 0$ and so from (2.34) we also have $\mathfrak{h}_n = 0$. Because we work in Einstein frame the only interactions coming from the Einstein-Hilbert term involve the metric and so for a scalar-only graph we can take $V_{EHn} = 0$. In this case eqs. (2.37) through (2.40) imply

$$\mathcal{A}_E(H) \propto f^2 H^2 \left(\frac{1}{f}\right)^{E_s} \left(\frac{H}{4\pi f}\right)^{2L} \prod_{d_n=0} \left[\lambda_n \left(\frac{v^4}{H^2 f^2}\right)\right]^{V_n} \prod_{d_n>2} \left[g_n \left(\frac{H}{f}\right)^2 \left(\frac{H}{M}\right)^{d_n-4}\right]^{V_n}. \tag{2.42}$$

Recall that the scalar potential depends on the coefficients λ_n through (2.19).

2.3 Lessons for cosmology

Nothing *requires* UV physics to provide new light fields for us to discover, but the fact that cosmology only makes sense if well-understood matter is supplemented by Dark Energy and

Dark Matter seem to suggest that new light fields might well be present in Nature. It is also true that the few theories we have for which gravitational interactions make sense at the quantum level *at high energies* (such as string theory) usually predict the existence of a host of new particles that couple very weakly – often only with gravitational strength – to ordinary matter, of which the graviton is only one example. If any of these were to be light they could also be around to be discovered in cosmology or tests of GR (and might indeed play a role in Dark Matter or Dark Energy). For such new fields the previous section sets the general stage for what should be expected for them at low energies.

It already tells us something interesting. First, it tells us is that what cosmologists like the most about scalar field phenomenology – the scalar potential – is also the thing that is the most dangerous at low-energies. The potential is important for several reasons. A key property for any field relevant to cosmology or tests of GR is that it is very light compared with other scales in particle physics, and for scalar fields this is encoded in the scalar potential. Similarly, we saw in §1.2.2 that for simple scalar models of Dark Energy the equation of state parameter tells us that the energy density is currently dominated by the scalar potential.

The scalar potential actually contains several types of dangers at low energies. One of these is the appearance of inverse powers of the low-energy scale H in expressions like (2.30), which work to undermine the low-energy expansion that underpins the entire strategy of analyzing the model semiclassically. This becomes more of a problem the stronger the zero-derivative scalar interactions are. But we've also seen how this need not be a problem for potentials with the structure $V = v^4 U(\phi)$ for generic order-unity functions $U(\phi)$, when the low-energy scale H is the Hubble scale dominated by V, since in this case the dangerous factor in (2.25) cancels out because $H \sim v^2/M_p$, as in eq. (2.28).

In this section we start, in §2.3.1, by assuming the scalar potential is suppressed to the point that it does not overwhelm the two-derivative terms at low energies, returning in §3 to how difficult this regime is to achieve and to discuss more broadly the problems raised at low energies by zero-derivative interactions.

2.3.1 Two-derivative interactions: more is different

If the potential is somehow suppressed at low energies then it is the sigma-model term $G_{ij}\,\partial\phi^i\,\partial\phi^j$ that provides the next most important interactions for scalars at low energies. These scale at low-energies in precisely the same was as do the interactions in the Einstein-Hilbert action of GR, which also involve precisely two derivatives. Two-derivative interactions can compete with one another at all energies without undermining the underlying derivative expansion, and with it the validity of semiclassical methods themselves.

But (as mentioned above) these cannot have physical implications if they can be removed by a field redefinition, and this can always be done if the target-space metric $G_{ij}(\phi)$ is flat. An important example where the target space is flat is the case of a single scalar field. For a single field it is always possible to redefine $\phi = \phi(\psi)$ such that $G(\phi)\,(\partial\phi)^2 = (\partial\psi)^2$, with $\phi(\psi)$ found by integrating $G(\phi)\,\mathrm{d}\phi = \mathrm{d}\psi$. Because of this, in the special case of a single scalar field the leading nonminimal couplings are the higher-derivative interactions appearing in the second line of (2.18).

A large effort has been made to characterize the kinds of higher derivative interactions that can arise for a single scalar field, and how these can contribute to classical observables (see for example [29]).[21] There is an important conceptual problem with these analyses, however: to the extent that an interaction with 4 or more derivatives starts to compete with GR the power-counting argument that culminated in eqs. (2.24) through (2.27) shows that it is necessarily true that the derivative expansion itself must be breaking down. But (2.24) also shows that when this is true there is no justification for analyzing these systems classically while neglecting loop effects.

But this unfortunate conclusion is specific to the restriction to a single scalar field (or to multiple fields with a flat target space metric since for these $G_{ij} = \delta_{ij}$ can be arranged using a field redefinition). There is a premium for exploring models with at least two low-energy scalar fields because these are the only ones that allow the nontrivial sigma-model interactions that scale at low energies in the same way as do the interactions of GR itself.

If we work within the class of effective theories with scalar potentials of the form $V = v^4 U(\theta)$, where $U(\theta)$ is a generic order-unity function, then there is also an argument why it is not especially unlikely to have more than one field be very light. After all, the scalar mass matrix implied by this form for the potential is

$$M^2_{ij} := \frac{\partial^2 V}{\partial\phi^i \partial\phi^j} = \left(\frac{v^4}{f^2}\right) \frac{\partial^2 U}{\partial\theta^i \partial\theta^j}\,, \tag{2.43}$$

so if all of the derivatives of U are order unity the mass eigenvalues are all of order $m \simeq v^2/f$. In particular, for gravitationally coupled scalars $f \sim M_p$ and so these masses are all the same order as $H \sim v^2/M_p$. As we shall see, the puzzle with potentials with this kind of structure

[21] It is sometimes argued that it is necessary to restrict to a subset of higher-derivative interactions in order to avoid the Ostrogradski instability [30] that is generic to theories with higher derivative interactions. This turns out not to be necessary when working with low-energy EFTs because of the perturbative nature of the $1/M$ expansion [31], although in practice it turns out that the distinction between these points of view only arises at relatively high order in $1/M$ [32].

is why v should be so small, but once it is any gravitationally coupled scalar appearing in it is necessarily very light.

Minimal multi-scalar models of this type will be encountered again in §4 below.

2.3.2 Lots of potential

Returning to the scalar potential, we next ask how generic it is to have a potential of the form $V = v^4 U(\theta)$ with small v. The problem is that this form seems not to be easy to obtain at low energies from theories like the ones we believe describe nature on much shorter length scales. The good news is that it is not completely impossible either, and trying to find how they can arise seems to be an important clue when searching for descriptions of cosmology.

The basic problem can already be seen in the lagrangian (2.18), in which it was argued that the various terms are generically generated by integrating out multiple heavy particles. This leads to the expectation that an interaction $g\mathcal{O} \in \mathcal{L}_{\rm eff}$ with operator, $\mathcal{O}$, with mass-dimension n arises with a coefficient g that has mass-dimension $4 - n$ (in 4 spacetime dimensions) so that $g\mathcal{O}$ has dimension (mass)^4, just like $\mathcal{L}_{\rm eff}$. For instance, an operator like $\mathcal{O} = R^3$ has mass-dimension $n = 6$ and so its coupling has dimension (mass)$^{-2}$, as indicated in (2.18).

Intuitively, integrating out a particle of mass M should generate a contribution to each effective operator proportional to the appropriate power of M dictated by dimensional analysis, so $\delta g \propto M^{4-n}$. If g is obtained by summing many such contributions then the dominant mass M controlling the size for each g depends on the sign of $4 - n$: it is the smallest mass that dominates in g when $n > 4$ and it is the largest mass that dominates when $n < 4$.

This intuition can be made more precise using the power-counting arguments of the previous sections by applying them to the graphs used to compute contributions to $\mathcal{L}_{\rm eff}$ when a light particle of mass m is integrated out. To this end suppose we have a scalar potential $V = m^4 U(\phi/m)$ and kinetic term $(\partial\phi)^2$, as might reasonably be chosen to describe a scalar with mass m. (The Higgs potential in the Standard Model, for example, has a potential and kinetic term of this type with $m \sim 100$ GeV.) Integrating out this type of particle gives a contribution to $\mathcal{L}_{\rm eff}$ and its size turns out to be the found by computing the generating functional for amputated 2-particle irreducible (2PI) correlation functions,[22] and so can be estimated by specializing the power-counting argument given above to the choices $f \sim v \sim m$.

A term in (2.18) involving E factors of fields ϕ^i depends on scales like M, M_p, $f = v$ in same way as does $\mathcal{A}_E$ where the low-energy scale H can either be m or an external momentum

[22] 2PI correlation functions are computed by evaluating amputated Feynman graphs that cannot be broken into two disconnected graphs by cutting a single internal line.

(or derivative) q. The power counting estimate for the size of the L-loop coefficient of ϕ^2 in the scalar potential is therefore $\delta m^2 \sim \mathcal{A}_2(m)$ (since there are no derivatives). An estimate of the size of the correction to the kinetic term $\mathcal{G}(\partial\phi)^2$ is order $(q/m)^2\mathcal{A}_2(m)$ because it must be precisely quadratic in q, and so $\delta\mathcal{G} \sim m^{-2}\mathcal{A}_2(m)$. The contribution to the coefficient of $\mathcal{H}\phi^2(\partial\phi)^2$ is similarly order $\delta\mathcal{H} \sim m^{-2}\mathcal{A}_4(m)$, and so on.

Using $f \sim v \sim H \sim m$ in (2.42) therefore gives (in order of magnitude)

$$\delta m^2 \sim m^2 \left(\frac{1}{4\pi}\right)^{2L} \left(\prod_{d_n=0} \lambda_n^{V_n}\right) \prod_{d_n>2} \left[g_n \left(\frac{m}{M}\right)^{d_n-4}\right]^{V_n} \tag{2.44}$$

and

$$\delta\mathcal{G} \sim \left(\frac{1}{4\pi}\right)^{2L} \left(\prod_{d_n=0} \lambda_n^{V_n}\right) \prod_{d_n>2} \left[g_n \left(\frac{m}{M}\right)^{d_n-4}\right]^{V_n} \tag{2.45}$$

and

$$\delta\mathcal{H} \sim \frac{1}{m^2} \left(\frac{1}{4\pi}\right)^{2L} \left(\prod_{d_n=0} \lambda_n^{V_n}\right) \prod_{d_n>2} \left[g_n \left(\frac{m}{M}\right)^{d_n-4}\right]^{V_n} \tag{2.46}$$

which show the expected dimensional dependence on the mass m of the particle that is integrated out (plus corrections in powers of m/M). What we considered above as a light field of mass m would also be regarded as a heavy field from the point of view of any other, even lighter, field. From the point of view of this lighter field the particle of mass m is just another heavy field and so m would be lumped among the M's, which after all were masses of particles that had previously been integrated out. Because $m \ll M$ the contribution (say) $\delta\mathcal{H} \sim m^{-2}$ is much bigger than $\delta\mathcal{H} \sim M^{-2}$, showing in more detail why in practice we can take M to be the mass of the lightest UV field that was integrated out.

What is important is that the mass appearing on the right-hand side of an expression like (2.44) need not be the same as the mass being corrected on the left-hand side. For instance if we had two scalars, a scalar σ that is massless and a scalar ϕ with mass m that are coupled to one another by a term in V like $g\sigma^2\phi^2$ then the potential would start off looking like

$$V = \tfrac{1}{2}m^2\phi^2 + g\,\sigma^2\phi^2 = m^4\left(\frac{\phi^2}{2m^2} + \frac{g\,\phi^2\sigma^2}{m^4}\right), \tag{2.47}$$

and the leading correction to the σ mass due to integrating out ϕ would still be given by (2.44), with $L = 1$ and $\lambda_n = g$, implying $\delta m_\sigma^2 \sim g\, m^2/(4\pi)^2$.

From this point of view, when a sequence of heavy particles is integrated out it seems reasonable to find an enormous coefficient like[23] M_p^2 in front of the Einstein-Hilbert term given that for it the mass-dimension is $n = 2$ and so loops of the heaviest fields should dominate. *But all other things being equal it is not generic to find small contributions to interactions with mass-dimension $n < 4$.* This is true in particular for the vacuum energy (a field-independent piece in V for which $n = 0$) and for scalar masses ($V = \frac{1}{2}m^2\sigma^2$ for which $n = 2$). Cosmological applications require both of these to be small, whereas they naively should appear with prefactors M_p^4 and M_p^2 respectively.

How can this be reasonable? As the rest of these notes argue: the key part of the above italicized phrase is 'all other things being equal'. All other things need *not* be equal and understanding how cosmology can emerge from the low-energy limit of something more fundamental provides a crucial clue for unravelling what is really going on.

3. Naturalness: Tomorrow's Hope or Yesterday's News?

One attitude to take about the corrections to $\mathcal{L}_{\rm eff}$ estimated above is that they do not matter. This section discusses this attitude and argues why – despite being a self-consistent point of view – it has not put the discussion to rest.

3.1 The issue

Why should we care if the coefficient of the Einstein-Hilbert term arises as a sum over the masses of heavy particles when it is in any case only their sum M_p^2 that we directly measure?

This question becomes even more pointed when it is recalled that the Feynman graphs performed when generating this sum actually diverge in the UV.[24] Such divergences represent

[23] One might ask why the above argument ever stops – *i.e.* why should the Planck scale emerge as the 'heaviest' scale rather than something even bigger? Heaviest here means heaviest until one of the assumptions going into this EFT power-counting fails. This can happen if one reaches an energy scale above which a tower of states emerges, such as Kaluza-Klein (KK) modes if the high-energy theory is higher-dimensional. EFT methods never capture the power-counting appropriate to infinite towers of states because there is no longer a hierarchy of scales to exploit. If the tower encountered is a KK tower the effective description transitions to a higher-dimensional field theory, in which similar estimates would apply up to a scale where this description also breaks down (such as the onset of a tower of string states). This is why calculations in string theory usually have the form given above but with the string scale playing the role of the largest mass scale (which becomes powers of M_p once factors of the extra-dimensional size and the string coupling are included).

[24] Recall that we ignored these divergences when making dimensional estimates because we chose to regularize them using dimensional regularization (rather than some sort of a UV cutoff). We imagine renormalizing using a dimensionless scheme like minimal subtraction, and all masses in this section are renormalized in this way (so our expressions are not UV divergent).

the contributions of the shortest wavelengths of the theory and we normally do not angst too much about their size since they are in the end of the day absorbed into a renormalization of otherwise unknown parameters in $\mathcal{L}_{\rm eff}$ (such as M_p^2). It is only the final renormalized combination that is measurable. If divergences can be ignored like this, why should finite-but-large contributions proportional to heavy masses be any more worrisome?

Another way to phrase this attitude is to observe that renormalized parameters in any lagrangian run as a function of scale, and are typically only measured at a particular scale. All that matters is the value of a parameter (like the mass of a light field) at the scale where it is measured to be small. Why do we care if it is not small at some other scale?

To answer this it helps to have a concrete example in mind. Suppose we have a fundamental theory with two types of massive particles involving a hierarchy of mass scales $m \ll M$. We imagine there being a fundamental theory describing physics at energies $E > M$ for which the scalar potential contains terms like

$$V_{UV} = \widetilde{V}_0 + \tfrac{1}{2}\Big(\widetilde{m}^2\phi^2 + \widetilde{M}^2\psi^2\Big) + g\,\phi^2\psi^2 + \lambda_\phi\phi^4 + \lambda_\psi\psi^4 + \cdots\,. \tag{3.1}$$

The physical masses of these particles (as measured by experiments say) are related to the parameters in the lagrangian (including the leading loop correction) by formulae like

$$m^2_{\rm phys} = \widetilde{m}^2 + (\psi\text{-loop}) + (\phi\text{-loop}) + \cdots = \widetilde{m}^2 + \frac{C_1\,g}{(4\pi)^2}\,\widetilde{M}^2 + \frac{C_2\lambda_\phi}{(4\pi)^2}\,\widetilde{m}^2 + \cdots\,, \tag{3.2}$$

and

$$M^2_{\rm phys} = \widetilde{M}^2 + (\psi\text{-loop}) + (\phi\text{-loop}) + \cdots = \widetilde{M}^2 + \frac{C_3\lambda_\psi}{(4\pi)^2}\,\widetilde{M}^2 + \frac{C_4\,g}{(4\pi)^2}\,\widetilde{m}^2 + \cdots\,, \tag{3.3}$$

in which C_i are order-unity dimensionless constants.

For applications to low energies we can integrate out ψ and work with the EFT involving only ϕ, whose scalar potential includes

$$V_{\rm eff} = V_0 + \tfrac{1}{2}m^2\phi^2 + \lambda_\phi\phi^4 + \cdots\,. \tag{3.4}$$

Explicitly performing the integral over ψ allows the new parameters to be computed in terms of the old ones, leading to

$$V_0 \simeq \widetilde{V}_0 + \frac{C_5}{(4\pi)^2}\,M_2^4\,, \qquad m^2 \simeq \widetilde{m}^2 + \frac{C_1\,g}{(4\pi)^2}\,M_2^2 \tag{3.5}$$

for another dimensionless constant C_5. The calculation of the loop-corrected physical mass in the low-energy theory then is

$$m^2_{\rm phys} = m^2 + \frac{C_2\lambda_\phi}{(4\pi)^2}\,m^2 + \cdots\,, \tag{3.6}$$

which agrees with (3.2) (to the order we work) because of (3.5).

Now comes the main point. Weak coupling and the absence of very heavy particles in the low-energy theory imply $m_{\rm phys}$ and m are approximately equal. But $m_{\rm phys}$ is an observable and so must have the same value for both the full theory or the low-energy EFT. But if $\widetilde{M}$ is sufficiently large then $\tilde{m}$ cannot be similar in size to $m_{\rm phys}$. Instead it must be large and approximately opposite in sign to the loop correction $C_2\, g\tilde{M}^2/(4\pi)^2$ so that these terms mostly cancel to leave a small value for m^2. The required cancellation can be extremely accurate. For instance if $g/(4\pi)^2 \sim 10^{-3}$, $\tilde{M} \sim 10^{15}$ GeV and $m_{\rm phys} \sim 10^2$ GeV the cancellation must occur to more than 23 decimal places.

3.2 Technical naturalness

The kind of cancellation described above is remarkable for several reasons. First, while it is true that parameters in the lagrangian run and it is true that parameters that are small at some scales need not be small at other scales what is odd about the above is not that the parameter $\widetilde{m}^2$ is much larger than m^2 in the high-energy theory. The odd thing is the extreme precision with which the value of $\widetilde{m}^2$ must be chosen. The basin of attraction for the flow of $\tilde{m}^2$ that leads to acceptably small values for m^2 is extraordinarily narrow. Couplings for which the high-energy couplings must take extraordinarily accurate values to reach acceptably small sizes at low energies are called 'fine-tuned'.

A second, more telling, remarkable feature of the fine-tuning described above is the fact that it is unusual. That is, there are many hierarchies of scale known in nature and none of the ones we understand actually work this way. Normally we never are required to deal with all of the degrees of freedom in the universe all at once (thank God) so our description is cast in terms of some sort of EFT. But there is not a unique choice of effective lagrangian since different EFTs apply at different scales. Normally whenever a system has a hierarchy of scales – like $m \ll M$ in the above example – the hierarchy can be understood in *any* of these EFTs, and not just in the EFT describing the scales where the smaller scale is measured.

A famous example of a hierarchy is the large size of atoms relative to nuclei: $a_B \sim 10^5 r_N$. We can describe this in the EFT below 100 MeV in which the basic particles are protons and electrons and in this EFT $r_N \sim m_N^{-1}$ is a given parameter (the structure of the proton has been integrated out) that is of order the inverse of the proton mass, while atomic radii are of order $a_B \sim (\alpha m_e)^{-1}$ where $\alpha \sim 0.01$ is the electromagnetic fine-structure constant and m_e is the electron mass. Atoms are larger than nuclei because α is small (electromagnetic interactions are weak) and the electron is much lighter than the proton.

But the same question can be asked in the effective theory (for instance the Standard Model) applicable at energies much larger than 100 MeV. In this theory protons are described as bound states of quarks and gluons and the size and mass of the nucleus are set by the size of the QCD scale Λ_{QCD} while atomic radii are again given by $(\alpha m_e)^{-1}$. In this theory one can in principle compute r_N in terms of Λ_{QCD} and can also compute how the values of parameters like α and m_e in the lagrangian change once the physics above 100 MeV is integrated out. In practice $r_N \sim \Lambda_{QCD}^{-1}$ and the change in α is order α^2 and the change in the electron mass is proportional to the electron mass, $\delta m_e \propto m_e$ and so these changes are not large. No fine-tuning is needed to ensure that nuclei are small compared to atoms; it is a question that has a direct answer within any EFT we choose to ask the question.

The atom/nucleus example is the rule not the exception: with the exception of the ones arising in cosmology (such as why the vacuum energy is small – more about which below) all the well-understood examples of scale hierarchy satsify two properties:

- There is an understanding within the fundamental theory at high scales (such as within the Standard Model) why the hierarchy holds in the first place (such as because some ratio of parameters like m/M is small).
- There is an understanding of why the parameter choices necessary for the hierarchy *stay* small as successive layers of physics are integrated out to reach the lower energies where the parameters are measured. The hierarchy has an understanding in *all* the EFTs describing scales in between.

A hierarchy that satisfies both of these criteria is called *technically natural*.[25] Our understanding of the relative size of nuclei and atoms is technically natural in this sense, while the understanding of why $m \ll M$ in the two-scalar model discussed above is not technically natural. The question of how to understand the small size of the vacuum energy density in a technically natural way is widely known as the *cosmological constant problem*.[26]

It is conservative to ask that our understandings of other more poorly understood hierarchies should also be technically natural, since this just extrapolates what we know to be true in all other instances we do understand. But it is not compulsory. It might be that Nature

[25]The qualifier 'technically' is needed here to distinguish from many other things that are often called 'natural', some of which are only aesthetic.

[26]This problem pre-dated the discovery of evidence for Dark Energy because it was equally puzzling why the vacuum energy could be consistent with zero, given the much larger scales arising in particle physics. The question of why the Dark Energy density takes precisely the value it is observed to have rather than being for some reason exactly zero is sometimes called the 'new' cosmological constant problem.

does not care and theories like the two-scalar model are self-consistent even if fine-tuned. Technical naturalness is a very useful clue however because the ingredients needed to make a theory technically natural cannot just involve particles at inaccessibly high energies, like the Planck scale. If this were true we could integrate them out at low energies and the problem with technical naturalness problem is back. The mechanism that keeps the small parameter small in a natural way usually has other consequences at low energies and this tends to make these theories easier to test than the alternatives.

The next few sections describe approaches that have been tried to make the scalar field theories of interest to cosmology technically natural. This is a restrictive criterion because it is not generic that the things that are appealing for cosmological models (like very small scalar masses, m^2, and small vacuum energies, V_0, in a scalar potential) are technically natural. One hopes in this way to identify a well-motivated subset of the very many cosmological models on the market. Although observations alone cannot yet distinguish amongst the many models the hope is that observations together with technical naturalness can be a much more efficient filter. As we shall see, it is much easier to understand why scalar masses can be small in a technically natural way than it is to do the same for the vacuum energy density.

3.3 The Usual Suspects (symmetries)

Experience with other (noncosmological) hierarchies teaches that there is an important general mechanism for making small parameters technically natural: symmetries. Symmetries are usually preserved by quantum corrections (anomalies are the exceptions) so symmetry breaking term in a lagrangian are not generated by loops if the initial theory doesn't have them (and so respects the symmetry).

This also means that if a symmetry is only approximate – *i.e.* is broken by interactions with small coupling parameters, $\epsilon_i \ll 1$ – then loop corrections to these parameters satisfy:

$$\delta\epsilon_i \simeq C_i{}^j \epsilon_j \,, \tag{3.7}$$

for some matrix of coefficients $C_i{}^j$ since the corrections must vanish if the parameters themselves vanish (because then the symmetry is then unbroken). Having $\delta\epsilon$ be proportional to ϵ forbids getting large contributions to otherwise small parameters (like $\delta m^2 \propto M^2$ where $M \gg m$) and so help understand why small parameters can be technically natural.

For example suppose we have two scalar fields, ϕ and ψ, and the action for them is invariant under a symmetry of the form

$$\delta \begin{pmatrix} \phi \\ \psi \end{pmatrix} = \begin{pmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{pmatrix} \begin{pmatrix} \phi \\ \psi \end{pmatrix} \tag{3.8}$$

for arbitrary constant ω. The mass term allowed by this symmetry is $\frac{1}{2}m^2(\phi^2 + \psi^2)$ and so the symmetry requires both particles to have equal masses. Imagine now the theory is supplemented by a symmetry-breaking term of the form $\delta V = \frac{1}{2}\mu^2(\phi^2 - \psi^2)$ that splits these masses, with $\mu^2 \ll m^2$ and all other interactions are invariant under (3.8). Then $\delta\mu^2$ must be proportional to μ^2 – as opposed to, say m^2 – and the parameter μ^2 is technically natural. Importantly this is true even though corrections to scalar masses are otherwise *generically* expected to be dominated by the largest mass scales. The largest mass still wins, but the symmetry argument only allows contributions from the largest mass that breaks the symmetry.

This is what actually happens for the electron mass in the Standard Model (the electron mass $m_e \simeq 0.5$ MeV is much smaller than are generic SM masses, which are more like 100 GeV). If the electron mass is set to zero then the Standard Model acquires a new 'accidental' symmetry, under which its left- and right-handed parts rotate differently – *i.e.* invariant under a chiral rotation $\delta\chi = i\gamma_5\chi$ of the electron field. This is why integrating out heavy degrees of freedom with mass $M \gg m_e$ only corrects the electron mass by $\delta m_e \propto m_e$ (as opposed to $\delta m_e \propto M$).

If a lagrangian has the property that it has more symmetry when a parameter is set to zero then corrections to that parameter tend to preserve its small size automatically. A parameter of this type is called *'t Hooft natural* [33]. If a small parameter is 't Hooft natural in this way it is also technically natural because the symmetry protects its corrections.[27]

Since symmetries can help understand why small scalar masses can be technically natural, we next list the symmetries that can do so. The first observation is that finding such a symmetry is harder for scalars than it is for fermions (like the electron example above). It is possible to have any nonzero value for a fermion be 't Hooft natural because fermion kinetic terms have a larger symmetry group than do fermion mass terms. That is, whereas the electron mass term $m_e\overline{\chi}\chi$ is invariant under a $U(1)$ transformation $\delta\chi = i\omega\chi$ for arbitrary ω the electron kinetic term $\overline{\chi}\partial\!\!\!/\chi$ is invariant under a $U(1) \times U(1)$ symmetry $\delta\chi = i\omega\chi + i\tilde{\omega}\gamma_5\chi$ where both ω and $\tilde{\omega}$ are arbitrary. It is this extra $\tilde{\omega}$ symmetry that protects a small electron mass. For N real scalars, ϕ^i, however, the kinetic term $\frac{1}{2}\partial_\mu\phi^T\partial^\mu\phi$ is invariant under arbitrary orthogonal $O(N)$ rotations amongst the scalars. But this is also the symmetry of a mass term $\frac{1}{2}m^2\phi^T\phi$ for any nonzero m^2. The kinetic term's $O(N)$ symmetry can force different scalars

[27]Although 't Hooft naturalness is sufficient for technical naturalness there are examples in supersymmetric theories that show that it is not strictly necessary. Supersymmetric theories (more about which below) have nonrenormalization theorems that can forbid quantum corrections in some circumstances even if the theory does not have a symmetry that makes it 't Hooft natural.

to have the same mass, but does not force their masses to be zero.

3.3.1 Shift symmetries

An example of a symmetry that *can* require a scalar mass to vanish is a 'shift' symmetry:

$$\phi \to \phi + \omega \tag{3.9}$$

where ω is an arbitrary constant. Although this transformation is a symmetry of the kinetic term $\frac{1}{2}(\partial\phi)^2$ the only scalar potential that is invariant under (3.9) is a constant $V = V_0$ (independent of ϕ). Scalars with a symmetry of this type are Goldstone bosons and their presence flags the existence of a spontaneously broken symmetry[28] (*i.e.* a symmetry of the action but not the ground state).

Evidently any scalar with this symmetry must be massless and cannot have any zero-derivative interactions at all (a special case of Goldstone's theorem). This does not mean such scalars do not interact at all, however. For instance they can couple to other fields ψ through derivative interactions like $\partial_\mu\phi\, J^\mu(\psi)$.

If there are multiple dimensionless scalars, θ^i, then they can also interact amongst themselves (even at the two derivative level) if (3.9) is generalized to

$$\delta\theta^i = \omega^\alpha \xi^i_\alpha(\theta)\,, \tag{3.10}$$

and there is no value θ^i_0 for which all of the $\delta\theta^i$'s vanish, then although it is still true that the only invariant scalar potential must be a constant, the two-derivative σ-model interactions of the form $G_{ij}(\theta)\,\partial\theta^i\,\partial\theta^j$ can be invariant if for each α the following equation is satisfied

$$D_i\xi_{\alpha j} + D_j\xi_{\alpha i} = 0 \qquad \text{where} \qquad \xi_{\alpha i} := G_{ij}\,\xi^j_\alpha \tag{3.11}$$

and $D_i\xi_j = \partial_i\xi_j - \Gamma^k_{ij}\xi_k$ is the covariant derivative built using the Christoffel symbol Γ^k_{ij} for the target-space metric G_{ij}. Any solution ξ^i_α to (3.11) is called a Killing vector field for the metric G_{ij} and corresponds to a symmetry direction of the metric G_{ij}. Although not all metrics have such symmetries there is a broad class of nonflat metrics that do: the metrics on coset spaces of the quotient of two Lie groups: G/H (such as spheres). These describe the interactions of Goldstone bosons for systems where the action has a symmetry group G but the ground state is only invariant under a subgroup H. Because these metrics are not flat they contain nontrivial two-derivative interactions, consistent with the symmetry (3.10).

[28]Invariance under shifts is the smoking gun for spontaneous symmetry breaking because it is impossible for any particular classical background, $\phi = \phi_0$, to be invariant under (3.9).

These shift symmetries are overkill if our goal is only to have a technically natural scalar mass that is small but not exactly zero. The way to protect small nonzero masses is to have the symmetry (3.9) or (3.10) only be an *approximate* symmetry of the action. Scalars transforming as (3.10) under an approximate symmetry are called pseudo-Goldstone bosons. They can be systematically light when the symmetry breaking in the action is small because they must become honest-to-God massless Goldstone bosons in the limit that the symmetry breaking terms go away.

The low-energy lagrangian for pseudo-Goldstone bosons has the form

$$\mathcal{L}_{pGB} = -\sqrt{-g}\Big\{V_0 + \epsilon V_1(\theta) + f^2\Big[G_{ij}(\theta) + \epsilon H_{ij}(\theta)\Big]\partial^\mu\theta^i\,\partial_\mu\theta^j + \cdots\Big\}\,, \tag{3.12}$$

where ϵ represents a small symmetry-breaking parameter, V_0 is a constant (the only potential invariant under the symmetry) and G_{ij} is an invariant target-space metric, but V_1 and H_{ij} are *not* restricted to be invariant (and so are why the symmetry is only approximate). The symmetry-breaking terms involving V_1 and H_{ij} are 't Hooft natural because any corrections to them must be proportional to the small symmetry breaking parameter ϵ.

Notice in particular that if $V_1(\theta) = m^4 U(\theta)$ for some dimensionless function $U(\theta)$ then all of the field-dependence in the scalar potential has the form $V(\theta) = v^4 U(\theta)$ assumed earlier when power-counting, with $v^4 = \epsilon m^4 \ll m^4$. Small symmetry breaking ϵ for a group of pseudo-Goldstone boson can provide a technically natural explanation for why v^4 can be systematically small compared with the other larger mass scales in the problem.

Notice also that the field-independent term V_0 is always allowed by the symmetry and so is not similarly suppressed by ϵ. So although shift symmetries provide a good way to make small scalar masses technically natural, it does not do the same for the vacuum energy V_0.

3.3.2 Supersymmetry

Supersymmetry is a symmetry that relates bosons to fermions, which (when not spontaneously broken) to be present requires a theory to have equal numbers of bosonic and fermionic degrees of freedom [36]. For example, in 4 dimensions a left-handed Weyl fermion χ_L (which describes the two fermionic spin states of a spin-half particle) can transform into a complex scalar Φ (which describes the two states of two spinless bosons) with a transformation of the schematic form

$$\delta\Phi = \overline{\varepsilon}\chi_L \qquad \text{and} \qquad \delta\chi_L = \gamma_L\gamma^\mu\varepsilon\,\partial_\mu\Phi \tag{3.13}$$

where ε – the symmetry parameter – is a fermionic spinor (rather than a bosonic scalar) that has dimension $(\text{length})^{1/2}$.

When not spontaneously broken the bosons and fermions related in this way have precisely equal masses and couplings. For example, a lagrangian describing the supersymmetric interactions of Φ and χ on flat space has the form

$$\mathcal{L}_{susy} = -\tfrac{1}{2}\overline{\chi}\partial\!\!\!/\chi - (\partial_\mu \Phi)^*(\partial^\mu \Phi) - \tfrac{1}{2}\left[\frac{\partial^2 W}{\partial \Phi^2}\left(\overline{\chi}\gamma_L \chi\right) + \text{c.c.}\right] - \left|\frac{\partial W}{\partial \Phi}\right|^2 , \tag{3.14}$$

where $W(\Phi)$ is an arbitrary holomorphic function of Φ (and not Φ^*) and $\gamma_L = \frac{1}{2}(1+\gamma_5)$ projects onto left-handed states. This choice $W = \frac{1}{2}m\Phi^2 + \frac{1}{6}g\,\Phi^3$ gives renormalizable interactions and gives a theory where both bosons and fermions have mass m and both the scalar-spinor Yukawa interactions and the cubic and quartic interaction terms in the scalar potential are controlled by the single coupling parameter g.

The reason this matters for a discussion of naturalness is this: fermions and bosons contribute to corrections to the lagrangian with opposite signs. For instance, if the vacuum energy obtained by integrating out χ is $\delta\rho_{\text{vac}\,\chi} = Cm^4$ for some function $C(g)$ then the vacuum energy obtained by integrating out the complex scalar Φ is $\delta\rho_{\text{vac}\,\Phi} = -Cm^4$ if Φ and χ have the equal masses and couplings dictated by (3.14). This makes their contributions to ρ_{vac} completely cancel. Integrating out a heavy pair (or supermultiplet) of supersymmetric particles similarly cancels out in the contributions to the mass terms of scalar fields in other light supermultiplets.

The supersymmetric dark

This is all very nice, but we know that if the world is supersymmetric it must be spontaneously broken because none of the known elementary particles (like the electron) has a scalar partner with precisely equal mass and coupling. Spontaneous breaking of supersymmetry can occur and when it does the bosons and fermions within a supermultiplet acquire different masses. But once the masses in a supermultiplet differ the cancellation in ρ_{vac} or in low-energy scalar masses no longer cancels.

This needn't stop supersymmetry from being part of the explanation of why the electroweak hierarchy: why can the scale of the Higgs mass, $m_H \sim 100$ GeV (and so also the mass of all Standard Model particles) be so small compared with $M_p \sim 10^{18}$ GeV? Supersymmetry can help provided the mass differences within supermultiplets are not too much larger than m_H, but it is less useful if the mass splittings are much larger than this. This makes it an attractive proposal because the assertion that supersymmetry helps understand the electroweak hierarchy comes with testable predictions: the super-partners required to enforce the cancellations cannot be too far out of reach of current accelerators. (Sadly these predictions have not described well what was actually seen in experiments to date, where there is no evidence for super-partners for ordinary particles.)

At face value it seems less useful as a proposal for understanding the small size, $v_{\rm eff} \sim 10^{-2}$ eV of the cosmological constant, since this is much smaller than the mass splittings that can exist between ordinary particles and their hypothetical super-partners. As we shall see, although this is true at face value it need not imply that supersymmetry has no role to play in the final story. In particular, although we do know that the particles we produce at colliders are not supersymmetric little is known about whether or not the gravitationally coupled dark sector is supersymmetric.

Indeed there are good reasons to believe that any low-energy gravitationally coupled dark sector arising in a fundamental theory with supersymmetry at very high energies (such as string theory) could well be much more supersymmetric than are the particles of everyday experience (see for instance [37]). This is because in supersymmetric theories the splitting of masses within any particular supermultiplet is given by an expression of the form

$$\Delta m^2 \sim g\, \mathcal{F}\,, \tag{3.15}$$

where $\mathcal{F}$ is the expectation of the field that breaks supersymmetry spontaneously and g is the coupling of that field to the supermultiplet whose mass splitting is of interest. A supermultiplet whose couplings are all gravitational in strength is usually among the most weakly coupled supermultiplets in the theory, so it is not uncommon for their masses to be split by much less than other more strongly interacting sectors (like those containing the ordinary particles we see around us).

Suppose, for example, the supersymmetry breaking mass scale M_s is set by $\mathcal{F} = M_s^2$ and that Standard Model particles couple to this with a strength $g_{SM} \sim \alpha \sim 0.01$ not unusual for ordinary particles but the Dark sector couples only with gravitational strength: $g_D \sim M_s/M_p$. Then observations require that ordinary particles must be split from their superpartners by at least 10 TeV or so. The relation $\Delta m^2_{SM} \sim g_{SM}\mathcal{F} \sim \alpha M_s^2$ then implies $M_s \gtrsim 100$ TeV. But for $M_s \sim 100$ TeV masses within a gravitationally coupled dark sector would be split by $\Delta m_D^2 \sim M_s^2/M_p \sim 0.1$ eV, not so different than the scale of Dark Energy density. In such a world the Dark sector could just include the graviton and gravitino, but it might equally well include a variety of other supermultiplets coupled to one another in an approximately supersymmetric way, and this need not contradict experience with colliders. Supersymmetric Large Extra Dimensional (SLED) scenarios [38, 39] provide concrete extra-dimensional realizations of this wherein ordinary particles are localized on a non-supersymmetric brane embedded in an otherwise supersymmetric bulk. (See [22] for discussions of this scenario in a previous iteration of this school.)

3.3.3 Classical scaling

There is a closely related type of transformation that can also (in some circumstances) suppress V_0 as well as scalar masses (for early attempts to exploit this see [34]). A simple example of a transformations that can do so is obtained when the transformation (3.9) is also accompanied by a rescaling of the metric:

$$\sigma \to \sigma + \omega \qquad \text{and} \qquad g_{\mu\nu} \to e^{\omega} g_{\mu\nu} \,, \tag{3.16}$$

for constant parameter ω. There is no loss of generality in choosing e^{ω} rather than $e^{2\omega}$ (or another power) because we can always rescale ϕ to make (3.16) true. Under this type of transformation we have $\sqrt{-g} \to e^{2\omega}\sqrt{-g}$ and $R^{\mu}{}_{\nu\lambda\rho} \to R^{\mu}{}_{\nu\lambda\rho}$ and so $R = g^{\mu\nu} R_{\mu\nu} \to e^{-\omega} R$ and so in particular the Einstein-Hilbert action scales as $S_{EH} \to e^{\omega} S_{EH}$.

Since the Einstein-Hilbert action is not invariant this type of transformation is not a *bona fide* symmetry in the usual sense. But if $S[\sigma + \omega, e^{\omega} g_{\mu\nu}] \to e^{c\,\omega} S[\sigma, g_{\mu\nu}]$ for some constant c and arbitrary constant ω then this transformation takes a stationary point of S to another stationary point of S and so is a symmetry of the equations of motion. This can be good enough inasmuch as the transformation becomes an approximate symmetry, at least within the semiclassical expansion. This makes the field σ a pseudo-Goldstone boson for an approximate scaling symmetry, which is why it is called the *dilaton*.

The kinetic energy $\sqrt{-g}\, g^{\mu\nu} \partial_\mu \sigma \, \partial_\nu \sigma$ scales in precisely the same way as does the Einstein-Hilbert action, and the same is true for the sigma-model interaction $\sqrt{-g}\, g^{\mu\nu} G_{ij}(\theta)\, \partial_\mu \theta^i \, \partial_\nu \theta^j$ provided the scalars appearing in G_{ij} do not also transform, *i.e.* $\theta^i \to \theta^i$, as σ and $g_{\mu\nu}$ are scaled.[29] The potential energy also scales the same way provided that the potential depends on σ in a specific way:

$$V(\sigma, \theta) = e^{-\sigma} U(\theta) \,. \tag{3.17}$$

We are led to an effective action of the following form, expanded to the 2-derivative level:

$$\mathcal{L}_{\text{eff}} = -\sqrt{-g} \Big[v^4 e^{-\sigma} U(\theta) + \tfrac{1}{2} M_p^2 R + \tfrac{1}{2} Z(\theta) (\partial\sigma)^2 + \tfrac{1}{2} f^2 G_{ij}(\theta)\, \partial\theta^i \partial\theta^j + \cdots \Big] \,, \tag{3.18}$$

by the requirement $\mathcal{L}_{\text{eff}}[\sigma + \omega, \theta^i, e^{\omega}\, g_{\mu\nu}] = e^{\omega} S[\sigma, \theta^i, g_{\mu\nu}]$. Notice in particular that the scalar potential is always minimized at $V = 0$ provided only that U is non-negative. Minimization

[29] If any of the θ fields also shift under the symmetry we can always redefine $\tilde{\theta} := \theta - p\sigma$ with p chosen to ensure $\tilde{\theta}$ does not shift. So there is no loss of generality in assuming σ is the only scalar field that transforms.

can happen in one of two ways. If $U(\theta)$ is minimized at some values $\bar{\theta}^i$ for which $\overline{U} = U(\bar{\theta})$ is nonzero then the minimum occurs for $\sigma \to \infty$. If $\overline{U} = U(\bar{\theta})$ is instead zero then the minimum is really a flat direction along which σ can take any value and for which the potential vanishes.

For nonzero $\overline{U}$ the potential can be made arbitrarily small just by making σ large enough, with the effective scale for the potential being

$$v_{\rm eff} = v\, e^{-\sigma/4} \,. \tag{3.19}$$

In particular there is always a value of σ that is large enough that V is the right size to be the Dark Energy density. But because V has no minimum for finite σ this field in general rolls down the potential. For kinetic term $Z \sim M_p^2$ the evolution of σ occurs over cosmological time scales since $v_{\rm eff}^2/M_p = v^2 e^{-\sigma/2}/M_p$ is of order the Hubble scale. Whether this is acceptable depends on whether this evolution can be consistent with what we know about cosmology (more about which below). The required value for σ to do so would be very large for any particle-physics value choices for v.

Large σ can also be a good thing from another point of view. Quantum corrections will not preserve the form of the lagrangian (3.18) and so the size of these corrections is important to estimate. Their dependence on σ can be determined quite generally because it is always possible to rescale the metric from Einstein frame to a Jordan frame $\hat{g}_{\mu\nu} = e^{-\sigma}\, g_{\mu\nu}$, defined so $\hat{g}_{\mu\nu}$ does not transform under the transformation (3.16). Once this is done the only place where σ appears undifferentiated in (3.18) is as an overall factor $\mathcal{L}_{\rm eff}[\sigma, g_{\mu\nu}] = e^{\sigma}\mathcal{L}_{\rm eff}[\partial\sigma, \hat{g}_{\mu\nu}]$, as is indeed required to ensure $\mathcal{L}_{\rm eff} \to e^{\omega}\mathcal{L}_{\rm eff}$ in these variables. But this means that $e^{-\sigma}$ appears in the path integral integrand $e^{iS/\hbar}$ in the same way as does $\hbar$, and so repeating the power-counting arguments of previous sections shows each loop comes with a factor of $e^{-\sigma}$.

The upshot is that loop corrections to the action come as a series of the form

$$\mathcal{L}_{\rm eff} = e^{\sigma}\mathcal{L}_{\rm tree}(\hat{g}_{\mu\nu}) + \mathcal{L}_{1-{\rm loop}}(\hat{g}_{\mu\nu}) + e^{-\sigma}\mathcal{L}_{2-{\rm loop}}(\hat{g}_{\mu\nu}) + \cdots \,, \tag{3.20}$$

where $\mathcal{L}_{\rm tree}$ is given by all terms in (3.18) plus any others with higher derivatives that scale like $\mathcal{L}_{\rm tree} \to e^{\omega}\mathcal{L}_{\rm tree}$.. Each of the $\mathcal{L}$'s here is a function only of $\partial_\mu\sigma$, θ^i, $\hat{g}_{\mu\nu}$ and other scale-invariant combinations of fields, so the L-loop contribution transforms as $\mathcal{L}_{L-{\rm loop}} \to e^{(1-L)\omega}\mathcal{L}_{L-{\rm loop}}$ under (3.16). Although the one-loop term seems to re-introduce a σ-independent potential – seemingly again allowing constant contributions to the potential, like V_0 – this is an illusion because (3.20) is written in terms of the Jordan-frame metric $\hat{g}_{\mu\nu}$ (which does not satisfy the usual Einstein equations). In terms of the Einstein-frame metric (which does) any potential appearing in the one-loop term becomes

$$-\sqrt{-\hat{g}}\; U_{1-{\rm loop}}(\theta) = -\sqrt{-g}\; e^{-2\sigma}U_{1-{\rm loop}}(\theta) \,, \tag{3.21}$$

as required for the one-loop term to be invariant under (3.16).

The widely read reader might notice that having a field play the role of a loop-counting parameter, as in (3.20), is reminiscent of the role played by the 10-dimensional dilaton field in string theory. This is not an accident. A strong motivation for exploring classical scaling symmetries like (3.16) is precisely that they are generic to all known string vacuua [35]. They arise ubiquitously there because in string theory there are no parameters, only fields. So any expansion – be it weak coupling or low energy – is always an expansion in powers of fields like in (3.20). Each term in such an expansion (and in particular the first term) by construction scales in a specific way under appropriate rescalings of the fields.

We return below to whether these quantum corrections can be acceptably small, but it is encouraging that they are smallest in the regime of most interest: large σ (which makes the scalar potential small)

3.4 Dual is different

yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada yada

3.5 Desperate measures

The program for finding a technically natural understanding for why the cosmological constant is so small has so far not borne convincing fruit. It seems very hard to reconcile the known particle content in the Standard Model with having a technically natural vacuum energy as small as the observed Dark Energy density. This has caused some to doubt the utility of the criterion of technical naturalness altogether. This section describes the three options on offer by those who do so, leaning heavily (for two of them) on the review [22].

Head in the Sand

The most common attitude about the cosmological constant problem is pragmatic despair. Since there are no good theories on offer, one works on other areas of physics forlornly hoping that whatever solves the cosmological constant problem is not going to be important to this other physics.

This is an easier point of view to adopt the further one's field is from cosmology, since in cosmology a commitment must be made as to whether the dark energy clusters or evolves with time.

Anthropic arguments

A more sophisticated point of view interprets the absence of a compelling solution to the cosmological constant problem as evidence that quantum corrections to the vacuum energy need not be small after all [20, 80]. That is, one denies that both of the questions given in §3.2 must be answered for the cosmological constant, and simply accepts that there is a very precise cancellation that occurs between the renormalized cosmological constant and the quantum contributions to $\rho_{\rm vac}$. As emphasized earlier, this is a logically consistent point of view, though it is radical in the sense that would be the first example at easily accessible energies where this occurs for a parameter in the Wilson action.[30]

There is a better face that can be put on this cancellation if the microscopic theory has three features. First, the microscopic theory could have an enormous number of candidate vacua, with the effective cosmological constant differing from vacuum to vacuum. (This is actually likely to be true of a UV complete theory of quantum gravity if string theory is any guide.) Second, the microscopic theory might have a reason to have sampled many of these vacua somewhere in space at some time over the history of the universe. (This is also not far-fetched in theories that allow long periods of cosmic inflation within a complicated potential energy landscape, such as seems likely for string theory.) Third, it might be true that observers can only exist in those parts of the universe for which the vacuum energy has a very small range, not much different from the observed dark energy density.

With these conditions in place one might expect the universe to be populated with an enormous number of independent regions, in each of which a particular vacuum (and cosmological constant) is selected. The vast majority of these vacua do not have observers within them whose story needs telling, but those that do can only have a small cosmological constant since this is required for the observers to exist in the first place. Since we live in such a world we should not be surprised to find evidence for dark energy in the range observed.

Although this may well be how things work, most (though not all) of its proponents would prefer to have a technically natural solution to the problem (satisfying the two criteria of §3.2) if only this were to exist. There are two dangers to adopting this kind of anthropic

[30] There are examples of coincidences of scale that do not require a fundamental explanation, such as the apparent sizes of the Sun and the Moon as seen from the Earth. However I do not know of any examples of this type that involve the smallness of parameters in a Wilson action.

approach. One is that it becomes a dogma that stops people searching for a more traditional solution to the problem. Another is that it is difficult to know how to falsify it, and what the precise rules are that one should use when making predictions. (Of course this is partly the point: it is not clear how one makes predictions more generally in theories having an enormous landscape of possible vacua, and it is important that this gets thought through to see if a sensible formulation can be found.)

My own view on this is to accept that there is an important issue to be resolved to do with making predictions in theories (like string theory) that have a complicated landscape. But (to my understanding) so far no unambiguous framework for making predictions and deciding which parameters must be understood anthropically has been found, so it is hard to assess how useful the new anthropic framework really is.

In practice the problem right now is not that we know of too many acceptable vacua of UV complete theories. The real issue is it is hard to find any good vacua at all given the large number that must be sorted through. Once we have two examples that include the Standard Model and everything else we find around us (and nothing else) we can start worrying about their statistics. One thing that might help in this search is to have 'modules' that build in features we know to be true of the world around us. These modules include the Standard Model particle content and symmetries, some candidate for dark matter, and hopefully could include a technically natural description of dark energy if this could be found.

Swampy *vs* **Solid ground**

The Swampland program [81] provides a much more recent form of naturalness denial. In essence this program asserts that there exist otherwise reasonable effective field theories for which no UV completion including gravity exists.[31] Effective theories for which UV completions cannot be found within the landscape of possible vacua are said instead to lie in the swampland. If this picture were correct there would be a great premium on knowing which EFTs are not in the swampland because only those would be embeddable into a sensible theory of all scales.

There is even some evidence that a swampland like this might exist, if we assume (as people do in practice) that the UV completion is a string theory. For instance there are good arguments that perturbative string theories cannot contain any global symmetries [82, 83] and if so then any EFT with an exact global symmetry must be in the swampland.

[31]Since nobody knows what the real UV completion is for gravity in practice this assertion is taken to mean that the EFT cannot be obtained as the low-energy limit of some sort of string vacuum.

But this example also exposes a real difficulty in actually using this observation: when using EFTs one only ever works to some finite order in $1/M$ and it is easy to arrange a global approximate symmetry that only appears to be exact at some fixed order in $1/M$. The Standard Model is the poster child for this: if the low-energy world consists only of Standard Model fields then the most general possible interactions allowed at zeroeth order in $1/M$ is the Standard Model itself. But the Standard Model famously has several accidental symmetries – like baryon number and lepton number – that are automatic conseqences of renormalizability and so are broken once nonrenormalizable interactions at nonzero order in $1/M$ are included.

At low energies it is in practice incredibly difficult to tell the difference between an exact global symmetry and a 'fake' accidental approximate global symmetry [83]. A similar observation seems also to apply to the other lines of reasoning that support the existence of the swampland: the more sure we are that a low energy property is really required by a UV completion the easier it seems to be to fake at a fixed order in $1/M$ and so the less useful it is in constraining our options when describing the low-energy world. The difficulty in finding a criterion for the swampland that is both reliable and useful has been called the *Principle of Swamplementarity*.

Another difficulty is that nobody really knows everything that is possible within string theory. This drives people instead to propose conjectures about what is possible and what is not, and then to see if these conjectures are informative. One such a conjecture is that de Sitter solutions should only be possible in EFTs that lie in the swampland [84]. This in turn has led to a preference for quintessence like models and (more recently) to a resurgence of interest in large extra dimensions [85] when trying to describe the Dark Energy density.

In my opinion a key challenge for these models is their awkward relationship to decoupling and the general utility of EFT methods at low energies. To the extent that EFTs not in the swampland obey the usual rules, it should be possible to understand the naturalness issues that come with *any* EFT description of UV physics. But a key part of the swampland program is that there is low-energy information that is *not* captured by standard EFT methods, and a full assessment of its value will require understanding when EFT rules can be dropped. With the present state of the art there seems to be no new insights on the cosmological constant problem, apart from a belief that consistency with some of the conjectured behaviour of UV physics must eventually save the day in a way that cannot yet be explicitly articulated.

4. Ways forward (naturally)

This last section closes on a more optimistic note, focussing on the main directions that I think are the most promising ways to achieve a technically natural Dark Energy. Although this is a difficult thing to achieve I do not think that enough avenues have yet been sufficiently thoroughly explored to justify despair. This section aims to explain why, and to describe the predictions these directions make that (some of which seem to be starting to bear fruit).

There are two main directions that I believe deserve further exploration, each of which is briefly described here. They differ on whether the focus is on electroweak and higher energy scales or the much lower energies relevant to cosmology. In truth the two directions are likely two sides of the same coin.

4.1 Above eV scales: Supersymmetric Extra Dimensions

Let us start with the higher energies: the electroweak scales of everyday particle physics. In this energy range extra-dimensions provide a uniquely promising approach to dynamically evading the cosmological constant problem. This section reviews why this is so and what the challenges are (leaning heavily on lectures given at earlier versions of this school [22]).

To motivate the relevance of extra dimensions for the cosmological constant problem, recall what the essence of the problem is: we believe quantum fluctuations generate a large vacuum energy density, and the vacuum's Lorentz invariance automatically gives this the $w = -1$ equation of state of a cosmological constant: $T_{\mu\nu} = -\rho_{\rm vac} g_{\mu\nu}$. But when cosmologists measure the acceleration of the universe's expansion they are essentially detecting a very small curvature for 4D spacetime. The conundrum is that these are directly equated in Einstein's equations – eqs. (1.1) – with the measured curvature much smaller than what would be expected for typical vacuum energies.

We wish to break this direct link between the vacuum energy and the curvatures measured in cosmology. Moreover, we must do so only for very slow processes (like the vacuum) and not also for fast ones (like atoms) [52]. Fast quantum fluctuations should gravitate in an unsuppressed way because we know that such fluctuations do contribute to atomic energy levels in atoms, as inferred from precision tests of the equivalence principle. The equivalence principle is tested to a part in 10^{15} or so [53, 54] and we know quantum fluctuations contribute to atomic energies by more than this (and so their absence would be missed if they did not contribute).

The extra-dimensional loophole

The good news is there is a loophole within which it is possible to break the link between vacuum energy and curvature, without doing violence to everything else we know at accessible energies. This loophole is based on the observation that Lorentz invariance plays an important role in formulating the problem, because it so severely restricts the form of the vacuum stress energy.

The situation would be different in an extra-dimensional world because then we would only know that the vacuum must be Lorentz invariant in the four dimensions that we can see. We also would only really know that the curvatures must be small in these same dimensions since these are the ones we access in cosmology. Although the vacuum stress energy must curve something, in extra-dimensional models it need not curve the dimensions we see.

The gravitational field of a cosmic string in four dimensions illustrates this loophole more concretely. Consider a string whose world-sheet sweeps out the $z - t$ plane, transverse to the x and y directions. The stress energy of a relativistic string is Lorentz invariant in the $z - t$ directions, $T_{ab} = -\mathcal{T} g_{ab}\, \delta^2(x)$, where $\mathcal{T}$ is the string's mass per unit length, and a, b denotes the $z - t$ directions parallel to the string world-sheet. The gravitational field sourced by this stress energy is known [55] and the spacetime away from the string's position is flat. More precisely, the two dimensions transverse to the string have the geometry of a cone whose apex is located at the string's position. The tension on the string gives rise to a curvature singularity, with the transverse 2D geometry having a curvature scalar $R \propto \kappa^2 \mathcal{T}\, \delta^2(x)$ that is singular at the string's position. What is important for the present purposes is that the geometry along the two Lorentz-invariant on-string directions remains perfectly flat,[32] regardless of the precise value of $\mathcal{T}$. The consistency of a large $\mathcal{T}$ with a small spacetime curvature in the $z-t$ directions might appear to be a 'cosmological constant problem' to a 2D cosmologist unable to see off the surface of the string.

This suggests trying similar examples having two more dimensions (six dimensions in total) with the 2-dimensional string world-sheet being replaced by the world-volume of a 4-dimensional Lorentz-invariant brane, situated at specific points within two compact extra dimensions. In the simplest examples the two extra dimensions have the geometry of a sphere and there is a brane located at both the sphere's north and south poles. The transverse curvature at these poles also has conical singularities, like for a cosmic string, and this gives the overall geometry more of a rugby ball (or American football) shape. All of the elementary particles we know are imagined to be confined to one of these branes, whose tension (*i.e.* vacuum energy per unit volume) is not particularly small relative to known particle-physics scales — of order (10 TeV)^4. The hope is that the geometry seen by an observer on the brane (us) can remain flat regardless of the size of the brane vacuum energy density.

[32]More precisely, it is flat if the string sits within an asymptotically flat geometry. It would not be flat if the string were sitting in a curved space like de Sitter space.

The simplest models try to do so by simply assuming away the extra-dimensional cosmological constant [56], though this simply moves the underlying cosmological constant problem into the higher-dimensional theory. There is a better chance if the extra-dimensional physics is supersymmetric [38], however, because in six (and higher) dimensions supersymmetry forbids a cosmological constant (much as would more than one supersymmetry in four dimensions). Interestingly, they do so because extra-dimensional supergravities all have scaling symmetries like (3.16) [57, 58] described above. The generic appearance of scaling symmetries in extra dimensions also turns out to have a plausible explanation in their generic presence in string theory [35].

Notice that we do *not* also require the physics on the brane to be supersymmetric, and one might simply choose only the Standard Model to live on the brane. Such a brane can nonetheless be coupled consistently to supergravity using the 'Stückelberg trick'; that is, promoting the non-supersymmetric brane to something supersymmetric, but with supersymmetry nonlinearly realized by coupling a Goldstone fermion — the Goldstino — in the appropriate way [36, 62]. It remains consistent to regard extra-dimensional fields to be supersymmetric despite them coupling to nonsupersymmetric matter on the brane because brane-bulk couplings are weak; they are gravitational in strength. From an extra-dimensional point of view the brane provides a non-supersymmetric boundary condition for bulk modes that splits bosons from fermions by the KK scale, $\Delta m \sim 1/L$ (where L is the linear size of the extra dimensions). Because the 4D and 6D Planck scales are related by[33] $M_p \sim M_6^2 L$ this shows that bulk mass splittings are Planck-suppressed: $\Delta m \sim M_6^2/M_p$.

It turns out that extra dimensions can be large enough to allow $1/L \sim$ eV without running into conflict with observations [59, 60] and so can allow bulk supersymmetry to play a role suppressing the vacuum energy right down the the Dark Energy scale [38, 39, 61]. Remarkably extra dimensions can only be this large if there are at most two of them, providing another reason for liking six dimensions.[34] In any such a framework the 6D Planck scale is not too far above the electroweak scale, so from the 6D point of view the tension on any brane involving standard model particles would not be particularly small relative to the Planck scale.

This picture leads to a novel kind of supersymmetric phenomenology [63, 64]: a very supersymmetric gravity (or extra-dimensional, bulk) sector whose supersymmetry breaking scale is of order $1/L \simeq 10^{-2}$ eV coupled to a particle (brane) sector that is not supersymmetric at all. In particular, the nonlinear realization of supersymmetry on the brane implies that

[33]The 6D gravitational coupling is $\kappa_6^2 = 8\pi G_6 = 1/M_6^2$ where G_6 is the 6D Newton constant.

[34]6D models with radii this large can also be ruled out if KK modes dominantly decay into observable particles like photons, but whether such decays dominate is a model-dependent issue.

a supersymmetry transformation of a brane particle like the electron gives the electron plus a Goldstino (or, equivalently, a gravitino) rather than a selectron. One does not expect to find a spectrum of superpartners for the Standard Model, despite the very supersymmetric gravity sector.[35]

Within this kind of picture the cosmological constant problem is a special case of the general problem of back-reaction: how does the spacetime geometry react to microscopic changes, such as to the vacuum energy. In a higher-dimensional context this requires also understanding what stabilizes the size of the extra dimensions, since this is also part of the general issue of gravitational back-reaction. To pin these issues down precisely it is useful to work within a concrete example, solving explicitly the equations of a specific higher-dimensional supergravity [38].

There are a variety of 6D supergravities from which to choose when formulating such an example, but a particularly convenient choice uses the Nishino-Sezgin chiral gauged supergravity [65], for which a simple stabilization mechanism for the extra dimensions has long been known [66]. This involves the following 6D bosonic fields: the metric, g_{MN}, a scalar dilaton, ϕ, and a specific $U(1)_R$ gauge potential, A_M and a Kalb-Ramond 2-form gauge field B_{MN} (with 3-form field strength H_{MNP}).

To lowest orders in the derivative expansions, the action is the sum of bulk and brane contributions, $S = S_B + \sum_b S_b$, with the supersymmetric bulk contribution being

$$S_B = -\int \mathrm{d}^6x\sqrt{-g}\;\left[\frac{1}{2\kappa_6^2}\, g^{MN}\left(\mathcal{R}_{MN} + \partial_M\phi\,\partial_N\phi\right) + \frac{2e_6^2}{\kappa_6^4}\,e^{\phi} + \frac{1}{12}\,e^{-2\phi}H_{MNP}H^{MNP} + \frac{1}{4}\,e^{-\phi}F_{MN}F^{MN}\right], \tag{4.1}$$

while the contribution of each brane is

$$S_b = -\int_{W_b} \mathrm{d}^4x\sqrt{-\gamma}\;\left(\mathcal{T}_b + \tfrac{1}{4!}\mathcal{A}_b\,\varepsilon^{\mu\nu\lambda\rho}\mathcal{F}_{\mu\nu\lambda\rho} + \cdots\right). \tag{4.2}$$

Here W_b denotes the brane's world-volume and $\varepsilon_{\mu\nu\lambda\rho}$ is the volume form built from its induced metric $\gamma_{\mu\nu}$. $\mathcal{F}_{MNPQ} = \frac{1}{2}\epsilon_{MNPQRS}\,e^{-\phi}F^{RS}$ is the 6D dual of the Maxwell field, where ϵ_{MNPQRS} is the volume form built from the 6D metric. e_6 is the gauge coupling for the field F_{MN} and the parameter $\mathcal{T}_b$ denotes the brane tension. The quantity $\mathcal{A}_b$ measures the amount of Maxwell flux that is localized on the brane [67] in a way that is made more precise below.

The simplest situation is when the two branes are identical, in which case there is a rugby ball solution to these field equations [66, 38] for which $H_{MNP} = 0$, $\phi = \phi_0$ and

$$\mathrm{d}s^2 = \hat{g}_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu + \ell^2\Big(\mathrm{d}\theta^2 + \beta^2\sin^2\theta\,\mathrm{d}\xi^2\Big)e^{-\phi_0} \quad \text{and} \quad F_{\theta\xi} = Q\beta\ell^2\sin\theta\,, \tag{4.3}$$

[35]This particular prediction was made [63] before the LHC results showed it to be a huge success.

where $\hat{g}_{\mu\nu}$ is a maximally symmetric 4D geometry with curvature scalar $\widehat{R}$ and ϕ_0, Q, β and ℓ are constants. With this ansatz the field equations boil down to

$$\frac{1}{\ell^2} = \kappa_6^2 Q^2 = \left(\frac{2e_6}{\kappa_6}\right)^2 , \quad 1 - \beta = \frac{\kappa_6^2 \mathcal{T}}{2\pi} \quad \text{and} \quad \widehat{R} = 0 . \tag{4.4}$$

Inspection of (4.3) shows that the physical radius of the extra dimensions is $L = \ell\, e^{-\phi_0/2}$, and so eqs. (4.4) imply

$$L^2 e^{\phi_0} = \ell^2 = \left(\frac{\kappa_6}{2e_6}\right)^2 , \tag{4.5}$$

is fixed in terms of parameters in the lagrangian. Several features of this solution are noteworthy:

Flat direction and scaling

The value of ϕ_0 is not determined by any of the field equations. This 'flat direction' is a consequence of a classical scale invariance of extra-dimensional supergravity, along the lines described in §3.3.3. The scaling symmetry in this case applies to the 6D lagrangian, with

$$g_{MN} \to \zeta\, g_{MN} \quad \text{and} \quad e^{-\phi} \to \zeta\, e^{-\phi} \quad \text{with} \quad H_{MNP} \text{ and } F_{MN} \text{ held fixed.} \tag{4.6}$$

(This is why ϕ is called the 6D dilaton.) Under this $S_B \to \zeta^2 S_B$ and S_b scales the same way, but only if both $\mathcal{T}_b$ and $\mathcal{A}_b$ are assumed to be ϕ-independent.

Localized flux and flux quantization

As mentioned above, the $\mathcal{A}_b$ term changes the Maxwell equation to become

$$\partial_m \Big[e^{-\phi} \Big(\sqrt{g_2}\, F^{mn} - \sum_b \mathcal{A}_b\, \epsilon^{mn}\, \delta^2(x - x_b) \Big) \Big] = 0 , \tag{4.7}$$

where ϵ_{mn} is the volume form for the extra-dimensional 2D geometry. This introduces localized flux into the solution at the position of each brane, and changes the flux quantization condition into

$$\int_{M_2} F_{mn} + \sum_b \mathcal{A}_b \frac{\epsilon_{mn}}{\sqrt{g_2}} = \frac{n}{e_6} , \tag{4.8}$$

where n is an integer. Notice that this condition does not break scale invariance (because F_{MN} doesn't scale) *provided* that $\mathcal{A}_b$ is independent of ϕ.

For the solution given in (4.3) and (4.4) the integer must be $n = \pm 1$ and the flux localized on the branes is

$$\Phi_{\text{tot}} := \sum_b \mathcal{A}_b = \pm \frac{1 - \beta}{e_6} . \tag{4.9}$$

Notice also that this equation would not have any solutions at all if the $\mathcal{A}_b$'s were all assumed to be zero, as was often done in early studies of this system [69].

In the scale-invariant case – when $\mathcal{A}_b$ and $\mathcal{T}_b$ are independent of ϕ – eq. (4.9) imposes a condition that relates the $\mathcal{A}_b$ and $\mathcal{T}_b$ ($\mathcal{T}_b$ enters through β using (4.4)) in order for rugby ball solutions to exist. If $\mathcal{A}_b$ were to depend on ϕ then scale invariance breaks and the flux quantization condition can be used to determine the value of ϕ_0 given arbitrary choices for the $\mathcal{T}_b$ or $\mathcal{A}_b$.

4D flatness and extra-dimensional relaxation

Most remarkably, the brane action is flat ($\hat{R} = 0$) for *any* choice of brane lagrangian and in particular regardless of the value of $\mathcal{T}_b$. For later purposes it is useful to see in more detail how the extra-dimensional solutions relax to achieve flat 4D geometries: $\hat{R} = 0$.

The simplest way to see what happens is first to ask why the curvature is flat in the solution in the absence of branes (*i.e.* with $\mathcal{T} = 0$ and so $\beta = 1$) [66]. The 4D scalar potential for the fields ϕ_0 and L is obtained by evaluating the action using the 2D scalar curvature $R = -2/L^2$ and the Maxwell field subject to the flux quantization condition, which implies $\int e_6 F = n$ for some integer n, and so $F_{mn}F^{mn} \propto n^2/L^4$. Combining the Einstein and Maxwell actions with the scalar potential[36] then gives the scalar potential for the fields ℓ and ϕ. In the case of unit flux, $n = \pm 1$, this turns out to be a perfect square:

$$
\begin{aligned}
V(L, \phi_0) &= \int \mathrm{d}^2x \sqrt{-g} \left(\frac{1}{2\kappa_6^2} R + \frac{1}{4} e^{-\phi_0} F_{mn}F^{mn} + \frac{2e_6^2}{\kappa_6^4} e^{\phi_0} \right) \\
&\propto \frac{e^{\phi_0}}{L^2} \left(1 - \frac{\kappa_6^2}{4e_6^2 L^2 e^{\phi_0}} \right)^2 ,
\end{aligned} \tag{4.10}
$$

which is minimized at a fixed value of $L^2 e^{\phi_0} = \ell^2$ for which there is a flat direction along which $V = 0$ and $e^{\phi_0}/L^2 = e^{2\phi_0}/\ell^2$ is not determined.

Adding branes to this solution changes the above in two ways. First the action now includes the brane tensions coming from S_b. Second, the brane's gravitational field introduces a conical singularity to the 2D curvature, $\sqrt{g_2}\; R_{\rm sing} = -2\kappa_6^2 \sum_b \mathcal{T}_b \; \delta^2(x - x_b)$ localized at the brane positions, where $\mathcal{T}_b$ is the brane tension. Using the curvature singularity in the Einstein action (and using the delta function to perform the extra-dimensional integral d^2x) then gives a contribution to the action of the form $(1/2\kappa_6^2) \int \mathrm{d}^2x \sqrt{g_2}\; R = -\sum_b \mathcal{T}_b$, which precisely cancels the direct contribution of the brane tensions themselves. The lesson from this story is that back-reaction is crucial: this cancellation can never be seen working purely within a 'probe' approximation where the brane does not perturb its environment.

[36] After first transforming to the 4D Einstein frame: $g_{\mu\nu} \to (1/L^2) g_{\mu\nu}$.

More general classical solutions

The existence of extra-dimensional solutions that allow flat 4D geometries to coexist with large 4D-lorentz-invariant energy densities does not in itself solve the cosmological constant problem. One must re-ask the cosmological constant question in the 6D context: first identify which features of the branes are required for flat brane geometries, and then ask whether *these* choices are stable against integrating out high-energy degrees of freedom.

At the classical level many more general explicit solutions to these field equations are known [57, 68], such as when the tensions on the two branes are not equal, and although the extra-dimensional geometry for these new solutions generically becomes warped the 4D brane geometries remain exactly flat. Nonflat solutions can also be constructed, for some of which the 4D geometry is de Sitter[37] rather than flat [70]. The nonflat solutions are obtained by allowing $\mathcal{T}_b$ and/or $\mathcal{A}_b$ to depend nontrivially on ϕ.

There is a very general argument why 4D curvature requires nontrivial dependence of $\mathcal{T}_b$ and $\mathcal{A}_b$ on ϕ, since it can be proven on very general grounds that any solution to the field equations for which the near-brane limit of $r\partial_r\phi$ vanishes for all branes as $r \to 0$ (where r is the proper distance from the brane) must have a flat 4D geometry. This is proven by exploiting the scale invariance shared by essentially all higher-dimensional supergravities [58]. But the near-brane limit of $r\partial_r\phi$ is on very general grounds proportional to $\delta S_b/\delta\phi$, where S_b is the brane action [73]. They are related for much the same reason as the charge of a point source in electromagnetism can either be determined by differentiating the action of the point source with respect to the electrostatic potential, or by evaluating (in 2D) $r\partial_r$ of the Coulomb potential itself near the source.

The upshot is this: for a general solution to the field equations for the action (4.1) and (4.2) with maximally symmetric geometries in 4D a sufficient condition for the 4D geometry to be flat is to have none of the branes couple to the 6D dilaton. In order for quantum corrections to generate a 4D curvature they must also generate dilaton couplings to the branes.

Flux quantization vs tuning

It is sometimes argued that in the case of scale invariant (dilaton-independent) branes the flux quantization condition (4.8) itself represents a fine-tuning that is ruined by quantum corrections. This is most often argued in simpler 5D models [74], where similar issues arise and for which back-reaction can also be computed explicitly. In this case closer inspection

[37] de Sitter solutions to these equations are interesting in their own right as a counter-example [71] to no-go theorems for the existence of de Sitter solutions in supergravity [72].

[75] showed that flat solution arise due to a cancelation with branes whose presence was not explicit but required to interpret singularities that were necessary on topological grounds.

A similar argument in 6D expresses the extra-dimensional Euler number as the sum of brane tensions plus an integral over extra dimensional curvature. For the rugby-ball geometries of interest here (with the topology of a sphere) this is equivalent to the relation between defect angle and tension given in (4.4). The situation in 5D is more similar to toroidal compactifications in 6D, for which the Euler number vanishes and so a topological condition states that the sum of brane tensions must vanish.

Nonetheless, topological conditions are not in themselves ever an obstruction to technical naturalness (even for tori). If a tension is changed in a toroidal compactification, the extra dimensions simply curve to satisfy the topological constraint [39]. Continuous changes cannot violate the topological condition once this is initially satisfied. For technical naturalness the real issue is to check whether the choices required for flat 4D geometries are stable against integrating out short-wavelength modes.

Robustness to quantum corrections

Considerable effort has been invested into integrating out high-energy modes in these geometries, both from loops of high-energy bulk fields [77] and from high-energy brane degrees of freedom [76].

The good news is that if a brane is initially chosen to have no dilaton coupling (in the 6D Einstein frame) then no loop purely involving only brane degrees of freedom can generate a dilaton coupling. This in particular means that if Standard Model particles are localized on such a brane then the curvature of the 4D geometry is completely stable against graphs involving only Standard Model fields (which is usually the hard part in the cosmological constant problem).

The dangerous loops are those involving the bulk fields – the extra-dimensional graviton and its friends – because these *must* both couple to the dilaton (because of supersymmetry) and couple to the brane. But for these loops supersymmetry is important in suppressing the size of quantum effects, leading to their general suppression [77]. The interested reader is referred to the review [22].

4.2 Below eV scales: Scaling the Supersymmetric Dark

We here instead ask these questions in the low-energy 4D theory that is directly relevant to cosmology that is valid at energies much lower than the KK scale. Rather than explicitly integrating out the higher-dimensional fields we instead work with a general 4D theory that includes the light fields and incorporates the underlying approximate symmetries.

One of the most important of the approximate symmetries is the scale invariance corresponding to shifts of the 6D dilaton, which descends into the effective 4D theory as an approximate scale invariance of the type outlined in §3.3.3 above. This symmetry dictates how the low-energy 4D dilaton field, σ – corresponding to the modulus ϕ_0 of the 6D solution – appears undifferentiated in the low-energy action. More precisely, given the 6D convention that weak coupling corresponds to small e^{ϕ} and the 4D convention of §3.3.3 that chooses $e^{-\sigma}$ to be small we abusively define $\sigma = -\phi_0$. With this convention the transformation law

$$e^{\sigma} \to \zeta \, e^{\sigma} \tag{4.11}$$

is inherited from the 6D transformation rule (4.6).

Demanding the 4D Einstein-Hilbert term $\mathcal{L}_{EH} = -\frac{1}{2} M_p^2 \sqrt{-g} \; g^{\mu\nu} R_{\mu\nu}$ scale as $\mathcal{L}_{EH} \to \zeta^2 \mathcal{L}_{EH}$ (*i.e.* the same way as does the 6D action does under (4.6)) implies the 4D Einstein-frame metric scales as[38]

$$g_{\mu\nu} \to \zeta^2 g_{\mu\nu} \, . \tag{4.12}$$

The kinetic energy of any other bulk fields – such as for σ itself or KK modes and any of their superpartners – has the same scaling and so appear in the low-energy EFT without any prefactors of e^{σ} in 4D Einstein frame.

These expressions show that the 4D and 6D Einstein-frame metrics are related by $g_{\mu\nu} = e^{\sigma} \, \tilde{g}_{\mu\nu}$, where the scaling relation $\tilde{g}_{MN} \to \zeta \, \tilde{g}_{MN}$ for the 6D metric $\tilde{g}_{\mu\nu}$ is as in (4.6). Comparing this to the direct dimensional reduction from 6D to 4D shows how

$$e^{\sigma} \sim (M_6 L)^2 \tag{4.13}$$

is related to the extra-dimensional size, showing that large σ incorporates the physics of large extra dimensions. If L taken as large as it can possibly be (say $1/L \sim 0.1$ eV) and taking $M_6 \sim 100$ TeV) gives $M_6 L \sim 10^{15}$ and so one can see how values as large $e^{\sigma} \sim (M_6 L)^2 \sim 10^{30}$ can arise.

The σ-dependence of brane-localized interactions can be obtained in a similar way. In 6D the brane tension and the localized flux terms given in (4.2) both scale in the same way as does the Einstein-Hilbert term, and so once written in 4D Einstein frame these must contribute to the 4D EFT in the form

$$\mathcal{L}_{\rm pot} = -\sqrt{-g} \; e^{-2\sigma} \, U(\theta) \, , \tag{4.14}$$

[38]Since the focus now shifts to the 4D EFT from here on we denote the 4D Einstein-frame metric by $g_{\mu\nu}$ and relabel the 6D Einstein-frame metric – including its 4D components – as $\tilde{g}_{MN}$.

where θ denotes any other dimensionless scalar fields. Notice the similarity with the first few terms of the lagrangian (3.18).

In what follows it is useful to write $\{\theta^i\} = \{\vartheta^u, \psi^a\}$, to keep separate track of scalars ψ^a that live on the brane and those ϑ^u (including σ) that live in the bulk. This is useful because the kinetic energies for each type depend differently on σ. The kinetic energy of a bulk scalar is independent of σ in 4D Einstein frame for the same reason no σ's appear in the Einstein-Hilbert action in this frame. But dimensionless brane-localized scalar fields ψ^a have kinetic energies of the form $\mathcal{L}_{\rm bkin} \propto \sqrt{-\tilde{g}_4}\, \tilde{g}^{\mu\nu}\partial_\mu\psi\, \partial_\nu\psi$ and so $\mathcal{L}_{\rm bkin} \to \zeta\, \mathcal{L}_{\rm bkin}$ scales differently than other terms like the brane tension or the bulk action. In 4D Einstein frame the corresponding term in the low-energy 4D EFT must therefore have the form $\mathcal{L}_{\rm bkin} \propto \sqrt{-g}\, e^{-\sigma} g^{\mu\nu}\partial_\mu\psi\partial_\nu\psi$.

These arguments lead to a lagrangian of the form (3.18),

$$\mathcal{L}_{\rm eff} = -\sqrt{-g}\Big[M_p^4 e^{-2\sigma} U(\vartheta,\psi) + \tfrac{1}{2} M_p^2 R + \tfrac{1}{2} M_p^2 Z_{uv}(\vartheta)\, \partial\vartheta^u\, \partial\vartheta^v + \tfrac{1}{2} M_p^2 e^{-\sigma}\, G_{ab}(\vartheta,\psi)\, \partial\psi^a\partial\psi^b + \cdots\Big]\,, \tag{4.15}$$

and its corrections. It is the Planck scale that naturally sets the dimensions in all of these terms, with other low-energy scales arising as a consequence of our currently living in the large-σ regime.

Electroweak and neutrino hierarchies

It is worth asking whether a choice of σ exists that is consistent with the hierarchies we see in nature (*e.g.* explains why the electroweak scale is much smaller than M_p). If one of the ψ fields is the Standard Model Higgs then the mass predicted for it by a ψ^2 term within $U(\vartheta,\psi)$ in (4.15) defines the electroweak scale and is of order $M_{EW} \sim M_p\, e^{-\sigma/2}$. By contrast, the mass predicted for a bulk scalar by a ϑ^2 term within $U(\vartheta,\psi)$ is instead of order $m_B \sim M_p\, e^{-\sigma} \sim M_{EW}^2/M_p$. These broadly reproduce the predictions of large extra-dimension models when $e^{\sigma/2} \sim (M_6 L) \sim 10^{14}$, with $M_{EW} \sim 10$ TeV and $m_B \sim 0.1$ eV being of order the KK scale.

Having a σ-dependence to the mass term for the Standard Model Higgs boson also implies the same σ-dependence for the Higgs expectation value, and this in turn implies the same sigma-dependence appears in the masses of all other Standard Model particles since these are all linear in the Higgs vev (so their mass *ratios* remain σ-independent). Whether this is a problem again requires appealing to whether a successful cosmology can be built with these choices (though having mass ratios be field independent helps evade constraints from tests of the equivalence principle).

Interestingly enough, the only ordinary particles not to have masses proportional to M_{EW} are neutrinos, which famously [41] can acquire mass through a dimension-5 effective interaction that is quadratic in the Higgs vev. If neutrinos acquire masses in this way (or by mixing with a KK fermion in the bulk[39] [78]) they would be expected to have masses of size $m_\nu \sim M_p \, e^{-2\sigma} \sim M_{EW}^2/M_p \sim m_B \sim 0.1$ eV (up to dimensionless couplings). In this framework the same value for σ can account for both the electroweak and neutrino mass scales in a unified way. This is not entirely a surprise given that the same consistency also happens in the underlying SLED models [63].

Finally, the overall size of the scalar potential is $v_{\rm eff}^4 U$ where $v_{\rm eff} \sim M_p e^{-\sigma/2} \sim M_{EW}$, and so the scalar potential is technically natural but not particularly small. Although the lagrangian (4.15) captures the correct scaling of the brane tension it does not yet contain the extra-dimensional physics that allows the 4D solutions remain flat in the presence of a weak-scale brane tension. What remains missing in (4.15) is the extra-dimensional relaxation wherein the geometry of the extra dimensions adjust and back-react to the properties of the branes.

4.3 Natural relaxation

How does extra-dimensional relaxation get communicated to the low-energy effective 4D EFT? Can this scaling picture help make headway on how a small Dark Energy density can be technically natural?

It is here that supersymmetry (of the dark sector only, as described above) might play a role. Supersymmetry is important because it provides reasons why the potential $U_{\rm tree}$ can be a perfect square, as it is for example in the lagrangian (3.14) where $V = |\partial W/\partial \Phi|^2$, or at least approximately so. If $U_{\rm tree}$ is a perfect square then it is non-negative definite and so any places where it vanishes is necessarily a minimum. When this is true other fields in the problem will tend to seek out $U_{\rm tree} = 0$ as they minimize their energy, perhaps explaining by doing so why the Dark Energy density is so low.

This kind of approach was explored in [42], where the implications for the scaling symmetry (3.16) was explored within a framework wherein the dark sector is supersymmetric and the particle-physics sector is not. As discussed in §3.3.2, this is a fairly generic situation in supersymmetric theories given that a gravitationally coupled sector couples to everything – and so in particular to supersymmetry breaking – more weakly than other sectors [37], and

[39]Extra-dimensional neutrino mixing models rely on there being very light fermions in the bulk but rarely tell you why these should be present. Supersymmetric extra dimensions provide a robust answer: they are part of a generic Dark Sector. They are present and light because they are superpartners for the graviton [79].

is realized in extra-dimensional scenarios when a supersymmetry breaking brane is localized within an otherwise supersymmetric bulk (as in SLED models [38]). Assuming all superpartners of Standard Model particles are heavy enough to be integrated out at presently accessible energies (below 10 TeV or so) the effective theory needed to analyze this kind of situation requires coupling supergravity to non-supersymmetric matter, which is (happily) a solved problem [43].

In this framework σ is in the gravity sector which is approximately supersymmetric, and so it is partnered with another scalar field that we call an axion, $\mathfrak{a}$, because it has an independent shift symmetry of the form (3.9). These combine into a complex scalar $T = \tau + i\mathfrak{a}$ (called the axio-dilaton) that transforms in the standard way for a chiral multiplet under supersymmetry, where we define $\tau := e^{\sigma/2}$.

For afficionados the model is as usual specified by a Kähler function $K(T, T^*, X, X^*, \Psi, \Psi^*)$ and a superpotential $W(T, X, \Psi)$ where X is a nilpotent field describing the Goldstone fermion for supersymmetry breaking and Ψ generically denotes all other (*i.e.* Standard Model) fields. Approximate invariance under (3.16) is implemented consistent with supersymmetry by demanding

$$e^{K/3} = T + T^* + A(X, X^*, \Psi, \Psi^*) + \frac{B(X, X^*, \Psi, \Psi^*)}{T + T^*} + \cdots , \tag{4.16}$$

where A and B are arbitrary functions and the ellipses denote higher powers of $1/(T + T^*)$.

The kinetic energy for this pair of scalars implied by the $T + T^*$ term in (4.16) is

$$\mathcal{L}_{\rm kin} = -3M_p^2\sqrt{-g}\, g^{\mu\nu}\, \frac{\partial_\mu T^* \partial_\nu T}{(T + T^*)^2} = -\frac{3M_p^2}{4\tau^2}\sqrt{-g}\left[(\partial\tau)^2 + (\partial\mathfrak{a})^2\right] \tag{4.17}$$

up to corrections that are suppressed by additional powers of $1/\tau$. If we read off the axion decay constant as the coefficient of its kinetic term – $\mathcal{L}_{\rm kin} = -\frac{1}{2}F^2\sqrt{-g}\,(\partial\mathfrak{a})^2$ – (as is often done in the particle literature) it would be very small: $F \sim M_p/\tau \sim M_p e^{-\sigma/2} \sim 0.1$ eV. The error in doing so is the assumption that σ does not also appear in any coupling terms, like $\sqrt{-g}\,\partial_\mu\mathfrak{a}J^\mu$. However both the coupling and the kinetic term share the same metric factor $\sqrt{-g}\, g^{\mu\nu}$ and so scale the same way under metric rescalings. The σ-dependence of the physical decay constant is the *relative* power of $e^{-\sigma}$ appearing between the kinetic and interaction terms, which ultimately depends on how J_μ scales under (3.16) and in which term in (**??**) this coupling appears. If, for example, J_μ scales in the same way as does $\partial_\mu\mathfrak{a}$ then $F \sim M_p$ would be σ-independent (and large).[40]

[40]This is what actually happens in extra-dimensional UV completions in which $\mathfrak{a}$ is a KK mode of the higher-dimensional metric supermultiplet and so couples with gravitational strength [47].

The scalar potential for this class of models indeed has the form expected from (**??**), which in the Einstein frame becomes

$$V_{\rm eff} = M_p^4 \left[\frac{U_0}{\tau^2} + \frac{U_{1/2}}{\tau^3} + \frac{U_1}{\tau^4} + \cdots \right] , \tag{4.18}$$

with the additional information that

$$U_0 = |w_X|^2 \qquad \text{and} \qquad U_{1/2} \propto w_X \, , \tag{4.19}$$

where w_X is a function of the other scalars in the problem, related to $\partial W/\partial X$. The coefficients $U_{1/2}$ and U_1 are calculable in terms of W, A and B can be fairly arbitrary for the present purposes. The coefficient U_0 is a perfect square because it is basically an auxiliary field for supersymmetry (which is also the reason the potential in (3.14) is a perfect square).

Now comes the main thought: because $U_0 = |w_X|^2$ is a perfect square it wants to be minimized at zero, though the other terms $U_{1/2}$, U_1 and so on can obstruct the potential vanishing perfectly. Imagine then that the fields collectively denoted Ψ above contain one non-Standard Model field, ϕ, whose role is to seek out this minimum.[41] All we need assume is that ϕ appears in all of the U_i's and all of their derivatives are order unity. In this case if only the first term of (4.18) were present then minimizing $V_{\rm eff}$ would lead to $\phi = \phi_0$ where $U_0(\phi_0) = 0$.

But the other terms *are* present, but subdominant in our regime of interest, where $\tau \gg 1$. This means instead we find the real minimum occurs when $\phi = \overline{\phi}$ where $\overline{\phi} - \phi_0 = \mathcal{O}(1/\tau)$ and so $w_X(\overline{\phi}) \sim \mathcal{O}(1/\tau)$. Evaluated at this minimum, (4.19) implies the first three terms of (4.18) are all order $1/\tau^4$ and so the Dark Energy density is predicted to be order

$$V_{\rm min} = V_{\rm eff}(\overline{\phi}) \simeq \frac{\overline{U} M_p^4}{\tau^4} = M_p^4 \, \overline{U} \, e^{-2\sigma} \, . \tag{4.20}$$

This is actually a very interesting size when written in terms of the electroweak scale, which the size of σ was chosen to explain relative the Planck mass: $m_{EW} \sim M_p \, e^{-\sigma/4}$. In terms of this the above potential minimum is

$$V_{\rm min} \sim \left(\frac{M_{EW}^2}{M_p} \right)^4 , \tag{4.21}$$

which is in the right ballpark to describe Dark Energy given that $M_{EW}^2/M_p \sim 0.1$ eV.

Suggestive as this is, there are a great many things that must go right for this to be a full solution and this remains a work in progress. Here are some of the things we know so far:

[41] This relaxation field ϕ very naturally also can play another role as the inflaton in the very early universe [44, 45], but this is another story.

- It is one thing to want a large value for a field like τ but if we can compute its potential we should also be able to compute its size. If the potential for σ is exactly as given by (4.20) then there is no minimum for any finite value of σ, so the present-day value of σ is a function of the initial conditions in cosmology (whose explanation requires a theory of the earlier epochs of cosmology such as from inflation). But it is also possible that (4.20) is only approximate and the corrections introduce a minimum for σ. In this case its late-time value can be computed by minimizing the potential.

  A simple situation that would generate one [42, 44] builds on the fact that in eq. (4.20) the function $\overline{U}$ can acquire a weak dependence on $\sigma \sim \ln \tau$. This can happen because loop effects generically introduce logarithms of particle mass ratios everywhere and in these models particle masses in turn depend on τ. So if the two particles whose masses appear in the ratio depend differently on σ then a dependence on $\ln(m_1/m_2)$ turns into a dependence like a polynomial dependence on $\ln \tau$.

  For instance if $\overline{U}$ were to be a quadratic function, $a + b \ln \tau + c \ln^2 \tau$, then the potential $V_{\rm min}$ can easily have a local minimum. Even better, to have this minimum give $\sigma \sim 60$ – and so also $\tau^{1/4} \sim 10^{14}$, as required for the electroweak hierarchy – requires only that the coefficients a, b and c in the quadratic function are themselves of order 50 or so.

- It is a bit of a cheat to compare the vacuum energy to the electroweak scale – as done in (4.21) – since the size of w_X is actually dictated by the scale of supersymmetry breaking *in the Standard Model sector*, which cannot be smaller than $\mathcal{F} \gtrsim (10\,\text{TeV})^2$. Consistency requires this to be larger than electroweak scales, since these particular superpartners were regarded as being already integrated out. Ref. [42] explores this constraint in more detail and shows that the lower bound on the size of the supersymmetry breaking scale (in the ordinary particle sector) puts a lower bound on $V_{\rm min}$ that is of order

$$V_{\rm min} \gtrsim \frac{\epsilon^5 \mathcal{F}^* \mathcal{F}}{\overline{\tau}}\,, \tag{4.22}$$

  where $\epsilon \sim 1/\ln \overline{\tau}$ and $\overline{\tau} \sim 10^{28}$ is the vacuum value of τ chosen above to achieve the electroweak hierarchy. For $\overline{\tau} \sim 10^{28}$ (as required to reproduce the proper electroweak hierarchy) we have $(\overline{\tau} \ln^5 \overline{\tau})^{-1} \sim 10^{-37}$ and so if $\sqrt{|\mathcal{F}|} \gtrsim 10$ TeV this gives $V_{\rm min} \gtrsim 10^{-93} M_p^4$. Although not as small as the value $10^{-120} M_p^4$ required for Dark Energy, this is better than any of the alternatives on the market (and is the result 'out-of-the-box' inasmuch as the various inputs have not yet been seriously optimized to try to achieve the smallest possible result).

- There is a good reason these parameters have not yet been optimized. Once the potential minimum falls below around $V_{\rm min} = v_{\rm eff}^4 \sim 10^{-80} M_p^4$ the mass of the σ field around any minimum becomes less than of order $m_\sigma \sim v_{\rm eff}^2/M_p \sim 10^{-40} M_p \sim 10^{-13}$ eV and so the σ Compton wavelength is longer than $m_\sigma^{-1} \sim 10^6$ m. In this regime σ mediates long-range forces that can show up as deviations from GR in precision tests of gravity. This is a generic problem for *any* successful proposal that gives a technically natural Dark Energy, and is a serious one. There are ways to evade such bounds, such as if macroscopic collections of atoms (like planets or stars) should couple to σ much more weakly than would be guessed by summing the coupling strength atom-by-atom. This can happen for nonlinear couplings (and is generically called 'screening' – see *e.g.* [46]) but the jury is out so far on whether it can be done successfully in this case (see [47, 48]).

- If the potential depends on other fields (such as the Higgs field, h, or an axion field, $\mathfrak{a}$) in addition to the relaxation of the field ϕ, then relaxation will happen locally for each value of the other fields h and $\mathfrak{a}$, so $\overline{\phi} = \overline{\phi}(h, \mathfrak{a})$. But this also suppresses their contribution to the energy, giving the overall potential a trough-like shape whose bottom is parameterized by $V[h, \mathfrak{a}, \overline{\phi}(h, \mathfrak{a})]$: what minimizes a constant vacuum energy also tries to flatten the entire scalar potential for these other fields. One might (correctly) worry that this should in particular make their masses much smaller than naively expected, which at face value is a problem (at least for the higgs) its mass is actually measured (and the predictions for this were right before relaxation).

  The reason this need not actually be a problem is the relaxation is dynamical and so responds differently depending on the speed of the probe. Rapid processes like higgs particle collisions or decays occur effectively instantaneously and so ϕ has no time to respond. So these processes are in the 'sudden' approximation and so tend just to see the curvature of the 'bare' potential in the direction of the probe. This gives the mass without relaxation (as is usually assumed when computing *e.g.* collider signals). But for slow processes like cosmology the evolution of ϕ is instead adiabatic and so has time to adapt as other fields change, leading the evolution to preferentially explore the bottom of the trough (where masses really are much smaller than their naive values).

Long-story short: this framework is promising but there are a lot of working parts that must be pinned down in order to claim real progress on the cosmological constant problem. What is interesting is that there are often superficial objections (like the ones listed above) that superficially seem to be problems but which disappear on their own when examined more carefully. So far the preliminary indications continue to look good and work is underway to see whether this can persist, leading to a detailed working model. Initial indications are that cosmology can be very interesting [42, 49] and there is a tantalizing prospect of it pointing to a unified picture of the origins of both Dark Energy and Dark Matter [50], though there are also many dangerous constraints [51].

In my own mind the main worry is whether the phenomenology of having ordinary particle masses depend on the values of very light scalars can be ruled out based on what we know, but this is itself progress inasmuch as we trade the cosmological constant problem (which is very hard) for possibly much easier phenomenological issues to do with tests of gravity. Perhaps this is really an opportunity; if this class of models is how Nature works, it provides many observable consequences that perhaps are about to be discovered.

4.4 Summary

It is a remarkable opportunity that the long-distance physics we see in the sky seems to depend on how things work at much smaller distances; an opportunity that it behooves us to exploit. When this clue is ignored we have so many theoretical options that cosmological observations alone are unlikely to narrow our choices down sufficiently. Once this clue is included – for the cosmological constant problem specifically – then so far no compelling options have yet emerged at all. This shows that reconciling cosmology with high-energy physics is difficult. Should it be accomplished successfully we are likely to find an important part of how nature actually works.

These lectures have tried to make the following points.

1. Technical naturalness matters and is a natural consequence of the modern understanding of how classical gravitational physics fits into a broader quantum picture. Effective field theory is the key concept, designed to capture the important physics relevant at low energies when there is a large hierarchy of energy scales.

   Technical naturalness emerges as a criterion because there can be a different effective theory for every new range of scales, it should be possible to ask why a parameter is small at any scale we choose. This has two parts: why is the parameter small in the ultraviolet-complete theory at very high energies, and why does it stay small as one integrates out the lower-energy modes. Although we may not understand the answer of the first question until we get access to very high energies, the second part has implications even at low energies (and this is what makes the criterion of technical naturalness useful).

2. Although a technically natural understanding of the small size of the Dark Energy density has proven elusive, it is argued that it is too early to despair about solving the cosmological constant problem and the rewards for doing so are very high: any such a solution is likely to have a great many low-energy tests.

3. Personally, my own money is on low-energy approximate scale invariance being responsible for the electroweak hierarchy within which a relaxation mechanism (perhaps along the lines of [42]) accounts for the small size of $\rho_{\rm vac}$. This is likely also to point to the existence of supersymmetric large extra dimensions at accessibly low energies as a UV completion. Both of these require the existence of very light dilaton and a supersymmetric dark sector, with a host of potentially observable implications for cosmology and tests of gravity.

In the end these lectures argue the cosmological constant problem is a hopeful challenge and not a message of despair. The Universe is a Big Place, and this fact alone may well be telling us that new physics is just around the corner, since this is required by *any* real solution to the cosmological constant problem. The search so far has been hard and unsuccessful, but not all avenues have been exhaustively explored and the rewards with success are very high. With luck the interplay between cosmology, gravity and fundamental physics will soon teach us what is really going on.

Acknowledgements

I would like to thank the organizers of this school for their kind invitation to present these lectures to such a talented group of students in such pleasant environs. Over a lifetime I have learned much from my mentors, students and collaborators about the opportunities offered by naturalness and cosmology. My research has been supported in part by the Natural Sciences and Engineering Research Council of Canada. Research at the Perimeter Institute is supported in part by the Government of Canada through Industry Canada, and by the Province of Ontario through the Ministry of Research and Information.

References

[1] P.J.E. Peebles, *Principles of Physical Cosmology*, Princeton University Press (1993)

[2] S. Weinberg, *Gravitation and Cosmology*, Wiley 1973.

[3] C. W. Misner, J. A. Wheeler and K. S. Thorne, *Gravitation*, W. H. Freeman & Company 1973.

[4] Alan H. Guth,"The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems", Phys. Rev. **D23**, (1981), 347-356;

A.D. Linde, "A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems", Phys. Lett. **B108**, (1982), 389-393;

A. Albrecht and P.J. Steinhardt, "Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking", Phys. Rev. Lett. **48** (1982), 1220-1223;

[5] V. F. Mukhanov and G. V. Chibisov, JETP Lett. 33, 532 (1981) [Pisma Zh. Eksp. Teor. Fiz. 33, 549 (1981)];

A. H. Guth and S. Y. Pi, Phys. Rev. Lett. 49, 1110 (1982);

A. A. Starobinsky, Phys . Lett. B 117, 175 (1982);

S. W. Hawking, Phys. Lett. B 115, 295 (1982);

V. N. Lukash, Pisma Zh. Eksp. Teor. Fiz. 31, 631 (1980); Sov. Phys. JETP 52, 807 (1980) [Zh. Eksp. Teor. Fiz. 79, (1980)];

W. Press, Phys. Scr. 21, 702 (1980);

K. Sato, Mon. Not. Roy. Astron. Soc. 195, 467 (1981).

[6] V. Mukhanov, *Physical Foundations of Cosmology*, Cambridge University Press (2005).

S. Weinberg, *Cosmology*, Oxford University Press (2008).

[7] T. Louis *et al.* [ACT], "The Atacama Cosmology Telescope: DR6 Power Spectra, Likelihoods and ΛCDM Parameters," [arXiv:2503.14452 [astro-ph.CO]].

[8] Y. S. Murakami, A. G. Riess, B. E. Stahl, W. D. Kenworthy, D. M. A. Pluck, A. Macoretta, D. Brout, D. O. Jones, D. M. Scolnic and A. V. Filippenko, "Leveraging SN Ia spectroscopic similarity to improve the measurement of H_0," JCAP **11** (2023), 046 [arXiv:2306.00070 [astro-ph.CO]].

[9] W. L. Freedman, "Measurements of the Hubble Constant: Tensions in Perspective," Astrophys. J. **919** (2021) no.1, 16 [arXiv:2106.15656 [astro-ph.CO]].

[10] E. Di Valentino *et al.* [CosmoVerse], "The CosmoVerse White Paper: Addressing observational tensions in cosmology with systematics and fundamental physics," [arXiv:2504.01669 [astro-ph.CO]].

[11] T. M. C. Abbott *et al.* [DES], "Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing," Phys. Rev. D **105** (2022) no.2, 023520 [arXiv:2105.13549 [astro-ph.CO]].

[12] E. Di Valentino, O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri, D. F. Mota, A. G. Riess and J. Silk, "In the realm of the Hubble tension—a review of solutions," Class. Quant. Grav. **38** (2021) no.15, 153001 [arXiv:2103.01183 [astro-ph.CO]].

[13] A. R. Khalife, M. B. Zanjani, S. Galli, S. Günther, J. Lesgourgues and K. Benabed, "Review of Hubble tension solutions with new SH0ES and SPT-3G data," JCAP **04** (2024), 059 [arXiv:2312.09814 [astro-ph.CO]].

[14] M. Abdul Karim *et al.* [DESI], "DESI DR2 Results II: Measurements of Baryon Acoustic Oscillations and Cosmological Constraints," [arXiv:2503.14738 [astro-ph.CO]].

[15] P. J. E. Peebles and B. Ratra, "The Cosmological Constant and Dark Energy," Rev. Mod. Phys. **75** (2003), 559-606 [arXiv:astro-ph/0207347 [astro-ph]].

E. J. Copeland, M. Sami and S. Tsujikawa, "Dynamics of dark energy," Int. J. Mod. Phys. D **15** (2006), 1753-1936 [arXiv:hep-th/0603057 [hep-th]].

R. R. Caldwell and M. Kamionkowski, "The Physics of Cosmic Acceleration," Ann. Rev. Nucl. Part. Sci. **59** (2009), 397-429 [arXiv:0903.0866 [astro-ph.CO]].

A. Silvestri and M. Trodden, "Approaches to Understanding Cosmic Acceleration," Rept. Prog. Phys. **72** (2009), 096901 [arXiv:0904.0024 [astro-ph.CO]].

S. Tsujikawa, "Dark energy: investigation and modeling," [arXiv:1004.1493 [astro-ph.CO]].

[16] S. Weinberg, "Phenomenological Lagrangians," Physica A **96** (1979) no.1-2, 327-340

[17] C. P. Burgess, "Quantum gravity in everyday life: General relativity as an effective field theory," Living Rev. Rel. **7** (2004), 5-56 [arXiv:gr-qc/0311082 [gr-qc]].

[18] J. F. Donoghue, "The effective field theory treatment of quantum gravity," AIP Conf. Proc. **1483** (2012) 73 [arXiv:1209.3511 [gr-qc]].

[19] C. P. Burgess, "Introduction to Effective Field Theory," Cambridge University Press, 2020, ISBN 978-1-139-04804-0, 978-0-521-19547-8 doi:10.1017/9781139048040

[20] S. Weinberg, "The Cosmological Constant Problem," Rev. Mod. Phys. **61** (1989), 1-23

[21] S. M. Carroll, "The Cosmological constant," Living Rev. Rel. **4** (2001), 1 [arXiv:astro-ph/0004075 [astro-ph]].

[22] C. P. Burgess, "The Cosmological Constant Problem: Why it's hard to get Dark Energy from Micro-physics," [arXiv:1309.4133 [hep-th]].

[23] For other points of view on the cosmological constant problem see:

S. M. Carroll, "The Cosmological constant," Living Rev. Rel. **4** (2001) 1 [astro-ph/0004075];

P. Binetruy, "Cosmological constant versus quintessence," Int. J. Theor. Phys. **39** (2000) 1859 [hep-ph/0005037];

T. Padmanabhan, "Cosmological constant: The Weight of the vacuum," Phys. Rept. **380** (2003) 235 [hep-th/0212290];

J. Frieman, M. Turner and D. Huterer, "Dark Energy and the Accelerating Universe," Ann. Rev. Astron. Astrophys. **46** (2008) 385 [arXiv:0803.0982 [astro-ph]].

[24] XXX CITE HERE ARTICLES ON POSITIVITY CONDITIONS XXX

[25] B. S. DeWitt, "Quantum Theory of Gravity. 3. Applications of the Covariant Theory," Phys. Rev. **162** (1967), 1239-1256

[26] C. P. Burgess, H. M. Lee and M. Trott, "Power-counting and the Validity of the Classical Approximation During Inflation," JHEP **09** (2009), 103 [arXiv:0902.4465 [hep-ph]].

[27] P. Adshead, C. P. Burgess, R. Holman and S. Shandera, "Power-counting during single-field slow-roll inflation," JCAP **02** (2018), 016 [arXiv:1708.07443 [hep-th]].

[28] C. P. Burgess, "Intro to Effective Field Theories and Inflation," in the proceedings of the Les Houches school *Effective Field Theory in Particle Physics and Cosmology* [arXiv:1711.10592 [hep-th]].

[29] T. Kobayashi, "Horndeski theory and beyond: a review," Rept. Prog. Phys. **82** (2019) no.8, 086901 [arXiv:1901.07183 [gr-qc]].

G. W. Horndeski and A. Silvestri, "50 Years of Horndeski Gravity: Past, Present and Future," Int. J. Theor. Phys. **63** (2024) no.2, 38 [arXiv:2402.07538 [gr-qc]].

[30] R. P. Woodard, "Ostrogradsky's theorem on Hamiltonian instability," Scholarpedia **10** (2015) no.8, 32243 [arXiv:1506.02210 [hep-th]].

[31] C. P. Burgess and M. Williams, "Who You Gonna Call? Runaway Ghosts, Higher Derivatives and Time-Dependence in EFTs," JHEP **08** (2014), 074 [arXiv:1404.2236 [gr-qc]].

[32] A. R. Solomon and M. Trodden, "Higher-derivative operators and effective field theory for general scalar-tensor theories," JCAP **02** (2018), 031 [arXiv:1709.09695 [hep-th]].

[33] G. 't Hooft, in *Cargese Summer Inst.* 1979:135 (QCD161:S77:1979) (reprinted in *'t Hooft, G. (ed.): Under the spell of the gauge principle* 352-374, and in *Farhi, E. (ed.), Jackiw, R. (ed.): Dynamical gauge symmetry breaking* 345-367).

[34] C. Wetterich, "Cosmology and the Fate of Dilatation Symmetry," Nucl. Phys. B **302** (1988), 668-696 [arXiv:1711.03844 [hep-th]].

R. D. Peccei, J. Sola and C. Wetterich, "Adjusting the Cosmological Constant Dynamically: Cosmons and a New Force Weaker Than Gravity," Phys. Lett. B **195** (1987), 183-190

[35] C. P. Burgess, M. Cicoli, D. Ciupke, S. Krippendorf and F. Quevedo, "UV Shadows in EFTs: Accidental Symmetries, Robustness and No-Scale Supergravity," Fortsch. Phys. **68** (2020) no.10, 2000076 [arXiv:2006.06694 [hep-th]].

[36] D. V. Volkov and V. P. Akulov, "Is the Neutrino a Goldstone Particle?," Phys. Lett. B **46** (1973) 109; "Goldstone fields with spin 1/2," Theor. Math. Phys. **18** (1974) 28 [Teor. Mat. Fiz. **18** (1974) 39];

B. Zumino, "Nonlinear Realization of Supersymmetry in de Sitter Space," Nucl. Phys. B **127** (1977) 189;

S. Samuel and J. Wess, "A Superfield Formulation Of The Nonlinear Realization Of Supersymmetry And Its Coupling To Supergravity," Nucl. Phys. B **221** (1983) 153;

J. Wess and J. Bagger, "Supersymmetry and supergravity," Princeton, USA: Univ. Pr. (1992) 259 pp.

[37] C. P. Burgess and F. Quevedo, "Who's Afraid of the Supersymmetric Dark? The Standard Model vs Low-Energy Supergravity," Fortsch. Phys. **70** (2022) no.7-8, 2200077 [arXiv:2110.13275 [hep-th]].

[38] Y. Aghababaie, C. P. Burgess, S. L. Parameswaran and F. Quevedo, "Towards a naturally small cosmological constant from branes in 6-D supergravity," Nucl. Phys. B **680** (2004) 389 [arXiv:hep-th/0304256].

M. Cicoli, C. P. Burgess and F. Quevedo, "Anisotropic Modulus Stabilisation: Strings at LHC Scales with Micron-sized Extra Dimensions," JHEP **10** (2011), 119 [arXiv:1105.2107 [hep-th]].

[39] For some (relatively old) reviews see

C.P. Burgess, "Supersymmetric Large Extra Dimensions and the Cosmological Constant: An Update," *Ann. Phys.* **313** (2004) 283-401 [arXiv:hep-th/0402200]; "Towards a natural theory of dark energy: Supersymmetric large extra dimensions," in the proceedings of the Texas A&M Workshop on String Cosmology, [arXiv:hep-th/0411140].

[40] XXX CITE HERE MODELS THAT SUGGEST THE YOGA CONSTRUCTION WHERE SM FIELDS ARE FIRST IN THE LOOP TERM. (EG MOTIVATED BY BRANEWORLD CONSTRUCTIONS) XXX

[41] S. Weinberg, "Baryon and Lepton Nonconserving Processes," Phys. Rev. Lett. **43** (1979), 1566-1570

[42] C. P. Burgess, D. Dineen and F. Quevedo, "Yoga Dark Energy: natural relaxation and other dark implications of a supersymmetric gravity sector," JCAP **03** (2022) no.03, 064 [arXiv:2111.07286 [hep-th]].

[43] Z. Komargodski and N. Seiberg, "From Linear SUSY to Constrained Superfields," JHEP **09** (2009), 066 doi:10.1088/1126-6708/2009/09/066 [arXiv:0907.2441 [hep-th]].

R. Kallosh, F. Quevedo and A. M. Uranga, JHEP **12** (2015), 039 doi:10.1007/JHEP12(2015)039 [arXiv:1507.07556 [hep-th]].

E. Dudas, S. Ferrara, A. Kehagias and A. Sagnotti, "Properties of Nilpotent Supergravity," JHEP **09** (2015), 217 [arXiv:1507.07842 [hep-th]].

E. A. Bergshoeff, D. Z. Freedman, R. Kallosh and A. Van Proeyen, "Pure de Sitter Supergravity," Phys. Rev. D **92** (2015) no.8, 085040 [erratum: Phys. Rev. D **93** (2016) no.6, 069901] [arXiv:1507.08264 [hep-th]].

[44] C. P. Burgess and F. Quevedo, "RG-induced modulus stabilization: perturbative de Sitter vacua and improved D3-$\overline{\mathrm{D3}}$ inflation," JHEP **06** (2022), 167 [arXiv:2202.05344 [hep-th]].

[45] M. Cicoli, C. Hughes, A. R. Kamal, F. Marino, F. Quevedo, M. Ramos-Hamud and G. Villa, "Back to the origins of brane–antibrane inflation," Eur. Phys. J. C **85** (2025) no.3, 315 [arXiv:2410.00097 [hep-th]].

[46] J. Khoury and A. Weltman, "Chameleon fields: Awaiting surprises for tests of gravity in space," Phys. Rev. Lett. **93** (2004), 171104 [arXiv:astro-ph/0309300 [astro-ph]]; "Chameleon cosmology," Phys. Rev. D **69** (2004), 044026 [arXiv:astro-ph/0309411 [astro-ph]].

[47] P. Brax, C. P. Burgess and F. Quevedo, "Light axiodilatons: matter couplings, weak-scale completions and long-distance tests of gravity," JCAP **08** (2023), 011 [arXiv:2212.14870 [hep-ph]].

[48] C. P. Burgess and F. Quevedo, "Axion homeopathy: screening dilaton interactions," JCAP **04** (2022) no.04, 007 [arXiv:2110.10352 [hep-th]].

O. Lacombe and S. Mukohyama, "Multi-scalar theories of gravity with direct matter couplings and their parametrized post-Newtonian parameters," JCAP **08** (2023), 054 [arXiv:2302.08941 [gr-qc]].

P. Brax, C. P. Burgess and F. Quevedo, "Axio-Chameleons: a novel string-friendly multi-field screening mechanism," JCAP **03** (2024), 015 [arXiv:2310.02092 [hep-th]].

[49] A. Smith, M. Mylova, P. Brax, C. van de Bruck, C. P. Burgess and A. C. Davis, "CMB implications of multi-field axio-dilaton cosmology," JCAP **12** (2024), 058 [arXiv:2408.10820 [hep-th]].

A. Smith, P. Brax, C. van de Bruck, C. P. Burgess and A. C. Davis, "Screened Axio-dilaton Cosmology: Novel Forms of Early Dark Energy," [arXiv:2505.05450 [hep-th]].

[50] A. Smith, M. Mylova, P. Brax, C. van de Bruck, C. P. Burgess and A. C. Davis, "A Minimal Axio-dilaton Dark Sector," [arXiv:2410.11099 [hep-th]].

[51] M. Baryakhtar, O. Simon and Z. J. Weiner, "Cosmology with varying fundamental constants from hyperlight, coupled scalars," Phys. Rev. D **110** (2024) no.8, 083505 [arXiv:2405.10358 [astro-ph.CO]].

[52] R. Sundrum, "Fat gravitons, the cosmological constant and submillimeter tests," Phys. Rev. D **69** (2004) 044014 [hep-th/0306106].

[53] J.G. Williams, S.G. Turyshev and D.H. Boggs, "Progress in Lunar Laser Ranging Tests of Relativistic Gravity," Phys. Rev. Lett. **93** (2004) 261101 [arXiv:gr-qc/0411113].

[54] C. M. Will, "The confrontation between general relativity and experiment," Living Rev. Rel. **4** (2001) 4 [arXiv:gr-qc/0103036]; Living Rev. Rel. **9** (2005) 3 [arXiv:gr-qc/0510072];

D. J. Kapner, T. S. Cook, E. G. Adelberger, J. H. Gundlach, B. R. Heckel, C. D. Hoyle and H. E. Swanson, "Tests of the gravitational inverse-square law below the dark-energy length scale," Phys. Rev. Lett. **98** (2007) 021101 [hep-ph/0611184];

E. G. Adelberger, J. H. Gundlach, B. R. Heckel, S. Hoedl and S. Schlamminger, "Torsion balance experiments: A low-energy frontier of particle physics," Prog. Part. Nucl. Phys. **62** (2009) 102.

[55] A. Vilenkin, "Gravitational Field of Vacuum Domain Walls and Strings," Phys. Rev. D **23** (1981) 852;

R. Gregory, "Gravitational Stability Of Local Strings," Phys. Rev. Lett. **59** (1987) 740;

A. G. Cohen and D. B. Kaplan, "The Exact Metric About Global Cosmic Strings," *Phys. Lett.* **B215**, 67 (1988);

A. Vilenkin and P. Shellard, *Cosmic Strings and other Topological Defects*, Cambridge University Press (1994);

R. Gregory and C. Santos, "Cosmic strings in dilaton gravity," Phys. Rev. D **56** (1997) 1194 [gr-qc/9701014].

[56] J. W. Chen, M. A. Luty and E. Ponton, "A Critical cosmological constant from millimeter extra dimensions" JHEP **0009** (2000) 012 [arXiv:hep-th/0003067];

S. M. Carroll and M. M. Guica, "Sidestepping the cosmological constant with football-shaped extra dimensions," [arXiv:hep-th/0302067];

I. Navarro, "Codimension two compactifications and the cosmological constant problem," JCAP **0309** (2003) 004 [arXiv:hep-th/0302129];

E. Papantonopoulos and A. Papazoglou, "Brane-bulk matter relation for a purely conical codimension-2 brane world," JCAP **0507** (2005) 004 [arXiv:hep-th/0501112].

[57] Y. Aghababaie, C. P. Burgess, J. M. Cline, H. Firouzjahi, S. L. Parameswaran, F. Quevedo, G. Tasinato and I. Zavala, "Warped brane worlds in six-dimensional supergravity," JHEP **0309** (2003) 037 [hep-th/0308064].

[58] C. P. Burgess, A. Maharana, L. van Nierop, A. A. Nizami and F. Quevedo, "On Brane Back-Reaction and de Sitter Solutions in Higher-Dimensional Supergravity," JHEP **04** (2012), 018 [arXiv:1109.0532 [hep-th]].

F. F. Gautason, D. Junghans and M. Zagermann, "Cosmological Constant, Near Brane Behavior and Singularities," JHEP **09** (2013), 123 [arXiv:1301.5647 [hep-th]].

[59] N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, "The hierarchy problem and new dimensions at a millimeter," Phys. Lett. B **429** (1998) 263 [arXiv:hep-ph/9803315];

I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, "New dimensions at a millimeter to a Fermi and superstrings at a TeV," Phys. Lett. B **436** (1998) 257 [arXiv:hep-ph/9804398].

[60] G. F. Giudice, R. Rattazzi, J. D. Wells, "Quantum gravity and extra dimensions at high-energy colliders," Nucl. Phys. **B544** (1999) 3 [arXiv:hep-ph/9811291];

T. Han, J. D. Lykken, R. -J. Zhang, "On Kaluza-Klein states from large extra dimensions," Phys. Rev. **D59** (1999) 105006 [arXiv:hep-ph/9811350];

J. L. Hewett, "Indirect collider signals for extra dimensions," Phys. Rev. Lett. **82** (1999) 4765 [arXiv:hep-ph/9811356];

D. Atwood, C. P. Burgess, E. Filotas, F. Leblond, D. London and I. Maksymyk, "Supersymmetric large extra dimensions are small and/or numerous," Phys. Rev. D **63** (2001) 025007 [arXiv:hep-ph/0007178];

G. F. Giudice and A. Strumia, "Constraints on extra dimensional theories from virtual graviton exchange," Nucl. Phys. B **663** (2003) 377 [arXiv:hep-ph/0301232].

[61] C. P. Burgess and L. van Nierop, "Technically Natural Cosmological Constant From Supersymmetric 6D Brane Backreaction," arXiv:1108.0345 [hep-th].

[62] C. P. Burgess, L. van Nierop, S. Parameswaran, A. Salvio and M. Williams, "Accidental SUSY: Enhanced Bulk Supersymmetry from Brane Back-reaction," arXiv:1210.5405 [hep-th].

[63] C. P. Burgess, J. Matias and F. Quevedo, "MSLED: A Minimal supersymmetric large extra dimensions scenario," Nucl. Phys. B **706** (2005) 71 [arXiv:hep-ph/0404135].

[64] G. Azuelos, P. H. Beauchemin, C. P. Burgess, "Phenomenological constraints on extra dimensional scalars," J. Phys. G **G31** (2005) 1-20 [hep-ph/0401125];

P. H. Beauchemin, G. Azuelos, C. P. Burgess, "Dimensionless coupling of bulk scalars at the LHC," J. Phys. G **G30** (2004) N17 [hep-ph/0407196];

M. Williams, C. P. Burgess, A. Maharana and F. Quevedo, "New Constraints (and Motivations) for Abelian Gauge Bosons in the MeV-TeV Mass Range," JHEP **1108** (2011) 106 [arXiv:1103.4556 [hep-ph]];

R. Diener and C. P. Burgess, "Bulk Stabilization, the Extra-Dimensional Higgs Portal and Missing Energy in Higgs Events," JHEP **1305** (2013) 078 [arXiv:1302.6486 [hep-ph]].

[65] H. Nishino and E. Sezgin, *Phys. Lett.* **144B** (1984) 187; "The Complete N=2, D = 6 Supergravity With Matter And Yang-Mills Couplings," Nucl. Phys. **B278** (1986) 353;

S. Randjbar-Daemi, A. Salam, E. Sezgin and J. Strathdee, "An Anomaly Free Model in Six-Dimensions" *Phys. Lett.* **B151** (1985) 351.

[66] A. Salam and E. Sezgin, "Chiral Compactification On Minkowski X S**2 Of N=2 Einstein-Maxwell Supergravity In Six-Dimensions," Phys. Lett. B **147** (1984) 47.

[67] C. P. Burgess and L. van Nierop, "Bulk Axions, Brane Back-reaction and Fluxes," arXiv:1012.2638 [hep-th];

C.P.Burgess and L. van Nierop, "Large Dimensions and Small Curvatures from Supersymmetric Brane Back-reaction," [arXiv:1101.0152 (hep-th)].

[68] G. W. Gibbons, R. Guven and C. N. Pope, "3-branes and uniqueness of the Salam-Sezgin vacuum," Phys. Lett. B **595** (2004) 498 [hep-th/0307238];

C. P. Burgess, F. Quevedo, G. Tasinato and I. Zavala, "General axisymmetric solutions and self-tuning in 6D chiral gauged supergravity," JHEP **0411** (2004) 069 [hep-th/0408109].

[69] H. -P. Nilles, A. Papazoglou and G. Tasinato, "Selftuning and its footprints," Nucl. Phys. B **677** (2004) 405 [hep-th/0309042];

J. Garriga and M. Porrati, "Football shaped extra dimensions and the absence of self-tuning," JHEP **0408** (2004) 028 [hep-th/0406158];

J. Vinet and J. M. Cline, "Codimension-two branes in six-dimensional supergravity and the cosmological constant problem," Phys. Rev. D **71** (2005) 064011 [hep-th/0501098].

[70] A. J. Tolley, C. P. Burgess, D. Hoover and Y. Aghababaie, "Bulk singularities and the effective cosmological constant for higher co-dimension branes," JHEP **0603** (2006) 091 [arXiv:hep-th/0512218];

A. J. Tolley, C. P. Burgess, C. de Rham and D. Hoover, "Scaling solutions to 6D gauged chiral supergravity," New J. Phys **8** (2006) 324 [arXiv:0608.083 [hep-th]];

C.P. Burgess and Leo van Nierop, "Sculpting the Extra Dimensions: Inflation from Codimension-2 Brane Back-reaction" [arXiv:1108.2553v1 [hep-th]].

[71] C. P. Burgess, A. Maharana, L. van Nierop, A. A. Nizami and F. Quevedo, "On Brane Back-Reaction and de Sitter Solutions in Higher-Dimensional Supergravity," JHEP **1204** (2012) 018 [arXiv:1109.0532 [hep-th]];

see also R. Koster and M. Postma, "A no-go for no-go theorems prohibiting cosmic acceleration in extra dimensional models," JCAP **1112** (2011) 015 [arXiv:1110.1492 [hep-th]].

[72] J. M. Maldacena and C. Nunez, "Supergravity description of field theories on curved manifolds and a no go theorem," Int. J. Mod. Phys. A **16** (2001) 822 [hep-th/0007018];

D. H. Wesley, "New no-go theorems for cosmic acceleration with extra dimensions," arXiv:0802.2106 [hep-th];

P. J. Steinhardt and D. Wesley, "Dark Energy, Inflation and Extra Dimensions," Phys. Rev. D **79** (2009) 104026 [arXiv:0811.1614 [hep-th]].

[73] C. P. Burgess, D. Hoover, C. de Rham and G. Tasinato, "Effective Field Theories and Matching for Codimension-2 Branes," JHEP **0903** (2009) 124 [arXiv:0812.3820 [hep-th]];

A. Bayntun, C.P. Burgess and L. van Nierop, "Codimension-2 Brane-Bulk Matching: Examples from Six and Ten Dimensions," New J. Phys. **12** (2010) 075015 [arXiv:0912.3039 [hep-th]].

[74] N. Arkani-Hamed, S. Dimopoulos, N. Kaloper and R. Sundrum, "A small cosmological constant from a large extra dimension," Phys. Lett. B **480** (2000) 193 [arXiv:hep-th/0001197];

S. Kachru, M. B. Schulz and E. Silverstein, "Self-tuning flat domain walls in 5d gravity and string theory," Phys. Rev. D **62** (2000) 045021 [arXiv:hep-th/0001206];

A. Kehagias and K. Tamvakis, "A selftuning solution of the cosmological constant problem," Mod. Phys. Lett. A **17** (2002) 1767 [hep-th/0011006].

[75] S. Forste, Z. Lalak, S. Lavignac and H. P. Nilles, "A Comment on selftuning and vanishing cosmological constant in the brane world," Phys. Lett. B **481** (2000) 360 [hep-th/0002164];

C. Csaki, J. Erlich, C. Grojean and T. J. Hollowood, "General properties of the selftuning domain wall approach to the cosmological constant problem," Nucl. Phys. B **584** (2000) 359 [hep-th/0004133];

P. Binetruy, J. M. Cline and C. Grojean, "Dynamical instability of brane solutions with a self-tuning cosmological constant," Phys. Lett. B **489** (2000) 403 [hep-th/0007029].

[76] C. P. Burgess, D. Hoover and G. Tasinato, "UV Caps and Modulus Stabilization for 6D Gauged Chiral Supergravity," JHEP **0709** (2007) 124 [arXiv:0705.3212 [hep-th]];

C. P. Burgess, D. Hoover, G. Tasinato, "Technical Naturalness on a Codimension-2 Brane," JHEP **0906** (2009) 014. [arXiv:0903.0402 [hep-th]].

[77] C. P. Burgess and D. Hoover, "UV sensitivity in supersymmetric large extra dimensions: The Ricci-flat case," Nucl. Phys. B **772** (2007) 175 [hep-th/0504004];

D. Hoover and C. P. Burgess, "Ultraviolet sensitivity in higher dimensions," JHEP **0601** (2006) 058 [hep-th/0507293];

C. P. Burgess, D. Hoover, G. Tasinato, "Technical Naturalness on a Codimension-2 Brane," JHEP **0906** (2009) 014. [arXiv:0903.0402 [hep-th]];

M. Williams, C. P. Burgess, L. van Nierop and A. Salvio, "Running with Rugby Balls: Bulk Renormalization of Codimension-2 Branes," [arXiv:1210.3753 [hep-th]];

[78] K. R. Dienes, E. Dudas and T. Gherghetta, "Neutrino oscillations without neutrino masses or heavy mass scales: A Higher dimensional seesaw mechanism," Nucl. Phys. B **557** (1999), 25 [arXiv:hep-ph/9811428 [hep-ph]].

[79] J. Matias and C. P. Burgess, "MSLED, neutrino oscillations and the cosmological constant," JHEP **0509** (2005) 052 [arXiv:hep-ph/0508156];

[80] B. J. Carr and M. J. Rees, "The anthropic principle and the structure of the physical world," Nature **278** (1979) 605;

S. Weinberg, "Anthropic Bound on the Cosmological Constant," Phys. Rev. Lett. **59** (1987) 2607;

R. Bousso and J. Polchinski, "Quantization of four form fluxes and dynamical neutralization of the cosmological constant," JHEP **0006** (2000) 006 [hep-th/0004134];

T. Banks, M. Dine and L. Motl, "On anthropic solutions of the cosmological constant problem," JHEP **0101** (2001) 031 [hep-th/0007206];

R. Kallosh and A. D. Linde, "M theory, cosmological constant and anthropic principle," Phys. Rev. D **67** (2003) 023510 [hep-th/0208157];

J. Garriga and A. Vilenkin, "Testable anthropic predictions for dark energy," Phys. Rev. D **67** (2003) 043503 [astro-ph/0210358];

L. Susskind, "The Anthropic landscape of string theory," In *Carr, Bernard (ed.): Universe or multiverse?* 247-266 [hep-th/0302219];

J. Polchinski, "The Cosmological Constant and the String Landscape," hep-th/0603249.

M. Hertzberg, M. Tegmark and F. Wilczek, "Axion Cosmology and the Energy Scale of Inflation," Phys. Rev. D **78** (2008) 083507 [arXiv:0807.1726 [astro-ph]].

[81] C. Vafa, "The String landscape and the swampland," [arXiv:hep-th/0509212 [hep-th]].

[82] T. Banks and L. J. Dixon, "Constraints on String Vacua with Space-Time Supersymmetry," Nucl. Phys. B **307** (1988), 93-108

[83] C. P. Burgess, J. P. Conlon, L. Y. Hung, C. H. Kom, A. Maharana and F. Quevedo, "Continuous Global Symmetries and Hyperweak Interactions in String Compactifications," JHEP **07** (2008), 073 [arXiv:0805.4037 [hep-th]].

[84] G. Obied, H. Ooguri, L. Spodyneiko and C. Vafa, "De Sitter Space and the Swampland," [arXiv:1806.08362 [hep-th]].

[85] M. Montero, C. Vafa and I. Valenzuela, "The dark dimension and the Swampland," JHEP **02** (2023), 022 [arXiv:2205.12293 [hep-th]].