MPA Blackboard Lectures **Dragan Huterer, University of Michigan**

Dark Energy and the Accelerating Universe

Recommended reading:

- "Dark Energy and the Accelerating Universe", J. Frieman, M. Turner and D. Huterer, Ann. Rev. Astron. Astrophys. **46**, 385 (2008), arXiv:0803.0982 — *a* ~ 40*-page review of DE for a general-practice physicist or astronomer.*
- "Observational Probes of Cosmic Acceleration", D. Weinberg, M. Mortonson, D.J. Eisenstein, C. Hirata, A. Riess, and E. Rozo, Phys. Rept. **530**, 87 (2013); arXiv:1201.2434 — *a super-detailed, definitive review on cosmological probes of DE.*
- "Dynamics of Dark Energy", E. Copeland, M. Sami and S. Tsujikawa, Int. J. Mod. Phys. D15, 1753 (2006), arXiv:hep-th/0603057 — *a detailed review strong on theoretical attempts to explain DE.*
- "Measurements of the cosmological parameters Ω_M and Λ from 42 high-redshift supernovae", S. Perlmutter et al. (The Supernova Cosmology Project), Astrophysical Journal **517**, 565 (1999) — *a classic.*
- "Supernovae, Dark Energy, and the Accelerating Universe", S. Perlmutter, Physics Today, April 2003 — *very nice popular account.*
- "The Extravagant Universe", R.P. Kirshner, Princeton Univ Press 2004 — *an entertaining book about the discovery of DE.*

History of the discovery of DE. Type Ia Supernovae.

Early signs of a new component. At least as far back as the 1980s (and probably earlier), there has emerged evidence that the matter density in the universe is low, roughly $\Omega_M \simeq 0.1$-0.3. While the precise value was hard to pin down due to various systematics, it was pretty clear that the matter density was far from the value predicted by inflation, $\Omega_M = 1.0$.

One line of argument for low Ω_M used measurements of the age of the universe, which was lower-bounded to $t_0 \gtrsim$ 12-15 Gyr using ages of globular clusters. In a flat, matter-dominated universe, the age is $t_0 = (2/3)H_0^{-1} \simeq 6.5h^{-1}\,\mathrm{Gyr}$, where $h = H_0/(100\,\mathrm{km/s/Mpc})$. Even for lowest values of the Hubble constant reasonably expected at the time, $h \simeq 0.5$, the age is uncomfortably low.

The other line of argument includes various probes of galaxy clustering, velocities, and structure formation in general. The simplest and most compelling argument is that the measured "shape parameter" $\Gamma \equiv \Omega_M h$ of the galaxy power spectrum, which determines the turnover that corresponds to the size of the universe at matter-radiation equality, is surprisingly low. $\Gamma \simeq 0.25$ has consistently been measured implying, for reasonably values of the Hubble constant, that Ω_M simply cannot be one.

Faced with these difficulties, cosmologists have proposed all kinds of theoretical explanations in the 80s and early-to-mid 90s, including actually the cosmological constant (vacuum energy).

Parallel to these developments, a group at UC Berkeley has been planning to develop strategies to use special property of type Ia supernovae – that of an (approximate) standard candle – in order to build the Hubble diagram and decide if the universe is open, flat or closed.

Standard Candles. A "standard candle" is a hypothetical object that has a fixed luminosity (that is, fixed intrinsic power that it radiates). Having a standard candle would be useful since

then you could infer distances from objects just by using the inverse square law. In fact, you don't even to know the luminosity of the standard candle to be able to infer *relative* distances to objects. In other words, if you know that L is constant, then by measuring the flux

$$f = \frac{L}{4\pi d_L^2} \tag{1}$$

you immediately get ratios of luminosity distances (but not the absolute scales unless if you actually know the value of L).

Type Ia supernovae. Type Ia supernovae are interesting objects. They have been studied extensively by the famous American-Swiss astronomer Fritz Zwicky (there is a notable paper by Baade and Zwicky from 1934); Zwicky gave them their name. They have been known to have nearly uniform luminosity; this feature is easily understood from the currently favored explanation for the physics of these events: these are white dwarf stars accreting matter from a companion, going over the Chandrasekhar limit, and undergoing explosion.

The best standard candles known today – also satisfying the requirement that they are bright enough so they can be observed to cosmological distances – are type Ia supernovae (SNIa). These are presumed (so, not 100% certain but very likely) to be binary systems where a white dwarf is accreting matter from a companion star, reaching the critical Chandrasekhar mass. The density gets high, the temperature gets high, and a thermonuclear flame starts. These are extremely powerful explosions seen to distant parts of the universe.

Cosmology with SNIa. However, a real problem was scheduling telescopes to detect and "follow-up" SN that are discovered. Basically, if you point a telescope at a galaxy and wait for the SN to go off, you will wait an average of 500 years. There has been a program in the 1980s to do that and it discovered only one SN, and after the peak!

Four crucial developments contributed to the discovery of dark energy:

1. The first major breakthrough came in the 1990s when two teams of SN researchers Supernova Cosmology Project (SCP; led by Saul Perlmutter and organized in the late 1980s) and High-z Supernova Search Team (Highz; organized in the mid 1990s and led, at the time, by Brian Schmidt) developed an efficient approach to use world's most powerful telescopes working in concert to discover and follow up high-redshift SN, and thus complement the existing efforts at lower redshift led by the Calán/Tololo collaboration. These teams had been able to essentially guarantee that they would find batches of SN in each run. [For popular reviews of these exciting developments, see Bob Kirshner's book "The Extravagant Universe" and the Perlmutter-Schmidt article "Measuring Cosmology with Supernovae".]

2. The second breakthrough came in 1993 by Mark Phillips, astronomer working in Chile. He noticed that the SN luminosity – or absolute magnitude – is correlated with the decay time of SN light curve. Phillips considered the quantity Δm_{15}, the attenuation of the flux of SN between the light maximum and 15 days past the maximum. He found that Δm_{15} is strongly correlated with the intrinsic brightness of SN; see the left panel of Fig. 1. The "Phillips relation" roughly goes as

   Broader is brighter.

   In other words, supernovae with broader light-curves have a larger intrinsic luminosity. One way to quantify this relation is to use a "stretch" factor which is a (calibration) parameter that measures width of a light curve; see the right panel of Fig. 1. By applying the correction based upon the Phillips relation, astronomers found that the intrinsic dispersion of SN, which is of order ~ 0.5 magnitudes, can be brought down to $\delta m \sim 0.2$ magnitudes

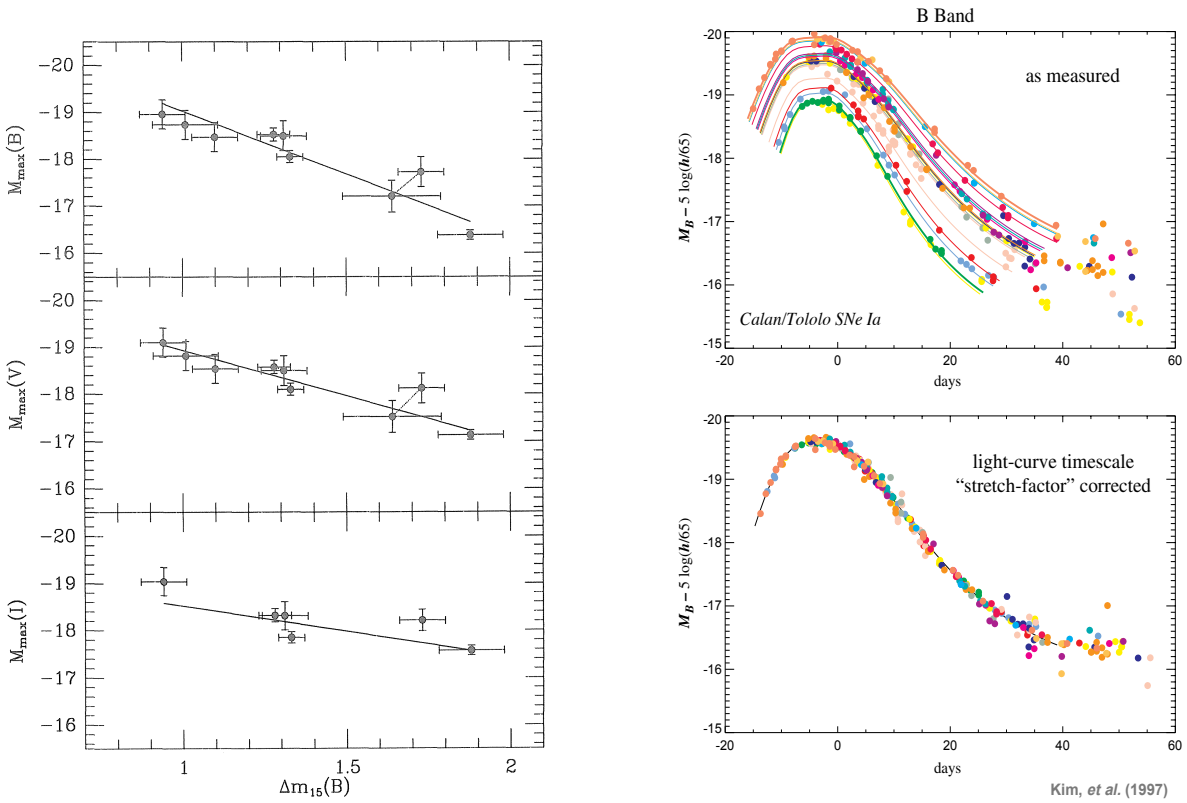

Figure 1: *Left panel:* Phillips relation, from his 1993 paper. The (apparent) magnitude of type Ia supernovae is correlated with Δm_{15}, the decay of the light curve 15 days after the maximum. *Right panel:* light curves of a sample of SNIa before correction for the Phillips relation (top), and after (bottom).

once we correct each SN luminosity using its stretch factor. Note that the final dispersion in magnitudes corresponds to the error in distance of

$$\delta d_L/d_L = \frac{\ln 10}{5}\,\delta m \simeq 0.5\,\delta m \simeq 0.1. \tag{2}$$

The Phillips relation was the second key ingredient that enabled SNIa to achieve precision needed to probe contents of the universe accurately.

3. The third key invention was the development of techniques to correct SN magnitudes for dimming by dust, or 'extinction', out of multi-color observation of SN light. Such corrections are an important part of SN cosmology to this day.

4. Finally, the fourth and perhaps most important ingredient for the discovery of dark energy was development and application of charge-coupled devices (CCDs) in observational astronomy. Both teams of SN hunters used the CCDs, which had originally been installed at telescopes at Kitt Peak and Cerro Tololo.

Some of the early results came out in 1997, which paradoxically indicated that the universe is matter-dominated and consistent with being flat. However the definitive results came out in 1998 (High-z team) and 1999 (SCP, though they had results earlier) and indicated that the universe is dominated by a component with negative pressure.

Observable and inferred quantities with SNIa. The astronomers use apparent magnitudes to measure apparent (measured) brightness of an object. The *apparent* magnitude is given

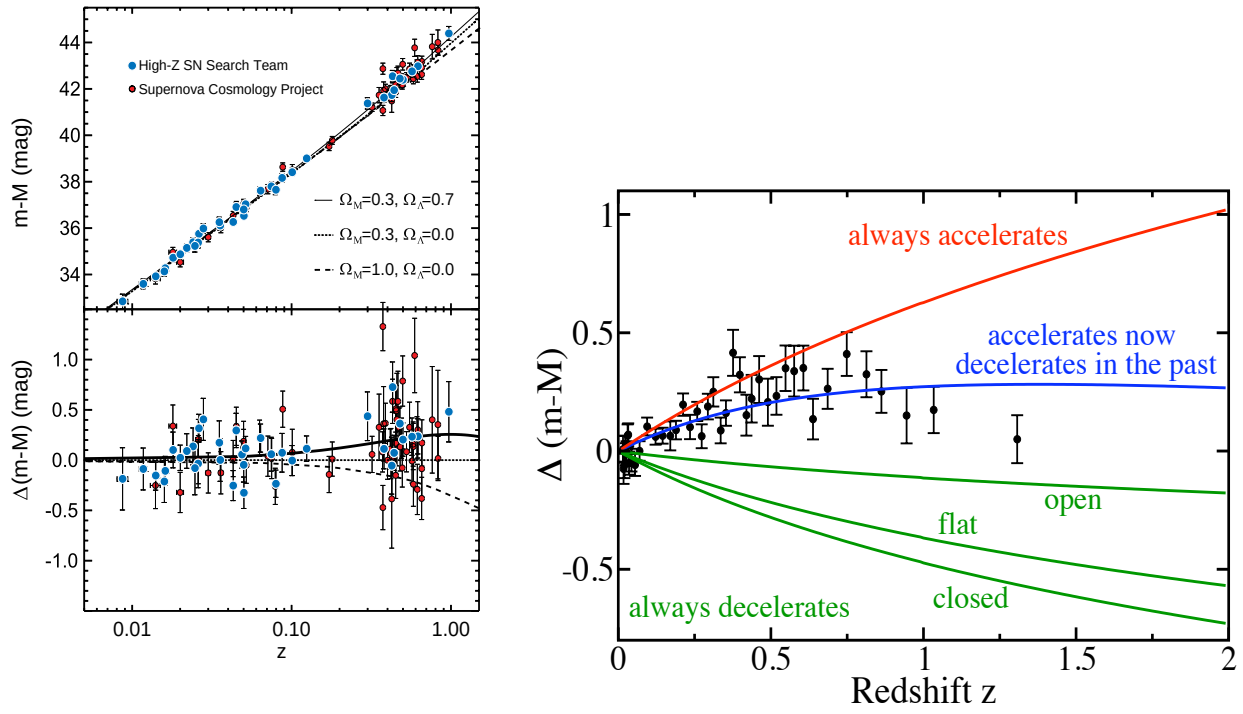

Figure 2: *Left panel:* Hubble diagram of two supernova teams indicating that distant SNIa are dimmer than expected in a matter-dominated universe. *Right panel:* same Hubble diagram, except with newer data, which now has been *binned* (grouped) in redshift for easier viewing. I also show several models to guide the eye; the model favored by the data is of course the "accelerates now, decelerated in the past" model with $\Omega_M \simeq 1 - \Omega_\Lambda \simeq 0.25$. In both cases the y-axis shows the distance modulus relative to some fiducial case (which happens to be the empty universe case, with all Omegas zero, but this is unimportant).

in terms of flux received from an object

$$m \equiv -2.5 \log_{10} \left(\frac{f}{f_x} \right) \tag{3}$$

where f_x is some normalization. Similarly, the *absolute* (intrinsic) magnitude is given by

$$M \equiv -2.5 \log_{10} \left(\frac{L}{L_x} \right) \tag{4}$$

where f_x is some normalization. f_x and L_x are defined so that the apparent and absolute magnitude agree for an object 10 pc away.

The difference between the apparent and absolute magnitude is the *distance modulus*. So

$$\mathrm{DM} \equiv m - M = 2.5 \log_{10} \left(\frac{L}{f} \right) + \mathrm{const} = 5 \log_{10} \left(\frac{d_L}{10\,\mathrm{pc}} \right) \tag{5}$$

where, note, we have taken care of the constants correctly, so that DM = 0 at 10 parsecs.

This equation can be re-written as

$$m = M + 5 \log_{10} \left(\frac{d_L}{1\,\mathrm{Mpc}} \right) + 25 \tag{6}$$

$$= M + 5 \log_{10}(H_0 d_L) - 5 \log_{10} \left(H_0 \times 1\,\mathrm{Mpc} \right) + 25 \tag{7}$$

$$\equiv 5 \log_{10}(H_0 d_L) + \mathcal{M} \tag{8}$$

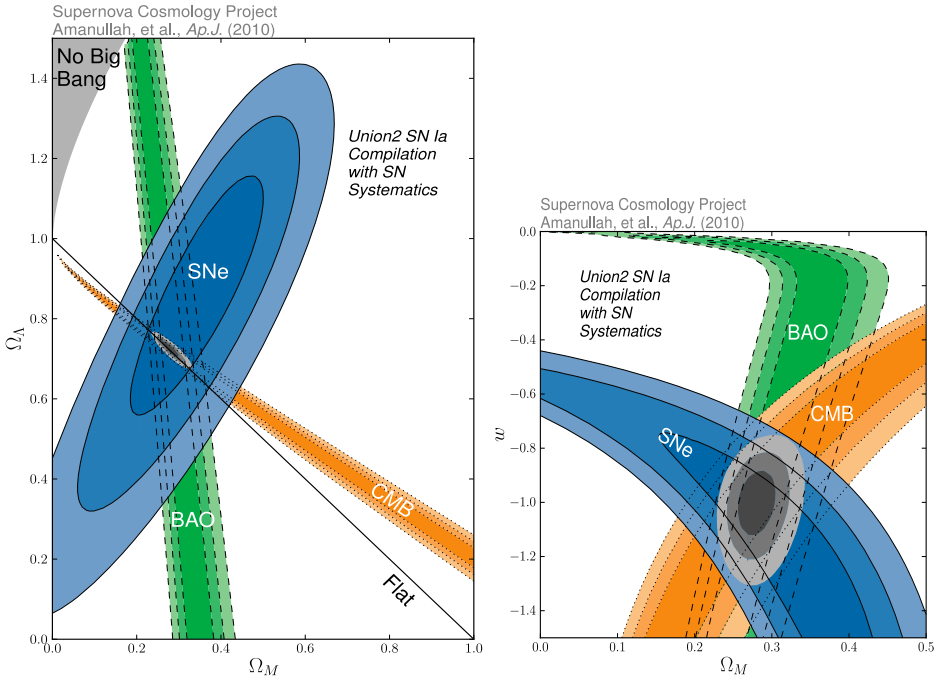

Figure 3: *Left panel*: Constraints upon $\Omega_{\rm M}$ and Ω_Λ in the consensus model using baryon acoustic oscillation (BAO), CMB, and SN measurements. *Right panel*: Constraints upon $\Omega_{\rm M}$ and constant w in the fiducial dark energy model using the same data sets. From Supernova Cosmology Project's Union2 compilation of 557 SN (Amanullah et al., 2010).

where the "script-M" factor is defined as

$$\mathcal{M} \equiv M - 5\log_{10}(H_0 \times 1\,\mathrm{Mpc}) + 25. \tag{9}$$

To summarize

$$\boxed{m = 5\log_{10}(H_0 d_L) + \mathcal{M}} \tag{10}$$

Note that $\mathcal{M}$ is a nuisance parameter that captures *two* uncertain quantities: the absolute magnitude (i.e. intrinsic luminosity) of a supernova, M, and the Hubble constant H_0. We typically do not know $\mathcal{M}$, and we need to marginalize (i.e. integrate) over this parameter in the cosmological analysis.

The situation is now clear: astronomers measure m, which could be peak magnitude of a SNIa. Then they measure the redshift of a SN. With a bunch of SN, they can marginalize over the parameter $\mathcal{M}$ and be left with, effectively, measurements of luminosity distance vs. redshift, $d_L(z)$. A plot of either $m(z)$ or $d_L(z)$ vs. z is called the Hubble diagram.

The discovery of dark energy. Two aforementioned teams, the Highz team and the SCP, published their findings in 1998 and 1999 respectively (the SCP team was a year late since they were notoriously slow with getting papers out; in fact the discoveries had been made around the same time). The results agreed and indicated that more distant SN are dimmer than would be expected in a matter-only universe. In other words, the universe is speeding up its expansion. This was a watershed event in modern cosmology, and these two papers are among the most cited physics papers ever.

These results have been greatly strengthened since, with many hundreds of SN Ia currently indicating same results, but with smaller errors, as in the original 1998-9 papers. Meanwhile, other cosmological probes have come in with results confirming the SN results (see Fig. 3).

Implication: the universe is accelerating. Using the second of the Friedmann equations – the acceleration equation

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) \tag{11}$$

one immediately sees the implication of a new component with $\Omega_{\rm DE} \simeq 0.75$ and $w \simeq -1$: it's that $\ddot{a} > 0$. Maybe it is easiest to rewrite this equation in terms of the *deceleration parameter* q_0:

$$q_0 \equiv -\frac{\ddot{a}a}{\dot{a}^2}\bigg|_{t=t_0} = -\frac{\ddot{a}}{aH^2}\bigg|_{t=t_0} \tag{12}$$

$$= \frac{4\pi G}{3H^2}\sum_i(\rho_i + 3p_i) = \frac{1}{2}\sum_i \Omega_i(1+3w_i). \tag{13}$$

Evaluating this for the familiar case of EdS universe ($\Omega_M = 1$, $w_M = 0$ and nothing else), we get $q_0 = +1/2$, a decelerating universe. Evaluating it, however, for $\Omega_M = 1 - \Omega_{\rm DE} = 0.25$ and $w_{\rm DE} = 1$, we get $q_0 \simeq -0.6$. The deceleration parameter (today) is negative — the universe is accelerating. Hooray!

Descriptions of DE. Theoretical ideas and challenges.

Parametrizations of dark energy: Introduction. The absence of a consensus model for cosmic acceleration presents a challenge in trying to connect theory with observations. For dark energy, the equation-of-state parameter w provides a useful phenomenological description Because it is the ratio of pressure to energy density, it is also closely connected to the underlying physics. However, w is not fundamentally a function of redshift, and if cosmic acceleration is due to new gravitational physics, the motivation for a description in terms of w disappears. On the practical side, determining a free function is more difficult than measuring parameters. We now review a variety of formalisms that have been used to describe and constrain dark energy.

First, let us recall some basics. From continuity equation

$$\dot{\rho} + 3H(p+\rho) = 0, \quad \text{or} \tag{14}$$

$$\frac{d\ln\rho}{d\ln a} = -3(1+w), \tag{15}$$

we can calculate the dark energy density

$$\rho_{\rm DE}(a) = \rho_{\rm DE,0}\,\exp\left(-3\int_1^a (1+w(a'))d\ln a'\right) \tag{16}$$

$$\rho_{\rm DE}(z) = \Omega_{\rm DE}\rho_{\rm crit,0}\,\exp\left(3\int_0^z (1+w(z'))d\ln(1+z')\right). \tag{17}$$

Let us also write possibly the most useful form

$$\boxed{\frac{\rho_{\rm DE}(z)}{\rho_{\rm DE}(0)} = \exp\left(3\int_0^z (1+w(z'))d\ln(1+z')\right).} \tag{18}$$

If the equation of state is constant, we of course get the familiar

$$\frac{\rho_{\rm DE}(z)}{\rho_{\rm DE}(0)} = (1+z)^{3(1+w)} \qquad (w = \text{const}) \tag{19}$$

Parametrizations. The simplest parameterization of dark energy is

$$w = \text{const}. \tag{20}$$

This form fully describes vacuum energy ($w = -1$) or topological defects ($w = -N/3$ with N an integer dimension of the defect – 0 for monopoles, 1 for strings, 2 for domain walls). Together

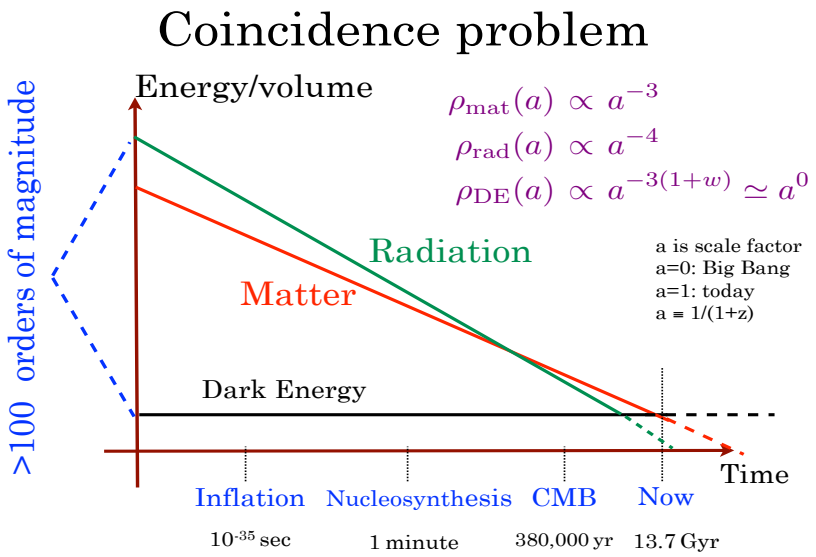

Figure 4: Coincidence problem illustrated: why do we live at an era when the two dominant components, DM and DE, are comparable to each other?

with $\Omega_{\rm DE}$ and $\Omega_{\rm M}$, w provides a 3-parameter description of the dark-energy sector (2 parameters if flatness is assumed). However, it does not describe scalar field or modified gravity models which generically have a *time-varying* w.

A number of two-parameter descriptions of w have been explored in the literature, e.g.,

$$w(z) \quad = \quad w_0 + w'z \tag{21}$$

For low redshift they are all essentially equivalent, but for large z, some lead to unrealistic behavior, e.g., $w \ll -1$ or $\gg 1$. The energy density with these parametrizations can easily be worked out starting with the continuity equation $\dot{\rho} + 3H(p + \rho) = 0$. The parametrization (Linder 2003)

$$\boxed{w(a) = w_0 + w_a(1 - a) = w_0 + w_a z/(1 + z)} \tag{22}$$

avoids this problem and leads to the most commonly used description of dark energy, namely $(\Omega_{\rm DE}, w_0, w_a)$.

The coincidence problem. Why do we live in an era when dark matter to dark energy ratio is of order unity (more precisely, about 1:3)? For $w \simeq -1$, if we lived in the early universe DE would have been completely negligible and unobservable, while in the distant future DE will presumably be completely dominant over DM. This conundrum is sometimes referred to as the "coincidence problem", or the "why now problem" of DE.

Opinions on this problem are divided: some think it is a problem, while others are not impressed, since it is hard to put a metric on being surprised by such an anomaly. What also complicates things is that baryonic matter is also comparable (being about 1/5th of the total matter budget), while radiation is not that far off, being about two magnitudes less important today.

Nevertheless, much effort has gone into trying to build models that naturally 'solve' the coincidence problem. For example, classes of quintessence models have the so-called attractor behavior, where at late times the scalar field's energy density becomes comparable and (with fine-tuning elsewhere, presumably!) comparable to the dominant energy component in the universe – first radiation, and then matter.

Cosmological constant problem. Dark energy could well be energy of vacuum, given that vacuum energy (also known by the greek letter Lambda; for example, ρ_Λ) has $p_\Lambda = -\rho_\Lambda$, or $w = -1$. However, there is a problem.

The energy density required to explain the accelerated expansion is about three quarters of the critical density or about

$$\boxed{\rho_\Lambda \sim \rho_{\rm crit,0} = \frac{3H_0^2}{8\pi G} \sim 10^{-29}\frac{\rm kg}{\rm cm^3} \sim 4\times 10^{-47}\,{\rm GeV}^4 \approx (3\times 10^{-3}{\rm eV})^4}. \tag{23}$$

This (see the last number) is tiny compared to energy scales in particle physics (with the exception of neutrino mass differences). Here is one way to see this. The zero-point energy of a quantum harmonic oscillator is $(1/2)\sqrt{k^2+m^2}$, where m is the mass of the particle and k is its 3-momentum. The total energy density of vacuum quanta is then obtained by summing over all momenta. Ignoring factors of order unity,

$$\rho_\Lambda = \frac{1}{2}\sum_{\rm fields} g_i \int_0^\infty \sqrt{k^2+m^2}\,\frac{d^3k}{(2\pi)^3} \sim \sum_{\rm fields}\frac{g_i k^4}{16\pi^2}\Big|_0^\infty = \infty \tag{24}$$

which diverges quartically ($\sim k^4$) in the ultraviolet. It's infinite, and that's bad!

Of course, infinities are a piece of cake for a seasoned QFT (Quantum Field Theory) using physicist. A better estimate is obtained if one makes a cut-off, corresponding to ignoring the scales at which new physics enters that changes the ultraviolet (high-energy) behavior of the theory. For example, at the energy corresponding to Planck scale, above which new physics has to come. Under those conditions, the expectation is

$$\rho_\Lambda \sim k_{\rm Pl}^4 = m_{\rm Pl}^4 \sim (10^{19}\,{\rm GeV})^4 = 10^{76}\,{\rm GeV}^4 \tag{25}$$

and this is... 120 orders of magnitude greater than the measured results (see Eq. 23). This is the greatest discrepancy between measurement and theory known to man, and is known as the *cosmological constant problem*.

Similar solutions don't work either. For example, a cutoff of 1 TeV (assuming new physics at that scale, perhaps due to supersymmetry) would leave a 58 orders-of-magnitude discrepancy. If supersymmetry were an unbroken symmetry, fermionic and bosonic zero-point contributions would cancel. However, if supersymmetry is broken at a scale of order M, one would expect that imperfect cancellations leave a finite vacuum energy of the order M^4, which for the favored value of $M \sim 100$ GeV to 1 TeV, would leave a discrepancy of 50 or 60 orders-of-magnitude.

The cosmological constant problem is one of the greatest and most famous problems in modern physics. Whoever solves it will become *very* famous. Note that, if the dark energy is *not* due to vacuum energy (or Lambda), then the cosmological constant problem remains, but in a milder form: why is the vacuum energy zero (as opposed to huge value predicted by theory)? If the dark energy is the C.C., then the problem is: why is the vacuum energy *small* (as opposed to huge value predicted by theory)?

Anthropic principle: a sign of desperation. Anthropic principle is very controversial, but is being applied in physics more and more. It basically says: things are the way they are, because if they were different, life as we know it wouldn't exist.

For example, why is Earth about 1 AU distant from the Sun? If it were closer, it would be too hot for life to form. If it were further, it would be too cold. You get the idea... and also why this type of reasoning is controversial.

One approach to the cosmological constant problem involves the idea that the value of the vacuum energy is a random variable which can take on different values in different causally disconnected pieces of the Universe. These corners of the multiverse are apparently predicted by string theory, and correspond to combination of different (generalized) fluxes that make up different vacua. Each vacuum has energy that takes value somewhere between $-{m_{\rm Pl}}^4$ and $+{m_{\rm Pl}}^4$. Because a value much larger than needed to explain the observed cosmic acceleration would preclude the formation of galaxies/entropy (and thus planets, people and Max Planck Institutes), we could not find ourselves in such a region *unless* we live in a patch with vacuum

value reasonably close to zero. [The bound was first worked out by Steven Weinberg, and it's asymmetric around zero and extends to roughly ±10 times the actual observed vacuum energy value.] This argument is sometimes phrased as a *selection effect*, in the spirit of the distance between the Earth and Sun example given above.

Our view? Let us quote a footnote in Kolb & Turner: "It is unclear to one of the authors how a concept so lame [as the anthropic principle] could have been elevated to a status of the principle."

Failure of theoretical models for DE. Theory just has not been keeping up with DE observations. No model to date successfully explains the cosmological constant problem. And dynamical models did make serious attempts to explain the coincidence problem but, as illustrated below, they suffer from their own shortcomings & fine-tunings.

Bottom line: there are no good models for DE.

Quintessence. That said, let us review perhaps the most popular class of dynamical DE models, those that include the evolving *scalar field*, or quintessence. [Original work on these goes back many decades, but the standard references that postulated this in modern setting are Ratra & Peebles 1988, and Wetterich 1988.]

The field is considered to be largely homogeneous, so that its gradient can be ignored – so $\nabla\phi \ll \dot{\phi}$; without this assumption there would be a $\nabla^2\phi/2$ extra term on the left-hand side in the eq below. Thus $\phi = \phi(t)$. By varying the action (and disregarding any couplings to other particles or self-couplings), we get the equation of motion of the field

$$\boxed{\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0} \tag{26}$$

where dots are derivatives with respect to physical time and prime is the derivative with respect to ϕ. This equation describes the evolution in time of the scalar field. It is the same as an equation for a ball rolling down a hill, with friction

$$\ddot{x} + \mu C\dot{x} + V'(x)/m = 0 \tag{27}$$

where x is the coordinate and $V' = d(mgh)/dx = mg\sin(\theta)$ is the change of potential energy per x (θ is the angle that the hill forms with respect to the horizontal).

Moreover, the pressure and energy density of the field are given in terms of its kinetic energy T and potential energy V:

$$\rho_\phi = T + V = \frac{\dot{\phi}^2}{2} + V(\phi) \tag{28}$$

$$p_\phi = T - V = \frac{\dot{\phi}^2}{2} - V(\phi). \tag{29}$$

The equation of state is therefore

$$w = \frac{p_\phi}{\rho_\phi} = \frac{\frac{\dot{\phi}^2}{2} - V(\phi)}{\frac{\dot{\phi}^2}{2} + V(\phi)} \tag{30}$$

Clearly, for a single non-interacting scalar field, $-1 \leq w(t) \leq +1$.

The reason that quintessence models enjoy(ed) a good amount of popularity is their *tracking behavior*. One can show, with some work, that certain classes of potentials – that is of functions $V(\phi)$ – lead to scaling (or, tracking) solutions, where quintessence, at late times, tracks (more or less) the dominant component – first radiation, then matter. In particular, the potentials with tracking capability include exponentials ($V(\phi) \propto \exp(-\lambda\phi)$) and positive and negative power laws in ϕ ($V(\phi) \propto \phi^{\pm n}$).

Note however that some fine-tuning and complication is nevertheless required for two fundamental reasons

- **Can't always track:** A straightforward exponential potential leads to quintessence DE that is *always* proportional to a fixed fraction of the dominant component; to satisfy the early-universe constraints (DE no more than a few percent of the total), DE in such simplest model couldn't be more than a few percent, and couldn't explain the 75% DE today. More complicated models are needed.
- **Mass is tiny:** Just on dimensional grounds from e.g. Eq. (26), the dynamical field's mass has to be comparable to the only mass scale in the (homogeneous) universe today – the Hubble rate H_0. Thus $m_\phi \simeq H_0 \simeq 10^{-33}\,\mathrm{eV}$. This is far below any known mass scales in the Standard Model of particle physics!

CMB and BAO.

CMB and BAO sensitivity to dark energy. We discussed how SNIa measure distances that are sensitive to DE. It turns out that the cosmic microwave background (CMB) anisotropies, and the baryon acoustic oscillations, are also largely measuring cosmmological distances.

CMB sensitivity to DE: peak locations. It turns out that the comoving *typical* separation between the hot and cold spots is given by the sound horizon – distance that sound can travel between the big bang and recombination. Recall that the speed of sound is given by $c_s = (\delta p/\delta\rho)^{1/2}$ in the early epoch when the photons and baryons are strongly coupled in a plasma. The energy density and pressure of the photons are

$$\rho_\gamma = \alpha T^4; \qquad p_\gamma = \frac{1}{3}\alpha T^4 \tag{31}$$

so that the familiar $p_\gamma/\rho_\gamma = 1/3$ obtains. Then

$$c_s = \sqrt{\frac{\delta p_\gamma}{\delta\rho_B + \delta\rho_\gamma}} = \sqrt{\frac{(1/3)\,\delta\rho_\gamma}{\delta\rho_B + \delta\rho_\gamma}} = \frac{1}{\sqrt{3(1+R)}} \tag{32}$$

where $R \equiv \delta\rho_B/\delta\rho_\gamma$. To calculate this last derivative, we assume that the fluctuations are *adiabatic*, that is, that each species density fluctuation separately obeys the continuity equation $\dot\rho_X + 3H(\rho_X + p_X) = 0$. Rearranging we get $\delta\rho_X/[\rho_X(1+w_X)] = -3H\delta t$, where the right-hand side is clearly species-independent, so that, for any two species

$$\frac{\delta\rho_X}{(1+w_X)\rho_X} = \frac{\delta\rho_Y}{(1+w_Y)\rho_Y} \quad \text{(adiabatic condition for any X, Y.)} \tag{33}$$

For baryons and matter we have

$$\frac{\delta\rho_B}{\rho_B} = \frac{\delta\rho_\gamma}{(4/3)\rho_\gamma} \quad \text{(adiabatic condition for photons, baryons)} \tag{34}$$

so that $\delta\rho_B/\delta\rho_\gamma = (3/4)\rho_B/\rho_\gamma$. Therefore,

$$R \equiv \frac{\delta\rho_B}{\delta\rho_\gamma} = \frac{3\rho_B}{4\rho_\gamma} = \frac{3\Omega_B}{4\Omega_\gamma}\,a \tag{35}$$

which, note, scales as a, or $(1+z)^{-1}$. Therefore, the speed of sound is $1/\sqrt{3}$ of the speed of light, corrected for the non-negligible presence of baryons. The sound horizon is just defined as the coordinate distance $r_s = \int dt/a(t) = \int da/(a^2 H(a))$, or

$$r_s = \frac{c}{\sqrt{3}}\int_0^{a_*} \frac{da}{a^2 H(a)\sqrt{1+\dfrac{3\Omega_B}{4\Omega_\gamma}\,a}} \simeq 100\,\mathrm{h}^{-1}\mathrm{Mpc} \quad \text{(sound horizon)} \tag{36}$$

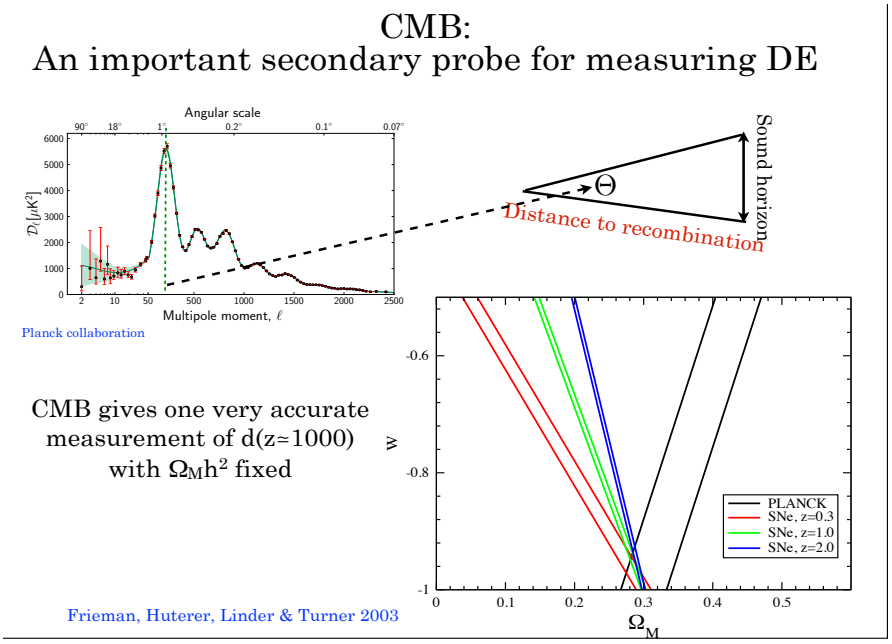

Figure 5: Illustration of how CMB measurements of (a) the peak in the angular power spectrum, and (b) the sound horizon determine the distance to recombination very accurately. As discussed in the text, this distance has $\Omega_M h^2$ fixed, and is thus complementary to other measurements, as illustrated in the bottom plot showing degeneracy directions in the $\Omega_M - w$ plane.

where a_* is the scale factor at recombination, and where the latter value is the predicted value from the standard ΛCDM model given cosmological parameter constraints.

Now the sound horizon determines the characteristic distance between the CMB hot and cold spots (we have not proven this; see e.g. Dodelson book for a lengthy derivation). From our vantage point, we observe the sound horizon as the "standard ruler" at the distance corresponding to recombination, $d_A(z_*)$. The angle of the first acoustic peak in the CMB power spectrum corresponds to this angle at which we observe the sound horizon. Therefore

$$\boxed{\ell_{\rm peak} \simeq \frac{\pi}{\theta_{\rm peak}} = \pi \frac{d_A(z_*)}{r_s}.} \tag{37}$$

The peak location can be determined to a fabulously good accuracy (small fraction of a percent) from the CMB data. The sound horizon can also be well determined from the CMB. Therefore, we get a very good measure of the angular diameter distance to recombination.

However, this is not just any distance – it is the distance *with the physical matter density* $\Omega_M h^2$ *nearly fixed*. The latter parameter is measured from the morphology of the CMB peaks. Therefore, the CMB really measures the so-called shift parameter $R_{\rm shift}$

$$\boxed{R_{\rm shift} \equiv \sqrt{\Omega_M h^2}\, d_A(z_*)}. \tag{38}$$

Figure 5 illustrates how the shift parameter information is actually very complementary to distance measurements at *any* redshift, since it traces out a different parameter degeneracy direction – e.g. a different direction in the $\Omega_M - w$ plane (assuming a flat universe in this example).

CMB sensitivity to DE: Integrated Sachs-Wolfe Effect. Large-angle CMB anisotropy has several contributions, the principal one being the Sachs-Wolfe effect, which is the intrinsic anisotropy at the last-scattering surface that is more than offset by photons from hotter areas climbing out the potential wells. See Fig. 6 for illustration.

A particularly important, though smallish, contribution to Sachs-Wolfe is the Integrated Sachs-Wolfe effect (ISW). This is an additional contribution that comes about from the *decay of*

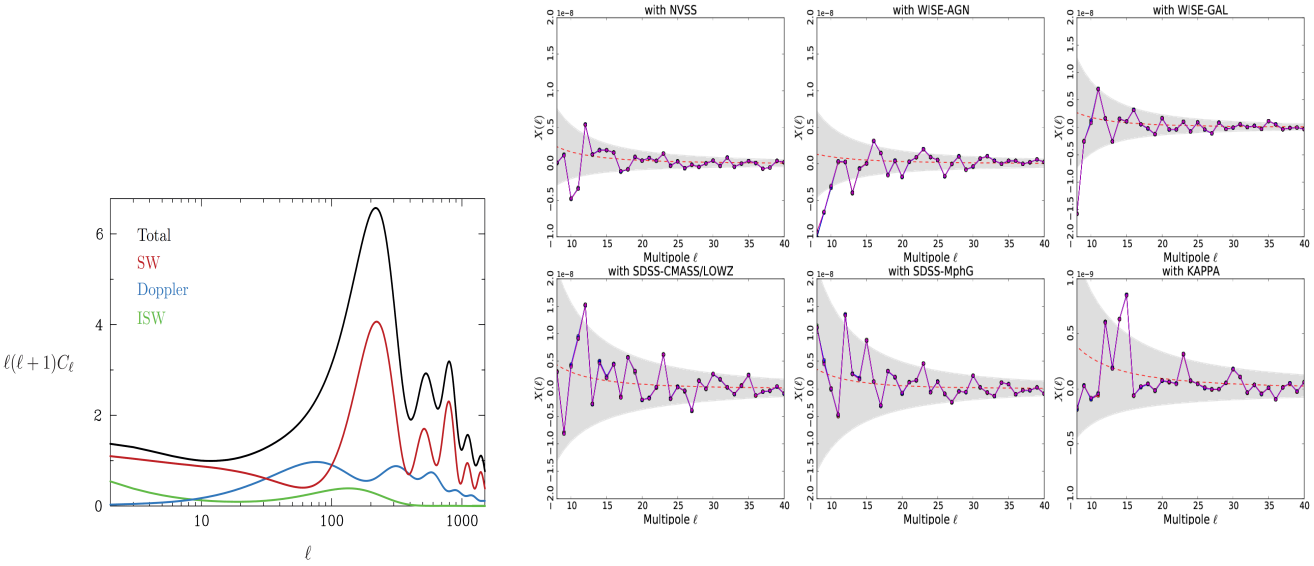

Figure 6: *Left panel:* Contributions to the CMB angular power spectrum, including the ISW effect. *Right panel:* Planck 2015 measurements of ISW via cross-correlations of CMB maps with galaxy surveys (arXiv:1502.01595). The red dashed line is a theoretical expectation given the depth of the galaxy survey, while the grey regions is the expected 68% cosmic variance error around the mean, and is obviously much larger than the measurement errors.

gravitational potential when the universe is not matter-dominated. A small contribution to ISW is at early times around recombination, when radiation still plays a small ($\sim 10\%$) role. A more important for us contribution is at late times, when dark energy takes over and the potential decays. The expression for the total CMB anisotropy at large scales, only taking into SW and ISW contributions for the moment, is

$$\frac{\delta T}{T} = \left(\frac{\delta T}{T}\right)_{\rm SW} + \left(\frac{\delta T}{T}\right)_{\rm ISW} \tag{39}$$

$$= -\frac{1}{3}\Phi_* + 2\int_{\tau_*}^{\tau_0} \frac{d\Phi}{d\tau}\, d\tau. \tag{40}$$

Because DE becomes important late (redshift $\lesssim 1$), the ISW leaves an imprint at very large angular scales, corresponding to the horizon scale at these late times. At these scales (multipoles $\lesssim 10$), the CMB cosmic variance is large, and the ISW signature is buried and cannot be detected more than at $\sim 1\sigma$ or so.

A much more promising way of detecting and utilizing the ISW is through *cross-correlations with galaxy surveys* (Crittenden & Turok 1996). In other words, one measures the cross-correlation between the CMB map and map of large-scale structure, which in real space looks like

$$C_{Tg}(\theta) \equiv \left\langle \frac{\delta T}{T}(\hat{n}) \frac{\delta n}{n}(\hat{n}') \right\rangle_{\hat{n}\cdot\hat{n}'=\cos\theta} \tag{41}$$

where T and n are the CMB temperature and galaxy number fluctuations, respectively. It turns out that this quantity is much more sensitive to the ISW signal. First detection using this methodology has been achieved about a decade ago (Boughn and Crittenden 2004). Current constraints from Planck are shown in the right panel of Fig. 6. Despite being wonderful as an independent (non-geometric) proof of the existence of DE and interesting probe of modified gravity models, one should note that the ISW detection using these cross-corr methods cannot be made at higher than something like $\sim 10\sigma$, even using futuristic Planck+LSST type surveys (Hu & Scranton 2004).

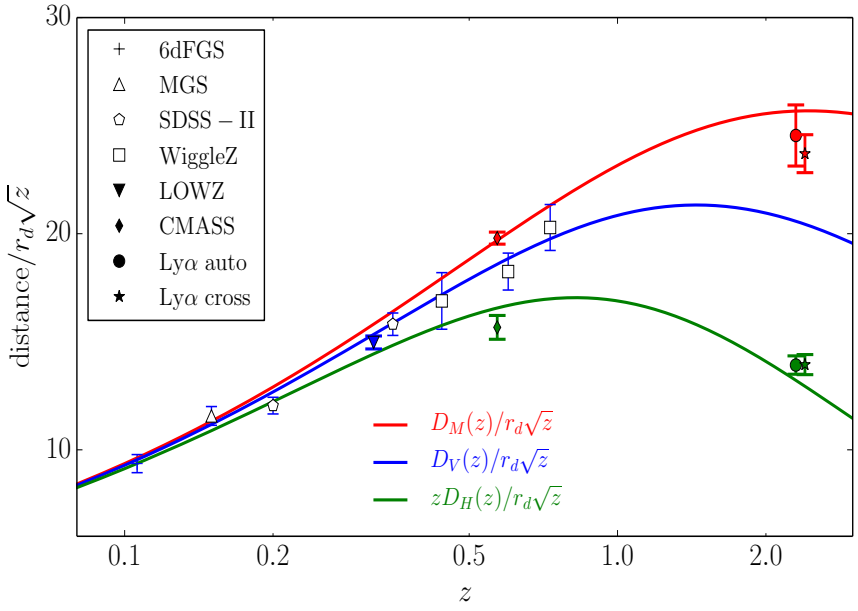

Figure 7: Constraints on the distance (to galaxies in corresponding survey samples) using the BAO. Here D_V is the combined 3D distance defined in our Eq. (42), D_M is just the comoving distance (r in our notation), while $D_H(z) \equiv 1/H(z)$. Adopted from Aubourg et al., arXiv:1411.1074.

Baryon acoustic oscillations (BAO) sensitivity to DE. Power spectrum of density perturbations in dark matter, $P(k)$, is mainly sensitive to the density in matter (relative to critical), Ω_M. We will discuss the full, "broad-band" power spectrum further below. However one particularly clean signature in the power spectrum are the baryon acoustic oscillations (BAO), which are sensitive to dark energy in a very similar way as the CMB peak locations – in fact, physics is the same.

The BAO probe a specific feature in galaxy clustering at a typical redshift of the survey (say, $z \sim 0.3$ and 0.6 for the two SDSS/BOSS survey samples). They encode a $\sim 10\%$ additional probability for galaxies being $\sim 100\,\mathrm{h}^{-1}$Mpc comoving away, where the latter distance is of course the comoving sound horizon. We essentially observe the angle at which the sound horizon is seen. Therefore, the BAO measure the angular diameter distance, $d_A(z)$, at some redshift z, in units of course of the sound horizon.

However, unlike in the CMB case, one can observe galaxy clustering not only in the plane-parallel (transverse) direction, but also in the direction perpendicular to the line of sight (the radial direction). Specifically, for two galaxies at the same redshift separated by comoving distance Δr and seen with separation angle $\Delta\theta$, we have $\Delta\theta = \Delta r/D_A(z)$ which enables measurement of the angular diameter distance given known separation between galaxies. Similarly, two galaxies at the same angular location but separated by redshift difference Δz are separated by comoving distance Δr, where $\Delta r = \Delta z/H(z)$. The information from these transverse and radial sensitivities can be conveniently combined into a single quantity, a generalized distance $D_V(z_{\rm eff})$ defined as (Eisenstein et al. 2005)

$$D_V(z) \equiv \left(\frac{(1+z)^2 d_A^2(z) cz}{H(z)} \right)^{1/3} . \qquad (42)$$

The BAO surveys measure $r_s/D_V(z_{\rm eff})$ (or its inverse), where r_s is the comoving sound horizon; see Fig. 7.

Key to successful application of baryon acoustic oscillations are redshift measurements of galaxies in the sample. You need the galaxy redshift in order to know at where radially to "put it", and thus to reconstruct the baryonic oscillations without bias.

Growth of cosmic structure.
Galaxy clustering, weak lensing and cluster counts.

Other than its effect on the distances (basically, making higher-z objects more distant than in a matter-only universe), dark energy has another very important effect on cosmology:

Dark Energy slows down the growth of cosmic structure.

To see this, consider the equation for the *linear* growth of density fluctuations:

$$\ddot{D} + 2H\dot{D} - 4\pi G\rho_M D = 0 \tag{43}$$

where $D(a)$ is the linear growth of density fluctuations, defined as (with $\delta \equiv \delta\rho_M/\rho_M$)

$$D(a) = \frac{\delta(a)}{\delta(a=1)}. \tag{44}$$

Note that, in linear theory (and assuming GR), the linear growth depends on cosmic time only and not scale – making things much simpler! Converting from proper time to scale factor, one easily derives a coding-friendly equivalent expression

$$\boxed{2\frac{d^2g}{d\ln a^2} + [5 - 3w(a)\Omega_{\rm DE}(a)]\,\frac{dg}{d\ln a} + 3\,[1 - w(a)]\,\Omega_{\rm DE}(a)g = 0} \tag{45}$$

where the *growth suppression* factor $g(a)$ is implicitly defined via

$$D(a) \equiv \frac{ag(a)}{g(1)}, \tag{46}$$

so that $D(1) = 1$. Here $g(a)$ is the growth rate relative to that in an EdS (flat, $\Omega_M = 1$) universe. For EdS case, $g(a) = 1$ by definition and at all times. For the ΛCDM universe, $g(a)$ is unity until DE domination, at which point it dips and reaches $g(a = 1) \simeq 0.76$. Figure 8 shows the linear growth (more precisely, our $D(a)g(1)$) for ΛCDM and EdS universes.

In non-linear theory, growth of cosmic structure is scale-dependent (with or without DE), and has to be calibrated in simulations. And of course, in modified theories of gravity, growth of structure may be scale-dependent even in linear theory!

Using growth to test for modified gravity. The Friedmann equation for a flat universe with dark matter and dark energy looks like

$$H^2 = \frac{8\pi G}{3}\left(\rho_m(z) + \rho_{\rm DE}(z)\right) \qquad \text{(DE with GR)} \tag{47}$$

Modified gravity explanations for the accelerating universe postulate that there is no DE fluid (just matter and radiation), but that the equations of GR are instead somehow modified. The equivalent MG Friedmann equation would look something like

$$H^2 - F(H) = \frac{8\pi G}{3}\rho_m(z) \qquad \text{(modified gravity)} \tag{48}$$

where $F(H(z)) \equiv (8\pi G/3)\rho_{\rm DE}(z)$. Note that, even with infinitely good measurements of $H(z)$, we cannot distinguish between these two mathematically equivalent possibilities.

However, by measuring the growth of structure, the separation is possible. For a given, say measured, expansion history $H(z)$, the growth (e.g. linear)

$$\ddot{D} + 2H\dot{D} - 4\pi G\rho_M D = 0 \tag{49}$$

is predicted given GR (one also needs measurements of total matter density $\rho_M(z)$). Then one can go on and *directly measure* the growth of structure and verify whether it satisfies this equation. If it does not, then in principle one has ruled out the assumption of standard GR.

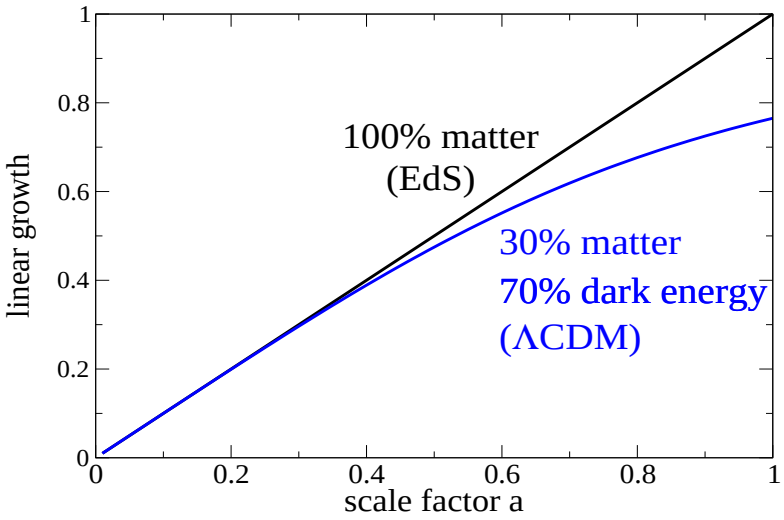

Figure 8: Linear growth in the currently favored ΛCDM model, as well as in the flat, matter-dominated Einstein-de-Sitter model. Note that the quantity plotted is actually $D(a)g(a)$, where $D(a)$ is the familiar linear growth, and $g(a)$ is the growth suppression factor.

Practical ways to use growth vary, but they mostly rely on the intuitive picture just described.

The broadband power spectrum. Intuitively, the most straightforward way to measure the growth of structure is to observe the evolution of the full, 'broadband' power spectrum $P(k, z)$ with cosmic time. In practice, this goal is complicated by the presence of the galaxy bias that has to be somehow concurrently determined as a function of scale *and* time.

Coding (that is, computer programming) the power spectrum is very useful. It is sometimes easiest to work with the dimensionless form of the power spectrum which is defined as

$$\Delta^2(k, a) \equiv \frac{k^3 P(k, a)}{2\pi^2} \tag{50}$$

which one can show to be the contribution to variance of density perturbations per log wavenumber. Without proof, we present the final formula for the power spectrum of dark matter density perturbations in standard FRW cosmology:

$$\boxed{\Delta^2(k, a) = A \, \frac{4}{25} \frac{1}{\Omega_M^2} \left(\frac{k}{k_{\rm piv}}\right)^{n-1} \left(\frac{k}{H_0}\right)^4 [ag(a)]^2 \, T^2(k) \, T_{\rm nl}(k, a)} \tag{51}$$

notice that $\Delta^2 \propto k^{n+3}$, and thus $P(k) \propto k^n$, was predicted by Harrison, Zeldovich and Peebles in the late 1960s (well before inflation *predicted* that $n \simeq 1$!). In this equation:

- A is the normalization of the power spectrum (for the concordance cosmology, $A \simeq 2.3 \times 10^{-9}$)
- $k_{\rm piv}$ is the "pivot" around which we compute the spectral index; for Planck and WMAP, the pivot confusingly can be either $k_{\rm piv} = 0.05\,{\rm Mpc}^{-1}$ or $k_{\rm piv} = 0.002\,{\rm Mpc}^{-1}$, depending on the paper. Also beware — this pivot is defined in ${\rm Mpc}^{-1}$, *not* $\rm h\,Mpc^{-1}$.
- $ag(a)$ is the linear growth of perturbations defined in Eq. (46)
- $T(k)$ is the transfer function: you can use fits (e.g. Hu & Eisenstein, 1997) or else the exact output out of computer codes that solve the coupled Einstein-Boltzmann equations, e.g. CAMB (`http://cosmologist.info`) or CLASS (`http://class-code.net/`)

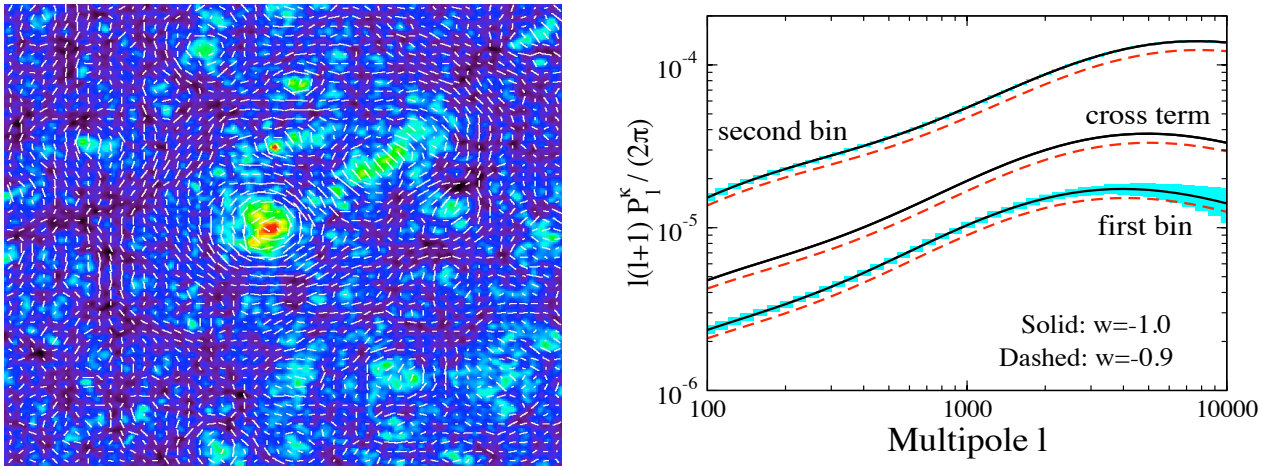

Figure 9: *Left panel:* Cosmic shear field (white ticks) superimposed on the projected mass distribution from a cosmological N-body simulation: overdense regions are bright, underdense regions are dark. Note how the shear field is correlated with the foreground mass distribution. Figure courtesy of T. Hamana. *Right panel:* Cosmic shear angular power spectrum and statistical errors expected for LSST for $w = -1$ and -0.9. For illustration, results are shown for source galaxies in two broad redshift bins, $z_s = 0-1$ (first bin) and $z_s = 1-3$ (second bin); the cross-power spectrum between the two bins (cross term) is shown without the statistical errors.

- $T_{\rm nl}$ is prescription for a *nonlinear* power spectrum. The nonlinearities are important, for example, today at scales $k \gtrsim 0.2\,{\rm h\,Mpc^{-1}}$. The nonlinearities add power in a complicated way, and are calibrated from numerical simulations, and given to theorists as fitting formulae. The most popular fitting formula is the 'halofit' prescription, motivated by the halo model and calibrated from simulations (R. Smith et al. 2003).

Equations (50) and (45) enable you to calculate the power spectrum of density perturbations at any redshift and any scale, and for any cosmological model.

Weak gravitational lensing. The gravitational bending of light by structures in the Universe distorts or shears the images of distant galaxies; see the left panel of Fig. 9. This distortion allows the distribution of dark matter and its evolution with time to be measured, thereby probing the influence of dark energy on the growth of structure (for detailed review, see e.g. Bartelmann & Schneider 2001). **The principal promise of weak lensing is in its complete independence of the galaxy-to-mass bias**.

Gravitational lensing produces distortions of images of background galaxies. These distortions can be described as mapping between the source plane (S) and image plane (I)

$$\delta x_i^S = A_{ij}\delta x_j^I \tag{52}$$

where $\delta\mathbf{x}$ are the displacement vectors in the two planes and A is the distortion matrix

$$A = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}. \tag{53}$$

The deformation is described by the convergence κ and complex shear (γ_1, γ_2); the total shear is defined as $|\gamma| = \sqrt{\gamma_1^2+\gamma_2^2}$. We are interested in the weak lensing limit, where κ, $|\gamma| \ll 1$. Magnification is also given in terms of κ and $\gamma_{1,2}$ as

$$\mu = \frac{1}{|1-\kappa|^2 - |\gamma|^2} \approx 1 + 2\kappa + O(\kappa^2, \gamma^2) \tag{54}$$

where the second approximate relation holds again in the weak lensing limit.

But how do you theoretically predict convergence and shear, given some source galaxies and the foreground distribution on the sky? The convergence in any particular direction on the sky $\hat{\mathbf{n}}$ is given by the integral along the line-of-sight

$$\kappa(\hat{\mathbf{n}}, \chi) = \int_0^\chi W(\chi')\, \delta(\chi')\, d\chi' \tag{55}$$

where δ is the relative perturbation in matter energy density and

$$W(\chi) = \frac{3}{2}\, \Omega_M\, H_0^2\, g(\chi)\, (1+z) \tag{56}$$

is the weight function that "assigns" lensing efficiency to foreground galaxies. Furthermore

$$g(\chi) = r(\chi) \int_\chi^\infty d\chi' n(\chi') \frac{r(\chi' - \chi)}{r(\chi')} \longrightarrow \frac{r(\chi) r(\chi_s - \chi)}{r(\chi_s)} \tag{57}$$

where $n(\chi)$ is the distribution of source galaxies in redshift (normalized so that $\int dz\, n(z) = 1$) and the expression after the arrow holds only if all sources are at a single redshift z_s. The function g peaks about halfway between the observer and the source (that is, at $\chi \sim \chi_s/2$. Therefore, *the most efficient lenses lie about half-way between us and the source galaxies.*

The statistical signal due to gravitational lensing by large-scale structure is termed "cosmic shear." The cosmic shear field at a point in the sky is estimated by locally averaging the shapes of large numbers of distant galaxies. The primary statistical measure of the cosmic shear is the shear angular power spectrum measured as a function of source-galaxy redshift z_s. (Additional information is obtained by measuring the correlations between shears at different redshifts or with foreground lensing galaxies.)

You can typically measure galaxy shear (from the galaxy shapes), but you can theoretically predict the convergence. Fortunately, in the weak lensing limit, convergence and shear are equal.

The convergence can be transformed into multipole space (e.g. Bartelmann & Schneider 2001)

$$\kappa_{lm} = \int d\hat{\mathbf{n}}\, \kappa(\hat{\mathbf{n}}, \chi)\, Y^*_{lm}(\hat{\mathbf{n}}), \tag{58}$$

and the power spectrum is defined as the two-point correlation function (of convergence, in this case)

$$\langle \kappa_{\ell m} \kappa_{\ell' m'} \rangle = \delta_{\ell\ell'}\, \delta_{mm'}\, P^\kappa_\ell. \tag{59}$$

The angular power spectrum is

$$\boxed{P^\gamma_\ell(z_s) \simeq P^\kappa_\ell(z_s) = \int_0^{z_s} \frac{dz}{H(z) d_A^2(z)} W(z)^2 P\left(k = \frac{\ell}{d_A(z)}; z\right)}, \tag{60}$$

where ℓ denotes the angular multipole, the weight function $W(z)$ is the efficiency for lensing a population of source galaxies and is determined by the distance distributions of the source and lens galaxies, and $P(k, z)$ is the usual power spectrum of density perturbations. Notice *integral* along the line of sight — essentially, weak lensing sums up contributions of lenses along the line of sight between us and the source galaxy.

The dark-energy sensitivity of the shear angular power spectrum comes from two factors:

- *geometry*—the Hubble parameter, the angular-diameter distance, and the weight function= $W(z)$; and
- *growth of structure*—through the redshift evolution of the power spectrum of density perturbations (really, function $G(a)$ from Eq. (51)).

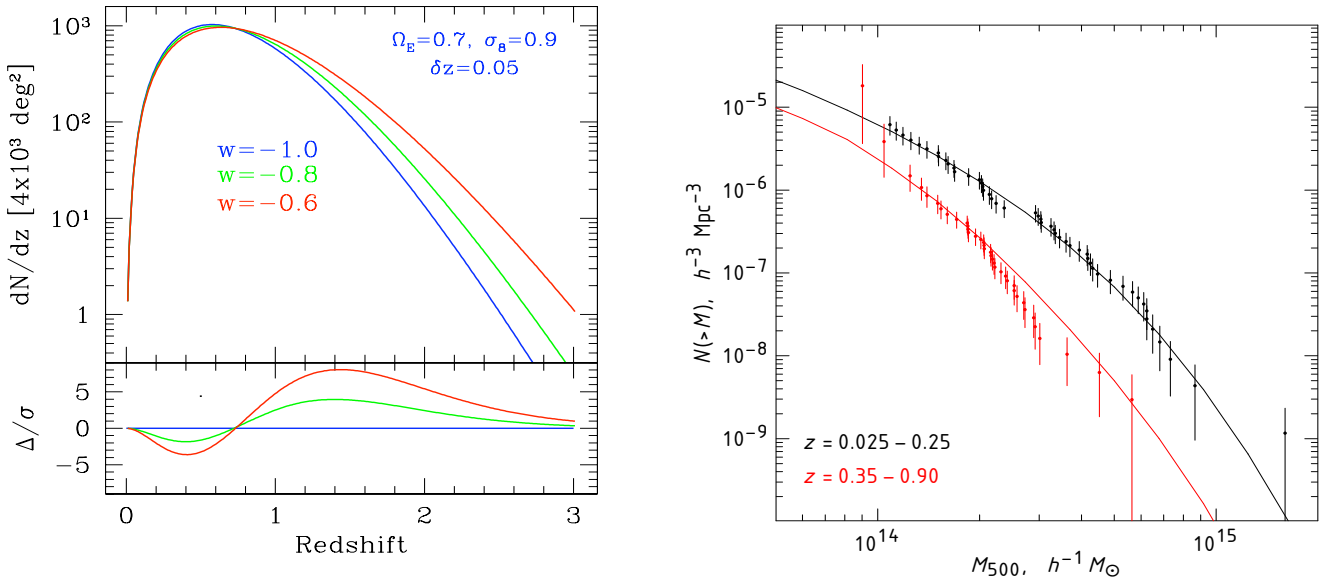

Figure 10: *Left panel:* Predicted cluster counts for a survey covering 4,000 sq. deg. that is sensitive to halos more massive than $2 \times 10^{14} M_\odot$, for 3 flat cosmological models with fixed $\Omega_{\rm M} = 0.3$ and $\sigma_8 = 0.9$. Lower panel shows fractional differences between the models in terms of the estimated Poisson errors. From Mohr (2005). *Right panel:* Measured mass function – $n(z, M_{\rm min}(z))$ – in our notation – from the 400 square degree survey of ROSAT clusters followed up by Chandra. Adopted from Vikhlinin et al. (2009).

It is also possible to use the *three*-point correlation function of cosmic shear is also sensitive to dark energy (power spectrum is the *two*-point correlation function.

The *statistical* uncertainty in measuring the shear power spectrum on large scales is

$$\Delta P_\ell^\gamma = \sqrt{\frac{2}{(2\ell+1) f_{\rm sky}}} \left[P_\ell^\gamma + \frac{\sigma^2(\gamma_i)}{n_{\rm eff}} \right] \quad , \tag{61}$$

where $f_{\rm sky}$ is the fraction of sky area covered by the survey (that is, $f_{\rm sky} = 0.5$ for half-sky, etc), $\sigma^2(\gamma_i)$ is the variance in a single component of the (two-component) shear (this number is ~ 0.2 for typical measurements), and $n_{\rm eff}$ is the effective number density per steradian of galaxies with well-measured shapes.

The first term in brackets dominates on *large scales*, and comes from *cosmic variance* of the mass distribution. The second term dominates on *small scales*, and represents the shot-noise from both the variance in galaxy ellipticities ("shape noise") combined with a finite number of galaxies (hence inverse proportionality to $n_{\rm eff}$).

The right panel of Fig. 9 shows the dependence on the dark energy of the shear power spectrum and an indication of the statistical errors expected for a survey such as LSST, assuming a survey area of 15,000 sq. deg. and effective source galaxy density of $n_{\rm eff} = 30$ galaxies per sq. arcmin. Current survey cover more modest hundreds of square degrees, with a comparable or slightly lower galaxy density. Note that the proportionality of statistical errors to $f_{\rm sky}^{-1/2}$ means that large sky coverage is at a premium, although addressing the systematic errors often drives you to a deeper (and not wider) survey.

Clusters of galaxies. Galaxy clusters are the largest virialized objects in the Universe. Therefore, not only can they be observed, but also their number density can be *predicted* quite reliably, both analytically, and from numerical simulations. Comparing these predictions to large-area cluster surveys that extend to high redshift ($z \gtrsim 1$) can provide precise constraints on the cosmic expansion history.

The absolute number of clusters in a survey of size $\Omega_{\rm survey}$ centered at redshift z and in the shell of thickness Δz is given by

$$N(z,\Delta z) = \Omega_{\rm survey} \int_{z-\Delta z/2}^{z+\Delta z/2} n(z, M_{\rm min}(z)) \, \frac{dV(z)}{d\Omega \, dz} \, dz. \tag{62}$$

where $M_{\rm min}$ is the minimal mass of clusters in the survey (it's usually of order $10^{14} M_\odot$). Note that knowledge of the minimal mass is extremely important, since the "mass function" $n(z, M_{\rm min}(z))$ is *exponentially* decreasing with M, so that most of the contribution comes from a small range of masses just above $M_{\rm min}$. The mass function is key to theoretical predictions, and it is usually obtained from a combination of analytic and numerical results; famous mass functions used in cosmology are the Press-Schechter mass function that dates back to 1970s, or the more recent Sheth-Tormen mass function (these are basically just different fitting function). The volume element can easily be related to comoving distance $r(z)$ and the expansion rate $H(z)$ via

$$\frac{dV(z)}{d\Omega \, dz} = \frac{r^2(z)}{H(z)} \tag{63}$$

and it is basically known exactly for a fixed cosmological model, with no theoretical uncertainty (unlike the mass function which is usually known to a few percent at best, at a given M and z).

The sensitivity of cluster counts to dark energy arises – as in the case of weak lensing – from two factors:

- *geometry*, the term $dV(z)/(d\Omega \, dz)$ in Eq. (62) is the comoving volume element
- *growth of structure*, $n(z, M_{\rm min}(z))$ depends on the evolution of density perturbations, cf. Eq. (43).

This last point is worth emphasizing further. Fitting functions for the cluster mass function (e.g. Press-Schechter) relate it to the primordial spectrum of density perturbations. The mass function's near-exponential dependence upon the power spectrum is at the root of the power of clusters to probe dark energy. In particular, the mass function explicitly depends on the *amplitude of mass fluctuations* smoothed on some scale R

$$\sigma^2(R) = \int_0^\infty \Delta^2_{\rm linear}(k) \left(\frac{3 j_1(kR)}{kR} \right)^2 d\ln k \tag{64}$$

where R is traditionally taken to be $\sim 8\,{\rm h}^{-1}{\rm Mpc}$, the characteristic size of galaxy clusters. Here of course $\Delta^2(k)$ is our dear power spectrum from Eq. (50), here evaluated in linear theory.

The left panel of Fig. 10 shows the sensitivity to the dark energy equation-of-state parameter of the expected cluster counts for the South Pole Telescope and the Dark Energy Survey. At modest redshift, $z < 0.6$, the differences are dominated by the volume element; at higher redshift, the counts are most sensitive to the growth rate of perturbations.

Advanced topics: characterizing expansion history.

Pivot w and pivot redshift. The two-parameter descriptions of $w(z)$ that are linear in the parameters entail the existence of a "pivot" redshift z_p at which the measurements of the two parameters are uncorrelated and the error in $w_p \equiv w(z_p)$ reaches a minimum; see the left panel of Fig. 11.

Here is how you compute the pivot redshift (assuming, for example, the (w_0, w_a) parametrization). Note that we are looking for the equation of state of the form as in Eq. (22):

$$w(a) = w_p + (a_p - a) w_a = w_p + \frac{z_p}{1+z_p} w_a \tag{65}$$

Then you proceed as follows:

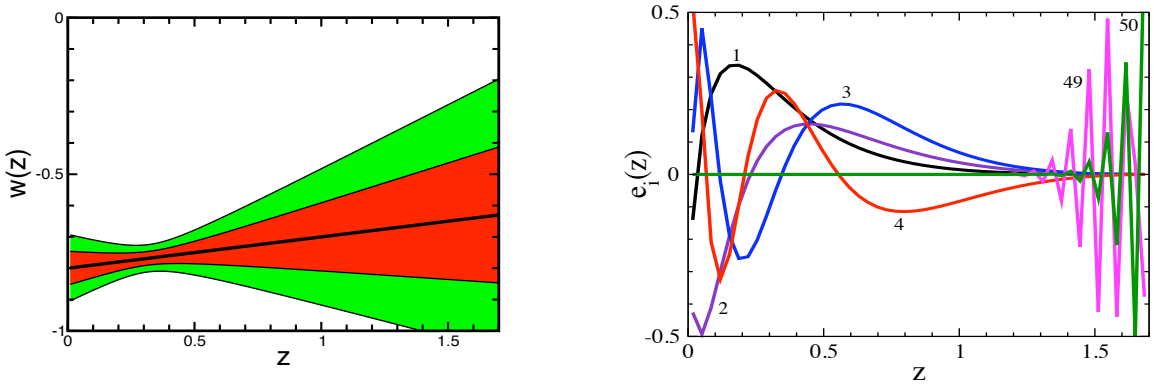

Figure 11: *Left panel:* Example of forecast constraints on $w(z)$, assuming $w(z) = w_0 + w'z$. The "pivot" redshift, $z_p \simeq 0.3$, is where $w(z)$ is best determined. Adopted from Huterer & Turner (2001). *Right panel:* The four best-determined (labeled $1-4$) and two worst-determined (labeled $49, 50$) principal components of $w(z)$ for a future SNIa survey such as SNAP, with several thousand SN in the redshift range $z = 0$ to $z = 1.7$. Adopted from Huterer & Starkman (2003).

- Compute the covariance matrix for all parameters (in the Fisher matrix formalism, this is just the inverse of the Fisher matrix). For two parameters p_i and p_j, the covariance is defined as
$$\mathrm{Cov}(p_i, p_j) \equiv \langle (p_i - \bar{p_i})(p_j - \bar{p_j}) \rangle \tag{66}$$
where this is just the likelihood-weighted average over all values of (p_i, p_j) and the overbar indicates the *mean* value of that parameter from the data.

- Marginalize (that is, integrate the likelihood) over all other parameters, leaving only the 2×2 matrix on w_0 and w_a

- Assuming $w_p = w_0 - (a_p - 1)w_a$, we have $0 = \mathrm{Cov}(w_p, w_a) = \mathrm{Cov}(w_0, w_a) - (a_p - 1)\mathrm{Cov}(w_a, w_a)$, implying that $a_p - 1 = \mathrm{Cov}(w_0, w_a)/\mathrm{Cov}(w_a, w_a)$. Therefore
$$a_p = 1 + \mathrm{Cov}(w_0, w_a)/\mathrm{Cov}(w_a, w_a) \tag{67}$$

- Finally, you can compute the error in w_p as (using in second line Eq. 67)
$$\mathrm{Cov}(w_p, w_p) = \mathrm{Cov}(w_0, w_0) - 2(a_p - 1)\mathrm{Cov}(w_0, w_a) + (a_p - 1)^2 \mathrm{Cov}(w_a, w_a) \tag{68}$$
$$= \mathrm{Cov}(w_0, w_0) - \frac{\mathrm{Cov}(w_0, w_a)^2}{\mathrm{Cov}(w_a, w_a)} \tag{69}$$

The redshift of this sweet spot varies with the cosmological probe and survey specifications; for example, for current SNIa surveys $z_p \approx 0.25$, while for weak lensing and baryon acoustic oscillations $z_p \sim 0.5$-0.7 because of the way in which dark energy enters the physical quantities featured in these probes.

Note that forecast constraints for a particular experiment on w_p are numerically equivalent to constraints one would derive on constant w. This is because almost all information is concentrated in w_p (which was optimally chosen precisely using that criterion).

Direct reconstruction. Another approach is to directly invert the redshift-distance relation $r(z)$ measured from SN data to obtain the redshift dependence of $w(z)$ in terms of the coordinate distance $r(z)$ that is measured, for example, by SN Ia. [SNIa, of course, measure the luminosity

distance $d_L(z)$, but in a flat universe the two are only offset by a factor of $1+z$.] Note that, since $r(z) = \int dz/H(z)$, we have

$$H(z) = \frac{1}{(dr/dz)} \tag{70}$$

Furthermore, it is easy to reconstruct the dark energy density, and the result is

$$\boxed{\frac{\rho_{\rm DE}(z)}{\rho_{\rm DE}(0)} = \frac{1}{1-\Omega_M}\left[\frac{1}{(d(H_0 r)/dz)^2} - \Omega_M(1+z)^3\right]} \tag{71}$$

As an exercise, you can show that the equation of state can also be reconstructed as

$$\boxed{1 + w(z) = \frac{1+z}{3}\,\frac{3H_0^2\Omega_{\rm M}(1+z)^2 + 2(d^2r/dz^2)/(dr/dz)^3}{H_0^2\Omega_{\rm M}(1+z)^3 - (dr/dz)^{-2}}\ .} \tag{72}$$

Note that H_0 only enters in the combination $\Omega_M H_0^2$ which is measured by the CMB extremely well. Therefore, the main sources of error should be in the derivatives of the comoving distance — more about that in a second.

Assuming that dark energy is due to a single rolling scalar field, the scalar potential can also be reconstructed,

$$V[\phi(z)] \;=\; \frac{1}{8\pi G}\left[\frac{3}{(dr/dz)^2} + (1+z)\frac{d^2r/dz^2}{(dr/dz)^3}\right] - \frac{3\Omega_{\rm M}H_0^2(1+z)^3}{16\pi G} \tag{73}$$

Direct reconstruction is the only approach that is truly model-independent. However, it comes at a price – taking derivatives of noisy data. In practice, one must fit the distance data with a smooth function — e.g., a polynomial, Padé approximant, or spline with tension; for example,

$$H_0 r(z) = \frac{z + a_2 z^2}{1 + b_1 z + b_2 z^2} \tag{74}$$

(which is a Padé approximant fit, made so that $H_0 r \to z$ when $z \ll 1$).

The fitting process introduces systematic biases – it is simply very hard to take second derivative of noisy data (if on a given day you want to be frustrated a little extra, try it!). While a variety of methods have been pursued it appears that direct reconstruction is too challenging and not robust even with SNIa data of excellent quality. Although the expression for $\rho_{\rm DE}(z)$ involves only first derivatives of $r(z)$, it contains little information about the nature of dark energy. For a review of dark energy reconstruction and related issues, see Sahni & Starobinsky (2006).

Principal components. The most general, and the only really non-parametric method, to reconstruct properties of DE is that of *principal components* of dark energy. The cosmological function that we are trying to determine — $w(z)$, $\rho_{\rm DE}(z)$, or $H(z)$ — can be expanded in terms of principal components, a set of functions that are uncorrelated and orthogonal by construction. In this approach, *the data determine* which aspects of a cosmological function — which linear combinations of $w(z_i)$, in this case — are measured best.

For example, suppose we parametrize $w(z)$ in terms of piecewise constant values w_i ($i = 1, \ldots, N$), each defined over a small redshift range (z_i, $z_i + \Delta z$). Then you get the principal components as follows

- Create the covariance matrix of all parameters, including the w_i
- Marginalize over all non-w parameters, and be left with the $N \times N$ covariance matrix in the w_i — call it C; also compute its inverse, C^{-1}
- Diagonalize the inverse covariance: $C^{-1} = W^T \Lambda W$, where Λ is diagonal and W is orthogonal

- Rows of W^T are the eigenvectors e_i — the "principal components" — while the elements of Λ give the accuracies in how well each PC is measured. The eigenvectors are of course orthonormal so that $\int e_i(z)e_j(z)dz = \delta_{ij}$.

In other words

$$\boxed{1 + w(z) = \sum_{i=1}^{N} \alpha_i \, e_i(z) \ ,} \tag{75}$$

with

$$\sigma(\alpha_i) = \lambda_i^{-1/2} \tag{76}$$

(note that we wisely added unity to w in its expansion, so that the fiducial expansion parameters are zero around ΛCDM).

In the limit of small Δz this recovers the shape of an arbitrary dark energy history (in practice, $N \gtrsim 20$ is sufficient), but the estimates of the w_i from a given dark energy probe will be very noisy for large N. Principal Component Analysis extracts from those noisy estimates the best-measured features of $w(z)$.

The coefficients α_i, which can be computed via the orthonormality condition

$$\alpha_i = \int (1 + w(z))e_i(z)dz \tag{77}$$

Examples of these components are shown for a future SN survey in the right panel of Fig. 11.

There are multiple advantages of using the PCs of dark energy (note you can also calculate PCs of $\rho_{\rm DE}(z)$, $H(z)$, etc):

- The method is as close as it gets to "model independent"
- Data tells you what you measure, and how well — no arbitrary parametrizations
- One can use this approach to design a survey that is most sensitive to the dark energy equation-of-state parameter in some specific redshift interval...
- ... or to study how many independent parameters are measured well by a combination of cosmological probes (i.e. how many PCs have $\sigma(\alpha_i)$, or $\sigma(\alpha_i)/\alpha_i$, less than some desired threshold value)

There are a variety of extensions of this method. A popular "relative" of the PCs is the approach where one parametrizes the equation of state in piecewise constant values, but then uses the diagonalization similar to that above to find linear, *nearly localized and 100% uncorrelated* linear combinations of these bins. A plot of these uncorrelated band-powers is easy to visualize in terms of checking of $w(z)$ varies with redshift. See Fig. 12 and Huterer & Cooray (2004) for details.

Figures of Merit. We finally discuss the so-called figures of merit (FoMs) for dark energy experiments. A FoM is a number, or collection of numbers, that serves as simple and quantifiable metrics by which to evaluate the accuracy of constraints on dark energy parameters from current and proposed experiments. For example, marginalized accuracy in the (constant) equation of state, w, could serve as a figure of merit – since a large FoM is "good", you would probably want something like FoM $= 1/\sigma_w$, or $1/\sigma_w^m$ where m is some positive power.

The most commonly discussed figure of merit is that proposed by the Dark Energy Task Force (Albrecht et al 2007, though this proposal goes back to Huterer & Turner 2001 paper), which is essentially inverse area in the $w_0 - w_a$ plane. For uncorrelated w_0 and w_a this would be $\propto 1/(\sigma_{w_0} \times \sigma_{w_a})$; because the two are typically correlated, the FoM can be defined as

$${\rm FoM}^{(w_0 - w_a)} \equiv (\det \mathbf{C})^{-1/2} \approx \frac{6.17\pi}{A_{95}}, \tag{78}$$

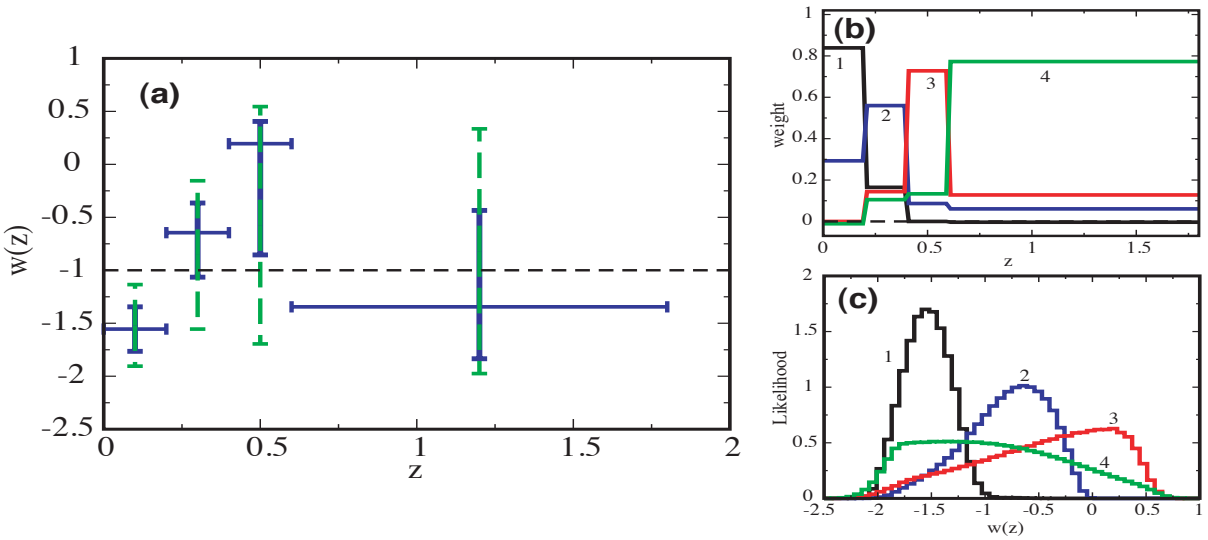

Figure 12: Uncorrelated band-power estimates of the equation of state $w(z)$ of dark energy are shown in panel (a), based on SN data available circa 2004. Vertical error bars show the 1 and 2-σ error bars (the full likelihoods are shown in panel (c)), while the horizontal error bars represent the approximate range over which each measurement applies. The full window functions in redshift space for each of these measurements are shown in panel (b); they have small leakage outside of the original redshift divisions. The window functions and the likelihoods are labeled in order of increasing redshift of the band powers in panel (a). Adopted from Huterer & Cooray (2004).

Note the constant of proportionality is really not that important, since typically you compare the FoM from different surveys, and the constant disappears when you take the ratio.

The standard "DETF FoM" defined in Eq. (78) keeps some information about the dynamics of DE (that is, the time variation of $w(z)$). But we can do better, and several more general FoMs have been proposed. My favorite one is, surprisingly, from a paper that I coauthored; Mortonson, Huterer & Hu (2010) proposed taking the FoM to be inversely proportional to the volume of the n-dimensional *ellipsoid* in the space of principal component parameters

$$\mathrm{FoM}_n^{(\mathrm{PC})} \equiv \left(\frac{\det \mathbf{C}_n}{\det \mathbf{C}_n^{(\mathrm{prior})}} \right)^{-1/2} . \tag{79}$$

where the prior covariance matrix is proportional to the (squares of) the product of prior ranges in the principal components. As with the multiplicative constant in Eq. (78), the prior matrix is unimportant since the $\det \mathbf{C}_n^{(\mathrm{prior})}$ term cancels out once you take the ratio of FoMs of two surveys you are comparing. See Fig. 13 for an illustration.

Summary of DE probes. Take-home list.

Summary of dark energy constraints. Figure 14 summarizes constraints in the $\Omega_M - w$ plane (assuming a flat universe) from SNIa, clusters, CMB, BAO, and redshift-space distortions, as of 2014.

The four principal probes of DE: systematics summary. In Table 1 we list the principal strengths and weaknesses of the four principal probes of DE. Control of systematic errors — observational, instrumental and theoretical — is crucial for these probes to realize their intrinsic power in constraining dark energy.

Secondary probes. There are a number of secondary probes of dark energy. Here we review a few of them. You are welcome to learn more about them at your leisure if you are interested.

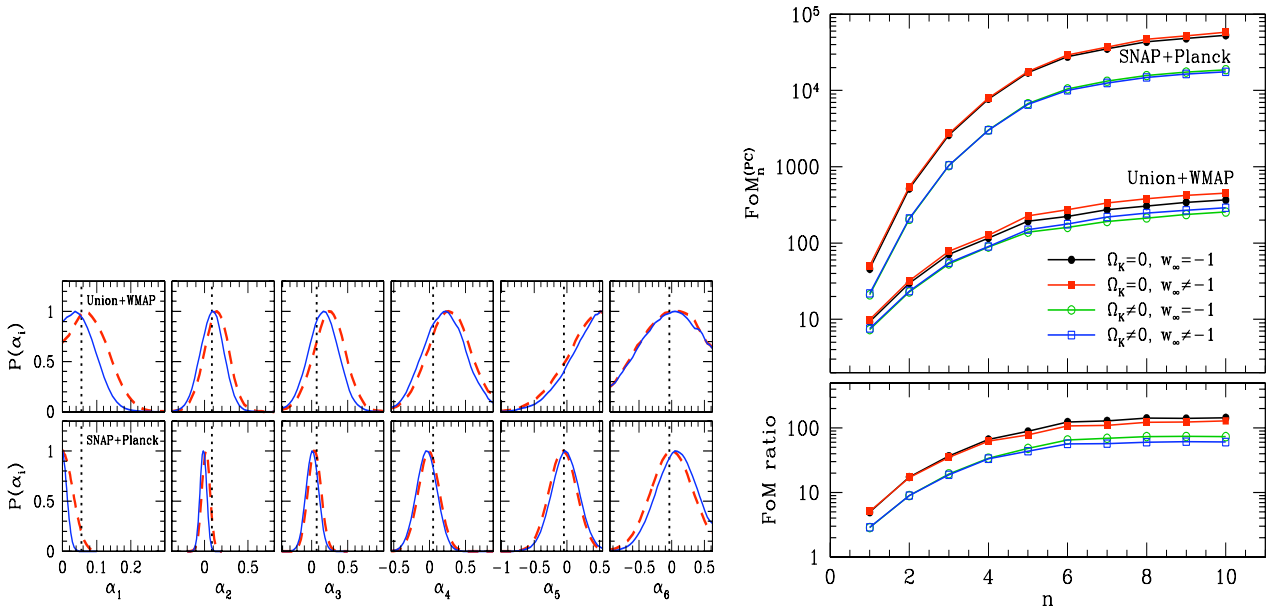

Figure 13: Figure of merit with PCs. *Left plot*: Marginalized 1D posterior distributions for the first 6 PCs of flat (solid blue curves) and nonflat (dashed red curves) quintessence models. Top row: current Union+WMAP data; bottom row: forecasts for SNAP+Planck assuming a realization of the data with $\alpha_i = 0$. *Right plot:* PC-based FoM comparisons. Top panel: PC figures of merit $\mathrm{FoM}_n^{(\mathrm{PC})}$ with forecasted uncertainties for SNAP+Planck and with measured uncertainties for Union+WMAP. Bottom panel: Ratios of $\mathrm{FoM}_n^{(\mathrm{PC})}$ forecasts to current values. In both panels, point types indicate different quintessence model classes: flat (solid points) or non-flat (open points), either with (squares) or without (circles) early dark energy. Adopted from Mortonson, Huterer & Hu (2010).

Table 1: Comparison of dark energy probes, adopted from Frieman, Turner and Huterer [2008]. CDM refers to Cold Dark Matter paradigm, FoM is the Figure-of-Merit for dark energy surveys defined in the Dark Energy Task Force (DETF) report, while SZ refers to Sunyaev-Zeldovich effect.

Method	Strengths	Weaknesses	Systematics
WL	growth+geometry, Large FoM	CDM assumptions	Shear systematics, Photo-z
SN	pure geometry, mature	complex physics	evolution, dust extinction
BAO	pure geometry, low systematics	coarse-grained information	bias, non-linearity, redshift distortions
CL	growth+geometry, X-ray+SZ+optical	CDM assumptions	mass-observable, selection function

- The Integrated Sachs-Wolfe (ISW) effect provided a confirmation of cosmic acceleration (e.g. Scranton et al. 2003 from the SDSS). ISW impacts the large-angle structure of the CMB anisotropy, but low-ℓ multipoles are subject to large cosmic variance, limiting their power. Nevertheless, ISW is of interest because it may be able to show the imprint of large-scale dark-energy perturbations (Hu & Scranton 2004).

- Gravitational radiation from inspiraling binary neutron stars or black holes can serve as "standard sirens" to measure absolute distances (Holz & Hughes 2005). If their redshifts can be determined, then they could be used to probe dark energy through the Hubble diagram (Dalal et al. 2006).

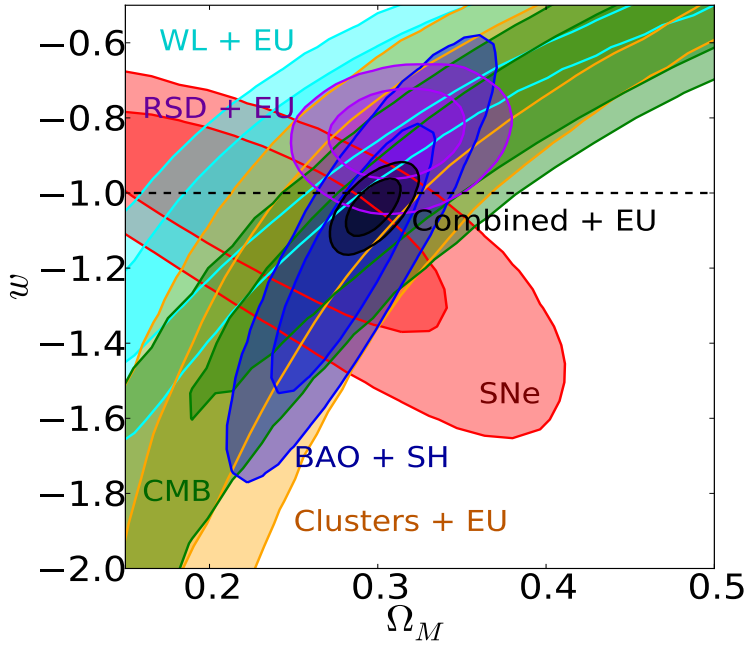

Figure 14: Summary of constraints in the $\Omega_M - w$ plane (assuming a flat universe) from various cosmological probes. Here, 'EU' refers to our early universe prior, while 'SH' refers to the sound horizon prior. Adopted from Ruiz & Huterer, arXiv:1410.5832

- Long-duration gamma-ray bursts have been proposed as standardizable candles (e.g. Schaefer 2003), but their utility as cosmological distance indicators that could be competitive with or complementary to SN Ia has yet to be established. The angular size-redshift relation for double radio galaxies has also been used to derive cosmological constraints that are consistent with dark energy.

- The optical depth for strong gravitational lensing (multiple imaging) of QSOs or radio sources has been proposed and used to provide independent evidence for dark energy, though these measurements depend on modeling the density profiles of lens galaxies.

- The Sandage-Loeb effect (Sandage 1962, Loeb 1998), the redshift change of an object measured using extremely high-resolution spectroscopy over a period of 10 years or more, will some day be useful in constraining the expansion history at higher redshift, $2 \lesssim z \lesssim 5$ (Corasaniti, Huterer & Melchiorri 2005).

- Polarization measurements from distant galaxy clusters in principle provide a sensitive probe of the growth function and hence dark energy (Cooray, Huterer & Baumann 2004).

- The relative ages of galaxies at different redshifts, if they can be determined reliably, provide a measurement of dz/dt and, from

$$t(z) = \int_0^{t(z)} dt' = \int_z^\infty \frac{dz'}{(1+z')H(z')}, \tag{80}$$

  measure the expansion history directly (Jimenez & Loeb 1998). Measurements of the abundance of lensed arcs in galaxy clusters, if calibrated accurately, provide a probe of dark energy.

Take-home messages. We end our brief review with our list of the ten most important facts about cosmic acceleration, adopted (and updated for numbers) from Turner & Huterer, arXiv:0706.2186:

1. Independent of general relativity and based solely upon the SN Hubble diagram, there is very strong evidence that the expansion of the Universe has accelerated recently.

2. Within the context of general relativity, cosmic acceleration cannot be explained by any known form of matter or energy, but can be accommodated by a nearly smooth and very elastic ($p \sim -\rho$) form of energy ("dark energy") that accounts for about 75% of the mass/energy content of the Universe.

3. Taken together, current data (SNe, galaxy clustering, CMB and galaxy clusters) provide strong evidence for the existence of dark energy and constrain the fraction of critical density contributed by dark energy to be $72 \pm 2\%$ and the equation-of-state to be $w \approx -1 \pm 0.05$ (stat) ± 0.05 (sys), with no evidence for variation in w. This implies that the Universe decelerated until $z \sim 0.5$ and age ~ 10 Gyr, when it began accelerating.

4. The simplest explanation for dark energy is the zero-point energy of the quantum vacuum, mathematically equivalent to a cosmological constant. In this case, w is precisely -1, exactly uniformly distributed and constant in time. All extant data are consistent with a cosmological constant; however, all attempts to compute the energy of the quantum vacuum yield a result that is many orders-of-magnitude too large (or is infinite).

5. There is no compelling model for dark energy. However there are many intriguing ideas including a new light scalar field, a tangled network of topological defects, or the influence of additional spatial dimensions. It has also been suggested that dark energy is related to cosmic inflation, dark matter and neutrino mass.

6. Cosmic acceleration could be a manifestation of gravitational physics beyond general relativity rather than dark energy. While there are intriguing ideas about corrections to the usual gravitational action or modifications to the Friedmann equation that can give rise to the observed accelerated expansion, there is no compelling, self-consistent model for the new gravitational physics that explains cosmic acceleration.

7. Even assuming the Universe has precisely the critical density and is spatially flat, the destiny of the Universe depends crucially upon the nature of the dark energy. All three fates — recollapse or continued expansion with and without slowing — are possible.

8. Cosmic acceleration is arguably the most profound puzzle in physics. Its solution could shed light on or be central to unraveling other important puzzles, including the cause of cosmic inflation, the vacuum-energy problem, supersymmetry and superstrings, neutrino mass, new gravitational physics, and dark matter.

9. Today, the two most pressing questions about dark energy and cosmic acceleration are: Is dark energy something other than vacuum energy? Does general relativity self consistently describe cosmic acceleration? Establishing that $w \neq -1$ or that it varies with time would rule out vacuum energy; establishing that the values of w determined by the kinematical and dynamical methods are not equal would indicate that GR cannot self consistently accommodate accelerated expansion.

10. Dark energy affects the expansion rate of the Universe, which in turn affects the growth of structure and the distances to objects. (In gravity theories other than GR, dark energy may have more direct effects on the growth of structure.) Upcoming ground- and space-based experiments should probe w at the percent level and its variation at the ten percent level. These measurements should dramatically improve our ability to discriminate between vacuum energy and something more exotic as well as testing the self consistency of general relativity. Laboratory- and accelerator-based experiments could also shed light on dark energy.