Tension in the Hubble Constant

Charles Stahl

April 27, 2021

Abstract

The Hubble constant H_0 measures the rate at which distant objects are receding away from us, as a function of their distance. In effect, it measures how fast our Universe is expanding. Although H_0 can be measured in our local region, it can also be inferred using the Cosmic Microwave Background (CMB) and the current standard model of cosmology, the Lambda-Cold Dark Matter (ΛCDM) model. Together, the measurement and the inference can provide a check of the validity of the model. Recently, errors bars on both methods of arriving at H_0 have grown small enough to put them in tension. Here we review the tension and some attempts at relieving it.

Contents

1 Introduction and History

When Einstein published the general theory of relativity in 1916 [1], it did not take long for the scientific community to realize it implied the Universe was evolving. The next year, he showed that the Universe could be static if the equations for GR included an extra constant, now called the cosmological constant Λ [2].

Throughout the 1920s scientists such as Friedmann and Lemaître continued to explore the possibility of an evolving universe [3, 4]. In 1929 Edwin Hubble published his observation that the Universe was in fact expanding [5], leading Einstein to call the introduction of the cosmological constant a blunder.[1] Einstein would never again use the cosmological constant in a published paper.

Hubble's measurements showed a linear relationship between an object's distance from Earth and its velocity, which implied that space itself is expanding. Although Hubble's data were far from perfect, observations have improved in the decades since. A comparison can be seen in Fig. 1. The slope of this graph, in units of $\mathrm{km\,s^{-1}\,Mpc^{-1}}$, is called the Hubble constant H_0.

In addition, our understanding of cosmology has improved. The current standard model of cosmology is still based on the fundamental work of Friedmann and Lemaître, along with Robertson and Walker. Called the Lambda-cold dark matter (ΛCDM) model, it describes a Universe that started with a Big Bang and evolved under general relativity with contents consisting of matter, radiation, and (back from the cosmological dustbin) some from of cosmological constant Λ.

A key confirmation of the Big Bang came in 1965 when Penzias and Wilson tried to make a really good radio antenna. They kept trying to remove some background in the microwave range that was present no matter where they looked in the sky. Finally, they realized this background was actually predicted by the Big Bang model, as leftover radiation from the time of recombination. [7].

[1] There is some controversy over George Gamow's claim that Einstein called it his "greatest blunder".

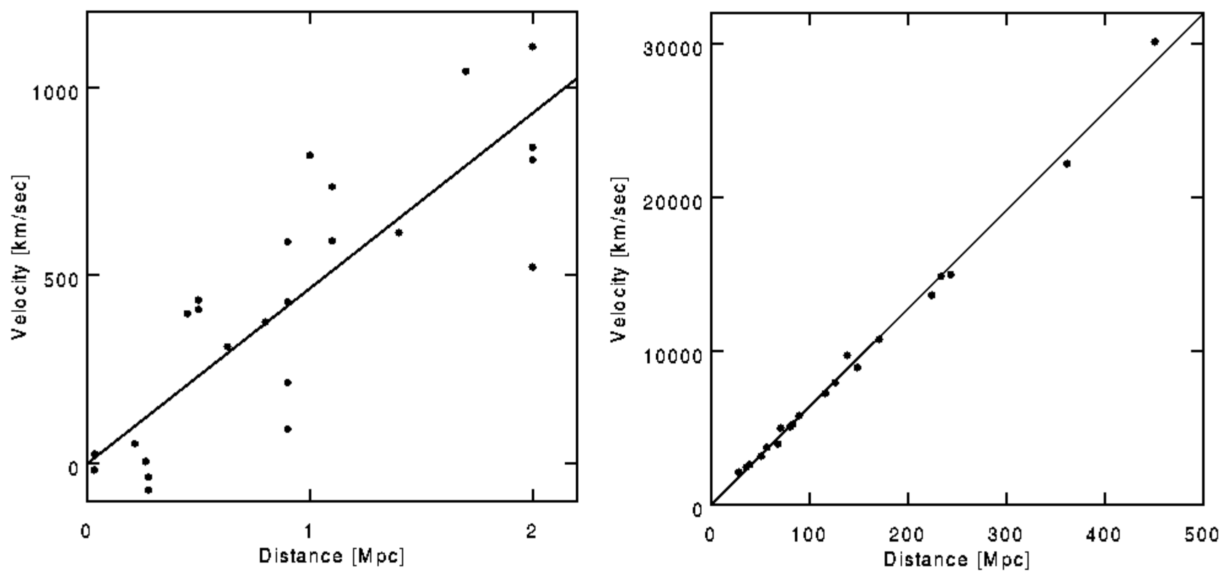

Figure 1: Comparison of Hubble's original data with more modern observations, from Ref. [6]. Note the much larger distance scale on the right, and the improved linear fit.

Soon after the Big Bang the universe was very hot. Eventually it reached a state in which the dominant form of matter was protons and electrons. At this point, the matter and was coupled to the radiation and the temperature was high enough to keep the matter ionized. Thus, the universe was a plasma, and the plasma was opaque. In this phase the entire universe was in thermal equilibrium. Note that this fact alone is surprising, since we know that there were points in thermal equilibrium that were outside each other's cosmic horizons. This is one of the motivations for inflation.

Eventually, the Universe cooled down enough for atoms to form, a process called recombination. At this point the matter and radiation lost contact and the photons were able to free-stream. The photons from recombination are still traveling through the now-transparent Universe, and we see them as a background microwave signal coming from every direction.

Although this cosmic microwave background (CMB) is extremely uniform, it has some anisotropies. In the 1990s cosmologists figured out how to use these perturbations to confirm the ΛCDM model and also make predictions. In particular, it is possible to use the CMB to infer the value of H_0. Initially, the value predicted from the CMB and the value measured by observations did not match, but had large enough error bars that the two values were compatible.

In recent years, both methods have drastically decreased their error bars without converging. The result is that today the two methods are in tension with each other with a statistical significance of $4 - 6\sigma$ [8]. It is possible that one method suffers from systematic bias, but if not then the ΛCDM model needs to be updated. This disagreement, called the Hubble tension, has been the topic of many recent conferences [9, 10]. In this paper we will describe how both methods arrive at a value for H_0 and summarize some possible resolutions to the tension.

The primary source for this paper will be Ref. [8], which reviews various classes of resolutions to the tension. We will refer to sections in that review as ([8] Sec. 1), etc, so that the interested reader can find more detailed information. We will start, however, by reviewing some helpful background material in order to provide the reader with a more gentle introduction to the Hubble tension. Throughout, we will set $c = 1$, but convert back to useful units when reporting values for H_0.

2 Background: Cosmology

In order to understand the tension in the Hubble constant, we will take a step back and review the history of the universe. Our tool here will be relativistic cosmology, which describes the evolution of the universe on the largest scales. All material in this section is taken from Ref. [6], unless otherwise noted. We will take only the bare bones for understanding the CMB prediction of H_0, but the interested reader can find more detail in that reference.

2.1 The Friedmann Equations

To a very good approximation, our Universe is homogeneous, isotropic, and flat. The most general metric that satisfies these conditions is the Robertson-Walker (or Friedmann-

Lemaître-Robertson-Walker or FLRW) metric,

$$ds^2 = -dt^2 + a^2(t)\left[dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right], \tag{1}$$

where $a(t)$ is called the scale factor. If the scale factor is constant, then the FLRW metric just describes Minkowski space. However, the ability for $a(t)$ to change with time allows the Universe to grow and shrink. In a universe with a Big Bang, the scale factor is $a = 0$ at $t = 0$.

From the scale factor we can define the Hubble parameter $H = \dot{a}/a$, where $\dot{a} = da/dt$. This value is constant in space but not in time. The current value, H_0, also called the Hubble constant, is the value of $H(t)$ at our time slice. This is the value that Hubble measured with his distance vs. velocity graphs. The Hubble constant is often reported in the literature using the standard but strange units $H_0 = 100h\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$.

In addition to the kinematic variables we need some dynamic variables. Usually, we assume the universe contains a collection of perfect fluids labeled by i, each defined by an energy ρ_i, a pressure p_i, and an equation of state $p_i = w_i\rho_i$ (no sum). The variable w_i is called the equation of state. As $a(t)$ changes, these fluids become more or less dense.

To understand how the universe expands and contracts in response to the fluids, we can use the Einstein equation of general relativity to arrive at the Friedmann equations. The first Friedmann equation is

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\sum_i \rho_i + \frac{\Lambda}{3}, \tag{2}$$

which constrains the expansion of the universe given its matter content. The famous cosmological constant Λ is just a constant of integration. It can be absorbed as a source of the equations if we let $\rho_\Lambda = \frac{\Lambda}{8\pi G}$. Under these conventions, the energy content of the universe is matter ρ_m (including baryonic matter ρ_b and cold dark matter ρ_c), radiation ρ_r, and dark energy ρ_Λ.

The second Friedmann equation is the evolution equation,

$$\dot{H} + \frac{3}{2}H^2 = \frac{\ddot{a}}{a} + \frac{1}{2}\left(\frac{\dot{a}}{a}\right)^2 = -4\pi G \sum_i p_i + \frac{\Lambda}{2}, \tag{3}$$

where we can again absorb Λ by letting $p_\Lambda = \frac{-\Lambda}{8\pi G}$. For the remainder of this paper we will absorb Λ as a source in both equations. In that case, the second Friedmann equation tells us $\sum_i \rho_i = \frac{3H^2}{8\pi G}$. We refer to this value as the critical density ρ_c and often report energy densities as $\Omega_i = \rho_i/\rho_c$ [11].

If we combine the Friedmann equations we can see that each fluid evolves as

$$\rho_i(t) \propto a(t)^{-3(1+w_i)}, \tag{4}$$

which reproduces some reasonable intuition. Matter dilutes as a^{-3} because it is in an expanding volume, radiation dilutes as a^{-4} because it cools in addition to being in an expanding volume, and the cosmological constant does not dilute (remains constant).

2.2 Distances and Horizons

Horizons and distance scales are of paramount importance in observational cosmology. To start with, we can define the redshift. Instead of time t, cosmologists usually use the redshift z to say when events occur. If light is emitted with wavelength λ when the scale factor is a, then if it is observed when the scale factor is a_0 it will have wavelength $\lambda_0 = \lambda\, a_0/a$. If we place the observer at our timeslice so that $a_0 = 1$, then we can define the redshift

$$z = \frac{1}{a} - 1, \tag{5}$$

so that $\Delta\lambda = z\lambda$. For example, recombination happened at $z \approx 1100$. The differentials

$$dt = \frac{da}{Ha} = -\frac{a\,dz}{H} \tag{6}$$

can be used to convert between time and redshift.

In the FLRW metric the speed of light is $1/a$. If an emitter emits a light signal at time t_e and an observer observes the signal at t_o, then the distance between them is

$$r_{oe} = \int_{t_e}^{t_o} \frac{dt}{a(t)} = \int_{z_o}^{z_e} \frac{dz}{H(z)}. \tag{7}$$

For example, in a universe with a Big Bang at $t = 0$, signals can only have traveled a distance of $\int_0^t dt'/a$ by time t. Eqn. 7 will be necessary for calculating horizons and distances. In fact, all we need for the CMB measurement of H_0 is this equation and an understanding of how the different fluid densities evolve in time.

3 Tension

Now that we have introduced the necessary cosmological prerequisites, we can understand what the tension in the Hubble constant really is. First we will see how measurements of the early universe through the CMB can "predict" the current value H_0. Then, we will review some methods of directly measuring the value of H_0 in our local universe, using methods that are more accurate than those that Hubble used.

3.1 CMB Prediction

As mentioned in the introduction, the CMB is very close to uniform, with perturbations a factor of 10^4 smaller than the uniform temperature. These perturbations have been measured by the European Space Agency's *Planck* satellite, along with other telescopes, to probe the evolution of the early universe. We will refer to the *Planck* 2013 results [11] and the *Planck* 2018 results [12] in this analysis.

The main idea here is that before the recombination time t_* there were some waves in the primordial plasma that created perturbations with scale r_s^*. Instead of measuring the linear size of these perturbations, *Planck* measures an angular scale θ_s^*. We can then use the relationship between these quantities to probe $H(t)$.

Before recombination, in the opaque plasma, the sound speed is

$$c_s = \frac{1}{\sqrt{3(1+\frac{3\rho_b}{\rho_\gamma})}}, \tag{8}$$

where ρ_b is the baryon density and ρ_γ is the density of photons. In analogy with Eqn. 7 we can define the sound horizon at recombination [11]

$$r_s^* = r_s(t_*) = \int_0^{t_*} \frac{c_s}{a} dt = \int_{z_*}^{\infty} \frac{c_s\, dz}{H}, \tag{9}$$

which defines how far density waves can have traveled by decoupling.

Since r_s^* depends only on baryon and photon densities, its value is independent of CMB measurements. From the CMB, we can instead extract the angular scale of the sound horizon at recombination, θ_s^*. In Euclidean geometry, an object of size r at distance $d >> r$ subtends an angle $\theta = r/d$. Similarly, the sound horizon and its angular scale are related by

$$\theta_s^* = \frac{r_s^*}{D_A^*}, \tag{10}$$

where D_A^* is the angular diameter distance. We can once again appeal to Eqn 7 to calculate the angular diameter distance as [13, 8]

$$D_A^* = \int_{t_*}^{t_0} \frac{dt}{a} = \int_0^{z_*} \frac{dz}{H(z)}, \tag{11}$$

so that a value of D_A^* can be used to infer $H(z)$. In particular, we can use the CMB pipeline to predict a value of $H(0) = H_0$.

Actually, the whole process is slightly more complicated than that. A more careful analysis shows that the above procedure most directly predicts the reduced density parameters $\omega_b = \Omega_b h^2$ and $\omega_c = \Omega_c h^2$. These predictions can then be combined with local measurements of Ω_b and Ω_c to predict the derived parameter h or H_0. Ref. [12] has nice graphs showing how h and other derived parameter co-vary with the different parameters like ω_b that are directly predicted by *Planck*.

Furthermore, the *Planck* predictions are based not only on the temperature anisotropy spectrum but also on polarization spectra. The gold standard result from *Planck* combines these spectra to produce a prediction of $H_0 = 67.27 \pm 0.6\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Different combinations can change the prediction, but only slightly.

3.2 Measurement

Of course, all that being said, we can follow Hubble's lead and just measure the distance and velocity of objects in our local region. This is complicated by the fact that nearby objects have proper motion, meaning that their velocity is not entirely determined by the expansion of the universe. That being said, if we look far enough away we should be able to see stars or other objects whose motion is dominated by the Hubble flow. These allow us to measure H_0 to high accuracy.

The goal then becomes to find many stars and measure both their velocity and their distance. It is relatively easy to measure velocity using the Doppler effect and the redshift in spectrum lines. For nearby stars it is possible to measure distance using parallax. For further distances we rely on types of stars that have very predictable intrinsic luminosities. Then, from the observed luminosity, we can infer the distance. The primary types of stars to use are pulsating Cepheid variables and exploding type Ia supernovae [8]. Together, all these measurements make up the so-called distance ladder.

Cepheid variable stars are stars that pulsate with a predictable relationship between luminosity and period. This relationship was discovered in 1908 by Henrietta Swan Leavitt [14] and allows observers to infer the intrinsic luminosity, and thus distance, from period. They are prevalent and bright enough to reach distances of $10 - 40\,\mathrm{Mpc}$. Supernovae, while brighter, are much rarer. They are used only to measure velocity at the furthest distances [8], and are the last rung on the ladder.

The SH0ES project [15] is the current gold standard for observational measurements of the Hubble constant, with a value of $73.2 \pm 1.3\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Other projects have varying estimates, but they generally agree with the stated value.

3.3 Summary

In order to quantify the tension between SH0ES and the *Planck* measurement, we want to calculate to what level of uncertainty the difference between the measurements vanishes [16]. Explicitly,[2] using the values for h to avoid units and letting subscripts distinguish the *Planck* and SH0ES data,

$$\begin{aligned} \Delta &= h_P - h_\mathrm{S} = .6727 - .732 = .0593, \\ \sigma &= \sqrt{\sigma_P^2 + \sigma_\mathrm{S}^2} = \sqrt{.0060^2 + .013^2} = .0143, \\ \frac{|\Delta|}{\sigma} &= 4.15, \end{aligned} \tag{12}$$

which is to say that the two measurements disagree at 4.2σ. Other transformations of the data from either the CMB or local measurements can result in slightly different tensions, but they generally lie in the 4σ to 6σ range [8]. Other ways of measuring H_0 are shown in Fig. 2. Note that many pairs of error bars do not overlap.

[2] This may be basic statistics, but the author was unable to find the calculation in any of the reference papers, and thought it might be useful to include.

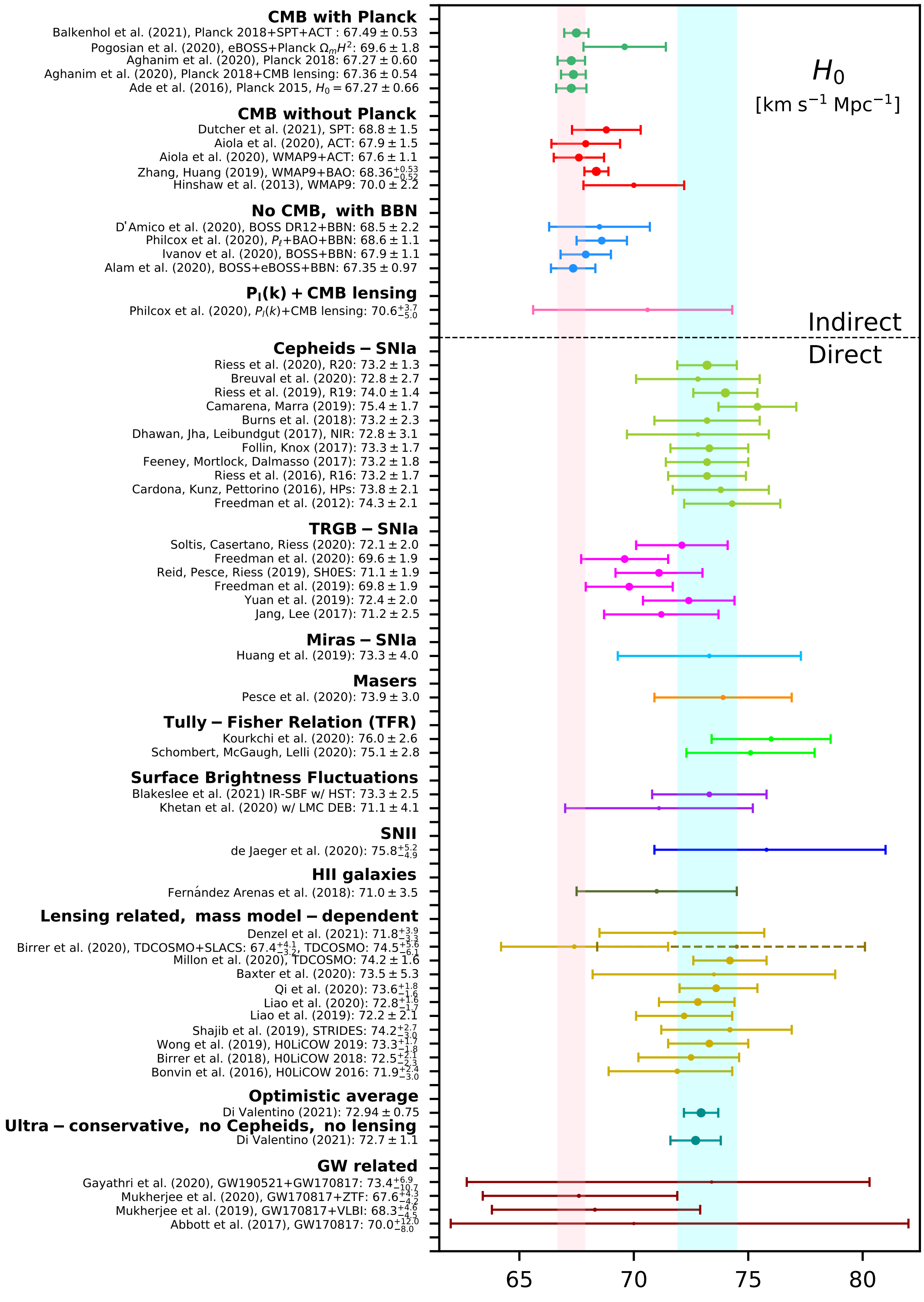

Figure 2: A summary of many different indirect and direct measurements of the Hubble constant, taken from Ref. [8]. The values quoted in the current paper are the third from the top and the first in the “indirect” section. This figure is included to show that there are many methods in addition to those described in the current paper for measuring H_0. Direct measurements in general favor a higher value, although sometimes with very large error bars.

4 Possible resolutions

The simplest possible resolution to the Hubble tension would be that we are still not measuring velocity far enough out in the universe, so that our local (distance ladder) value of H_0 is not the correct value for our entire timeslice. This possibility is highly unlikely. It would require that our local Universe is about 20 times more underdense than statistical fluctuations suggest is likely [8]. There are many more proposed ways to resolve the Hubble tension.

Recall that the ingredients of the ΛCDM model are General Relativity with the FLRW metric, some form of dark energy Λ, cold dark matter, the Standard Model of matter we can see, and a history of the Universe. In some sense, the most phenomenological element of the ΛCDM model is dark energy. If we claim that ρ_Λ is equal to the vacuum energy of the effective field theory of the Standard Model, it is off by a stunning 120 orders of magnitude [6]. Futhermore, recall that Einstein originally did not include Λ, then did include it, then again did not. Thus, it is tempting to modify its contribution to the Friedmann equations before any other component.

One way to modify dark energy is to allow it to have an equation of state $w_{\rm DE} \neq -1$. This possibility, called the wCDM model, would fall under the category of late dark energy (LDE, [8] Sec. 5) because it changes how dark energy behaves right now. A wCDM model can completely alleviate the Hubble tension, at the cost of introducing phantom-like dark energy, $w_{\rm DE} < -1$. From Eqn. 4 we can see that such a flavor of dark energy would have an energy density that increases as the Universe expands. Then the first Friedmann equation tells us that the scale factor a will diverge in finite time, a situation called the Big Rip.

Another way to resolve the tension is to allow the various perfect fluids to interconvert. We already know this happens, for example when an electron and a positron annihilate to create photons. Additional mixing could come from early dark energy (EDE, [8] Sec. 4) that initially behaves like dark energy but transforms into something more like matter or radiation at late time. Models of this type can decrease the Hubble tension to $2 - 3\sigma$.

Alternatively, we can consider extra interactions between dark energy and dark matter or between either of the two dark fluids and ordinary matter ([8] Sec. 8). These are, in effect, extensions of the Standard Model of particle physics. A fun example of an extra interaction model is the $L_\mu - L_\tau$ model [17], which includes a massive Z' boson that decays into neutrinos. The authors claim that this form of extension to the Standard Model, while largely unconstrained, is able to solve both the Hubble tension and the longstanding and recently-confirmed tension in measurements of the muon magnetic moment [18].

There are, of course, many other modifications to any of the other components listed at the beginning of this section. The interested reader is encouraged to consult Ref. [8].

5 Conclusion

The goal of this paper has been to introduce the Hubble tension, while splitting the difference between cosmology reviews that do not explain how to measure H_0 and Hubble tension reviews that do not introduce the relevant cosmology. To that goal, we have tried to not shy away from some of the complicated aspects. That being said, the story here has still been highly simplified. For example, the actual way that H_0 is inferred from CMB data involves heavy numerical modeling and statistical work.

In addition there are many confounding factors, from lensing of the CMB to foreground sources to instrumental noise. All of these are discussed in detail in the many *Planck* data releases.

Furthermore, there are other cosmological sources of data such as Baryon Acoustic Oscillations (BAO) that can provide alternative methods of measuring H_0. These methods can be combined with those mentioned in this paper to paint a picture of the Universe for which we do not have a single dominant consistent model. Observational cosmology will certainly be a useful testbed of fundamental physics for decades to come.

References

[1] Albert Einstein. The foundation of the general theory of relativity. *Annalen der Physik*, 4(49):769–822, 1916.

[2] Albert Einstein. Cosmological considerations in the general theory of relativity. *Preussische Akademie der Wissenschaften*, 1917:142–152, 1917.

[3] A. Friedman. Über die krümmung des raumes. *Zeitschrift für Physik*, 10(1):377–386, December 1922.

[4] Georges Lemaître. A homogeneous universe of constant mass and increasing radius accounting for the radial speed of extra-galactic nebulae. *Annales de la Société Scientifique de Bruxelles*, A47:49–56, 1927.

[5] E. Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences*, 15(3):168–173, March 1929.

[6] Mark Trodden and Sean M. Carroll. Tasi lectures: Introduction to cosmology, 2004.

[7] A. A. Penzias and R. W. Wilson. A measurement of excess antenna temperature at 4080 mc/s. *The Astrophysical Journal*, 142:419, July 1965.

[8] Eleonora Di Valentino, Olga Mena, Supriya Pan, et al. In the realm of the hubble tension − a review of solutions, 2021.

[9] Licia Verde, Tommaso Treu, and Adam G. Riess. Tensions between the early and late universe. *Nature Astronomy*, 3(10):891–895, September 2019.

[10] Eleonora Di Valentino, Luis A. Anchordoqui, Ozgur Akarsu, et al. Cosmology intertwined ii: The hubble constant tension, 2020.

[11] P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, et al. Planck2013 results. XVI. cosmological parameters. *Astronomy & Astrophysics*, 571:A16, October 2014.

[12] N. Aghanim, Y. Akrami, M. Ashdown, et al. Planck 2018 results. *Astronomy & Astrophysics*, 641:A6, September 2020.

[13] Dorothea Samtleben, Suzanne Staggs, and Bruce Winstein. The cosmic microwave background for pedestrians: A review for particle and nuclear physicists. *Annual Review of Nuclear and Particle Science*, 57(1):245–283, November 2007.

[14] Henrietta S. Leavitt and Edward C. Pickering. Periods of 25 variable stars in the small magellanic cloud. *Harvard College Observatory Circular*, 173:1–3, March 1912.

[15] Adam G. Riess, Stefano Casertano, Wenlong Yuan, et al. Cosmic distances calibrated to 1% precision with gaia EDR3 parallaxes and hubble space telescope photometry of 75 milky way cepheids confirm tension with ΛCDM. *The Astrophysical Journal*, 908(1):L6, February 2021.

[16] Jörg Dietrich. How Much Do My Measurements Disagree?, 2014.

[17] Miguel Escudero, Dan Hooper, Gordan Krnjaic, and Mathias Pierre. Cosmology with a very light $L_\mu - L_\tau$ gauge boson. *Journal of High Energy Physics*, 2019(3), March 2019.

[18] B. Abi, T. Albahri, S. Al-Kilani, et al. Measurement of the positive muon anomalous magnetic moment to 0.46 ppm. *Phys. Rev. Lett.*, 126:141801, Apr 2021.