Planck and the local Universe: Quantifying the tension

Licia Verde[a,b,*], Pavlos Protopapas[c,d], Raul Jimenez[a,b]

[a]ICREA & ICC, Universitat de Barcelona, UB-IEEC, Marti i Franques 1, 08034 Barcelona, Spain
[b]Theory Group, Physics Department, CERN, CH-1211 Geneva 23, Switzerland
[c]Harvard Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA
[d]Institute for Applied Computational Science, School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA

ARTICLE INFO

Keywords:
Cosmology
Hubble parameter
Age of the Universe
Cosmic Microwave Background
Bayesian methods

ABSTRACT

We use the latest *Planck* constraints, and in particular constraints on the derived parameters (Hubble constant and age of the Universe) for the local universe and compare them with local measurements of the same quantities. We propose a way to quantify whether cosmological parameters constraints from two different experiments are in tension or not. Our statistic, $\mathcal{T}$, is an evidence ratio and therefore can be interpreted with the widely used Jeffrey's scale. We find that in the framework of the ΛCDM model, the *Planck* inferred two dimensional, joint, posterior distribution for the Hubble constant and age of the Universe is in "*strong*" tension with the local measurements; the odds being $\sim$1:50. We explore several possibilities for explaining this tension and examine the consequences both in terms of unknown errors and deviations from the ΛCDM model. In some one-parameter ΛCDM model extensions, tension is reduced whereas in other extensions, tension is instead increased. In particular, small total neutrino masses are favored and a total neutrino mass above 0.15 eV makes the tension "*highly significant*" (odds $\sim$1:150). A consequence of accepting this interpretation of the tension is that the degenerate neutrino hierarchy is highly disfavored by cosmological data and the direct hierarchy is slightly favored over the inverse.

1. Introduction

Cosmic Microwave Background (CMB) data have been crucial to define and confirm the currently favored cosmological model: a flat cosmological constant-dominated, cold dark matter model, ΛCDM. It is important to keep in mind that CMB observations predominantly probe the physics of the early Universe ($z \gtrsim 1100$). When these observations are interpreted in terms of the standard cosmological parameters, defined at $z = 0$, an extrapolation is needed, which is done within a given cosmological model. In our previous work [1], we argued that local, model-independent measurements of cosmologically-relevant quantities can be used to test the self-consistency of the currently favored cosmological model and to constrain deviations from it. Direct measurements of the Hubble constant and age of the Universe are especially suited to this aim. In fact, these measurements have a long history and have now reached a level of precision and accuracy that make them competitive with other cosmological observations. If local and high-redshift measurements are to be combined to constrain cosmological parameters within a given model, the well established framework of Bayesian parameter inference can be used. However, if the two sets of measurements are to be used to distinguish between models then Bayesian model selection and Bayesian Evidence[1] (model-averaged likelihood) should be used [2]. Methods in cosmology have been used for almost two decades now (see [3–11] and references therein). We refer the reader to these references for explanation and probabilistic interpretation of the Evidence.

Ref. [1] used pre-*Planck* state of the art data, so it is a natural extension to that work to consider the post-*Planck* [9] state-of-the-art cosmological data. The *Planck* team official analysis pointed out that local direct measurements of the Hubble constant seem to be at odds with CMB data when interpreted within the context of the ΛCDM model. Post-*Planck* CMB data are however not at odds with other cosmological measurements (e.g., Baryon Acoustic Oscillations). This conclusion arises from a parameter-estimation analysis, where the best fit value of the Hubble constant extrapolated from CMB data is $\sim$2.5σ away from the direct measurement.

We will examine these findings in the framework of Bayesian model selection and discuss their implications for cosmology. We will also consider measurements of the age of the Universe and investigate whether, from a model selection point of view, the tension between local and high redshift measurements disfavors the ΛCDM model in favor of a more complex model.

---

* Corresponding author at: ICREA & ICC, Universitat de Barcelona, UB-IEEC, Marti i Franques 1, 08034 Barcelona Spain. Tel.: +34 934031328.
E-mail address: liciaverde@gmail.com (L. Verde).

---

[1] Here we use capital "Evidence" for the Bayesian quantity to distinguish it from the colloquial *evidence*.

This paper is organized as follows: In Section 2, we present the data sets and combinations of data sets we will use. In Section 3, we review different statistics that measure the "distance" or the "difference" between two distributions. We present our statistic of choice that stems from the Bayesian evidence ratio and is suited to assess whether two posterior distributions from two different experiments are in tension or not. We present our results in Section 4, where we examine the tension between local and high redshift cosmological measurements. We then explore several possibilities for this tension and examine the consequences both in terms of unknown errors and deviations from the ΛCDM model. Finally, we conclude in Section 5. In Appendix A, we report the Kullback–Leibler divergence between Planck and WMAP for selected models and parameters.

2. Data

We consider the *Planck* mission CMB data with local measurements of the Hubble parameter H_0 and of the age of the Universe t_U.

2.1. CMB data

The Planck collaboration, along with the nominal mission temperature data, has also released the outputs of the Markov Chains Monte Carlo (MCMC) used to sample from the space of possible cosmological parameters. These MCMCs have been used by the team to generate estimates of the posterior mean of cosmological parameters, along with their confidence intervals [12]. Here, we use the publicly available outputs of the *Planck* team MCMCs. The *Planck* data have been analyzed by the team in several ways: using only Planck temperature data on scales corresponding to multipoles $\ell < 2500$ (*Planck*), using Planck data in conjunction with WMAP polarization data at low ℓ (*WP*), and using Planck data, WMAP data and also include the measurement of the lensing potential, which was reconstructed from the Planck temperature maps themselves through the measurement of the four-point function (*lensing*). In some cases, data from high-resolution but partial sky ground-based CMB experiments (ACT and SPT) have also been included (*highL*). In combination with Planck these data better constrain the foreground-model parameters. Here however, we concentrate on the *Planck + WP* data combination. Except in the Appendix A, we will not consider the *lensing* information in line with the considerations presented in [13]; the combination *Planck + WP + highL* predicts a value for the lensing amplitude which is about 2σ higher than the value measured from the convergence power spectrum [12,14] and the origin of this tension is not yet fully understood.

2.2. Local measures: H_0 and t_U

For the local data we follow exactly Ref. [1]. We combine the values reported in Refs. [15,16] in a "world average" where the central value is given by the variance-weighted mean and the error conservatively given by the average of the errors: $H_0 = 74.08 \pm 2.25$ Km s^{-1} Mpc^{-1}.

Estimates of the age of the Universe can be obtained from the ages of the oldest objects and in particular from the ages of the oldest stars since these objects form very shortly after the Big Bang. Accurate dating of globular clusters has been the subject of active investigation for decades but the error bars were, for a long time, large. It is only relatively recently that error-bars have become smaller due to better estimates of their distances [17] or the use of distance independent methods [18]. Recently, it has become possible to use single stars to estimate t_U; accurate distances using direct parallax measurements were obtained for nearby sub-giant stars. In particular the star HD 140283 is a sub-giant moving off the main sequence, so its luminosity is a very good age-indicator.

Consequently, in order to constrain the current age of the universe, we use recent determinations of the ages of the oldest stars in the Milky Way. We use two kind of measurements: the age of the nearby sub-giant HD-140283 and the ages of the oldest globular clusters. The age of HD-140283 has been accurately measured by [19] using HST parallaxes and spectroscopic determinations of its chemical abundance. In addition, they have used state of the art stellar evolutionary models and carried out a careful and extensive error budget. The age of HD-140283 is determined to be 14.5 ± 0.8 Gyr (including systematic errors, which dominate the error bar). Additionally, the ages for some of the most metal poor Milky Way globular clusters (NGC 6397, NGC 6752, and 47 Tuc) have been determined by Ref. [20]. Taking into account the revised nuclear reaction rate for $^{14}N(p, \gamma)^{15}O$ of the CNO burning cycle [21], we obtain an age 14.2 ± 0.6 (± 0.8 systematics) Gyr. In what follows we will linearly add these (random and statistics) two sources of errors.

It is remarkable that age determinations for such different systems are in such a good agreement. There are three main ingredients that dominate the error budget of stellar ages: distance, chemical composition and theory of stellar evolution. The last one has been studied with extreme care in the last decade and major improvements have been made. It is very unlikely that stellar evolution theory needs any further significant revisions; the contribution to the age error budget from stellar evolution theory is now negligible contributing to about 1% to the total age uncertainty. The two dominant error sources remain distance and chemical composition.

The distance uncertainty can be efficiently removed by obtaining trigonometric parallaxes to the oldest stars or globular clusters. This has been the case for HD-140283. However, for the globular clusters it is not yet possible to obtain trigonometric parallaxes and distances have to be obtained by indirect methods. Currently the uncertainty budget in the age determination of galactic globular clusters is dominated by its distance estimation, which is done via indirect methods (sub-dwarf fitting).

On the other hand, the dominant source of error of nearby sub-giants is the chemical composition of the stars. Current 10-m class telescope observations provide abundance with an accuracy of 0.1 dex, which translates into a ~5% uncertainty when estimating the age of the oldest stars.

The ages determinations of nearby sub-giants and globular clusters are dominated by different and independent systematics (chemical abundance and distance respectively) and we therefore combine the two above measurements by inverse variance weighting. Recall that the ages of the oldest objects at $z = 0$ only provide a lower estimate of t_U. However, we know that the first generation of stars formed at $z > 20$ [22]. At this redshift the age of the universe is only ~0.1–0.2 Gyr, which is much smaller than the current age estimation errors. This formation time-lag will only play a role when age estimations uncertainties are reduced by at least a factor of five, to the 0.7% level. In our analysis we convert the age of the star to the age of the Universe by adding to the stellar age 0.15 ± 0.05 Gyr, assuming a Gaussian distribution (cutting negative tails when appropriate). We obtain the following estimate for the age of the Universe: $t_U = 14.4 \pm 0.7$. Here we assume Gaussian distributions for both H_0 and t_U measurements.

It is important to note that the error-budget in the age determination will change dramatically with the launch of the GAIA satellite in October 2013. GAIA will obtain trigonometric parallaxes to globular clusters and identify a large number of metal-poor sub-giants. Therefore, the only dominant source of error remaining will be the chemical composition of the stars. A way to improve on this will be to obtain longer integrations or use different line diagnostic for the metallicity in the infrared. Thus the dominant uncertainty in the error budget for the age of the local universe, will be the chemical composition reaching an error-floor of ~2–3%; see [1] for discussion.

3. Methods

The questions we wish to address are the following. (*a*) We have two measurements of cosmologically interesting quantities in the form of a – two or higher dimensional – posterior distribution. In our application these two measurements are the high-redshift, model-dependent joint distribution of H_0 and t_U, and the local, cosmological-model-independent one. In the Bayesian framework, how would one quantify whether these two measurements are or not in agreement (tension)? In other words, if the null hypothesis is that the two measurements are "sampled" from the base model adopted, when should the null hypothesis be abandoned? (*b*) If the answer to (*a*) is that the two measurements are in tension, then Bayesian model selection can be used to study extensions to the base model adopted and select which is the favored model. Alternatively the detected tension might indicate the presence of unaccounted for, residual systematic errors. Possible options at this point are: discredit the measurement most likely affected by systematics or artificially increase its errors. If instead no tension is detected the measurements can be combined to perform, for example, joint parameter estimation.

Clearly to address this type of question a measure of distance or difference between two distributions should be used. There are many statistics that quantify the difference between two distributions. The most widely used are the Kullback–Leibler divergence [23], the Jeffrey's divergence [26], and Jensen–Shannon divergence [28]. These are well rooted in information theory and they measure the difference between two probability distributions, say $\mathcal{P}$ and $\mathcal{Q}$, and therefore are suited to quantify how well $\mathcal{P}$ approximates $\mathcal{Q}$ or the information content that $\mathcal{Q}$ adds to $\mathcal{P}$. In fact the Kullback–Leibler divergence is not symmetric (clearly, the information content that $\mathcal{Q}$ adds to $\mathcal{P}$ is not the information content that $\mathcal{P}$ adds to $\mathcal{Q}$) and the Jeffrey's divergence and Jensen–Shannon divergence are two approaches to symmetrize it. We will explore the application to cosmology of the entropy-based Kullback–Leibler divergence measure in the Appendix A. For our question (*a*) at hand we know *a priori* that we will be comparing two different distributions (in fact we are comparing different experiments). We want to know whether the best fit values are consistent given the shape of the respective distributions, in other words we need the Bayesian (multi-dimensional) parallel of the standard equivalency test.[2] We argue that this is given by the Evidence ratio as follows.

3.1. Evidence for tension

Imagine we have performed two experiments: *A*, *B* and for each experiment we produce a posterior $P_{A,B}(\theta|D_{A,B})$ where θ represents the parameters of the model and $D_{A,B}$ represents the data from experiments *A*, *B* respectively. Let us also assume that for producing both posteriors we have used the same, uniform priors over the same "support", *x*, i.e., $\pi_A = \pi_B = \pi$, $\pi = 1$ or 0 and therefore $\pi_A \pi_B = \pi$.

Let H_1 be the (null) hypothesis that both experiments measure the same quantity, the models are correct and there are no unaccountable errors. In this case, the two experiments will produce two posteriors, which, although can have different (co)variances, and different distributions, have means that are in agreement. The alternative hypothesis, H_1 is when the two experiments, for some unknown reason, do not agree, either because of systematic errors or because they are effectively measuring different things or the model (parameterization) is incorrect. In this case, the two experiments will produce two posteriors with two different means and different variances.

To distinguish the two hypothesis we use the Bayes factor, that is the ratio of the Evidences. The Bayes factor has been extensively used

---

Table 1
The slightly modified Jeffreys' scale we use for interpreting the tension $\mathcal{T}$.

Ln $\mathcal{T}$	Interpretation	Betting odds
<1	not worth a bare mention, not significant	<3:1
1−2.5	substantial	∼3:1
2.5−5	strong	>12:1
>5	highly significant	>150:1

in cosmology, predominantly to perform model comparison between models with different number of parameters, mostly nested models. However, it has been used in a similar context before to quantify the consistency of different data sets in [24,25].

In any practical application, the absolute normalization of the posteriors is often unknown, but we can still work as follows. We define Table 1:

$$\int P_A P_B dx = \lambda \int \mathcal{L}_A \mathcal{L}_B \pi_A \pi_B dx = \lambda \int \mathcal{L}_A \mathcal{L}_B \pi dx = \lambda E = \mathcal{E}, \qquad (1)$$

where $\mathcal{L}$ denotes the likelihood and $\lambda^{-1} = \int \mathcal{L}_A \pi_A dx \int \mathcal{L}_B \pi_B dx'$. *E* is the Bayesian Evidence for the joint distribution, thus $\mathcal{E}$ is akin to an unnormalized Evidence. This quantity per se is therefore of limited use. However we are interested in an evidence ratio between the null hypothesis and the alternative hypothesis, and the (difficult to compute) normalization factor will cancel if a suitable ratio is taken. Operationally, let us now imagine that we can perform a *translation* (shift) of (one or both of) the distributions in *x* and let us define $\overline{P}_A$ the shifted distribution. This translation changes the location of the maximum but does not change the shape or the width of the distribution. If the maxima of the two distributions coincide then

$$\int \overline{P}_A \overline{P}_B dx = \overline{\mathcal{E}}|_{\max A = \max B} . \qquad (2)$$

This can be considered our "straw man" null hypothesis (H_1). As the distance between the maxima increases (but the shape of the distributions remains the same),

$$\int \overline{P}_A \overline{P}_B dx = e < \overline{\mathcal{E}}, \qquad (3)$$

and eventually $e \longrightarrow 0$ as the two distributions diverge. Clearly the Evidence ratio for the (null) hypothesis E_1 is $\mathcal{E}/\overline{\mathcal{E}}|_{\max A = \max B}$, as the normalization factors λ cancel out, and the Evidence ratio for the alternative H_1 is its reciprocal. We therefore introduce:

$$\mathcal{T} = \frac{\overline{\mathcal{E}}|_{\max A = \max B}}{\mathcal{E}}, \qquad (4)$$

which denotes the degree of tension that can be interpreted in the widely used (slightly modified, [27]) Jeffrey's [26] scale (Table 1). $\mathcal{T}$ indicates the odds: $1 : \mathcal{T}$ are the chances for the null hypothesis. In other words, a large tension mens that the null hypothesis (max*A* = max*B*) is unlikely.

In this scale, $\ln \mathcal{T} < 1$ is *not significant*, if $1 < \ln \mathcal{T} < 2.5$ the evidence is *substantial*, becomes *strong* only if $2.5 < \ln \mathcal{T} < 5$ and *highly significant* if $\ln \mathcal{T} > 5$.

This scale is empirically calibrated, and should be used only as a guide as it introduces sharp decision-making boundaries. Here we use the boundaries as a loose classification of the degree of tension; we also use Jeffreys' nomenclature.

To give an intuition about the meaning of $\ln \mathcal{T}$ values, consider two Gaussian distributions with unit variance: a shift of the central value of one of the two distributions of 2σ would give $\ln \mathcal{T} = 1$ which is the threshold between *not significant* and *substantial*; a shift of more than 3σ would yield 'strong' tension and of more than $>4.5\sigma$ to give *highly significant* tension.

E_1 can be seen as the Evidence for the joint distribution i.e., the joint likelihood integrated over the – uniform – prior and thus $\mathcal{T}$ as its

reciprocal. As such, it addresses the question: "of all the values of the parameters allowed by the prior in a given model, how well on average the model fits the (combined) data?" Of course, if we now interpret the likelihood of one data set (say, the data set from experiment B) times the uniform prior as the (new) prior, this Evidence would say: how well on average the parameters allowed by this (new) prior within the model, fit the data from experiment A?" If the fit is bad then the Tension is high. As such it gives additional information compared to the more widely used Bayes factor between two cosmological models: Model 1 could be favored over Model 2 by the standard Bayes factor, but the Bayes factor will not tell if one (or both) models are bad fit to the (combined) data. A high value of the Tension on the other hand would indicate that the model does not fit well the data. As such it gives additional information compared to the more widely used Bayes factor between two cosmological models: Model 1 could be favored over Model 2 by the standard Bayes factor, but the Bayes factor will not tell if one (or both) models are bad fit to the (combined) data. A high value of the Tension on the other hand would indicate that the model in question does not fit well the data.

In some practical applications the shift needed to compute the numerator of $\mathcal{T}$ may be slightly incorrect or at least misleading. For example, we know that for CMB data the covariance matrix depends on the assumed cosmology because of cosmic variance. Therefore a rigid translation of the distribution is strictly incorrect. One could in principle imagine an extreme case where shifting P_A so that the maximum coincides with that of P_B gives a highly significant Evidence but instead shifting P_B does not.

However, for the practical applications we can think of, in the era of precision cosmology, this effect is small, or if it is large it means that the shift is large and the Evidence for tension will be highly significant anyway. Therefore this effect will not drastically change the interpretation of the $\mathcal{T}$ value. Nevertheless, in what follows, we shift the local measurements distribution, which is not affected by cosmic variance and thus does not depend (too strongly) on cosmology.

When computing $\mathcal{T}$ in a practical application, there is a delicate point to bear in mind: the above relies on having a uniform prior on H_0 and t_U. Here we wish to use the output of MCMCs which were performed with uniform priors on other parameters[3] (not H_0 and t_U). The relation between these parameters and H_0, t_U is non-linear, thus in the MCMCs the prior in not uniform in H_0, t_U. As a consequence, the MCMC outputs in principle, cannot simply be importance-sampled and then used to perform the above integrals by Monte Carlo integration i.e., simply adding up the MCMC (updated) weights. Of course, a change of variables can be made by re-weighting the chains outputs by the Jacobian of the transformation. This effect for the data sets considered here is, however, small. Alternatively the posterior surface in the H_0, t_U sampled by the MCMC can be fit by a smooth surface using the likelihood values instead of the weights. This surface provides then a functional form for the posterior which can be integrated to compute $\mathcal{E}$ and $\mathcal{T}$. We use the latter approach.

With this type of analysis we are entering the regime of "meta analysis" which has an extensive literature mostly in the medical field.

Logically, if $\mathcal{T}$ is small and the tension is not significant the results of the experiments can be combined and a joint analysis can be safely

performed. However if $\mathcal{T}$ is large (e.g., *strong* or *highly significant*), then it is an indication that either (*a*) one of the two experiments is affected by errors (systematic or statistical) that are unaccounted for or that (*b*) the underlying model used is incorrect and must be extended.

The situation described in case (*a*) has been considered before in the cosmology literature [4,29]. The authors advocate introducing "meta parameters" describing possible systematic shifts or statistical errors and marginalize over them. The detailed implementation in the Bayesian context is numerically very heavy, but the upshot is that in practice such an approach leads to down-weighting the discrepant measurements (i.e., effectively increasing the corresponding error-bars) but still combining them. A simplified approach (which is the one we will pursue here) is to increase the error-bars of the discrepant measurement(s) until the tension $\mathcal{T}$ is not significant. After that we can analyze jointly the data sets. We will refer to this case as "blame the measurements".

The situation described in (*b*) is similar to what is being done extensively in cosmology when implementing model selection in the Bayesian framework. There is a simpler model (typically the 6 parameters, flat, ΛCDM – "base" – model) and simple extensions of it, where one or two quantities, which have fixed values in the "base" model, are promoted to parameters of the model. In this case model selection is carried out by computing the Bayes factor, or evidence ratio, between the two models. The question this approach addresses is the model selection question: "is the introduction of the extra parameter(s) warranted by data?". However, there is also another question we can ask (and we are interested in addressing here): "Does the introduction of the extra parameter reduces the tension as defined in Eq. (4)?" And further, "what are the fixed values (if any) of the extra parameter that would make the tension not significant?". In this paper we will concentrate on the last two questions. We will refer to this case as "blame the model".

4. Results

We begin by repeating some of the key steps of the analysis of Ref. [1] using the updated state-of-the-art data. Fig. 1 shows constraints on the t_U–H_0 plane from local measurements and from CMB in the framework of the standard ΛCDM model; both WMAP and *Planck* constraints are shown. In Appendix A we quantify how much information Planck has added to WMAP for this particular parameter combination within the ΛCDM model. For now, we can appreciate that the *Planck* central value has shifted compared to WMAP's. This shift is well within the WMAP 1σ confidence region, but the reduced *Planck* error-bars mean that now the 1σ confidence regions of CMB and local measures do not overlap (only the 2σ joint still do). This represents the above mentioned "tension". We will return on this in Section 4.1 below.

The smallness of the *Planck* allowed region on this plane is due to the assumption of the ΛCDM model. In Fig. 2 we show how this changes for simple (one or two parameters) extensions to the ΛCDM model. Among the extensions considered, non-standard effective neutrino species and non-standard equation of state parameter for dark energy, bring the CMB and the local measures closer.

This is further illustrated in Fig. 3 (top panels) for the effective number of neutrino species (N_{eff}) extension to ΛCDM. A N_{eff} value larger than the standard 3.046 brings in better agreement the H_0 determinations, but the agreement worsens for t_U.

Similarly the bottom left panel of Fig. 3 shows the effect for the (total) neutrino mass, M_ν, extension and the bottom right panel for the non-standard equation of state parameter, w, extension of ΛCDM. Clearly the constraints on M_ν obtained using the local H_0 determination are very tight thanks to the (local) H_0 central value. Also values of $w < -1$ bring the two H_0 estimates in better agreement, in this case, the t_U constraint is not useful given the direction of the CMB

---

[3] We use the MCMCs made public by the Planck collaboration: a detailed discussion on the priors used on cosmological parameters can be found in [12]. Uniform priors were used on the physical densities of baryons and cold dark matter, on the angular size distance to the last scattering surface, on the logarithm of the amplitude of primordial perturbations, on the primordial power spectrum spectral slope, on the logarithm of the integrated optical depth to the last scattering surface, on the density parameter of dark energy, on the total neutrino mass, on the equation of state parameter of dark energy and on the number of effective neutrino species. The width of these priors do not matter in the Tension calculation because of the empirical normalization choice and because the prior ranges always encompass the region where the likelihood is significantly non-zero.
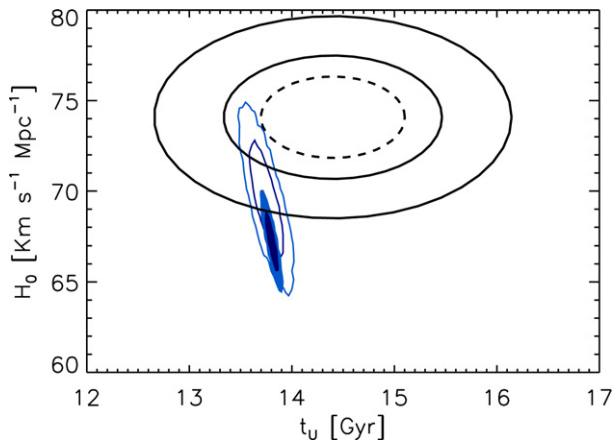
Fig. 1. Constraints (1 and 2σ joint) in the t_U–H_0 plane from local measurements (black solid contours, the dashed contours corresponds to the single parameter, marginalized $1 - \sigma$ constraint; $\triangle \ln \mathcal{L} = 0.5$) and CMB data (blue). The transparent set of contours correspond to WMAP and the filled contours to Planck.

degeneracy.

4.1. Is there Evidence for tension?

We can use the method outlined in Section 3.1 to quantify if there is Evidence for tension between the local and CMB measurements in the H_0–t_U plane. For the "base" ΛCDM model we obtain $\mathcal{T} = 53$, $\ln \mathcal{T} = 3.95$, i.e. that the null hypothesis, H_1, is disfavored with odds of roughly 1 to 50. This indicates a *strong* evidence for tension. We estimate[4] that the numerical error on $\ln \mathcal{T}$, due to the fact that the CMB posterior is not known with infinite precision but sampled by MCMC, is 0.1.

To gain a physical intuition about this result, let us assume that the *Planck's* one-dimensional posterior distribution for H_0 (marginalized over all other parameters) and that for t_U are Gaussian. This is a good approximation for the ΛCDM model. The local age determination is in good agreement with *Planck's*: $\ln \mathcal{T} = 0.34$. However for H_0 we obtain $\ln \mathcal{T} = 3.5$[5] Thus the tension between CMB and local measurements is entirely due to H_0.

This *strong* evidence for tension is a signal that caution must be exercised if the results of the two experiments are to be combined, and the result of this combination should be interpreted with care. We will proceed examining the two cases (*a*) and (*b*) outlined above in turn.

It is important to note that this measurement of tension is model-dependent and the value reported here applies only to the ΛCDM model. Even in simple extensions of the models, $\mathcal{T}$ can differ widely. We will return to this in Section 4.3.

4.2. Option "Blame the measurement": Interpretation in terms of unknown errors

Discarding measurements (or combining measurements) can be seen as two special cases of the use of hyper-parameters, as described in [29]. The Planck team noticed some tension between Planck data and the H_0 measurement and argued that the local measurement is more likely to be affected by some unknown systematic than the CMB. For this reason they discard H_0 when combining Planck with other measurements. They nevertheless also provided results and MCMC

outputs for data combinations that include H_0. There could however be a spectra of intermediate possibilities where the H_0 measurement is combined but downweighted by a factor α (or equivalently its error increased by $1/\alpha$). Clearly $\alpha = 1$ corresponds to doing the standard joint analysis and $\alpha = 0$ to excluding the measurement.

Here we follow this train of thoughts and increase the H_0 errors to find out what correction would be needed to reduce the tension $\ln \mathcal{T}$ in the H_0–t_U plane, to a more "comfortable" level. This is shown in Fig. 4. Of course we could have decided to downweight the CMB and/or the age measurements. The age measurements and the CMB are in good agreement with each other – but t_U has still large error-bars – so we would have had to downweight both. We decided here to follow the reasoning of the Planck team, downweight H_0 and interpret the consequences. Another possibility is to model an unknown systematic error by shifting the H_0 measurement by ΔH_0 and see what shift is needed to significantly reduce the tension. This is also shown in Fig. 4.

This figure shows that an increase in the H_0 error by 30–40% would be needed to reduce $\ln \mathcal{T}$ to *substantial* but a factor 2.5 would be needed for *not significant*. On the other hand a shift of the central value by little less than 1σ would bring $\ln \mathcal{T}$ to *substantial* but a shift of almost $\sim 2\sigma$ would be needed for it to become *not significant*.

4.3. Option "blame the model": Extending the "base" model

The values of the tension obtained so far are valid within the ΛCDM model. There may be simple extensions of this model that reduce or remove tension.

4.3.1. Does tension decrease in ΛCDM extensions?

In the first line of Table 2 we report $\ln \mathcal{T}$ for simple extensions of the ΛCDM model. We see that including the curvature, Ω_K as a parameter or including a parameter for non-zero neutrino mass, M_ν, increases $\ln \mathcal{T}$, thus disfavoring these model extensions. The two model extensions, however, should not be considered on the same footing: inflation strongly motivates the assumption that the Universe is flat (and therefore there is no need for the extra curvature parameter). On the other hand neutrino oscillations indicate that neutrinos have a non-zero mass, M_ν must be larger than about 0.05 eV and the cosmologically-independent upper limit is $\sim$ 2eV. Therefore assuming a ΛCDM model with massless neutrinos, or with M_ν fixed at a value close to its lower limit, is not really motivated and one could argue that ΛCDM $+ M_\nu$ model should be the "base" model.

The addition of the effective number of neutrino species N_{eff} as a parameter brings the tension down to *substantial*, and the addition of the equation of state for dark energy parameter w brings it to *not significant*. This indicates that these model extensions are particularly interesting and warrant more investigation (under case (*b*), of course). Contrary to the ΛCDM $+ M_\nu$ case, these model extensions do not have other strong experimental motivations.

It is interesting to compare these findings with the standard Bayesian model selection.

4.3.2. Bayesian model selection

We carry out the standard Bayesian model selection computing the Bayes factor (i.e. the Evidence ratio) between the "base" ΛCDM model and its extensions. We start by producing versions of the relevant MCMC chains importance-sampled with the local constraints. We then follow Ref. [11] to compute the Evidence ratio between two nested models from an MCMC output via the Savage–Dickey density ratio. The Evidence ratio for ΛCDM extensions involving neutrino properties differ from those in Ref. [11] because here we also include the t_U determination. The results are reported in Table 2 (second line). We see that even combining *Planck* with local universe measurements, there is never substantial or strong evidence for the model
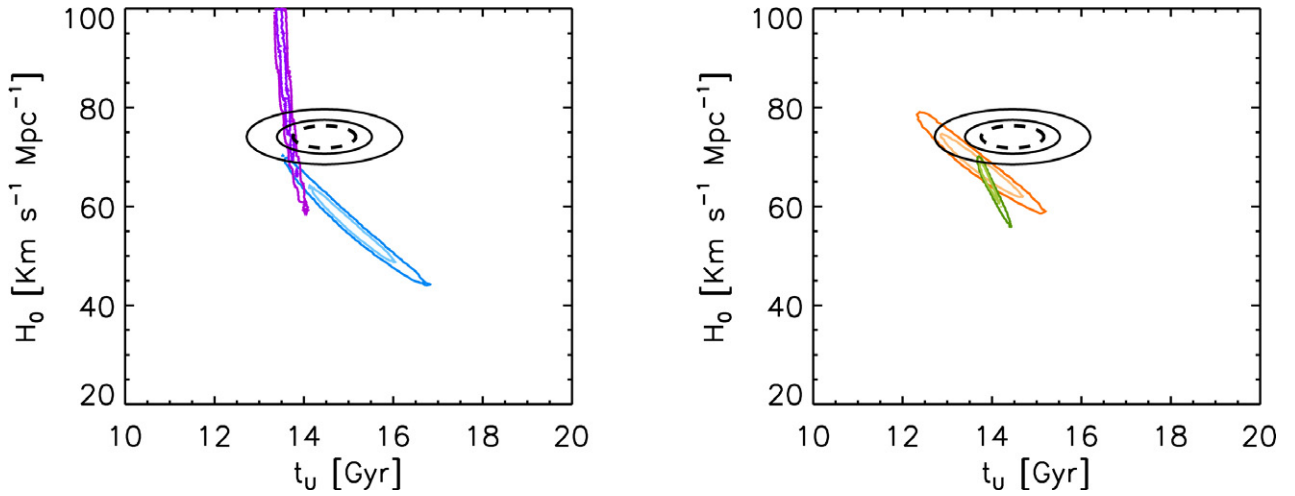
---

[4] We have performed several tests: We have both performed integrals on fits to the surface, on Gaussian approximations and using directly the MCMC outputs. We have done so both using the full MCMC's samples of the posterior surface and splitting the MCMC's.

[5] $\text{Ln} \mathcal{T} = 3.5$

Fig. 2. Left panel: Blue: curvature extension to the ΛCDM model, magenta: equation of state parameter for the dark energy w extension to the ΛCDM model. Right panel: Non standard neutrino properties. Green: neutrino mass and primordial helium content extension and Orange: number of effective neutrino species and primordial helium content extension. The plot range and color scheme have been chosen so these figures can be compared directly, at a glance, with Fig. 3 of Ref. [1] for a direct comparison with WMAP.
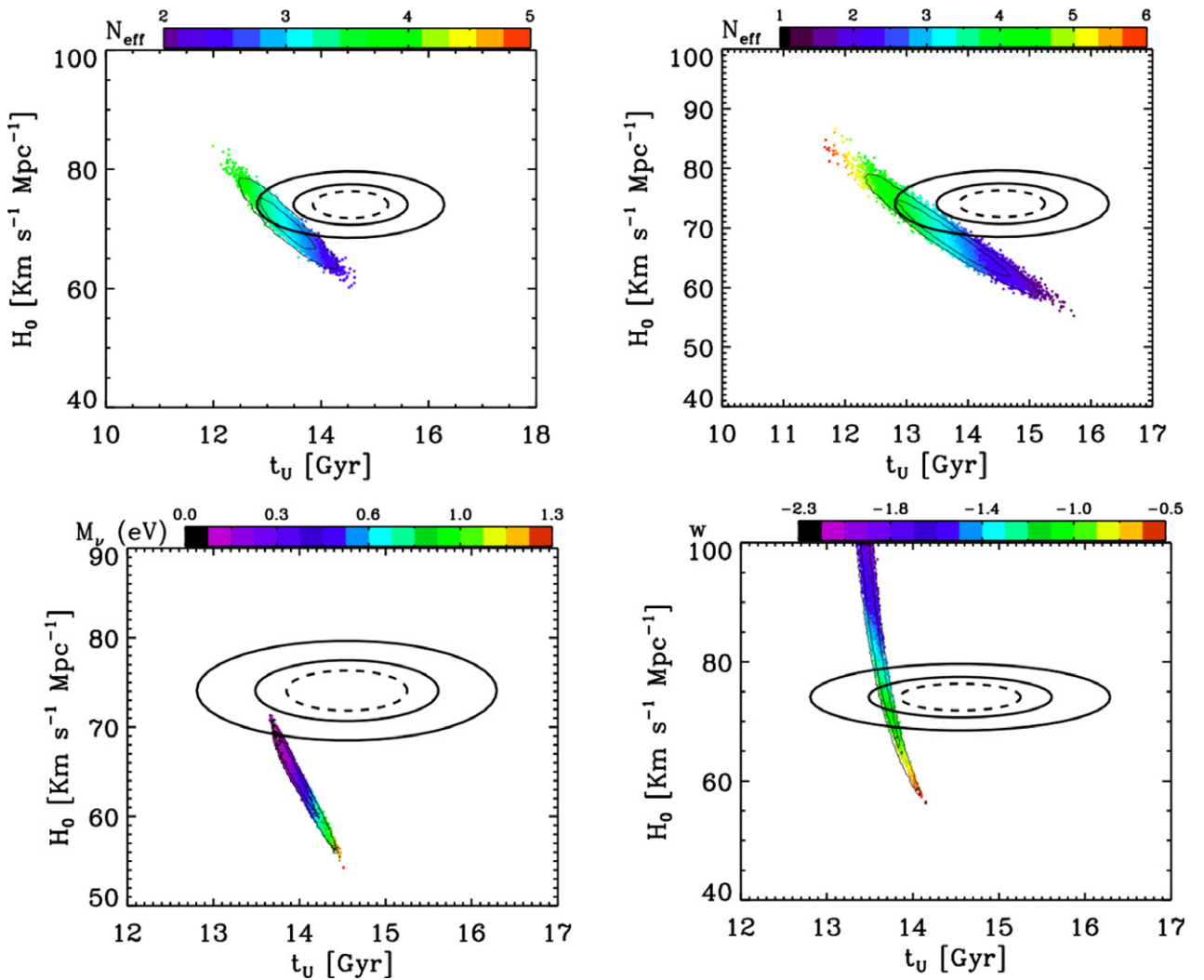
Fig. 3. Posterior distributions in the t_U–H_0 plane for local and CMB measurements. A random sub-sample of the CMB MCMC points has been shown color-coded by the value of the effective number of neutrino species N_{eff} on the top panels. In the left panel the primordial helium fraction is kept fixed at the nucleosynthesis value while on the right it is left as a parameter which is then marginalized. A larger N_{eff} value brings in better agreement the H_0 determinations (but the agreement worsens for t_U). On the bottom panels we show the M_ν (left) and w (right) extension to the ΛCDM model.
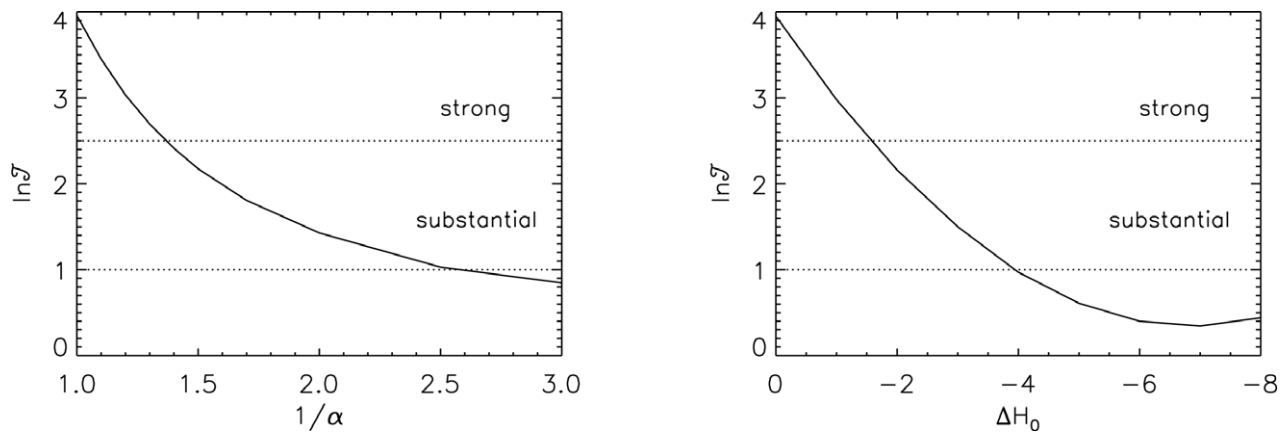
Fig. 4. Ln $\mathcal{T}$ for Planck and the local measurements as a function of the factor downweighting the measurement of H_0 or as a function of a shift in H_0.

Table 2
Tension (ln $\mathcal{T}$) between Planck and the local measurements in ΛCDM extensions and Evidence ratio between the ΛCDM model and its extensions. For the tension, numbers above 2.5 signify *strong* tension and above 5 *highly significant* tension. The last column reports the ΛCDM numbers for reference. For the Evidence ratio positive numbers mean that the simpler model is preferred, negative numbers that the more-complicated model is preferred; $|\Delta \ln E|$ should be at least 2.5 to be *strong*. The errors on the evidence ratios are in most cases about ± 0.02, reaching ± 0.1 for thinned, post-processed chains (indicated by P).

Model extension	w	Ω_k	N_{eff}	M_ν	$N_{\text{eff}} + Y_P$	ΛCDM
Ln $\mathcal{T}$	0.74	5.24	1.94	4.5	2.2	3.95
ln $E_{\Lambda \text{CDM}}/E_{\text{extension}}$	-0.72	3.70	−0.27(P)	3.45	1.93(P)	0

extension and in some cases the simpler model ΛCDM is strongly favored.

This is not necessarily in contradiction with the findings of Section 4.3.1 as the two quantities measure different things and address different questions. In particular the Bayesian Evidence is not concerned with whether at least one of the two models considered provides a good fit to all the data simultaneously. But this is exactly what ln $\mathcal{T}$ does.

Qualitatively however the trends of the two statistics are similar: the models more disfavored by the Bayesian Evidence are those for which the tension increases with respect to the ΛCDM case. The models that decrease the tension most are those that favor (only ever so slightly) the more complex model.

4.3.3. Tension and its implications for models and parameters

The fact that for some ΛCDM extensions tension is reduced (or even eliminated) but not for others, suggest that we can use the quantity ln $\mathcal{T}$ for deriving constraints on the parameters of the model extension.

For example we can ask: "if we had a compelling reason to fix the extra parameter to a given value, what would the tension be?" Clearly, parameters values that yield "very strong" tension should be disfavored. In other words, if the ΛCDM model showed negligible or no tension, this plot would indicate what offset from the LCDM value of the parameter in question would give tension. This is what we show in Figs. 5 and 6.

In Fig. 5 we show results for ΛCDM model extensions involving non-standard neutrino properties. In the two panels we fix M_ν or N_{eff} at a different fiducial value than the standard ΛCDM model then treat the resulting model as a modified ΛCDM for the purpose of computing the tension. We find that the lower the neutrino mass the lower ln $\mathcal{T}$. For values of M_ν higher than 0.15 eV, the tension between local measurements and Planck derived values increases to *highly significant* (odds $\sim$ 1:150). This indicates that the degenerate hierarchy for the neutrino mass spectrum is highly disfavored and that normal hierarchy is preferred over the inverted one (although no value for M_ν yield *not significant* or even *substantial* tension).

Values of N_{eff} larger than the standard value ($3.4 < N_{\text{eff}} < 4.1$)

reduce ln $\mathcal{T}$, but values $N_{\text{eff}} > 4.6$ make it *highly significant*. Note that for no range of N_{eff} values the tension is *not significant*. This is because among all the N_{eff} values that are a good fit to Planck data, values higher than the fiducial improve the fit to H_0 data but worsen the fit to t_U.

In Fig. 6 we show results for the dark energy equation of state parameter w, and curvature Ω_k, ΛCDM model extensions. Both these model extensions have parameter values that make the tension *not significant*. A slightly "phantom" ($w < -1$) value for the equation of state parameter brings the tension to more comfortable *substantial* and values around $w = -1.2$ to *not significant*. Alternatively a slightly positive curvature $5 \times 10^{-3} < \Omega_k < 1.5 \times 10^{-2}$ brings the tension to a more comfortable (*substantial*) level.

5. Discussion and conclusions

We have analised cosmology independent, local measurements of H_0 and t_U and the Planck derived values for these quanities within the ΛCDM model and in simple extensions of it. We started by introducing and developing a statistic, the tension $\mathcal{T}$, to determine when two posterior distributions of parameters are in tension and to quantify the level of tension. We argue that the tension is based on the Bayesian evidence and the Bayes factor and, as such, it can be interpreted using the popular Jeffreys' scale. To give an intuition about the meaning of ln $\mathcal{T}$ values we should bear in mind that if we have two 1-dimensional Gaussian distributions with unit variance, a shift of the central value of one of the two distributions of 2σ would give ln $\mathcal{T} = 1$ which is the threshold between *not significant* and *substantial*; a shift of more than 3σ would yield *strong* tension and of more than 4.5σ to give *highly significant* tension.

We then have shown that, in agreement with the(recently) commonly accepted wisdom, in the framework of the ΛCDM model, these two determinations are in *strong* tension (odds $\sim$ 1 : 50). In our naive Gaussian interpretation above these odds corresponds to a shift of one of the two unit-variance Gaussians if $\sim 4\sigma$.

We recognize that two broad classes of explanations for this result are possible. "Blame the measurement": one (or more) of the measurements have errors that are unaccounted for or "blame the
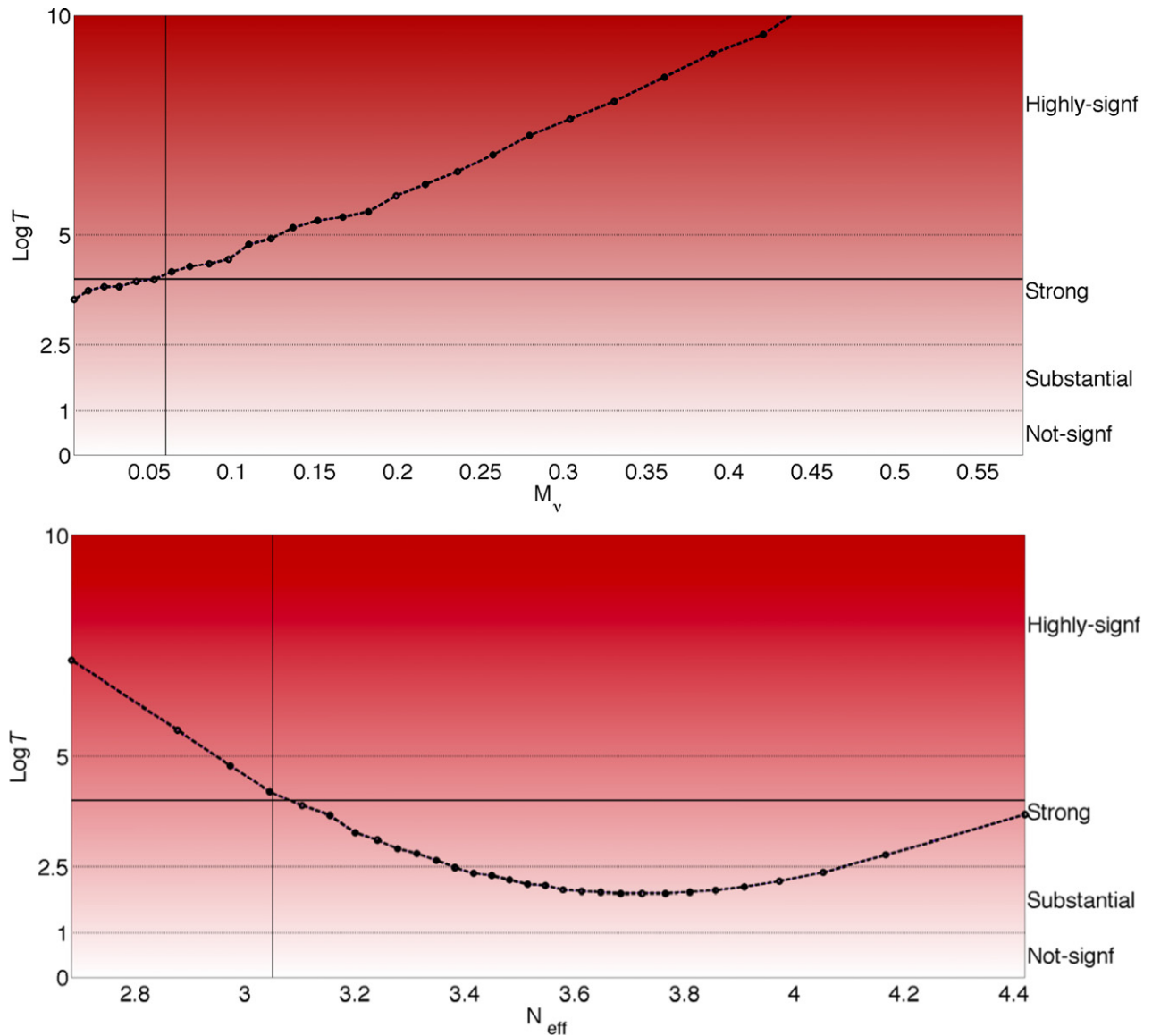
Fig. 5. $\mathrm{Ln}\,\mathcal{T}$ as a function of M_ν (top panel) and as a function of N_{eff} (bottom panel). The black solid line corresponds to the ΛCDM value, odds ~ 1: 50. Note that for values of M_ν higher than 0.15 eV, the tension between local measurements and Planck derived values increases to *highly significant* (odds $\sim$ 1:150). This indicates that the degenerate hierarchy for the neutrino mass spectrum is highly disfavored and that normal hierarchy is preferred over the inverted one. However no value of M_ν yields *non-significant'* or even *substantial* tension.

model": extensions of base ΛCDM model should be considered.

Following the interpretation suggested by the *Planck* team we have then explored how to alleviate the tension by exploring by how much the errors and/or the central value of H_0 needs to be changed in order to alleviate the tension to tolerable levels, odds > 1:3 (>1:10). We found that this can be achieved if the error bars have been underestimated by a factor 2.5 (30–40%) or the central value is wrong by 4 ($\sim$2) km s^{-1} Mpc^{-1}. Having quantified this, we leave it to the experts in the field of the H_0 measurements to judge whether this is a realistic possibility, keeping in mind that two independent measurements, in excellent agreement with each other, were suitably combined to obtain the adopted H_0 constraint.

We have then explored extensions to the ΛCDM that can alleviate this tension. We found that several extra parameters can achieve this: allowing for phantom values of the equation of state of dark energy ($w \sim -1.2$) or allowing a small positive curvature ($7 \times 10^{-3} < \Omega_k < 1.5 \times 10^{-2}$).

No values of the effective number of species reduces the tension to not significant, but values around 3.6–3.8 get close to that. An interesting finding is that when using the neutrino mass as an additional parameter for the ΛCDM model, although no values for the total neutrino mass reduce the tension to *not significant* (odds 1 : 3), a total mass above 0.15 eV makes the tension *highly significant* (odds $\sim$ 1:150). A consequence of accepting this interpretation of the tension is that the degenerate neutrino hierarchy is highly disfavored by cosmological data and the direct hierarchy is slightly favored over the inverse one. Of course, if we accept this interpretation of the "blame the model" option, this could be the first indication from cosmology for a neutrino hierarchy (e.g., [30] and references therein).

There is one possible explanation for the tension that lies in-between "blame the data" and "blame the model" options, that is that we are in a local underdensity. Ref. [31] shows that measurements of the local Hubble constant are subject to a cosmic variance error with can be as high as 2%. If, simply by cosmic variance, we happen to live inside an underdensity, the local values of H_0 could be higher than the cosmological value by this amount. As we have seen
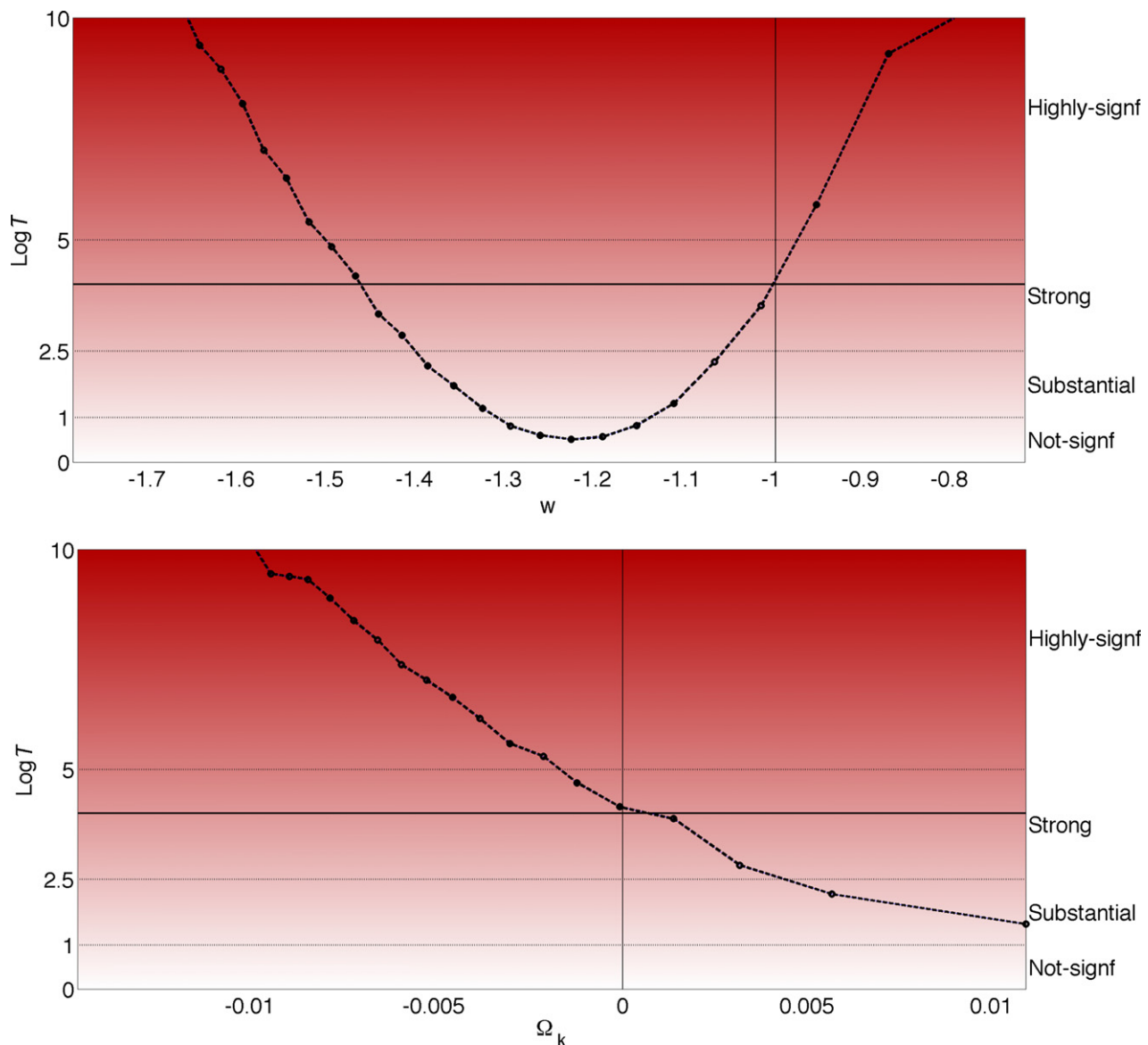
Fig. 6. Same as Fig. 5 but now changing the equation of state parameter for dark energy w (top panel) and the curvature parameter Ω_k (bottom panel). A slightly "phantom" ($w < -1$) value for the equation of state parameter bring the tension to more comfortable "substantial" and and values around $w = -1.2$ to "not significant". Alternatively a slightly positive curvature $\Omega_k > 8 \times 10^{-3}$ brings the tension to "not significant".

in Fig. 4, such shift would certainly reduce the tension but not remove it. The larger shift needed for this, the authors of [31] argue, would however require a very rare (i.e., unlikely) fluctuation for the model.

Tantalizingly, observations of the luminosity density as function of redshift [32] suggest that we might be located in a ∼300 Mpc/h local under density. It remains to be assessed whether such under density is due to cosmic variance, can be accommodated in the ΛCDM scenario and thus only adds a (small) systematic correction to the local H_0 measurement (case "blame the measurement"). On the other hand, if the underdensity is large enough, it could eliminate the tension but at the expense of requiring indeed an (interesting) extension to the ΛCDM model (case "blame the model").

We have focused our analysis only on local, cosmology independent, measurements and the CMB derived local universe within a given model. We have chosen not to use other cosmological probes as BAO or Supernova as they are not local measurements and not as mature from a theoretical point of view (and not as powerful when used

alone) as the CMB to be used as high-redshift measurements. Moreover we have not used the high-ℓ CMB data (*highL*) nor the CMB lensing information because of possible internal tension with the *Planck* temperature data (although we have checked that adding *highL* data does not change significantly or qualitatively our findings). The *Planck* experiment statistical power will improve drastically when polarization data will be included in the analysis over the next year. The inclusion of these extra *Planck* data will cement the CMB constraints (strengthening or weakening the above findings).

After that, if tension remains, a way forward is to improve local data to the % level. The ages of the oldest stars show no tension with Planck CMB data, but error-bars are still relatively large; thus much more accurate ages from the GAIA mission will help elucidate the situation.

Appendix A. How many bits of information has Planck added to WMAP?

The Kullback–Leibler divergence [23] $D_{KL}(\mathcal{P} \parallel \mathcal{W})$, as mentioned in Section 3, quantifies the information content that one distribution ($\mathcal{P}$) adds to the other one ($\mathcal{W}$); in other words, how many bits of information are needed if we are given the second distribution ($\mathcal{W}$) and we want to recover the first one ($\mathcal{P}$), or, how many bits are lost if one were to use $\mathcal{W}$ to approximate $\mathcal{P}$. Here we are interested in using this statistic to quantify how much new information has Planck added to WMAP. We will use $D_{KL}(\mathcal{P} \parallel \mathcal{W})$, where $\mathcal{P}$, $\mathcal{W}$ denote the Planck(+ WP) and WMAP posterior distributions,

$$D_{KL}(\mathcal{P} \parallel \mathcal{W}) = \int_x \log_2 \left(\frac{\mathcal{P}(x)}{\mathcal{W}(x)} \right) \mathcal{P}(x)\, dx. \qquad (5)$$

The logarithm in base 2 means that the information is expressed in bits. The above equation can be interpreted and applied in several ways for example it can be applied to the one dimensional distribution for one parameter marginalized over all the other ones, or as the joint distribution of two parameters or as the joint, multivariate distribution of all the cosmological parameters. Here we present only few selected cases, see Table A.1. In the first block we report the D_{KL} parameter by parameter in the "base" ΛCDM model. The one dimensional posterior used has been marginalized over all other parameters. The second block includes the D_{KL} for simple extensions of the ΛCDM model. We report the one dimensional posterior of the extra parameter marginalized over all other parameters. First we report the results for Planck + WP, then the addition of highL and the further addition of lensing.

These numbers should be interpreted as follows: they represent the extra bits of information added, which do not need to be whole numbers, by Planck over WMAP in bits in base 2. So, for example, if 1 is added, the information is multiplied by a factor 2^1. In the case of N_{eff}, Planck has increased the information of WMAP by a factor $2^{1.7} = 3.2$. This is the parameter for which most information has been added, while for the other parameters Planck has increased information by about a factor 2. If we include Planck's highL, the one-parameter ΛCDM numbers are mostly unchanged except for the (scalar) power spectrum spectral slope, n_s, which becomes 1.36. Including also lensing does not add significant information for this model. This is not the case for the ΛCDM model extensions. The effect of the previously mentioned "tension" between Planck + WP + highL and lensing can be seen in the M_ν column.

Table A.1

Kullback–Leibler divergence of Planck from WMAP, $\mathcal{D}_{KL}(\mathcal{P} \parallel \mathcal{W})$. Here w_b denotes the physical density of baryons, w_c the physical density of cold dark matter, n_s the primordial matter power spectrum spectral slope, H_0 the Hubble constant, t_U the age of the Universe, Ω_k the curvature parameter, w the dark energy equation of state parameter, N_{eff} the effective number of neutrino species and M_ν the total neutrino mass. The "base" data set corresponds to Planck + WP; the addition of highL and lensing are also considered.

ΛCDM	w_b	w_c	n_s	H_0	t_U
1D for parameter	1.35	1.63	1.09	1.21	0.83

ΛCDM Extension	Ω_k	w	N_{eff}	M_ν
1D for parameter	0.67	1.05	1.70	0.39
+ highL	0.82	1.18	1.90	0.91
+ lensing	1.07	1.15	1.94	0.38

Table A.1 should not be interpreted as that Planck has improved over WMAP by a factor 2. In fact the reported numbers are for individual parameters, but the experiment measures all the parameters of the model (typically 6 or 7) as well as non-cosmological but nevertheless astrophysically interesting parameters, and all are improved.

References

[1] L. Verde, R. Jimenez, S. Feeney, Phys. Dark Universe (2013).
[2] R.T. Cox, Am. J. Th. Phys 14 (1946) 1.
[3] A.H. Jaffe, Astrophys. J. 471 (1996) 24, (astro-ph/9501070).
[4] M.P. Hobson, S.L. Bridle, O. Lahav, MNRAS 335 (2002) 377.
[5] A.R. Liddle, Mon, Not. Roy. Astron. Soc. 351 (2004) L49, (astro-ph/0401198).
[6] D. Parkinson, P. Mukherjee, A.R Liddle, Phys. Rev. D 73 (2006) 123523, (astro-ph/0605003).
[7] R. Trotta, A. Melchiorri, Phys. Rev. Lett. 95 (2005) 011305, (astro-ph/0412066).
[8] R. Trotta, Mon. Not. Roy. Astron. Soc. 378 (2007) 72, (astro-ph/0504022).
[9] P.A.R. Ade et al., Planck collaboration, arXiv:1303.5062 (astro-ph.CO).
[10] S.M. Feeney, H.V. Peiris, L. Verde, JCAP 1304 (2013) 036, .
[11] L. Verde, S.M. Feeney, D.J. Mortlock, H.V. Peiris, JCAP 1309 (2013) 013..
[12] P. A. R. Ade et al., Planck collaboration, arXiv:1303.5076 (astro-ph.CO).
[13] S. Feeney, H.V. Peiris, L. Verde, in preparation (2013).
[14] P. A. R. Ade et al., Planck collaboration, arXiv:1303.5077 (astro-ph.CO).
[15] A.G. Riess, L. Macri, S. Casertano, ApJ 730 (2011) 119.
[16] W.L. Freedman, B.F. Madore, V. Scowcroft, ApJ 758 (2012) 24.
[17] L.M. Krauss, B. Chaboyer, Science 299 (2003) 65.
[18] R. Jimenez, P. Thejll, U. Jorgensen, J. MacDonald, B. Pagel, Mon. Not. Roy. Astron. Soc. 282 (1996) 926, (astro-ph/9602132).
[19] H.E. Bond, E.P. Nelan, D.A. VandenBerg, G.H. Schaefer, D. Harmer, ApJL 765 (2013) L12.
[20] R.G. Gratton, A. Bragaglia, E. Carretta, G. Clementini, S. Desidera, F. Grundahl, S. Lucatello, Distances and ages of NGC 6397, NGC 6752 and 47 Tuc Astron. Astrophys. 408 (2003) 529–543.
[21] G. Imbriani, The bottleneck of CNO burning and the age of globular clusters Astron. Astrophys. 420 (2004) 625–629.
[22] Z. Haiman, Cosmology: a smoother end to the dark ages Nature 472 (2011) 47–48.
[23] S. Kullback, R.A. Leibler, On information and sufficiency Ann. Math. Stat. 22 (49) (1951) 86.
[24] H. Jeffreys. Scientific Inference. Cambridge University Press, 1973.
[25] J. Lin, Divergence measures based on the Shannon entropy IEEE Trans. Inf. Theory 37 (1991) 145151.
[26] P. Marshall, N. Rajguru, A. Slosar, Phys. Rev. D 73 (2006) 067302.
[27] M.C. March, R. Trotta, L. Amendola, D. Huterer, Mon, Not. Roy. Astron. Soc. 415 (2011) 143.
[28] R.E. Kass, A.E. Raftery, Bayes factors JASA 90 (430) (1995) 773–795.
[29] O. Lahav, S.L. Bridle, M.P. Hobson, A.N. Lasenby, L. Sodré, MNRAS 315 (2000) L45.
[30] R. Jimenez, T. Kitching, C. Pena-Garay, L. Verde, JCAP 1005 (2010) 035.
[31] V. Marra, L. Amendola, I. Sawicki, W. Valkenburg, Phys. Rev. Lett. 110 (2013) 241305.
[32] R.C. Keenan, A.J. Barger, L.L. Cowie, 2013, arXiv:1304.2884.