Early recombination as a solution to the $H_0$ tension

Toyokazu Sekiguchi$^1$ and Tomo Takahashi$^2$

$^1$Theory Center, IPNS, KEK, Tsukuba 305-0801, Japan
$^2$Department of Physics, Saga University, Saga 840-8502, Japan

(Dated: July 8, 2020)

We show that the $H_0$ tension can be resolved by making recombination earlier, keeping the fit to cosmic microwave background (CMB) data almost intact. We provide a suite of general necessary conditions to give a good fit to CMB data while realizing a high value of $H_0$ suggested by local measurements. As a concrete example for a successful scenario with early recombination, we demonstrate that a model with time-varying $m_e$ can indeed satisfy all the conditions. We further show that such a model can also be well fitted to low-$z$ distance measurements of baryon acoustic oscillation (BAO) and type-IIa supernovae (SNeIa) with a simple extension of the model. Time-varying $m_e$ in the framework of $\Omega_0\Lambda$CDM is found to be a sufficient and excellent example as a solution to the $H_0$ tension, yielding $H_0 = 72.3^{+2.7}_{-2.8}$ km/sec/Mpc from the combination of CMB, BAO and SNeIa data even without incorporating any direct local $H_0$ measurements. Apart from the $H_0$ tension, this model is also favored from the viewpoint of the CMB lensing anomaly.

Introduction.— The Hubble-Lemaître constant $H_0$ is one of the most relevant cosmological parameters characterizing the Universe. It has long been studied by the distance ladder, which now utilizes Cepheids and type Ia supernovae (SNeIa) as standard candles [1]. Meanwhile, many other means have been devised. For instance, gravitational lens time-delay now rivals the distance ladder in local (almost direct) measurements of $H_0$ [2]. Moreover, the cosmic microwave background (CMB) and the baryon acoustic oscillations (BAO), allow us to measure cosmic distances to very different redshifts ($z \lesssim 10^3$ and $z \lesssim 2$) based on the scales of the sound horizon of photon-baryon fluid, $r_s$, at recombination $z = z_s$ and the drag epoch $z = z_{drag}$, respectively. Consistency in the cosmic expansion history over such a huge range of redshifts enables us to infer $H_0$.

However, as measurements of $H_0$ become more precise, disagreements grow apparent between local direct measurements and other indirect ones such as CMB. The value of $H_0$ from local measurements, $H_0 = (73.8 \pm 1.0)$ km/sec/Mpc [3], is about 10% larger than that from CMB, $H_0 = (67.36 \pm 0.54)$ km/sec/Mpc [4], assuming the canonical flat $\Lambda$CDM ($\Lambda$CDM hereafter) model. Significance of the $H_0$ tension is now more than 5σ. Interestingly, different and independent measurements appear consistent within either local or indirect measurements (For recent review, see [3]). This indicates that single systematic error alone cannot remove the tension.

A number of cosmological solutions have been proposed already. However, it seems extremely difficult to solve the tension when one combines various observations such as CMB, BAO and SNeIa. The reason for the difficulty has been clarified in [3] [4]. SNeIa and distance ladder jointly measure luminosity distance seamlessly at $z \lesssim 2$. This gives the transverse distance $D_M(z)$ at the redshifts of BAO measurements, where $D_M(z)$ is given by

$$D_M(z) = \begin{cases} 
\sin \left( \frac{\sqrt{-\Omega_k}H_0\chi(z)}{\sqrt{-\Omega_k}H_0} \right) & \text{for } \Omega_k < 0 \text{ (closed)} \\
\frac{\sqrt{-\Omega_k}H_0\chi(z)}{\sqrt{-\Omega_k}H_0} & \text{for } \Omega_k = 0 \text{ (flat)} \\
\sinh \left( \frac{\sqrt{-\Omega_k}H_0\chi(z)}{\sqrt{-\Omega_k}H_0} \right) & \text{for } \Omega_k > 0 \text{ (open)}
\end{cases}$$

with $\chi(z) = \int_0^z dz \frac{dz}{\Omega(z)}$ being the comoving distance to $z$. This enables model-independent estimation of $r_s(z_{drag})$. Enhancing $H_0$ by 10% requires decreasing $r_s(z_s) \propto r_s(z_{drag})$ by the same rate, which is very difficult keeping a reasonable fit to CMB. This also explains why models modifying only late-time expansion can increase $H_0$ only marginally.

Considerations above lead to following four necessary conditions which successful cosmological solutions to the $H_0$ tension should satisfy:

1. In order not to spoil the successful fit achieved by $\Lambda$CDM, CMB power spectra should be left almost intact except at low-$\ell$, where cosmic variance is large.
2. $r_s(z_s) \propto r_s(z_{drag})$ is reduced by $\simeq 10\%$.
3. $D_M(z_s)$ is reduced, so that $\theta_s(z_s) = r_s(z_s)/D_M(z_s)$ is kept constant (this is somewhat redundant with the condition 1).

1 Given the baryon drag at $z_s$, $R(z_s) = 3\rho_b(z_s)/4\rho_r(z_s)$, which is very precisely determined by CMB power spectra, specifying either of $r_s(z_s)$ or $r_s(z_{drag})$ virtually determines the other.
4. BAO, SNeIa, and other low-z distance measurements should be satisfied.

With the first condition being met, the second condition is quite difficult to be satisfied. Many attempts have tried to increase the expansion rate by e.g., adding extra energy components (See [2] for review). However, these modifications have some limitations since the relative scale between the sound horizon and the Silk scale, or the photon diffusion length, also varies, which inevitably violates the first condition [6]. This is the reason why those attempts can mitigate the $H_0$ tension only somewhat partially.

We in this Letter pursue a cosmological solution to the $H_0$ tension, in particular focusing on modified recombination (See earlier studies [7,8] but without concrete models). We first argue how one can modify recombination epoch keeping CMB power spectra almost unchanged. Then as a working example, we discuss a model with time-varying electron mass $m_e$ (for possible models of time-varying $m_e$, see, e.g., a recent review [9])\(^2\), which can shift $z_s$ and $z_{\text{drag}}$ from the baseline model sizeably without affecting CMB power spectra much.

In the following, we often refer to the Planck 2018 best-fit $\Lambda$CDM model [4] as the baseline. The reduced Hubble constant and density parameters are given by $h = H_0/[100 \text{ km/sec/Mpc}]$ and e.g. $\omega_i = \Omega_i h^2$ for component $i$, respectively. Let $\Delta_x$ denote the fractional variation in a quantity $x$ from the baseline value [e.g. $\Delta_m = \log(m_e/m_{e,\text{baseline}})$].

Effects of early recombination on CMB. Let us discuss effects of early recombination on CMB and how to cancel those effects by varying cosmological parameters. In the analytical argument below, we utilize the scale factor at recombination $a = a_*$, which is useful since it can well capture effects of modified recombination on CMB.

CMB observations tightly constrain the following two quantities at the recombination $a = a_*$:

\begin{align}
R(x) &= \frac{3\omega_b a_*}{4\omega_r} x, \\
[a^2H]_x &= \frac{1}{L \sqrt{\omega_m a_* x + \omega_r}},
\end{align}

where $x \equiv a/a_*$ is the scale factor normalized to unity at recombination and $L = (H_0/h)^{-1} \approx 2998 \text{ Mpc}$ is a constant length. The former gives the baryon drag, which is measured by the relative heights of even and odd acoustic peaks. The latter determines the early integrated Sachs-Wolfe (ISW) effect, which is measured by the heights of acoustic peaks relative to the SW plateau. From Eqs. (2) and (3), we can leave both $R$ and $a^2H$ unaffected as functions of $x$ by varying $\omega_b$ and $\omega_m$ inversely proportionally to $a_*$:

$$\Delta \omega_b = \Delta \omega_m = -\Delta a_*.$$  \hspace{1cm} (4)

Now we consider the sound horizon at the recombination epoch:

$$r_s(z_*) = \frac{a_*}{\sqrt{3}} \int_0^1 \frac{1}{\sqrt{1 + R(x)}} \frac{dx}{[a^2H](x)}.$$  \hspace{1cm} (5)

from which we can immediately see $r_s(z_*) \propto a_*$ when we vary $\omega_b$ and $\omega_m$ in accord with Eq. (4) (i.e., $R$ and $a^2H$ remain unchanged as functions of $x$). Not to change CMB power spectra, the relative scale of the Silk scale, $1/d_{D_*}$, to $r_s$, should be kept unchanged, where

$$\frac{1}{k_D(z_*)^2} = \frac{a_*^2}{6} \int_0^1 \frac{R^2 + \frac{1}{2}(1 + R)}{(1 + R)^2} \frac{1}{a_*^2 m_e \sigma T [a^2H]} \frac{dx}{x}.$$  \hspace{1cm} (6)

This requires

$$1/k_D(z_*) \propto a_*.$$  \hspace{1cm} (7)

Finally, the viewing angle of the sound horizon, $r_s(z_*)/D_M(z_*)$, should be kept constant, which means $D_M(z_*)$ should vary proportionally to $a_*$. Within $\Lambda$CDM background, we find that

$$\Delta h \approx -3.23 \Delta a_*.$$  \hspace{1cm} (8)

approximately realizes $D_M(z_*) \propto a_*$, where the numerical coefficient is evaluated around the baseline.

Conditions (4) and (8) can be easily satisfied by varying standard cosmological parameters. Contrastively, Eq. (7) is non-trivial. As we will show below, varying $m_e$ models can satisfy this non-trivial condition while the other ones are also held.

Varying $m_e$ and CMB power spectra. As a working example, here we consider a model with time-varying $m_e$. The electron mass $m_e$ affects physics of CMB at recombination through the following aspects:

- Energy levels of hydrogen: $E \propto m_e$
- Thomson scattering cross section: $\sigma_T \propto m_e^{-2}$
- Others (two-photon decay rate, photo-ionization cross section, recombination coefficients etc.)

If recombination proceeds in thermal equilibrium, the third effects can be omitted. Although non-equilibrium processes are evident in observed CMB power spectra, their impact is indeed relatively minor as long as $m_e$ alone is varied [10]. Neglecting the third effects to simplify discussion, $a_*$ is inversely proportional to $m_e$ through the first effect:

$$\Delta m_e = \Delta T_s(z_*) = -\Delta a_*.$$  \hspace{1cm} (9)

\(^2\) In the following discussion, we assume a different value of $m_e$ for the present time and the recombination epoch. In this sense, $m_e$ is time-varying, however, we make a simplified assumption where $m_e$ is constant until some time after recombination, then at some epoch, $m_e$ becomes the present value.
To see how $1/k_D$ in Eq. (6) is modified in response to $m_e$, let us consider the following factor:

$$a_e^2m_e\sigma_T = x_e \frac{1 - Y_p \rho_{\text{crit}}}{m_H \hbar^2} (\omega_b a_s) \left( \frac{\sigma_T}{a_e^2} \right) \frac{1}{x_e}, \quad (10)$$

where $x_e$, $Y_p$, $m_H$ and $\rho_{\text{crit}} = \frac{3\pi^2}{20\hbar^2}$ are the ionization fraction, the mass fraction of $^4\text{He}$, the hydrogen mass, and the critical density, respectively. When we vary $\omega_b$ according to Eq. (4), Eq. (10) does not change as a function of $x$. Thus, the integral in Eq. (6) is kept constant, which means $1/k_D(a_*) \propto a_*$ and Eq. (7) is satisfied.

Fig. 1 demonstrates the parameter degeneracy in CMB power spectra, which are nothing but the scales of BAO measured along the transverse and line-of-sight directions, respectively. Except for low-$\ell$ in $C_T^T$, where the late-ISW effect is significant, CMB power spectra remain remarkably unchanged. This manifests that varying $m_e$ satisfies all the first three conditions we raised in Introduction.

Low-$z$ distances. — While CMB spectra are almost conserved, the parameter modification Eqs. (4) and (8) in general also modifies late-time expansion and geometric distances, which is severely constrained by BAO and SNeIa. To see this, we plot the late-time distance and the expansion history in Fig. 2. Here we have introduced two quantities:

$$\theta_T(z) \equiv \frac{r_s(z_{\text{drag}})}{D_M(z)}, \quad \theta_L(z) \equiv r_s(z_{\text{drag}})[aH](z) \quad (11)$$

which are nothing but the scales of BAO measured along the transverse and line-of-sight directions, respectively. In addition, recent BAO [13,15] and SNeIa data [16] are overlaid in the same figure. When $\Lambda$CDM background is assumed (left panel of Fig. 2), the model becomes effectively one-parameter model according to Eqs. (4) and (8). One can see that late-time geometry changes as $m_e$ varies from the baseline. Since the baseline model fits with BAO data well, $m_e$ is tightly constrained when CMB is combined with BAO, and somewhat less tightly but also with SNeIa.

In the $\Lambda$CDM background, there're no more degrees of freedom to tune the late-time geometry while $\Delta_{m_e}$ is kept nonzero, and hence it is impossible to solve the $H_0$ tension just with varying $m_e$.

Low-$z$ distances. — While CMB spectra are almost conserved, the parameter modification Eqs. (4) and (8) in general also modifies late-time expansion and geometric distances, which is severely constrained by BAO and SNeIa. To see this, we plot the late-time distance and the expansion history in Fig. 2. Here we have introduced two quantities:

$$\theta_T(z) \equiv \frac{r_s(z_{\text{drag}})}{D_M(z)}, \quad \theta_L(z) \equiv r_s(z_{\text{drag}})[aH](z) \quad (11)$$

which are nothing but the scales of BAO measured along the transverse and line-of-sight directions, respectively. In addition, recent BAO [13,15] and SNeIa data [16] are overlaid in the same figure. When $\Lambda$CDM background is assumed (left panel of Fig. 2), the model becomes effectively one-parameter model according to Eqs. (4) and (8). One can see that late-time geometry changes as $m_e$ varies from the baseline. Since the baseline model fits with BAO data well, $m_e$ is tightly constrained when CMB is combined with BAO, and somewhat less tightly but also with SNeIa.

In the $\Lambda$CDM background, there're no more degrees of freedom to tune the late-time geometry while $\Delta_{m_e}$ is kept nonzero, and hence it is impossible to solve the $H_0$ tension just with varying $m_e$. However, it easily becomes possible when the background model is extended appropriately. In the right panel of Fig. 2, the background cosmology is extended to allow a non-flat Universe ($\Omega_k \Lambda$CDM hereafter) and plotted is the late-time geometry along a parameter direction

$$\Delta_h = 1.5\Delta_{m_e}, \quad \omega_k = -0.125\Delta_{m_e}, \quad (12)$$

instead of Eq. (8). This realizes a good fit to the low-$z$ distance observations even with $\Delta_{m_e}$ as large as 5%. Curvature of the Universe is playing an essential role here. As can be read from Eq. (11), deviations from flatness grow as $\chi(z)\sqrt{\omega_k}/L$ increases. Therefore, the curvature selectively affects only the angular diameter distance to CMB and offers the freedom for low-$z$ and CMB distances to be fitted well simultaneously even with large $\Delta_{m_e}$. Therefore all the four conditions in Introduction are satisfied in $\Omega_k \Lambda$CDM background with varying $m_e$.

**MCMC parameter estimation.** — We perform Markov chain Monte Carlo (MCMC) analysis using CosmoMC [17] modified to incorporate varying $m_e$. We adopt the Planck 2018 reference CMB likelihood TT,TE,EE+lowE [18] in combination with the BAO [13,15] and SNeIa data [16]. We checked that our analysis is consistent with a previous work [19] when a $\Lambda$CDM background is assumed.

---

3 If we neglect the non-equilibrium nature of recombination, $x_e$ does not change as function of $x$. We also have omitted the marginal dependence of $Y_p$ on $\omega_b$ in the Big Bang nucleosynthesis prediction.

4 Precisely speaking, $\theta_L(z)$ is the separation of the BAO scale along the line-of-sight in $\log(1+z)$.

5 We have normalized the SNeIa luminosity distances to give $D_M(z)$ consistent with BAO at $z \simeq 0.5$. 
Fig. 3 shows the posterior distribution of $H_0$ in models with varying $m_c$ in different background including $\Lambda$CDM, $\Omega_b$CDM, $w$CDM, where dark energy (DE) equation of state (EoS) $w$ is assumed to be constant, and $w_0w_a$CDM models, where DE EoS is parametrized as in [20] [21]. For reference, $\Lambda$CDM model without varying $m_c$ (“reference” model hereafter) is also plotted. We also compare those posterior distributions with the direct measurements $H_0 = 74.1 \pm 1.3$ km/sec/Mpc (Hereafter H0) [3], which is not incorporated in the default parameter estimation.

From the figure, one can immediately see that the varying $m_c$ in $\Omega_b$CDM model gives a posterior distribution matching well with the direct measurements. As expected from the parameter degeneracies discussed above, the $\Omega_b$CDM background allows substantially broader distributions compared to the reference model.

Besides, it is remarkable that the distribution peak coincides with the direct $H_0$ measurements. The preference for higher $H_0$ in association with $\Omega_k < 0$ is brought about by the Planck data at $\ell > 30$, which is known to favor larger lens amplitude, $A_L > 1$ [4]. Indeed, we found that the posterior mean values in our analysis, which are consistent with local $H_0$ measurements, yield $C_{\ell}^{TT}$ at $\ell > 800$ similar to that from the baseline but with $A_L = 1.1$. Although a closed Universe enhances the CMB lensing effect, in general the fit to BAO and SNeIa gets worse [22]. However, varying $m_c$ in $\Omega_b$CDM model can keep the fit to BAO and SNeIa excellent.

While $H_0$ tension is relaxed with varying $m_c$ in other backgrounds too, as posterior distributions become broader compared to the reference model, their peaks are still displaced from the direct measurements. This is because these models lack the freedom to fit with the CMB and low-$z$ distances simultaneously.

Table I summarizes mean values and $68\%$ intervals of $H_0$ from the default data set, CMB+BAO+SNeIa and an extended one, CMB+BAO+SNeIa+H0 as well as the effective $\Delta \chi^2$ for CMB+BAO+SNeIa+H0 against the reference model. As expected, varying $m_c$ in $\Omega_b$CDM model yields a significant improvement in data fits with $\Delta \chi^2_{\text{eff}} < -23$. This is caused not only by resolving the $H_0$ tension, but also by improving the fit to CMB data, which alone reduces $\Delta \chi^2_{\text{eff}}$ by 6.2. Our results clearly prove that varying $m_c$ in $\Omega_b$CDM is preferred by data over the reference model.

Discussion.— The parameter degeneracy in Eq. [4] is not perfect and varying $m_c$ distorts CMB power spectra through non-equilibrium nature of recombination. Therefore, CMB-S4 [23] may be able to constrain/verify our examples. Substantial deviations from the baseline in low-$z$ distances are also predicted. For instance, in varying $m_c$ with $\Omega_b$CDM background, $\Delta \chi^2_{\text{eff}}(z_{\text{drag}}) \approx -0.05$ and $\Omega_k \approx -0.01$ are required to solve the $H_0$ tension. Future distance measurements will be able to test such deviations from the baseline [24].

Acknowledgments

This work is supported by JSPS KAKENHI Grant Numbers 18H04339 (TS), 18K03640 (TS), 17H01131 (TT, TS), 19K03874 (TT) and MEXT KAKENHI Grant Number 19H05110 (TT). This research was conducted using the Fujitsu PRIMEargy CX600M1/CX1640M1 (Oakforest-PACS) in the Information Technology Center, The University of Tokyo.
TABLE I: Summary of estimation of $H_0$ and $\Delta \chi^2_{\text{eff}}$.  

| $H_0$ [km/sec/Mpc] (mean with 68\% errors) | $\Lambda\text{CDM}$ | $\Omega_b\Lambda\text{CDM}$ | $w\text{CDM}$ | $w_{\text{CDM}}\Lambda\text{CDM}$ | $\Lambda\text{CDM}$ (reference) |
|---------------------------------------------|----------------------|-----------------------------|--------------|-------------------------------|-------------------------------|
| based on CMB+BAO+SNeIa                   | $68.7^{+1.2}_{-1.0}$ | $72.3^{+2.7}_{-2.8}$      | $68.7^{+1.1}_{-1.2}$ | $67.5^{+1.3}_{-1.6}$ | $67.1^{+0.4}_{-0.4}$ |
| based on CMB+BAO+SNeIa+H0               | $71.2^{+0.9}_{-1.0}$ | $72.9^{+1.0}_{-1.0}$     | $71.0^{+1.0}_{-1.0}$ | $71.5^{+1.1}_{-0.9}$ | $68.4^{+0.4}_{-0.4}$ |
| $\Delta \chi^2_{\text{eff}}$ relative to the reference based on CMB+BAO+SNeIa+H0 | $-12.2$ | $-23.5$ | $-12.5$ | $-13.2$ | $0$ |

![FIG. 3: Posterior distributions of $H_0$ for varying $m_e$ with different background models and the reference model. Gray band shows the direct $H_0$ measurement $H_0 = 74.1 \pm 1.3$ km/sec/Mpc without SNeIa [3]. Solid and dashed lines are obtained from the combination CMB+BAO+SNeIa. In order for demonstration, we also depicted the posterior distribution from CMB+BAO+SNeIa+H0 only for varying $m_e$ model with $\Omega_k\Lambda\text{CDM}$ background (orange dotted line).](image)

[1] A. G. Riess, S. Casertano, W. Yuan, L. M. Macri, and D. Scolnic, Astrophys. J. 876, 85 (2019), 1903.07603.
[2] K. C. Wong et al. (2019), 1907.04869.
[3] A. G. Riess, Nature Rev. Phys. 2, 10 (2019), 2001.03624.
[4] N. Aghanim et al. (Planck) (2018), 1807.06209.
[5] K. Aylor, M. Joy, L. Knox, M. Millea, S. Raghunathan, and W. K. Wu, Astrophys. J. 874, 4 (2019), 1811.00537.
[6] L. Knox and M. Millea, Phys. Rev. D 101, 043533 (2020), 1908.03663.
[7] C.-T. Chiang and A. z. Slosar (2018), 1811.03624.
[8] M. Liu, Z. Huang, X. Luo, H. Miao, N. K. Singh, and L. Huang, Sci. China Phys. Mech. Astron. 63, 290405 (2020), 1912.00190.
[9] C. Martins (2017), 1709.02923.
[10] P. Ade et al. (Planck), Astron. Astrophys. 580, A22 (2015), 1406.7482.
[11] A. Lewis, A. Challinor, and A. Lasenby, Astrophys. J. 538, 473 (2000), astro-ph/9911177.
[12] Y. Ali-Haimoud and C. M. Hirata, Phys. Rev. D 83, 043513 (2011), 1011.3758.
[13] S. Alam et al. (BOSS), Mon. Not. Roy. Astron. Soc. 470, 2617 (2017), 1607.03155.
[14] P. Zarrouk et al., Mon. Not. Roy. Astron. Soc. 477, 1639 (2018), 1801.03062.
[15] J. E. Bautista et al., Astron. Astrophys. 603, A12 (2017), 1702.00176.
[16] D. Scolnic et al., Astrophys. J. 859, 101 (2018), 1710.00845.
[17] A. Lewis and S. Bridle, Phys. Rev. D 66, 103511 (2002), astro-ph/0205436.
[18] N. Aghanim et al. (Planck) (2019), 1907.12875.
[19] L. Hart and J. Chluba, Mon. Not. Roy. Astron. Soc. 493, 3255 (2020), 1912.03986.
[20] M. Chevallier and D. Polarski, Int. J. Mod. Phys. D 10, 213 (2001), gr-qc/0009008.
[21] E. V. Linder, Phys. Rev. Lett. 90, 091301 (2003), astro-ph/0208512.
[22] E. Di Valentino, A. Melchiorri, and J. Silk, Nature Astron. 4, 196 (2019), 1911.02087.
[23] K. N. Abazajian et al. (CMB-S4) (2016), 1610.02743.
[24] M. Denissenya, E. V. Linder, and A. Shafieloo, JCAP \textbf{03}, 041 (2018), 1802.04816.