HyRec: A fast and highly accurate primordial hydrogen and helium recombination code

Yacine Ali-Haïmoud and Christopher M. Hirata

California Institute of Technology, Mail Code 350-17, Pasadena, California 91125, USA
(Received 20 November 2010; published 15 February 2011)

We present a state-of-the-art primordial recombination code, HYREC, including all the physical effects that have been shown to significantly affect recombination. The computation of helium recombination includes simple analytic treatments of hydrogen continuum opacity in the He I $2^1P^o - 1^1S$ line, the He I] $2^3P^o - 1^1S$ line, and treats feedback between these lines within the on-the-spot approximation. Hydrogen recombination is computed using the effective multilevel atom method, virtually accounting for an infinite number of excited states. We account for two-photon transitions from $2s$ and higher levels as well as frequency diffusion in Lyman-α with a full radiative transfer calculation. We present a new method to evolve the radiation field simultaneously with the level populations and the free electron fraction. These computations are sped up by taking advantage of the particular sparseness pattern of the equations describing the radiative transfer. The computation time for a full recombination history is ~ 2 seconds. This makes our code well suited for inclusion in Monte Carlo Markov chains for cosmological parameter estimation from upcoming high-precision cosmic microwave background anisotropy measurements.

I. INTRODUCTION

Until recently, primordial recombination was considered one of the few solved problems in astrophysics and cosmology. The seminal works of Peebles [1] and Zeldovich *et al.* [2] in the 1960s established that hydrogen recombination did not proceed in Saha equilibrium, and that two-photon decays from $2s$ are critical to the recombination dynamics because of the very low escape rate of Lyman-α photons. They provided a simple effective three-level atom model to compute primordial hydrogen recombination histories. With the advent of high-precision cosmic microwave background (CMB) experiments such as *WMAP* [3] and *Planck* [4], it has become clear that these early calculations are not sufficiently accurate for an unbiased estimate of cosmological parameters [5–7]. Uncertainties in the recombination history indeed propagate to the visibility function and ultimately to the predicted CMB temperature and polarization anisotropies.

These considerations have motivated Seager *et al.* [8,9] to extend Peebles' effective three-level atom model to a multilevel atom calculation. They showed that accounting for excited states of hydrogen leads to a speed up of recombination at late times. Their commonly used recombination code RECFAST approximately reproduces these results by solving an effective three-level atom model with an artificially enhanced recombination coefficient. While this model is sufficiently accurate for current CMB data analysis, it does not meet the $\sim 0.1\%$ accuracy target for *Planck*.

Several physical effects have since then been shown to significantly affect hydrogen and helium recombination. First, the multilevel computations of Seager *et al.* assumed statistical equilibrium among the angular momentum substates of a given energy shell of hydrogen. At late times, this assumption breaks down, and an accurate multilevel atom computation should resolve the angular momentum substates [10,11]. Whereas it is a straightforward conceptual generalization of previous works, this problem can be computationally challenging. Several works have tackled the "high-n problem", including increasingly larger numbers of excited states of hydrogen to reach sufficient accuracy [10–13]. The "standard" multilevel atom (MLA) method used in these works requires solving for the abundances of all the excited states at every time step, which makes the computation very time-consuming. In a recent paper (Ref. [14], hereafter "Paper I"), we have introduced a new method of solution for the multilevel atom problem, which allows to only solve for the populations of a few excited states (typically, $2s$ and the low-lying p states), provided one uses precomputed *effective* bound-bound and bound-free transition rates, which contain all the information about the highly excited states of hydrogen. This method alleviates the computational difficulty associated with the highly excited states of hydrogen and is key to the speed of the recombination code presented here.

Another important aspect of the recombination problem is that of radiative transfer in the vicinity of the Lyman-α line. In its early stages, hydrogen recombination is mostly controlled by the slow escape (via redshifting) of photons from the Lyman-α line and the rate of two-photon decays from the $2s$ state. Accurate values for these rates require treatments of the radiation field that go beyond the simple Sobolev approximation [8,15]. Important corrections include feedback from higher-order lines [16,17], time-dependent effects in Ly$-\alpha$ [18], and frequency diffusion due to resonant scattering [19–21]. An accurate $2s - 1s$ two-photon decay rate also requires following the radiation field to account for stimulated decays [22] and absorption of nonthermal photons [23,24]. Dubrovich & Grachev [25]

suggested that two-photon transitions from higher levels may have a significant effect on the recombination history. Later computations confirmed this idea [26], and provided an accurate treatment of radiative transfer in the presence of two-photon transitions, as well as a solution for the double-counting problem (which arises for resonant two-photon transitions, already included in the one-photon treatment as "1 + 1" transitions) [24,27].

The accuracy requirement is less stringent for primordial helium recombination, as it is completed by $z \sim 1700$, much earlier than the peak of the visibility function. Corrections at the percent level are still important, and several works have been devoted to the problem [5,8,25,28–35]. The most important effect is continuum opacity in the He I $2^1P^o - 1^1S$ line due to photoionization of neutral hydrogen, which requires a detailed line + continuum radiative transfer analysis [30,31,34]. The inclusion of the intercombination line He I] $2^3P^o - 1^1S$ is also significant.

Several other processes have been investigated and shown not to be significant for CMB anisotropies, for example, the effects of the isotopes D and ^{3}He [15,17,33], lithium recombination [36], quadrupole transitions [12], high-order Lyman line overlap [15], and Thomson scattering [15,21]. Collisional processes are negligible for helium recombination [33]; for hydrogen recombination, collisional corrections appear to be small [37], but whether they are truly negligible is still under investigation.

Previous works have all concentrated on one or a few aspects of the primordial recombination problem. Producing a complete and fast recombination code has so far been hindered by the computational burden previously associated with the high-n problem. Given that this problem is now solved, and that it seems that the main radiative transfer effects have now all been identified, it is timely to deliver a single code that computes an accurate hydrogen and helium recombination history and incorporates all the relevant physics. The purpose of this paper is to introduce our new recombination code, HYREC, which is publicly available,[1] and can compute a highly accurate recombination history (with errors at the level of a few times 10^{-3} for helium recombination and a few times 10^{-4} for hydrogen recombination) in only $\sim$2 seconds on a standard laptop. Our code does not account for collisional transitions in hydrogen, as their rates are poorly known. When accurate rates are available and if collisional transitions are shown to significantly impact recombination, we will update our code with the appropriate effective rates.

Recently, a similar work has been carried out by Chluba & Thomas [38], also relying on the effective MLA method presented in Paper I. The code they present includes the

same physics as ours. The main difference is the treatment of radiative transfer. In Ref. [38], an "order zero" recombination history is first computed, with a simple treatment of radiative transfer. The radiative transfer equation is then solved, given this order zero history. Corrections to the net decay rates to the ground state are then evaluated, and used to compute a corrected recombination history. This procedure can in principle be iterated, but because the corrections are small, it is essentially converged after one iteration. Our solution, on the other hand, is nonperturbative, in the sense that we solve *simultaneously* for the radiation field and the recombination history. A detailed code comparison is in progress, and a full error budget will be presented once it is completed.

This paper is organized as follows. In Sec. II, we review the effective three-level atom model and discuss its limitations. We then review the standard MLA computation and describe the effective MLA method in Sec. III. We show that weak transitions to the ground state from excited states with $n \geq 3$ can in fact be accounted for almost exactly with an effective four-level atom model. Two-photon processes and frequency diffusion are formally described in Sec. IV. In Sec. V, we present our numerical solution for the radiative transfer equation. We use a new method of solution, extending that of Ref. [24] to account for frequency diffusion, that allows to solve for the atomic populations and the radiation field simultaneously. We describe our treatment of helium recombination in Sec. VI. We conclude in Sec. VII. Appendix A demonstrates some relations satisfied by the effective rates, Appendix B describes how we extrapolate the effective rates to an infinite number of excited states, Appendix C describes our ordinary differential equation (ODE) integrator, and Appendix D derives an analytic expression for the post-Saha expansion used at early times in hydrogen recombination.

Throughout this paper we use a flat background ΛCDM cosmology with $T_0 = 2.728$ K, $\Omega_b h^2 = 0.022$, $\Omega_m h^2 = 0.13$, $\Omega_\Lambda h^2 = 0.343$, $Y_{\mathrm{He}} = 0.24$ and $N_{\nu,\mathrm{eff}} = 3.04$.

II. HYDROGEN RECOMBINATION: OVERVIEW

A. The effective three-level atom model

The basic process of primordial hydrogen recombination was already well understood in the late sixties. The seminal papers by Peebles [1] and Zeldovich *et al.* [2] established the following picture. Direct recombinations to the ground state are highly inefficient, as they produce photons that can immediately ionize another hydrogen atom. Electrons and protons can therefore recombine efficiently only to the excited states of hydrogen. This situation is familiar in the study of the interstellar medium: it is referred to as "case-B" recombination (see e.g. Ref. [39]). Once they have recombined to one of the excited states of hydrogen, electrons "cascade down" to the $n = 2$ shell, on a much shorter time scale than the overall recombination

---

[1] HYREC is available for download at the following url: http://www.tapir.caltech.edu/~yacine/hyrec/hyrec.html.

time scale. Denoting n_H the total number density of hydrogen, $x_e = n_e/n_H$ the free electron fraction and $x_2 = n_{H(n=2)}/n_H$ the fraction of hydrogen in the excited state, the effective rate of recombinations to $n = 2$ shell can be written as

$$\dot{x}_2|_{\text{rec}} = -\dot{x}_e = n_H x_e^2 \alpha_B(T_m) - x_2 \beta_B(T_r), \qquad (1)$$

where T_m is the matter temperature, locked to the radiation temperature T_r by Thomson scattering at most times during recombination, α_B is the case-B recombination coefficient, and β_B is the corresponding photoionization rate, which can be obtained from α_B by the principle of detailed balance:

$$\beta_B(T_r) = \frac{g_e}{4} e^{E_2/T_r} n_H \alpha_B(T_m = T_r), \qquad (2)$$

where we have defined

$$g_e \equiv \frac{(2\pi \mu_e T_r)^{3/2}}{h^3 n_H}, \qquad (3)$$

where μ_e is the reduced mass of the electron-proton system.

Once they have reached the $n = 2$ shell, electrons can reach the ground state by emitting a Lyman-α photon from the $2p$ state. Because of the very high optical depth of the Lyman-α transition, emitted photons will however almost certainly be reabsorbed by another atom. The way out of this bottleneck is for photons to redshift below the Lyα resonant frequency due to cosmological expansion. The *net* rate of decays to the ground state from the $2p$ state is then just the rate at which photons redshift across the line and escape reabsorption:

$$\dot{x}_{1s}|_{2p} = -\dot{x}_{2p}|_{1s} = R_{\text{Ly}\alpha}(x_{2p} - 3x_{1s}e^{-E_{21}/T_r}), \qquad (4)$$

where x_{1s} is the fraction of hydrogen in the ground state and the second term accounts for Lyα absorptions and is obtained by detailed balance. The rate of *escape* of Lyα photons is given by

$$R_{\text{Ly}\alpha} \equiv \frac{8\pi H}{3 n_H x_{1s} \lambda_{\text{Ly}\alpha}^3}. \qquad (5)$$

Equations (4) and (5) can be derived in the Sobolev approximation, in the limit of large Sobolev optical depth (see, for example, Ref. [15]).

The escape rate of Lyα photons is comparable to the rate of the slow two-photon decays from the $2s$ state, $\Lambda_{2s,1s} \approx 8.22$ s^{-1}, and the latter process must therefore be accounted for. The net rate of two-photon decays from the $2s$ state is

$$\dot{x}_{1s}|_{2s} = -\dot{x}_{2s}|_{1s} = \Lambda_{2s,1s}(x_{2s} - x_{1s}e^{-E_{21}/T_r}), \qquad (6)$$

where the second term accounts for two-photon absorptions and can be obtained by a detailed balance argument. Because of the strong thermal radiation bath, the excited

states of hydrogen are near Boltzmann equilibrium with each other, $x_{2p} = 3x_{2s} = (3/4)x_2$. The rate of change of the population of the $n = 2$ shell due to decays to the ground state is therefore

$$\dot{x}_2|_{1s} = \left(\frac{3}{4}R_{\text{Ly}\alpha} + \frac{1}{4}\Lambda_{2s,1s}\right)(4x_{1s}e^{-E_{21}/T_r} - x_2). \qquad (7)$$

The last step is to realize that the atomic rates, even for the slow $2s \to 1s$ decays or the slow escape out of the Lyα resonance, are many orders of magnitude larger than the overall recombination rate, which is of the order of (10 times) the Hubble expansion rate, that is $\sim 10^{-13} - 10^{-12}$ s^{-1}. The population of the $n = 2$ shell can therefore be obtained to high accuracy in the steady state approximation, i.e., assuming that the rate of recombinations to the $n = 2$ shell equals the rate of transitions to the ground state:

$$\dot{x}_2 = \dot{x}_2|_{\text{rec}} + \dot{x}_2|_{1s} \approx 0. \qquad (8)$$

We can therefore solve for x_2 and obtain

$$x_2 = \frac{n_H x_e^2 \alpha_B + (3R_{\text{Ly}\alpha} + \Lambda_{2s,1s})x_{1s}e^{-E_{21}/T_r}}{\beta_B + \frac{3}{4}R_{\text{Ly}\alpha} + \frac{1}{4}\Lambda_{2s,1s}}. \qquad (9)$$

From Eq. (1) we then obtain the rate of change of the free electron fraction:

$$\dot{x}_e = -C(n_H x_e^2 \alpha_B - 4x_{1s}\beta_B e^{-E_{21}/T_r}), \qquad (10)$$

where the Peebles C factor is given by

$$C \equiv \frac{\frac{3}{4}R_{\text{Ly}\alpha} + \frac{1}{4}\Lambda_{2s,1s}}{\beta_B + \frac{3}{4}R_{\text{Ly}\alpha} + \frac{1}{4}\Lambda_{2s,1s}}. \qquad (11)$$

As noted by Peebles, this factor represents the probability that an atom initially the $n = 2$ shell reaches the ground state before being photoionized. Note that we could have obtained the same equation starting from $\dot{x}_e = -\dot{x}_{1s} = -(\dot{x}_{1s}|_{2p} + \dot{x}_{1s}|_{2s})$ (this is because we have set $\dot{x}_2 = 0$).

At all relevant times during the epoch of hydrogen recombination, $x_2 \ll 1$, and therefore $x_{1s} = 1 - x_e$. If matter and radiation temperatures are set to be equal, Eq. (10) is therefore a simple ordinary differential equation for x_e, that can be easily integrated. A simple improvement is to also explicitly follow the matter temperature evolution, which is determined by the Compton evolution equation:

$$\dot{T}_m = -2HT_m + \frac{8\sigma_T a_r T_r^4 x_e (T_r - T_m)}{3(1 + f_{\text{He}} + x_e)m_e c}, \qquad (12)$$

where σ_T is the Thomson cross section, a_r is the radiation constant, m_e is the electron mass and f_{He} is the He:H ratio by number of nuclei.

The simple yet insightful picture presented here is known as the effective three-level atom model. It provides a good approximation for the recombination problem.

It is however not sufficiently accurate for high-precision cosmology.

B. Hydrogen recombination phenomenology

We show in Fig. 1 the evolution of the Peebles C factor and the population of the $n = 2$ shell relative to its value in Saha equilibrium with the continuum, as a function of redshift, for a standard recombination history. We can see that there are two distinct regimes.

At early times ($z \gtrsim 1000$), electrons in the $n = 2$ shell have a high probability of being photoionized, and the C factor is much smaller than unity, $C \ll 1$. As a consequence, the population of the $n = 2$ shell is very close to Saha equilibrium with the continuum

$$x_2 \approx x_2|_{\text{Saha}} \equiv \frac{4}{g_e} e^{-E_2/T_{\text{m}}} x_e^2. \qquad (13)$$

The rate of change of the free electron fraction is then approximately equal to the rate of decays from the $n = 2$ shell:

$$\dot{x}_e(z \gtrsim 1000) \approx \dot{x}_2|_{1s}(x_2 = x_2|_{\text{Saha}}). \qquad (14)$$

During that period, the recombination rate is therefore virtually independent of the exact value of the recombination coefficient, but is strongly dependent on the small net decay rate from the $n = 2$ shell to the ground state. This is usually referred to as the "$n = 2$ bottleneck" and has motivated abundant work on radiative transfer in the vicinity of the Lyman transitions [15–18,20–24,26,27]. The result from this series of papers is that to the level of accuracy required by *Planck*, Lyman transitions up to Lyγ must be included, properly accounting for feedback between them. In addition, the radiation field must be solved for with a radiative transfer calculation, accounting for two-photon transitions and frequency diffusion in the Lyman-α line. We consider all these effects in Sec. IV and V.

At late times ($z \lesssim 700$), $C \approx 1$, and the $n = 2$ shell is no longer in Saha equilibrium with the continuum (note that it
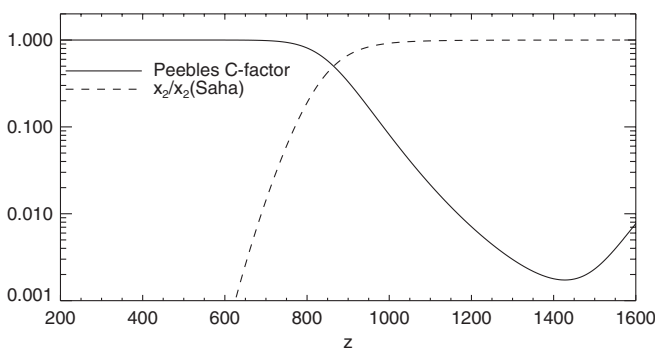

FIG. 1. Peebles C factor [Eq. (11)] and ratio of the population of the $n = 2$ shell to its value in Saha equilibrium with the continuum, as a function of redshift.

is *not* in Boltzmann equilibrium with the ground state either, as the rate of recombinations to the $n = 2$ shell dominates over the net rate of two-photon or Ly $- \alpha$ absorptions from the ground state). The free electron fraction is many orders of magnitude above the value it would have in Saha equilibrium because of the slow recombination rate. In that case, the second term in Eq. (10) is negligible and the evolution of the free electron fraction becomes

$$\dot{x}_e(z \lesssim 700) \approx -n_{\text{H}} x_e^2 \alpha_{\text{B}}. \qquad (15)$$

As we can see, the evolution of the free electron fraction is then virtually independent of the rate of decays to the ground state from the $n = 2$ shell, but is highly sensitive to the exact value of the effective recombination coefficient. Moreover, the assumption that the excited states are in Boltzmann equilibrium with each other breaks down at late times because of the decrease in the radiation temperature. An accurate recombination rate at late times can only be obtained in a full multilevel atom calculation, that accounts for (possibly stimulated) bound-bound and bound-free transitions between all—at least, a large number of—the excited states of hydrogen [8,11–13]. We will review the multilevel atom calculations in Sec. III.

Of course, at intermediate redshifts $700 \lesssim z \lesssim 1000$ both the exact recombination rate and the rate of decays to the ground state are important and should be carefully accounted for.

III. THE MULTILEVEL ATOM

In this section we first present the "standard" multilevel atom (MLA) method [8], then review the effective MLA (hereafter EMLA) method of solution that we presented in Paper I and that allows for a fast computation of recombination histories.

A. The standard multilevel atom method

The effective three-level atom equations may be easily generalized to account for an arbitrarily large number of excited states of hydrogen (in practice, one must of course impose a cutoff). We denote x_{nl} the fractional abundance of hydrogen atoms in the excited state with principal quantum number n and angular momentum quantum number l. The generalization of Eq. (1) is then, for $n \geq 2$:

$$\dot{x}_{nl}|_{\text{rec}} = n_{\text{H}} x_e^2 \alpha_{nl}(T_{\text{m}}, T_{\text{r}}) - x_{nl} \beta_{nl}(T_{\text{r}}), \qquad (16)$$

where $\alpha_{nl}(T_{\text{m}}, T_{\text{r}})$ is the recombination coefficient to the excited state nl, including stimulated recombinations, and $\beta_{nl}(T_{\text{r}})$ is rate of photoionizations from nl by blackbody photons.

The effective three-level atom model does not account for bound-bound transitions between excited states (except for instantaneous spontaneous decays that ultimately lead

to the $n = 2$ shell). Transitions between the nl and $n'l'$ states (with $n, n' \geq 2$) change their populations at the rate

$$\dot{x}_{nl}|_{n'l'} = -\dot{x}_{n'l'}|_{nl} = x_{n'l'}R_{n'l',nl}(T_r) - x_{nl}R_{nl,n'l'}(T_r),$$ (17)

where the bound-bound transition rate from nl to $n'l'$, $R_{nl,n'l'}(T_r)$, is the rate of absorptions of blackbody photons resonant with the transition if $n < n'$ and the rate of spontaneous and stimulated decays if $n > n'$. We give explicit expressions of the bound-bound and bound-free rates and explain how we compute them in Paper I.

Finally, the rate of decays to the ground state from the $2s$ state are given by Eq. (6) (we will see how to make this rate more accurate in Sec. IVA). In the Sobolev approximation, the net decay rate from the np states to the ground state is given by a generalization of Eq. (4), accounting for feedback between optically thick Lyman lines:

$$\dot{x}_{1s}|_{np} = -\dot{x}_{np}|_{1s} = R_{\mathrm{Ly}n}[x_{np} - 3x_{1s}f_{np}^+],$$ (18)

where $R_{\mathrm{Ly}n} \equiv (\lambda_{\mathrm{Ly}\alpha}/\lambda_{\mathrm{Ly}n})^3 R_{\mathrm{Ly}\alpha}$ is the rate at which photons redshift out of the Lyman-n line (with the convention that Ly $-$ 2 is Lyα), and f_{np}^+ is the photon occupation number incoming on the blue side of the Ly $- n$ transition. If no radiative processes affect the radiation field between neighboring Lyman lines, then

$$f_{np}^+(z) = f_{n+1,p}^-(z'),$$ (19)

where the earlier redshift z' is given by

$$z' = \frac{\lambda_{\mathrm{Ly}n}}{\lambda_{\mathrm{Ly}(n+1)}}(1 + z) - 1.$$ (20)

In the optically thick limit which is valid here, the photon occupation number on the red side of the Ly $- n$ line is given by $f_{np}^- = x_{np}/(3x_{1s})$.

For excited states nl other than $2s$ and np, not radiatively connected to the ground state, we have $\dot{x}_{nl}|_{1s} = 0$.

The recombination history can then be computed by evolving simultaneously the system of differential equations:

$$\dot{x}_{nl} = \dot{x}_{nl}|_{\mathrm{rec}} + \sum_{n' \geq 2, l'} \dot{x}_{nl}|_{n'l'} + \dot{x}_{nl}|_{1s}$$ (21)

$$\dot{x}_e = -\dot{x}_{1s} = \dot{x}_{2s}|_{1s} + \sum_{n \geq 2} \dot{x}_{np}|_{1s},$$ (22)

where in the last equation we used $x_e = 1 - x_{1s}$, valid as the fractional abundance of hydrogen in the excited states is always much less than unity. The atomic transition rates are many orders of magnitude larger than the overall recombination rate (of the order of the Hubble rate). A highly accurate approximation therefore consists in first solving for the populations of the excited states in the steady state approximation. This first step amounts to solving a large system of linear algebraic equations. The

populations x_{2s}, x_{np} can then be used in Eq. (22) to evolve the free electron fraction.

This generalization of the three-level atom model is relatively straightforward conceptually. Its practical implementation is, however, very time-consuming. It requires solving a very large system of algebraic equations at each time step (or evolving the same number of stiff differential equations). If one accounts for excited states up to principal quantum number n_{max}, then the number of equations is $N = n_{\mathrm{max}}(n_{\mathrm{max}} + 1)/2$, which, for $n_{\mathrm{max}} \gtrsim 100$, exceeds several thousands. Furthermore, the computational cost of an exact linear system solution scales as $\mathcal{O}(N^3)$, although this can be significantly sped up by using the sparseness of the system due to selection rules [12] and iterative solution techniques [37]. In the following section, we review the much more efficient yet exactly equivalent effective MLA method.

B. The effective multilevel atom method

1. General description

The effective multilevel atom method, described in Paper I, relies on three aspects of the primordial recombination problem. First, the time scales for transitions out of the excited states are much shorter than the overall recombination time scale—this property is used when solving for the populations of the excited states in the steady state approximation in the standard MLA method. This allows us to factor all the *nearly instantaneous* transitions involving the "interior" excited states (which are not radiatively connected to the ground state) into effective transitions into and out of the smaller set of "interface" states which are radiatively connected to the ground state. Second, all bound-bound and bound-free transitions for which the lower state is an excited state are *optically thin*, and therefore do not distort the ambient blackbody radiation field in the vicinity of the corresponding frequencies. All the transition rates between "interior" states therefore only depend on the radiation temperature T_r (as well as atomic physics constants). This translates into a simple dependence for the effective rates, which are functions of the matter (through recombinations of thermal electrons and protons) and radiation temperatures only. If collisional transitions are included, they will depend additionally on the free electron (of equivalently the free proton) abundance. Finally, the set of "interface" states that need to be considered is small. In principle, $2s$ and all the p states need to be considered as "interface" states; however, in practice only the lowest order Lyman transitions significantly affect the recombination rate, explicitly only Lyα, Lyβ and Lyγ [14,15]. In addition, when considering two-photon transitions from higher levels, one should in principle add the ns and nd states as "interface" states. However, only two-photon transitions from $2s$, $3s$ and $3d$, are important (and $4s$, $4d$ at the level of a few 10^{-4}) [24,27]. This means that one needs to pretabulate only *a*

few functions of temperature that fully account for the multilevel structure of hydrogen. These tabulated effective rates can then be interpolated when computing a recombination history with an effective few-level atom model. As we will show in Sec. III B 2, we can in fact further simplify the problem to an effective four-level atom model with virtually no loss of accuracy.

Following Paper I, we denote $\mathcal{A}_i(T_m, T_r)$ the effective recombination coefficient to the "interface" state i, $\mathcal{B}_i(T_r)$ the effective photoionization rate from this state, and $\mathcal{R}_{i,j}(T_r)$ the transfer rate from the interface state i to the interface state j (the dependences are valid in the purely radiative case; when collisions are included, all effective rates depend on T_m, T_r, n_e). They are obtained as follows: for the effective recombination coefficients,

$$\mathcal{A}_i = \alpha_i + \sum_K \alpha_K P_K^i; \tag{23}$$

for the effective ionization coefficient,

$$\mathcal{B}_i = \beta_i + \sum_K R_{i,K} P_K^e; \tag{24}$$

and for the effective interstate transition rates,

$$\mathcal{R}_{i,j} = R_{i,j} + \sum_K R_{i,K} P_K^j. \tag{25}$$

Here K is a general index for "interior" states, P_K^i is the probability that an electron initially in the "interior" state K ultimately reaches the "interface" state i, and P_K^e is the probability that an atom initially in the state K ultimately gets photoionized. These probabilities are the solutions of the linear systems [14]:

$$P_K^i = \sum_L \frac{R_{K,L}}{\Gamma_K} P_L^i + \frac{R_{K,i}}{\Gamma_K} \tag{26}$$

and

$$P_K^e = \sum_L \frac{R_{K,L}}{\Gamma_K} P_L^e + \frac{\beta_K}{\Gamma_K}, \tag{27}$$

where

$$\Gamma_K \equiv \sum_L R_{K,L} + \sum_i R_{K,i} + \beta_K \tag{28}$$

is the inverse lifetime of the state K.

The *net* decay rate from the interface state i to the ground state can always be expressed as a linear function of x_i:

$$\dot{x}_{1s}|_i = -\dot{x}_i|_{1s} = x_i \tilde{R}_{i,1s} - x_{1s} \tilde{R}_{1s,i}, \tag{29}$$

where we emphasize with this notation that in general the net decay rates may depend in a complicated way on the current and past values of the free electron fraction, as well as on cosmological parameters—see, for example, Eqs. (18) and (5) for the net $np \rightarrow 1s$ decay rate. This contrasts with the bound-bound rates between excited states, which only depend on atomic constants and the radiation temperature.

Once the effective bound-bound and bound-free rates for the interface states are tabulated, the recombination history can be computed by evolving the small set of ordinary differential equations:

$$\dot{x}_i = x_e^2 n_H \mathcal{A}_i + \sum_{j \neq i} x_j \mathcal{R}_{j,i} + x_{1s} \tilde{R}_{1s,i}$$
$$- x_i \left(\mathcal{B}_i + \sum_{j \neq i} \mathcal{R}_{i,j} + \tilde{R}_{i,1s} \right) \tag{30}$$

and

$$\dot{x}_e = -\sum_i (n_H x_e^2 \mathcal{A}_i - x_i \mathcal{B}_i) \tag{31}$$

$$= -\dot{x}_{1s} = \sum_i (x_{1s} \tilde{R}_{1s,i} - x_i \tilde{R}_{i,1s}), \tag{32}$$

where Eqs. (31) and (32) are equivalent as the fractional abundance of excited hydrogen is very small. In practice, the population of the effective states can once again be solved in the steady state approximation. Equation (30) becomes a system of a few algebraic linear equations, and in that case Eqs. (31) and (32) are mathematically equivalent since we set $\dot{x}_i = 0$ [this can be seen by summing Eq. (30) over i].

Equations (30) and (31), along with the definitions for the effective coefficients Eqs. (23)–(28), are strictly equivalent to the standard MLA equations presented in the previous section, as was derived in Paper I [Eq. (31) was not derived in that paper and we give a proof in Appendix A 2]. The advantage of the new method is that the system of equations that need to be solved at each time step is much smaller, as only the low-lying s, p, d states need to be followed. In the following section we show that we can actually further reduce the problem to an effective four-level atom model.

2. Further simplification: the effective four-level atom

If we wish to follow n_* "interface" states, then the system of Eqs. (30), in the steady state approximation, is a $n_* \times n_*$ system. Moreover, one needs to interpolate n_* functions of 2 variables (the effective recombination coefficients—the effective photoionization rates are obtained by detailed balance), and $n_*(n_* - 1)/2$ functions of 1 variable (half of the effective bound-bound rates, the other half being obtained by detailed balance). Here we show how this system can be further reduced to a 2×2 system involving only $2s$ and $2p$, requiring only 2 functions of 2 variables and $2n_* - 3$ functions of one variable, with virtually no loss of accuracy.

From now on we use the general index K for all states with principal quantum number $n \geq 3$. Even if some of the states with $n \geq 3$ are radiatively connected to the ground state, one can still formally define the effective transition

rates Eqs. (23)–(25) for the $2s$ and $2p$ states because of the near-instantaneity of transitions out of the excited states. However, these coefficients do not have a simple temperature dependence anymore, and are therefore not well suited for fast interpolation. Indeed, the probabilities P_K^i and P_K^e, where $i = 2s, 2p$, are still defined by Eqs. (26) and (27), but the inverse lifetime of the state K, Eq. (28), should now account for the net downward transition rate to the ground state, and one should make the replacement:

$$\Gamma_K \to \tilde{\Gamma}_K \equiv \Gamma_K(T_r) + \tilde{R}_{K,1s}, \qquad (33)$$

where we define $\Gamma_K(T_r)$ is the inverse lifetime of the K-th interface state when transitions to the ground state are not included. In addition, we need to define additional effective transition rates with the ground state. For $i = 2s, 2p$, we define

$$\tilde{\mathcal{R}}_{i,1s} \equiv \tilde{R}_{i,1s} + \sum_K R_{i,K} \tilde{P}_K^{1s} \qquad (34)$$

and

$$\tilde{\mathcal{R}}_{1s,i} \equiv \tilde{R}_{1s,i} + \sum_K \tilde{R}_{1s,K} P_K^i, \qquad (35)$$

where the probabilities $\tilde{P}_K^{1s}$ must satisfy the self-consistency relations

$$\tilde{P}_K^{1s} = \sum_L \frac{R_{K,L}}{\tilde{\Gamma}_K} \tilde{P}_L^{1s} + \frac{\tilde{R}_{K,1s}}{\tilde{\Gamma}_K}. \qquad (36)$$

The standard MLA equations, in the steady state approximation for the excited states with $n \geq 3$, can then be shown to be exactly equivalent to the following set of equations: for the net rate of production of $2s$,

$$\dot{x}_{2s} = x_e^2 n_H \mathcal{A}_{2s} + x_{2p} \mathcal{R}_{2p,2s} + x_{1s} \tilde{\mathcal{R}}_{1s,2s}$$
$$- x_{2s}(\mathcal{B}_{2s} + \mathcal{R}_{2s,2p} + \tilde{\mathcal{R}}_{2s,1s}); \qquad (37)$$

for the net rate of production of $2p$,

$$\dot{x}_{2p} = x_e^2 n_H \mathcal{A}_{2p} + x_{2s} \mathcal{R}_{2s,2p} + x_{1s} \tilde{\mathcal{R}}_{1s,2p}$$
$$- x_{2p}(\mathcal{B}_{2p} + \mathcal{R}_{2p,2s} + \tilde{\mathcal{R}}_{2p,1s}); \qquad (38)$$

and for either the net recombination rate or the net rate of generation of H(1s),

$$\dot{x}_e = \sum_{i=2s,2p} [x_i \mathcal{B}_i - x_e^2 n_H \mathcal{A}_i] \qquad (39)$$

$$= \sum_{i=2s,2p} \left[x_{1s} \tilde{\mathcal{R}}_{1s,i} - x_i \tilde{\mathcal{R}}_{i,1s} \right]. \qquad (40)$$

The proof of equivalence is a simple generalization of that of Paper I and we do not reproduce it here. Note that we use the same notation for the effective rates independently of the number of "interface" states considered, but they obviously have a different meaning that should be clear from the context.

The coefficients in the above system are in principle not simple functions of temperature anymore if one wishes to account for the transitions to the ground state from excited states with $n \geq 3$. We can nevertheless simplify their expressions with some minimal approximations. We start by noticing that for excited states nl with $n \geq 3$, the rate of spontaneous decays to the $n', l \pm 1$ states with $1 < n' < n$ is much larger than the net decay rate to the ground state. For the $3p$ state, for example, we find that in the Sobolev approximation for Lyβ decays, $\tilde{R}_{3p,1s}/A_{3p,2s} < 8 \times 10^{-6}$ for $200 < z < 1600$, where $A_{3p,2s}$ is the Einstein A-coefficient of the $3p \to 2s$ transition. Therefore, to an excellent accuracy (with relative errors of order $\tilde{R}_{K,1s}/\Gamma_K$), one can neglect $\tilde{R}_{K,1s}$ in Eq. (33), and simply use $\Gamma_K(T_r)$ instead of $\tilde{\Gamma}_K$ wherever the latter appears. With this approximation, the rate coefficients $\mathcal{A}_{2s/2p}$, $\mathcal{B}_{2s/2p}$ and $\mathcal{R}_{2s,2p}$ are simply the usual effective rates computed in the case that only $2s$ and $2p$ are considered as interface states, and depend only on matter and radiation temperatures. We explain in Appendix B how we obtain effective rates extrapolated to $n_{max} = \infty$.

We show in Appendix A 3 that using $\tilde{\Gamma}_K \approx \Gamma_K(T_r)$ in Eq. (36), we can rewrite Eq. (34) as

$$\tilde{\mathcal{R}}_{i,1s} = \tilde{R}_{i,1s} + \sum_K \tilde{R}_{K,1s} \frac{g_K}{g_i} e^{-E_{K2}/T_r} P_K^i(T_r), \qquad (41)$$

where g_K, g_i are the statistical weights of the states K, i and $E_{K2} \equiv E_K - E_2$ is the energy difference between the state K and the $n = 2$ shell.

In addition, the populations of the "weak" interface states X_K are sometimes required—for example the photon occupation number depends on the populations of the s and d states, see Sec. V B. We show in Appendix A 4 that the following relation is verified for $T_m = T_r$:

$$X_K|_{T_m=T_r} = \frac{g_K}{g_e} e^{-E_K/T_r} P_K^e(T_r) x_e^2$$
$$+ \sum_{i=2s,2p} \frac{g_K}{g_i} e^{-E_K/T_r} P_K^i(T_r) x_i. \qquad (42)$$

In fact, $T_m < T_r$ and the coefficient of x_e^2 in the above equation should be slightly higher. In practice though, $|T_m/T_r - 1| < 1\%$ for $z \gtrsim 500$, and for lower redshifts $P_K^e \ll 1$ and transitions to the ground state are unimportant anyway, so the above equation is very accurate.

We therefore only need to tabulate the additional $2(n_* - 2)$ functions $P_K^{2s}(T_r)$ and $P_K^{2p}(T_r)$ to account for $n_* - 2$ "weak" interface states in addition to $2s$ and $2p$ (note that $P_K^e = 1 - P_K^{2s} - P_K^{2p}$).

In practice, we can further reduce the computational load by simply using $P_{ns}^{2s} = P_{nd}^{2s} = P_{np}^{2p} = 0$, $P_{ns}^{2p} = P_{nd}^{2p} = P_{np}^{2s} = 1$. This amounts to assuming that for $n \geq 3$, ns and nd states are in Boltzmann equilibrium with $2p$, whereas np states are in Boltzmann equilibrium with $2s$—see Eq. (42), and rewriting transitions from $n \geq 3$

states as transitions from the $n = 2$ state to which they are tightly coupled. This is extremely accurate at all times when two-photon transitions significantly affect recombination. The validity of this statement is somewhat weaker for the $n = 4$ states, which are somewhat in between equilibrium with the $2s$ state and the $2p$ state (because allowed transitions connect them to both states). However, as decays from $4p$, $4s$ and $4d$ only marginally affect recombination anyway (at the level of a few 10^{-4}), and $2s$ and $2p$ are very close to equilibrium at the relevant times, this approximation is still very accurate. We explicitly checked that using the approximate values for the P_K^i instead of their exact values in Eqs. (35), (41), and (42) leads to maximum errors on the recombination history $|\Delta x_e|/x_e < 3 \times 10^{-5}$.

We illustrate the formulation adopted in this paper schematically in Fig. 2.

The system of Eqs. (37)–(39) is just the extension of Peebles effective three-level atom to an effective four-level atom, properly accounting for the nonzero radiation field, the nearly instantaneous multiple transitions between excited states, the fact that $2s$ and $2p$ are out of Boltzmann equilibrium, and possibly additional radiative transfer effects and decays from higher shells through the appropriate coefficients $\tilde{\mathcal{R}}$. Making the usual steady state assumption for the excited states, we can first solve for x_{2s} and x_{2p} and then evolve x_e. When using the simple $2p \leftrightarrow 1s$ and $2s \leftrightarrow 1s$ transition rates of Sec. II A, the system is simple enough that we may write the function $\dot{x}_e$ explicitly as an illustration:

$$\dot{x}_e = -C_{2s}(n_H x_e^2 \mathcal{A}_{2s} - x_{1s}\mathcal{B}_{2s}e^{-E_{21}/T_r})$$
$$- C_{2p}(n_H x_e^2 \mathcal{A}_{2p} - 3x_{1s}\mathcal{B}_{2p}e^{-E_{21}/T_r}), \quad (43)$$

where the C factors are given by

$$C_{2s} \equiv \frac{\Lambda_{2s,1s} + \mathcal{R}_{2s \to 2p}\frac{R_{\mathrm{Ly}\alpha}}{\Gamma_{2p}}}{\Gamma_{2s} - \mathcal{R}_{2s \to 2p}\frac{\mathcal{R}_{2p \to 2s}}{\Gamma_{2p}}}, \quad (44)$$

and

$$C_{2p} \equiv \frac{R_{\mathrm{Ly}\alpha} + \mathcal{R}_{2p \to 2s}\frac{\Lambda_{2s,1s}}{\Gamma_{2s}}}{\Gamma_{2p} - \mathcal{R}_{2p \to 2s}\frac{\mathcal{R}_{2s \to 2p}}{\Gamma_{2s}}}, \quad (45)$$

and where we have used the effective inverse lifetimes

$$\Gamma_{2s} \equiv \mathcal{B}_{2s} + \mathcal{R}_{2s,2p} + \Lambda_{2s,1s} \quad \text{and}$$
$$\Gamma_{2p} \equiv \mathcal{B}_{2p} + \mathcal{R}_{2p,2s} + R_{\mathrm{Ly}\alpha}. \quad (46)$$

In Fig. 3 we show the changes to the recombination history resulting from an accurate effective multilevel computation, as compared to the effective three-level atom computation, using in both cases the simple decay rates to the ground state described in Sec. II A. For comparison, we also show the resulting changes when using an effective three-level atom model with a fudge factor $F = 1.14$ as in the code RECFAST [9]. We checked that it is not possible to reproduce the correct effective MLA computation with a constant fudge factor in an effective three-level atom. We find that the best fitting fudge factor would be $F = 1.126$, with relative errors reaching 0.2%. In any case, the effective MLA computation is so simple and computationally efficient that the need for nonphysical fudge factors does not arise.

In Fig. 4 we show the effect of adding higher-order Lyman transitions and feedback between them. It is sufficient to include Lyman transitions up to $\mathrm{Ly}\gamma$, as neglecting higher transitions leads to relative changes of 10^{-5} only [15]. This initially speeds up recombination by adding more decay paths to the ground state, then slows it down due to delayed reabsorptions of $\mathrm{Ly}\beta$ photons in the $\mathrm{Ly}\alpha$ line. Our results are similar to those of Ref. [17].
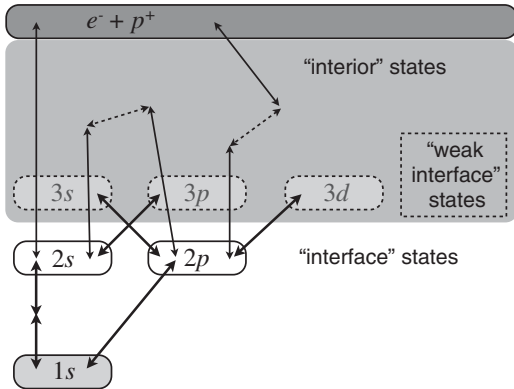

FIG. 2. Schematic representation of the hydrogen atom, with the nomenclature used in this paper. Slow transitions from the "weak interface" states to the ground state are counted as transitions from the $n = 2$ state with which they are in equilibrium.
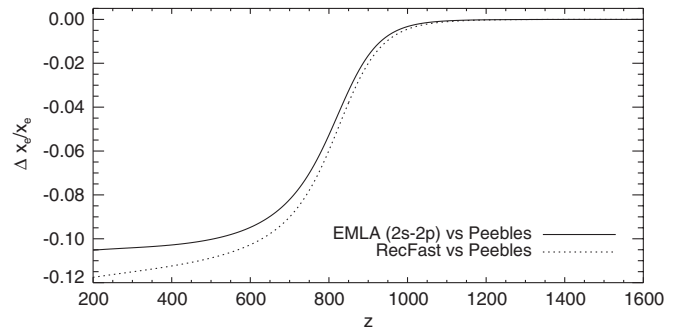

FIG. 3. Fractional changes in the ionization history relative to the effective three-level atom model. The "RecFast" model is an effective three-level atom with the case-B recombination coefficient multiplied by a fudge factor F = 1.14. The same prescription for the evolution of the matter temperature is used in all cases, see Sec. V E.
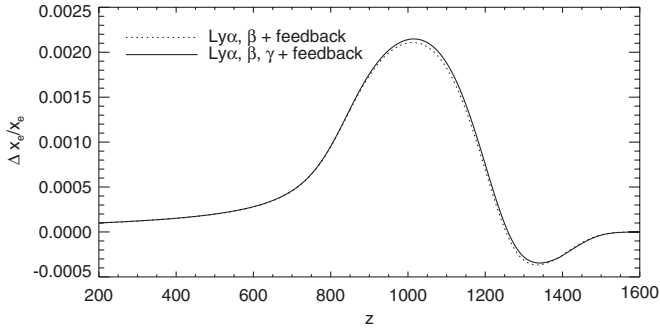
FIG. 4.   Fractional changes in the ionization history when including higher-order Lyman transitions and feedback between them, compared to the effective multilevel atom model with $2s$ and $2p$ only.

For now on, our "base" model will be the effective multilevel atom model with Lyman-α, β and γ transitions and feedback between them (assuming a blackbody radiation field incoming on Lyγ). In the next sections we graft two-photon processes onto this base model.

IV. TWO-PHOTON PROCESSES: FORMAL DESCRIPTION

A. Overview

It is well known since the first works on primordial recombination that $2s \rightarrow 1s$ two-photon decays significantly contribute to the recombination dynamics [1,2]. Even with a relatively low decay rate, the forbidden $2s \rightarrow 1s$ decays are indeed comparable in efficiency to the highly self-absorbed Lyman-α transition; in fact, more than half of hydrogen atoms have formed through the $2s \rightarrow 1s$ channel [40]. This process was traditionally accounted for with the total $2s \rightarrow 1s$ decay rate in a vacuum, $\Lambda_{2s,1s} \approx 8.22$ s^{-1}, with a two-photon absorption rate obtained by detailed balance considerations. For the level of accuracy required for future CMB experiments, one needs to account for stimulated two-photon decays [22] and nonthermal absorptions [23,24].

Recently, it was suggested that two-photon decays from higher lying ns and nd states may also lead to percent level corrections to the recombination history [25]. Inclusion of such decays presents an additional conceptual difficulty which was not present for the $2s \rightarrow 1s$ decays: the problem of double-counting. Indeed, there is no fundamental difference between a sequence of two allowed one-photon transitions $nl \rightarrow n'p$, $n'p \rightarrow 1s$, with $1 < n' < n$, and a two-photon decay from the nl state near resonance (i.e., where the energy of the two photons are near $E_{nn'}$ and $E_{n'1}$, respectively). Approximate solutions were presented in Refs. [25,26,41] (for a review, see Ref. [24]). The double-counting problem as well as the reabsorption problem were resolved with a numerical approach, solving the radiative transfer equations for the photon field in Ref. [24], which also provided analytic approximations

to check the validity of the numerical result. In this work, we will use the same numerical method as in Ref. [24], which we then extend to account for frequency diffusion near the Lyα line. In this section, we review the formalism presented in Ref. [24] and how to solve the double-counting problem. In Sec. V we will describe our numerical method for solving simultaneously the radiative transfer equation and the evolution of the atomic level populations.

B. Two-photon decays and Raman scattering

We start by defining the coefficient:

$$\frac{d\Lambda_{nl}}{d\nu} \equiv \frac{\alpha_{\text{fs}}^6 \nu^3 \nu'^3}{108(2l+1)E_I^6} |\mathcal{M}(\nu)|^2, \qquad (47)$$

where the matrix element $\mathcal{M}(\nu)$ is given by Eq. (B5) of Ref. [24], and $\nu' \equiv |\nu - \nu_{n1}|$, where ν_{n1} is the frequency of the Ly $-n$ transition. For $\nu < \nu_{n1}$, $d\Lambda_{nl}/d\nu$ is the rate of spontaneous two-photon decays from nl per frequency interval. For $\nu > \nu_{n1}$, $d\Lambda_{nl}/d\nu \times f_{\nu'}$ (where f_ν is the photon occupation number at frequency ν) is the rate of spontaneous Raman scatterings per frequency interval per atom initially in nl (in the notation of Ref. [24], $d\Lambda_{nl}/d\nu = dK_{nl}/d\nu$ for $\nu > \nu_{n1}$). The function $d\Lambda_{nl}/d\nu$ is continuous across $\nu = \nu_{n1}$, where it vanishes.

We can now write the net rate of $nl \leftrightarrow 1s$ two-photon transitions per frequency interval per hydrogen atom, for which the highest energy photon has frequency $\nu < \nu_{n1}$:

$$\Delta_{nl}(\nu < \nu_{n1}) = \frac{d\Lambda_{nl}}{d\nu}\left[x_{nl}(1+f_{\nu'})(1+f_\nu) - \frac{g_{nl}}{g_{1s}}x_{1s}f_{\nu'}f_\nu\right]. \qquad (48)$$

For $\nu > \nu_{n1}$, the appropriate rate is that of Raman scattering events:

$$\Delta_{nl}(\nu > \nu_{n1}) = \frac{d\Lambda_{nl}}{d\nu}\left[x_{nl}f_{\nu'}(1+f_\nu) - \frac{g_{nl}}{g_{1s}}x_{1s}(1+f_{\nu'})f_\nu\right], \qquad (49)$$

In both cases, we can assume that the photon occupation number for the low-energy photons is that of a blackbody, since the optical depth for two-photon absorption of the low-energy photons is tiny (for a discussion, see Ref. [24]). We therefore set $f_{\nu'} = (e^{h\nu'/T_r} - 1)^{-1}$. Moreover, the photon occupation number for frequencies $\nu > \nu_{\text{Ly}\alpha}/2$ is much smaller than unity: $f_\nu \ll 1$. This means that we can neglect stimulated emission by the high-energy photons in Eqs. (48) and (49). Given these considerations, the net rate of two-photon transitions can be written in the following form, valid for both $\nu < \nu_{n1}$ and $\nu > \nu_{n1}$:

$$\Delta_{nl}(\nu) = \frac{d\Lambda_{nl}}{d\nu}|e^{h(\nu-\nu_{n1})/T_r} - 1|^{-1}$$
$$\times \left[x_{nl} - \frac{g_{nl}}{g_{1s}}x_{1s}e^{h(\nu-\nu_{n1})/T_r}f_\nu\right]. \qquad (50)$$

C. Resonant scattering in Lyman-α

We now consider pure scattering events:

$$H(1s) + \gamma \rightarrow H(1s) + \gamma. \tag{51}$$

In the low-frequency limit this corresponds to the familiar Rayleigh scattering phenomenon; the cross section however has resonances at the Lyman-series lines, which correspond to resonant Rayleigh scattering.

Rayleigh scattering events conserve the photon frequency in the atom's rest frame. In the comoving frame (frame in which the CMB appears isotropic), the frequency of the scattered photon appears shifted due to the thermal motions of the scatterers. The frequencies of the incoming and outgoing photons are however statistically correlated. Mathematically, there is a definite probability distribution $p(\nu, \nu')$, such that $p(\nu, \nu')d\nu'$ is the probability that the outgoing photon has frequency in $[\nu', \nu' + d\nu']$ given that the incoming photon had frequency ν, and this function generally depends on both ν and ν'. For $T_{\rm m} \ll h\nu$, which is the case near the Lyman lines, the variance of the frequency shifts imparted by thermally moving atoms is given by

$$\langle \delta \nu^2 \rangle \equiv \int (\nu' - \nu)^2 p(\nu, \nu') d\nu' = \frac{2T_{\rm m}}{m_{\rm H} c^2} \nu^2. \tag{52}$$

The rate of injection of photons per frequency interval at frequency ν, due to resonant Rayleigh scattering in $Ly - \alpha$, can be written in the general form (neglecting stimulated scatterings):

$$\Delta_{1s}(\nu) = x_{1s}\left[\int f_{\nu'} R(\nu', \nu) d\nu' - \int f_\nu R(\nu, \nu') d\nu'\right], \tag{53}$$

where $R(\nu, \nu') = \frac{d\Lambda_{1s}}{d\nu} p(\nu, \nu')$ is the differential rate of scatterings per hydrogen atom in the ground state, per unit frequency interval for both the incoming and outgoing photons (it has units of s^{-1} Hz^{-2}). The scattering kernel must respect detailed balance:

$$R(\nu, \nu')e^{-h\nu/T_{\rm m}} = R(\nu', \nu)e^{-h\nu'/T_{\rm m}}. \tag{54}$$

To be fully general one should compute the scattering kernel from first principles. However, simplifications can be easily made in various regimes.

Far from any resonance, the rate of redshifting due to the Hubble expansion is much larger than the rate of frequency diffusion due to scattering (see, for example, the discussion in Ref. [15]). We can neglect Rayleigh scattering there, and set $\Delta_{1s}(\nu) = 0$.

Near Lyman resonances, we have

$$\frac{d\Lambda_{1s}}{d\nu}(\nu \approx \nu_{n1}) \approx 3A_{np,1s}p_{\rm sc}^n \phi_{V,n}(\nu), \tag{55}$$

where $p_{\rm sc}^n = A_{np,1s}/\Gamma_{np}$ is the scattering probability in the Lyman-n line (the complementary events being two-photon absorptions and two-photon photoionizations), and $\phi_{V,n}(\nu)$ is the Voigt profile for the $Ly - n$ line. In the Doppler core, we can approximate the partial redistribution induced by scattering events by a complete redistribution, i.e., approximate $p(\nu, \nu') \approx \phi_{V,n}(\nu) \approx \phi_{D,n}(\nu)$, where ϕ_D is the Doppler profile. This approximation is valid because in the Doppler core, complete redistribution recovers the correct rms frequency shift during scattering events, Eq. (52) (if one averages over the frequencies of absorbed photons).

In the damping wings of Lyman resonances above $Ly\alpha$, the rate of scatterings is of the same order as the rate of two-photon absorptions. Each scattering event shifts the photon frequency by a very small amount compared to the width over which the radiation field varies ($\delta\nu_{\rm rms}/\nu \sim 2.5 \times 10^{-5}$). Partial redistribution is therefore essentially coherent in the comoving frame, i.e., $p(\nu, \nu') \approx \delta(\nu' - \nu)$, which implies $\Delta_{1s}(\nu) \approx 0$. For a more quantitative argument, see Ref. [15].

The only frequency regime where Rayleigh scattering affects the radiation field in a nontrivial way is in the damping wings of $Ly\alpha$. In this line, indeed, scattering events are much more frequent than two-photon absorption events (by a factor of $\sim 10^4$). Resonant scattering therefore leads to a significant diffusion in frequency. Because the frequency shifts are small compared to the width over which the radiation field varies, the integral scattering operator can be approximated by a second order differential operator—a Fokker-Planck operator [20,21,42,43]. For the purpose of numerical implementation, the relevant properties are (i) the fact that this operator is nearly local (it only connects neighboring bins in frequency) (ii) it must respect detailed balance and (iii) the diffusion rate must be correct. We will explain our numerical method for the implementation of Lyman-α diffusion in Sec. VA.

We note that a number of analytic treatments of Lyman-α scattering in the recombination epoch have been proposed in the past [43–47]. However, since two-photon emission and absorption act on the same region of frequency space, and since both processes involve high optical depth, an accurate recombination history can only be obtained by considering all processes simultaneously.

D. The radiative transfer equation

The radiative transfer equation for the photon occupation number is

$$\frac{\partial f_\nu}{\partial t} - H\nu \frac{\partial f_\nu}{\partial \nu} = \frac{c^3 n_{\rm H}}{8\pi\nu^2}\left[\sum_{n\geq 2, l} \Delta_{nl}(\nu) + \Delta_{1s}(\nu)\right], \tag{56}$$

where the left-hand-side is the derivative of the photon occupation number along a photon trajectory in the expanding universe, and the prefactor on the right-hand-side converts the number of photons per unit frequency per hydrogen atom to the photon occupation number.

E. Inclusion in the effective multilevel atom rate equations

1. Formal two-photon decay rates

As mentioned earlier, including two-photon decays from states with $n > 2$ and Raman scattering events poses a double-counting problem. In principle, to avoid this double-counting issue, one should discard "1 + 1" decays (or decays following an absorption event, which is like a Raman scattering event on resonance) altogether. If one were to pursue this idea, one should not consider the p states at all anymore (as they are formally only intermediate states in two-photon processes), but consider all s and d states as "interface states" and allow for two-photon recombinations to the ground state. The two-photon $nl \leftrightarrow 1s$ transition rates would then become

$$\dot{x}_{nl}|_{1s}^{(2\gamma)} = -\dot{x}_{1s}|_{nl}^{(2\gamma)} = x_{1s}\tilde{R}_{1s,nl}^{\text{total}} - x_{nl}\tilde{R}_{nl,1s}^{\text{total}}, \quad (57)$$

where the formal transition rates are given by

$$\tilde{R}_{1s,nl}^{\text{total}} \equiv \int \frac{d\Lambda_{nl}}{d\nu} \frac{g_{nl}}{g_{1s}} |e^{h(\nu_{\text{n1}}-\nu)/T_{\text{r}}} - 1|^{-1} f_\nu d\nu \quad (58)$$

and

$$\tilde{R}_{nl,1s}^{\text{total}} \equiv \int \frac{d\Lambda_{nl}}{d\nu} |e^{h(\nu-\nu_{\text{n1}})/T_{\text{r}}} - 1|^{-1} d\nu, \quad (59)$$

where the integrals run from $\nu_{n1}/2$ to ν_c. In principle Eq. (57)–(59) can be included in a standard or effective multilevel atom code, provided one solves simultaneously for the radiation field, using the radiative transfer equation Eq. (56).

2. Decomposition into "1 + 1" transitions and nonresonant contributions

Two-photon decays from higher excited states constitute, however, a correction to the recombination history computed in the standard "1 + 1" picture, and we would like to implement it as such. We start by formally separating the integrals in Eqs. (58) and (59) in two contributions: the resonant pieces, for $\nu \approx \nu_{n'1}$, and a nonresonant piece, for frequencies far enough from any resonance. We therefore rewrite, formally,

$$\tilde{R}_{1s,nl}^{\text{total}} = \sum_{n'} \tilde{R}_{1s,nl}^{(n'p)} + \tilde{R}_{1s,nl} \quad \text{and}$$

$$\tilde{R}_{nl,1s}^{\text{total}} = \sum_{n'} \tilde{R}_{nl,1s}^{(n'p)} + \tilde{R}_{nl,1s}, \quad (60)$$

where the resonant contributions $\tilde{R}_{1s,nl}^{(n'p)}$ and $\tilde{R}_{nl,1s}^{(n'p)}$ are defined in a similar manner as in Eqs. (58) and (59), but with the integration being carried over a narrow range $\Delta\nu$ near $\nu_{n'1}$, and $\tilde{R}_{1s,nl}$ and $\tilde{R}_{nl,1s}$ are the nonresonant pieces required to complete the total rates. So far the separation is just formal and we have not made any approximation.

3. "1 + 1" Resonant contribution

We now notice that near a resonance $\nu \approx \nu_{n'1}$, the two-photon differential decay rate $d\Lambda_{nl}/d\nu$ takes on the following form (if $n > n'$):

$$\frac{d\Lambda_{nl}}{d\nu}\bigg|_{\nu\approx\nu_{n'1}} \approx \frac{1}{4\pi^2} \frac{A_{nl,n'p}A_{n'p,1s}}{(\nu - \nu_{n'1})^2 + (\Gamma_{n'p}/4\pi)^2}$$

$$= A_{nl,n'p} \frac{A_{n'p,1s}}{\Gamma_{n'p}} \phi_L(\nu - \nu_{n'1}; \Gamma_{n'p}), \quad (61)$$

where $\Gamma_{n'p}$ is the total inverse lifetime of the state $n'p$, and the Lorentzian profile is given by

$$\phi_L(\Delta\nu; \Gamma) \equiv \frac{\Gamma/(4\pi^2)}{\Delta\nu^2 + (\Gamma/4\pi)^2}. \quad (62)$$

For $n < n'$, the first coefficient in Eq. (61) should be $g_{n'p}/g_{nl} \times A_{n'p,nl}$ instead of $A_{nl,n'p}$. When accounting for the thermal motions of atoms, the Lorentzian profile should be replaced by a Voigt profile. We can now approximate the resonant pieces with the following expressions, valid for both $n < n'$ and $n > n'$:

$$\tilde{R}_{1s,nl}^{(n'p)} \approx 3A_{n'p,1s}\bar{f}_{\nu_{n'1}} \frac{R_{n'p,nl}}{\Gamma_{n'p}} \quad (63)$$

and

$$\tilde{R}_{nl,1s}^{(n'p)} \approx R_{nl,n'p} \frac{A_{n'p,1s}}{\Gamma_{n'p}}, \quad (64)$$

where $\bar{f}_{\nu_{n'1}}$ is the photon occupation number averaged over the Voigt profile near the resonance $\nu \approx \nu_{n'1}$. Equations (63) and (64) are exactly what one would obtain in the "1 + 1" picture after "factoring out" the p states (with a procedure similar to what is used to get rid of the "interior" states in the EMLA method). Having these resonant rates is exactly equivalent to having optically thin one-photon transitions between the nl and $n'p$ states, with rates $R_{nl,n'p}(T_{\text{r}})$ and $R_{n'p,nl}(T_{\text{r}})$, and optically thick Lyman transitions, with net rate

$$\dot{x}_{n'p}|_{1s} = -\dot{x}_{1s}|_{n'p} = A_{n'p,1s}(3x_{1s}\bar{f}_{\nu_{n'p}} - x_{n'p}). \quad (65)$$

To obtain the net decay rates in the Lyman transitions, one then needs to solve for the radiation field in the immediate vicinity of Lyman resonances. If the frequency region for which two-photon transitions are considered as "resonant" is narrow enough, this can be done in the Sobolev approximation. Indeed, all the relevant conditions are met (see also discussion in Ref. [24]; for more details on the Sobolev approximation, see, for example, Ref. [15]):

First, the two-photon absorption and emission profiles can both be approximated by the same resonance profile Eq. (61). This relies on the assumption that the blackbody radiation field varies little across the "resonant" region, and requires for its width to satisfy $\Delta\nu \ll T_{\text{r}}/h$.

Second, we argued in Sec. IV C that one could assume complete frequency redistribution for resonant scattering near the Doppler core of Lyman resonances. The "resonant" region should therefore not exceed a few Doppler widths.

Finally, if we consider regions in frequency narrow enough around the resonances, we can use the steady state approximation. This requires $\Delta \nu / \nu \ll 1$.

We can see that considering the "resonant" region around each Lyman resonance to be a few Doppler widths wide meets all the requirements.

An additional assumption required here is that excited states are near Boltzmann equilibrium, which is very accurate at redshifts for which two-photon processes are important. In the Sobolev approximation, and in the limit of large Sobolev optical depth, Eq. (65) becomes the standard Lyman decay rate Eq. (18), where f_{np}^+ is the photon occupation number incoming on the resonance, preprocessed by two-photon processes and diffusion in the blue damping wing of the line.

The Sobolev approximation is probably the least accurate for Lyα decays where partial redistribution due to resonant scattering is important. However, the large optical depth to two-photon absorptions in the Lyman-α blue damping wing, in conjunction with frequent scatterings, drive the radiation field to the equilibrium value $f_\nu = x_{n'p}/(3x_{1s})e^{((-h(\nu-\nu_{n'1}))/T_m)}$ over several Doppler widths (of the order of 40 Doppler widths, see Ref. [15]). As a consequence the net decay rate in the core of the resonance is very small anyway. We checked that in the presence of two-photon transitions and frequency diffusion, even setting $\dot{x}_{n'p}|_{1s} = 0$ instead of the expression given by Eq. (65) leads to relative changes to the recombination history of at most 7×10^{-4}. Given that frequency diffusion leads to corrections of a few percent at most to the decay rate in Lyα when radiative transfer is treated carefully even at the line center [20], we can be confident that using the Sobolev approximation for the resonant contributions of two-photon decays is accurate to better than 10^{-4}.

4. "Pure two-photon" nonresonant contribution

In the previous section we discussed how two-photon decays within a few Doppler widths of Lyman resonances can in fact be accounted for in the standard "1 + 1" picture. To evaluate the nonresonant pieces, $\tilde{R}_{1s,nl}$ and $\tilde{R}_{nl,1s}$, we need to solve the radiative transfer equation, Eq. (56), to obtain the photon occupation number. The subject of Sec. V is to describe our numerical method of solution.

Note that choosing the "resonant" regions to be a few Doppler widths has an additional advantage. Since a Doppler width is $\sim 10^3$ times wider than the natural width of Lyman lines, it is not necessary to account for the pole displacements in the computation of the differential two-photon decay rates in the nonresonant region. In addition,

the fraction of two-photon decays that are considered non-resonant will be small (of the order of $\Gamma_{np}/(4\pi^2)/\Delta\nu$, where $\Delta\nu$ is the width of the "resonant" region). For $\Delta\nu$ of a few Doppler widths, this fraction is $\sim 10^{-4}$. This means that the "pure" two-photon decay rates $\tilde{R}_{nl,1s}$ are much smaller than the total inverse lifetime of the nl state, Γ_{nl}, which is required to simplify the effective MLA model to an effective four-level atom model as we discussed in Sec. III B 2.

As a final note, we want to emphasize why the final result is independent of the exact boundary between "resonant" and "nonresonant" regions, so long as the resonant regions are a few Doppler widths wide. If one were to increase the width of the "resonant" region, then the "pure" two-photon transition rates $\tilde{R}_{nl,1s}$ and $\tilde{R}_{1s,nl}$ would decrease, mainly because of the change of the integration region in the blue wings of the resonance—in the red wing, the radiation field has reached near equilibrium with the line and the net rate of decays immediately blueward of line center is very small anyway. This decrease would be nearly exactly compensated by the increase of what is considered as "1 + 1" decays, as the photon occupation number incoming on the Lyman resonances, f_{np}^+, would be decreased due to the smaller optical depth due to "pure" two-photon absorptions in the blue wing. Hirata (2008) checked the independence of the result form the exact value chosen for the width of the "resonant" region, and found that even changing this width by a factor of 9 lead to relative changes of at most 4×10^{-4} in the recombination history.

V. NUMERICAL SOLUTION OF THE RADIATIVE TRANSFER EQUATION

A. Discretization of the radiative transfer equation

To solve the radiative transfer equation [Eq. (56)] numerically in the "nonresonant" frequency region, we follow the method of Hirata (2008), and extend it to also account for frequency diffusion.

We will consider the radiation field in the vicinity of N frequency "spikes" ν_b, for $b = 1, 2, \ldots N$. Each spike has an associated width $\Delta\nu_b$ (which is just the separation between consecutive spikes if they are linearly spaced, for example).

We use the discretized differential two-photon rate

$$\frac{d\Lambda_{nl}}{d\nu}\bigg|_{\text{used}} = \sum_b A_{nl,b}\delta_\epsilon(\nu - \nu_b), \qquad (66)$$

where we use the coefficients

$$A_{nl,b} \equiv \int_{\Delta\nu_b} \frac{d\Lambda_{nl}}{d\nu}d\nu, \qquad (67)$$

where the integral is carried over the frequency region associated with the spike $\Delta\nu_b$. The function $\delta_\epsilon(\nu - \nu_b)$ in Eq. (66) should be understood as a sharp profile centered

at ν_b, which integrates to unity, and has support in $[\nu_b - \epsilon, \nu_b + \epsilon]$. The solution we derive is in the limit $\epsilon \to 0$, for which $\delta_\epsilon \to \delta$, the Dirac delta function. Doing so, we are simply approximating the optical depth as concentrated in discrete frequencies instead of being a smooth function.

The main new contribution of the present work is the discretization method for the scattering operator. We use the discretized scattering kernel

$$R(\nu, \nu')|_{\text{used}} = \sum_{b,b'} R_{b,b'} \delta_\epsilon(\nu - \nu_b)\delta_\epsilon(\nu' - \nu_{b'}). \quad (68)$$

We enforce detailed balance:

$$R_{b,b'}e^{((-h\nu_b)/T_m)} = R_{b',b}e^{((-h\nu_{b'})/T_m)}. \quad (69)$$

Moreover, we use the diffusion approximation for resonant scattering. This allows us to assume that the numerical scattering kernel $R_{b,b'}$ is nonvanishing only for neighboring bins, $b' = b \pm 1$. In order to obtain the correct diffusion rate, we set

$$(\nu_{b+1} - \nu_b)^2 R_{b,b+1} + (\nu_{b-1} - \nu_b)^2 R_{b,b-1}$$
$$= 3\frac{A_{2p,1s}^2}{4\pi^2(\nu - \nu_{\text{Ly}\alpha})^2}\Delta\nu_b\frac{2T_m}{m_H c^2}\nu_{\text{Ly}\alpha}^2, \quad (70)$$

where we used the damping wing approximation for the absorption profile (and approximate $\nu_b^2 \approx \nu_{\text{Ly}\alpha}^2$ in the multiplicative factor).

As boundary conditions, we assume a vanishing photon flux due to diffusion at the boundaries of our domain, i.e., formally, $R_{1,0} = R_{N,N+1} = 0$ (in fact we set these conditions at the boundaries of the diffusion domain, smaller than the entire frequency domain considered). Using Eq. (70), we then obtain $R_{1,2}$ and $R_{N,N-1}$. Using iteratively Eqs. (69) and (70), we can then obtain all the coefficients of the numerical diffusion kernel, starting from the boundaries, and up to line center. Denoting b_1 the highest bin below $\text{Ly} - \alpha$ and $b_1 + 1$ the first bin above $\text{Ly} - \alpha$, we obtain all coefficients up to $R_{b_{\text{Ly}\alpha},b_1}$ on the red side of $\text{Ly} - \alpha$, and up to $R_{b_{\text{Ly}\alpha},b_1+1}$ on the blue side (we do not follow the radiation field at the central bin $b_{\text{Ly}\alpha}$ but can still define these coefficients). Note that with this method we cannot ensure that the diffusion rate at the central bin is correct. However, the exact value of the diffusion rate at line center does not matter, as long as it is high enough to ensure that the photon occupation number reaches the equilibrium spectrum $f_\nu \propto e^{((-h\nu)/T_m)}$.

B. Solution of the discretized radiative transfer equation

To simplify the notation, we define the following rate coefficients:

$$R_{nl,b} \equiv \frac{d\Lambda_{nl}}{d\nu}\Big|_{\nu_b}|e^{h(\nu_b - \nu_{n1})/T_r} - 1|^{-1}\Delta\nu_b \quad \text{and}$$

$$R_{b,nl} \equiv \frac{g_{nl}}{g_{1s}}e^{h(\nu_b - \nu_{n1})/T_r}R_{nl,b}. \quad (71)$$

This coefficients can be thought of as transition rates between bound states and a set of "virtual" levels with associated energies $E_b = h\nu_b$ [24].

We define the total Sobolev optical depth in the b-th frequency spike:

$$\Delta\tau_b \equiv \frac{c^3 n_H x_{1s}}{8\pi\nu_b^3 H}\left(\sum_{nl} R_{b,nl} + \sum_{b'=b\pm 1} R_{b,b'}\right). \quad (72)$$

We also define the average photon occupation number near ν_b:

$$\bar{f}_{\nu_b} \equiv \int_{\nu_b-\epsilon}^{\nu_b+\epsilon}\delta_\epsilon(\nu - \nu_b)f_\nu d\nu. \quad (73)$$

Finally, we define the equilibrium photon occupation number at the b-th frequency spike:

$$f_{\nu_b}^{\text{eq}} \equiv \frac{\sum_{nl} x_{nl} R_{nl,b} + x_{1s}\sum_{b'}\bar{f}_{\nu_{b'}}R_{b',b}}{x_{1s}(\sum_{nl} R_{b,nl} + \sum_{b'}R_{b,b'})}. \quad (74)$$

In the vicinity of ν_b, the discretized radiative transfer equation becomes

$$\frac{1}{H\nu_b}\frac{\partial f_\nu}{\partial t} - \frac{\partial f_\nu}{\partial\nu} = \Delta\tau_b\delta_\epsilon(\nu - \nu_b)[f_{\nu_b}^{\text{eq}} - f_\nu]. \quad (75)$$

In the limit that the support of the delta function becomes vanishingly small, $\epsilon \to 0$, the discretized radiative transfer equation can be solved in the steady state approximation, and one can neglect the time derivative. This is similar to the commonly used Sobolev approximation, except that we are now making this approximation in the vicinity of an artificially introduced spike (as opposed to a true resonance line), for the purposes of numerical resolution. Another conceptual difference is that the equilibrium photon occupation number also depends on the averaged value of the radiation field at neighboring bins, because of frequency diffusion. Given the photon occupation number at the blue edge of the b-th spike, $f_{\nu_b+\epsilon}$, this equation has a well-known solution f_ν. The quantities of interest for us are the photon occupation number at the red edge of the spike $f_{\nu_b-\epsilon}$ and the average photon occupation number in the spike $\bar{f}_{\nu_b}$. They are given by the following expressions (for a derivation, see, for example, Refs. [15,24]):

$$f_{\nu_b-\epsilon} = f_{\nu_b+\epsilon}e^{-\Delta\tau_b} + f_{\nu_b}^{\text{eq}}(1 - e^{-\Delta\tau_b}), \quad (76)$$

and

$$\bar{f}_{\nu_b} = \Pi_b f_{\nu_b+\epsilon} + (1 - \Pi_b)f_{\nu_b}^{\text{eq}}, \quad (77)$$

where Π_b is the Sobolev escape probability from the b-th spike:

$$\Pi_b \equiv \frac{1 - e^{-\Delta\tau_b}}{\Delta\tau_b}. \tag{78}$$

We now use the variables

$$x_b \equiv x_{1s}\bar{f}_{\nu_b}. \tag{79}$$

As explained in Ref. [24], x_b can be interpreted as the population of the virtual level b. One should however keep in mind that this is just a convenient rewording for the radiation field intensity.

Using the definition of $f_{\nu_b}^{\mathrm{eq}}$, Eq. (74), we can rewrite Eq. (77) in the form

$$T_{b,b}x_b = \sum_{nl}x_{nl}R_{nl,b} + \sum_{b'=b\pm1}x_{b'}R_{b',b} + s_b, \tag{80}$$

where we have defined

$$T_{b,b} \equiv \frac{1}{1-\Pi_b}\left(\sum_{nl}R_{b,nl} + \sum_{b'=b\pm1}R_{b,b'}\right) \quad \text{and}$$

$$s_b \equiv \Pi_b x_{1s}f_{\nu_b+\epsilon}T_{b,b}. \tag{81}$$

We only follow two-photon decays in the damping wings of resonances, but we should still account for frequency diffusion between line center and the neighboring bins. At the Lyman-α line center, the radiation field is in equilibrium with the $2p/1s$ ratio: $f_{\nu_{\mathrm{Ly}\alpha}} = x_{2p}/(3x_{1s})$. If b_1 is the highest frequency bin below Ly $-$ α, (and $b_1 + 1$ is the first bin above Ly $-$ α), we therefore define the transition rates with $2p$:

$$R_{2p,b_1} = \frac{1}{3}R_{b_{\mathrm{Ly}\alpha},b_1} \quad \text{and} \quad R_{2p,b_1+1} = \frac{1}{3}R_{b_{\mathrm{Ly}\alpha},b_1+1}. \tag{82}$$

Provided that we set $R_{b_1,b_1+1} = R_{b_1+1,b_1} = 0$, Eq. (80) remains valid for $b = b_1$, $b_1 + 1$. Adding these transitions with the central frequency bin will ensure that the photon occupation number is driven to its equilibrium value near line center, $f_\nu^{\mathrm{eq}} = x_{2p}/(3x_{1s})e^{((-h(\nu-\nu_{\mathrm{Ly}\alpha}))/\mathrm{T_m})}$.

We now use Eq. (42) for the populations x_{nl} with $n \geq 3$. We define the coefficients, for $i = 2s, 2p$:

$$T_{b,i} \equiv -R_{i,b} - \sum_{n\geq3,l}\frac{g_{nl}}{g_i}e^{-E_{n2}/T_r}P_{nl}^i(T_r)R_{nl,b}. \tag{83}$$

We define the new source vector:

$$S_b \equiv s_b + x_e^2\sum_{n\geq3,l}\frac{g_{nl}}{g_e}e^{-E_n/T_r}P_{nl}^e(T_r)R_{nl,b}, \tag{84}$$

We also define the coefficients

$$T_{b,b\pm1} \equiv -R_{b\pm1,b}. \tag{85}$$

The discretized radiative transfer equation then takes the final form:

$$T_{b,2s}x_{2s} + T_{b,2p}x_{2p} + \sum_{b'=b-1}^{b+1}T_{b,b'}x_{b'} = S_b. \tag{86}$$

C. Populations of the excited states

Given the radiation field, we can now compute the two-photon transition rates. Using Eqs. (58) and (59) limited to the "nonresonant" frequency region, we obtain, after discretization,

$$\tilde{R}_{nl,1s} = \sum_b R_{nl,b} \quad \text{and} \quad \tilde{R}_{1s,nl} = \sum_b \bar{f}_{\nu_b}R_{b,nl}. \tag{87}$$

The effective transition rates from the $i = 2s$ and $2p$ states to the ground state are therefore, according to the discussion in Sec. III B 2, and using the definition of $T_{b,i}$ Eq. (83),

$$\tilde{\mathcal{R}}_{i,1s} = -\sum_b T_{b,i} + \sum_n \frac{g_{np}}{g_i}R_{\mathrm{Ly}n}e^{-E_{n2}/T_r}P_{np}^i(T_r), \tag{88}$$

where the first term accounts for two-photon transitions and the second term for escape from Lyman lines (it is understood that $P_{2p}^{2s} = 0$ and $P_{2p}^{2p} = 1$). The effective transition rate for the reverse process is given by

$$\tilde{\mathcal{R}}_{1s,i} = -\sum_b T_{i,b}\bar{f}_{\nu_b} + 3\sum_n R_{\mathrm{Ly}n}P_{np}^i(T_r)f_{np}^+, \tag{89}$$

where we have defined the coefficients, for $i = 2s, 2p$:

$$T_{i,b} \equiv \frac{g_i}{g_{1s}}e^{h(\nu_b-\nu_{21})/T_r}T_{b,i}. \tag{90}$$

For a given radiation field and free electron fraction, we can now obtain an equation for the populations of the excited states $2s, 2p$. We do so in the steady state approximation, i.e., setting $\dot{x}_i = 0$ for $i = 2s, 2p$ in Eqs. (37) and (38). We first define the 2×2 matrix of elements

$$T_{i,i} \equiv \mathcal{B}_i + \mathcal{R}_{i,j} + \tilde{\mathcal{R}}_{i,1s} \quad \text{and} \quad T_{i,j} \equiv -\mathcal{R}_{i,j}. \tag{91}$$

We also define the source vector of elements

$$S_i \equiv n_{\mathrm{H}}x_e^2\mathcal{A}_i + 3\sum_{n\geq2}R_{\mathrm{Ly}n}P_{np}^i(T_r)f_{np}^+. \tag{92}$$

The steady state equation for each state i translates into the linear equation:

$$\sum_{j=2s,2p}T_{i,j}x_j + \sum_b T_{i,b}x_b = S_i, \qquad i = 2s, 2p. \tag{93}$$

D. Evolution of the coupled system of level populations and radiation field

We now have all the necessary pieces to evolve simultaneously the level populations and the radiation field, and compute the free electron fraction. In this section we summarize the procedure and recall the main equations. We start with an initially thermal radiation field, $f_\nu = e^{((-h\nu/T_r))}$. At each time step, we do the following computations:

(1) We obtain the photon occupation number incoming on each bin b assuming free streaming between frequency spikes:

$$f_{\nu_b+\epsilon}(z) = f_{\nu_{b+1}-\epsilon}\left(z' = (1+z)\frac{E_{b+1}}{E_b} - 1\right). \quad (94)$$

We also obtain in the same way the incoming photon occupation number at the Ly $-n$ transitions, f_{np}^+.

(2) We solve for the populations of the $2s$ and $2p$ states and the average photon occupation number at each frequency spike $\bar{f}_{\nu_b} = x_b/x_{1s}$ *simultaneously* by solving the coupled linear system given by Eqs. (86) and (93). Even with a large number of bins for the radiation field ($N = 311$ in our fiducial case), this system is easily solved because the matrix of coefficients $T_{b,b'}$ is tridiagonal and the overall system has the particular sparsity pattern shown in Fig. 5. Such a sparse system can be solved in $\mathcal{O}(N)$ operations (specifically, we can solve the system in $\sim 16N$ operations).

(3) We update the photon occupation number at the red side of each spike, $f_{\nu_b-\epsilon}$, using Eq. (76). At the red side of Lyman resonances, we use $f_{np}^- = x_{np}/(3x_{1s})$, valid in the optically thick limit, where x_{np} is given by Eq. (42) for $n \geq 3$.

(4) After step #2, we can obtain the function $\dot{x}_e(z, x_e)$ through Eq. (39) or[2] (40). This allows us to evolve the free electron fraction to the next time step.

E. Implementation, convergence tests and results

We evolve the free electron fraction during hydrogen recombination in several phases. We use even steps in $\ln a$ (where $a = (1+z)^{-1}$ is the scale factor), with $\Delta \ln a = 8.5 \times 10^{-5}$. We describe our ODE integrator in Appendix C.

(i) We checked that hydrogen and helium recombination never overlap and can be followed separately (to an accuracy of a few times 10^{-4}). We therefore start computing the hydrogen recombination history once helium is completely recombined. Quantitatively, we start hydrogen recombination once the fractional abundance of He$^+$ ions is less than 10^{-4} relative to hydrogen. If this criterion is met earlier than $z = 1650$ we only switch on the hydrogen recombination computation at $z = 1650$. We checked that at this redshift the exact free electron fraction differs from the Saha equilibrium value by no more than a few times 10^{-4} anyway.

(ii) In the first phases of hydrogen recombination, we use the post-Saha expansion described in Appendix D. We do so as long as the free electron
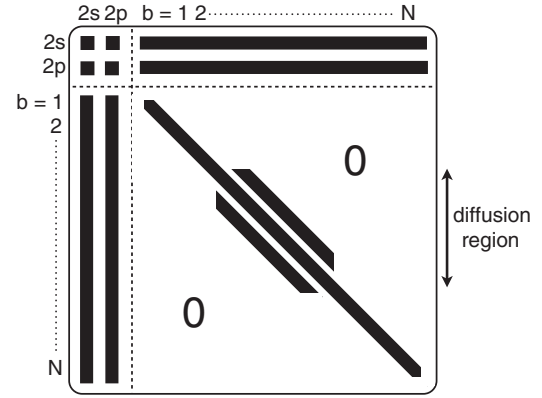

FIG. 5.   Sparsity pattern of the linear system solved for evolving simultaneously the level populations and the radiation field, in the presence of two-photon transitions and frequency diffusion.

fraction differs from the Saha value by $\Delta x_e < 5 \times 10^{-5}$. We checked that explicitly integrating the ODE for x_e instead (with a much smaller time step as the ODE is stiff at early times) leads to maximum changes of $\Delta x_e/x_e \lesssim 3 \times 10^{-4}$.

(iii) From then on and until $z = 700$ we solve simultaneously for the level populations and radiative transfer with two-photon processes and diffusion as described in this section.

(iv) For $z < 700$, we use the simple EMLA equations, with simple decay rates from $2s$ and $2p$ only (i.e., not accounting for higher-order Lyman lines and radiative transfer effects). We checked that moving the last switch to $z = 400$ instead of 700 leads to maximum changes $|\Delta x_e|/x_e < 10^{-4}$.

(v) For the matter temperature evolution, we use the asymptotic solution of Hirata [24] [it can be obtained by setting $\dot{T}_{\rm m} = -HT_{\rm m}$ in Eq. (12)] as long as $1 - T_{\rm m}/T_{\rm r} < 5 \times 10^{-4}$. Depending on cosmology, this corresponds to $750 < z < 950$. After that we switch to solving for x_e and $T_{\rm m}$ simultaneously by using Eq. (12).

All the checks mentioned above were made for a wide range of cosmological parameters.

Our fiducial parameters for the numerical solution of radiative transfer are $N = 311$ frequency bins extending from $\nu_{\rm Ly\alpha}/2$ to $\nu_{\rm Ly\gamma}$, and a diffusion region with 80 bins extending to $\Delta\nu/\nu_{\rm Ly\alpha} = \pm 1.7 \times 10^{-2}$. The minimal spacing between bins is $\min[\ln(\nu_{b+1}/\nu_b)] = 8.5 \times 10^{-5}$, which sets the largest step in $\ln a$ that we can take (this is also half of the width $\Delta\nu/\nu$ of the "resonant" region around Lyα). We checked (for the fiducial cosmology only) that reducing the diffusion region to $\Delta\nu/\nu_{\rm Ly\alpha} = \pm 1 \times 10^{-2}$ leads to changes $|\Delta x_e|/x_e < 6 \times 10^{-6}$. Reducing the diffusion region to $\Delta\nu/\nu_{\rm Ly\alpha} = \pm 5 \times 10^{-3}$ leads to changes $|\Delta x_e|/x_e < 4 \times 10^{-5}$. We checked that

---

[2]Equation (39) contains near exact cancellations but Eq. (40) contains a large number of terms for which numerical round-off errors can add up. We checked that both equations give the same result within numerical round-off errors. We use Eq. (39) in the final code simply because it is more compact.

using a 10 times finer frequency grid in the diffusion region (and a 10 times smaller time step) leads to maximum changes $|\Delta x_e|/x_e \approx 1.5 \times 10^{-4}$ at $z \approx 900$.

We are therefore confident that our numerical treatment is converged at the level of a few parts in 10^4.

We show in Fig. 6 the changes in the recombination history due to two-photon processes. We find that including two-photon transitions from the initial states $2s$, $3s$, $3d$, $4s$ and $4d$ is sufficient for the level of accuracy required—we checked that including two-photon transitions from $5s$ and $5d$ leads to a maximum change $\Delta x_e/x_e \sim 8 \times 10^{-5}$ at $z \sim 1200$ and can therefore be neglected. The effect of frequency diffusion in the Lyα line is shown in Fig. 7.

We compared our results to the two-photon MLA code of Hirata [24], as well as to the results of Hirata & Forbes [20] for frequency diffusion. For this comparison, we use $n_{\max} = 30$. The result of the comparison is shown in Fig. 8. The maximum difference between the codes for $700 < z < 1600$ is $|\Delta x_e|/x_e = 0.0005$. The increase of the relative difference at late times is most likely due to small differences in the bound-free rates, which are computed with different methods (we use the recursion relations of Ref. [48] whereas Hirata (2008) directly integrates the products of wave functions to compute matrix elements). This difference remains even when switching off two-photon processes. The $\sim 3 \times 10^{-5}$ kink at $z \sim 1570$ is a startup transient due to switching from the post-Saha solution to solving the full ODE. The kink at $z = 1350$ is due to the Lyβ photons emitted at $z = 1600$ starting to redshift into Lyα. Overall, the agreement is excellent ($|\Delta x_e|/x_e < 10^{-4}$ for $z > 900$), even though the codes use different methods to compute atomic rates and different approaches for solving the MLA problem and treating Lyα frequency diffusion.

FIG. 7. Changes in the recombination history when including frequency diffusion in Lyman-α, compared to a model with two-photon transitions but no diffusion.

FIG. 8. Comparison of HYREC with the MLA + two-photon code of Hirata (2008) and to the results of Hirata & Forbes (2009) who account for Lyα diffusion. See text for comments.
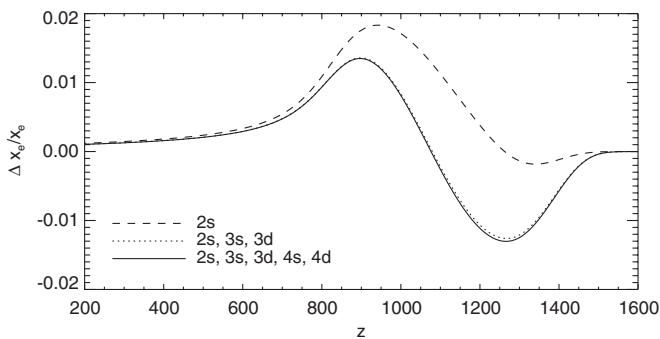

FIG. 6. Changes in the recombination history when including two-photon decays and Raman scattering (no diffusion), compared to our "base" model. The line labeled "$2s$" shows the changes in x_e when one properly accounts for stimulated $2s \rightarrow 1s$ decays, as well as absorptions of distortion photons and Raman scattering form $2s$. The other lines show the cumulative correction when adding two-photon transitions from higher levels.
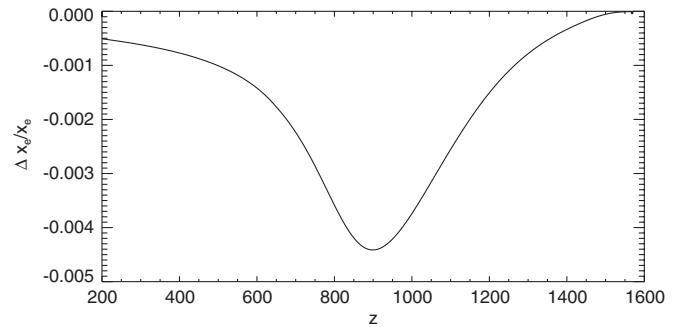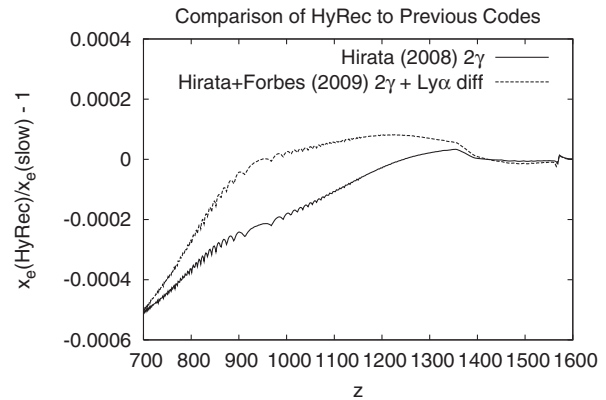
VI. HELIUM RECOMBINATION

Roughly 14% of the electrons in the Universe are associated with helium rather than hydrogen, so it is critical to model helium recombination as well [5,33–35]. Because the ionization energy of helium is greater than that of hydrogen ($E_{I_1} = 24.6$ eV for He I and $E_{I_2} = 54.4$ eV for He II), helium recombines earlier than hydrogen, well before the epoch of last scattering. Therefore, we do not observe helium recombination directly: rather it affects the diffusion of photons at early times and hence controls the Silk damping length [5]. A faster helium recombination (smaller x_e) leads to a longer photon mean free path and hence a larger damping length. The net effect is then to reduce the high-ℓ multipoles of the CMB temperature and polarization power spectra [33]. However, since most of the Silk damping occurs at later times, we do not need extraordinary accuracy in following helium recombination: the corrections identified by Refs. [31–33], which changed x_e by up to 3% at $z \approx 1800$, amounted to a $\sim 1\sigma$ correction for *Planck* and $\sim 8\sigma$ for a hypothetical cosmic variance limited experiment to $\ell = 3000$. Therefore, a helium recombination code accurate to

~0.3% should reduce any residual errors to the point of being negligible for *Planck*. Here we describe our "fast" helium recombination code.

Throughout the helium recombination process, we may take a single temperature $T \equiv T_m = T_r$. We make several other crude approximations, as detailed below. While analytically motivated, their quantitative justification rests on the comparison to the "full" version of our calculations [33]. For the equilibrium calculations, the He I level energies and are obtained from the NIST database (itself based on the compilation of Ref. [49]). The bound-bound Einstein coefficients are obtained from Ref. [50]: $A[2^1P^o - 1^1S] = 1.7989 \times 10^9$ s^{-1} and $A[2^3P^o - 1^1S] = 177.58$ s^{-1}. The bound-free rates are not required here since the excited levels of He I remain in equilibrium with the continuum (i.e., these rates may simply be taken to be "fast").

A. He III → II recombination

The He III → II recombination has previously been found to be in Saha equilibrium to high accuracy [31,34] and is followed using the Saha equation

$$\frac{q(1 + f_{He} + q)}{f_{He} - q} = s \equiv g_e e^{-E_{I_2}/T}, \qquad (95)$$

where $q \equiv x_{HeIII}$ and we assume that the rest of the helium is singly ionized, $x_{HeII} = f_{He} - q$, and all hydrogen is ionized, $x_e = 1 + f_{He} + q$. In Eq. (95), E_{I_2} is the second ionization energy of helium and g_e is given by Eq. (3) but using the appropriate reduced mass of the electron-He III system. We first solve Eq. (95) for q:

$$q = \frac{2sf_{He}}{1 + f_{He} + s}\left(1 + \sqrt{1 + \frac{4sf_{He}}{(1 + f_{He} + s)^2}}\right)^{-1}. \quad (96)$$

We then obtain the free electron fraction from $x_e = 1 + f_{He} + q$.

This equation is used to obtain x_e until $q = x_{HeIII} = 10^{-9}$ (which corresponds to $z \sim 4000$).

B. He II → I recombination: Near-equilibrium stage

In contrast to He III → II recombination, the He II → I recombination is a highly nonequilibrium process (it occurs at much lower density and in a weaker radiation field, with a much slower 2γ decay process, and the excited levels of He I are much closer to the continuum as a fraction of the ionization energy than those of He II). Therefore, it must be followed in several stages.

The first is the near-equilibrium stage, before the main He I resonance line ($2^1P^o - 1^1S$ at 584 Å) develops sufficient optical depth to push helium recombination out of equilibrium. In this stage, the free electron fraction is close to the Saha solution $x_e^{Saha} = 1 + q$, where $q \equiv x_{HeII}$ satisfies

$$\frac{q(1 + q)}{f_{He} - q} = s \equiv 4g_e e^{-E_{I_1}/T}, \qquad (97)$$

and we assume that all hydrogen is ionized and that the helium is distributed between He^0 and He$^+$. In Eq. (97), E_{I_1} is the first ionization energy of Helium, and one should use the appropriate reduced mass of the electron-He II system in g_e. The additional factor of 4 relative to Eq. (95) is due to the lower spin degeneracy of He I. Again, we first solve Eq. (97) for q:

$$q = \frac{2sf_{He}}{1 + s}\left(1 + \sqrt{1 + \frac{4sf_{He}}{(1 + s)^2}}\right)^{-1}, \qquad (98)$$

from which we get $x_e^{Saha} = 1 + q$. We then obtain the post-Saha expansion for the free electron fraction $x_e = x_e^{Saha} + \Delta x_e$ as described in Appendix D. We do so until Δx_e reaches 5×10^{-4}, which corresponds to $2500 < z < 3000$ depending on cosmology. We checked that using the post-Saha expansion until $\Delta x_e = 10^{-5}$ instead and then numerically integrating the ODE for x_e (described in the next section) leads to maximum changes of $\sim 3 \times 10^{-4}$ in the free electron fraction.

C. He II → I recombination: Nonequilibrium stage

At lower redshifts, one must follow helium recombination carefully, via a Peebles-style ODE [1], i.e., an equation of the form

$$\dot{x}_e = -\mathcal{F}(x_e, z; \mathbf{p}), \qquad (99)$$

where $\mathbf{p}$ is the vector of cosmological parameters. Despite the complicated radiative transfer physics in the helium problem, a single ODE turns out to suffice because the portion of the ultraviolet spectrum that is relevant is relatively narrow and may be treated as in steady state (it encompasses the 584 and 591 Å lines). The construction of the $\mathcal{F}$ function is however more complicated than the Peebles ODE for hydrogen.

The excited levels of He I ($n \geq 2$) have been found to remain in Saha equilibrium with the continuum throughout the process since the strong CMB blackbody ionizes them far faster than they can reach the ground state [31]. The significant processes for net decays to the ground state are fourfold:

(i) The two-photon process, He(2^1S) → He(1^1S) + $\gamma + \gamma$.

(ii) Emission of photons via the main resonance line $2^1P^o - 1^1S$ with $\lambda = 584$ Å, followed by redshifting out of the line (the He I analogue of the H I Lyman-α escape process).

(iii) The absorption of He I 584 Å photons by H(1s) atoms via photoionization, H(1s) + $\gamma \rightarrow$ H$^+ + e^-$. The electron rapidly thermalizes its energy, leading to a loss of a resonance line photon and a net decay of He I.

(iv) Emission of photons in the intercombination line He I] $2^3P^o - 1^1S$ with $\lambda = 591$ Å. This line has Sobolev optical depth of order unity during He I recombination so the full Sobolev escape probability formula must be used.

In constructing the function $\mathcal{F}$ above, we take several steps. First, we compute the abundance of the species H I, H II, He I and He II. The formulas for the latter are $x_{\text{HeII}} = x_e - 1 + x_{\text{HI}}$ and $x_{\text{HeI}} = f_{\text{He}} - x_{\text{HeII}}$, but these require knowledge of the H I fraction. This cannot be completely ignored, but it is small and so the Saha equation for H (Eq. (D2)) suffices for this purpose.[3]

Next is the determination of the excited level (2^1S and 2^1P) abundances assuming equilibrium with the continuum,

$$x_{[2^1S]} = \frac{1}{4g_e} e^{-E_{[2^1S]}/T} x_e (x_e - 1) \qquad (100)$$

and $x_{[2^1P]} = 3x_{[2^1S]} \exp[-(E_{[2^1P]} - E_{[2^1S]})/T]$.

Armed with this information, we proceed to investigate each decay mechanism.

1. The two-photon process

The downward rate for the He I 2γ decay is $\Lambda = 50.94$ s^{-1} [51], so we find a downward decay rate

$$\mathcal{F}^{(2\gamma)}(x_e, z; \mathbf{p}) = \Lambda(x_{[2^1S]} - e^{-(E_{[2^1S]} - E_{[1^1S]})/T} x_{\text{HeI}}). \quad (101)$$

2. The 584 Å line

The rate of emission of photons in the 584 Å line is complicated because photons in this line experience multiple processes: (i) "true" absorption and emission, (ii) resonant scattering (which redistributes photons in frequency since in the comoving frame the ingoing and outgoing photons need not have the same frequency), (iii) H I photoionization opacity (which has absorption and emission), and (iv) redshifting due to Hubble expansion. We construct here a highly simplified model of these processes; the "full" treatment can be found in Ref. [31]. Ref. [31] also showed that the resonant scatterings can be neglected. It should be noted that in the case of the 584 Å line, the nontrivial physics takes place in the damping wings: the line center is very optically thick, and the reaction

$$\text{He}(2^1P^o) \leftrightarrow \text{He}(1^1S) + \gamma \qquad (102)$$
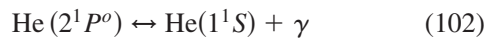
reaches equilibrium there (analogous to the H I Lyman-α line).

---

[3]Equation (D2) technically neglects the contribution of helium to the electron abundance, but this correction is small in the very latest stages of He recombination when the H I correction becomes significant.

In this situation, the radiative transfer equation for the photon occupation number f_ν in the vicinity of the 584 Å line can be simplified to [Eq. (36) of Ref. [31], with individual expressions substituted in]

$$\dot{f}_\nu = H\nu \frac{\partial f_\nu}{\partial \nu} + H\nu \eta_c (e^{-h\nu/T} - f_\nu)$$
$$+ H\nu \tau_{\text{abs}} \phi(\nu)(f_{584}^{\text{eq}} - f_\nu), \qquad (103)$$

where η_c is the H I opacity in units of absorption optical depth per unit frequency (i.e., $H\nu\eta_c$ is the absorption optical depth per unit time), and $f_{584}^{\text{eq}} \equiv x_{[2^1P^o]}/(3x_{[1^1S]})$ is the equilibrium photon occupation number in the 584 Å line. In Eq. (103), the $e^{((-h\nu)/T)}$ term corresponds to emission from direct H I recombinations to the ground state (i.e., the inverse process of photoionization); τ_{abs} is the Sobolev optical depth to true absorption in the 584 Å line; $\phi(\nu)$ is the 584 Å line profile normalized by $\int \phi(\nu)d\nu = 1$. We neglect stimulated emission processes at 584 Å, since the photon phase space density is $\ll 1$. In steady state, the left-hand side of Eq. (103) is approximated as zero. The remaining contributions are as follows.

The H I continuum optical depth is, assuming an H I abundance in Saha equilibrium with H II (assumed to be nearly all hydrogen, i.e., $x_p \approx 1$),

$$\eta_c = \frac{e^{E_I/T}}{g_e} \frac{\sigma_{\text{pi}} c n_H x_e}{H\nu}. \qquad (104)$$

[This comes from combining Eqs. (27) and (28) of Ref. [31].] Here σ_{pi} is the photoionization cross section of H I at $\lambda = 584$ Å and E_I is the hydrogen ionization energy.

The line profile is in principle a Voigt profile. However near line center where $\phi(\nu)$ becomes large, we have $f_\nu \rightarrow f_{584}^{\text{eq}}$ irrespective of the details in order to keep Eq. (103) finite. Therefore, we approximate it by the damping wing approximation,

$$\phi(\nu) \approx \frac{\Gamma}{4\pi^2(\nu - \nu_0)^2}, \qquad (105)$$

where ν_0 is the central frequency of the line and Γ is the intrinsic width. The latter is the sum of the rates for all processes that depopulate 2^1P^o (the width of the 1^1S state is negligible), i.e., we may write $\Gamma = A_{584} + \Gamma_{\text{other}}$, where A_{584} is the contribution from $2^1P^o \rightarrow 1^1S$ and Γ_{other} is the contribution from all other states:

$$\Gamma_{\text{other}} = \frac{A_{[2^1P^o],[2^1S]}}{1 - e^{-(E_{[2^1P^o]} - E_{[2^1S]})/T}} + \sum_i \frac{g_i}{3} \frac{A_{i,[2^1P^o]}}{e^{(E_i - E_{[2^1P^o]})/T} - 1}, \qquad (106)$$

where the first term corresponds to decays to 2^11S (supplemented by stimulated transitions) and the second term corresponds to absorptions from 2^1P^o to a higher level i. We include the levels n^1S and n^1D for $3 \leq n \leq 5$. In principle we should include higher n and the continuum

levels, but in practice the first few levels dominate the sum because of the exponential factor.

We next need the optical depth to true absorption, which is the Sobolev optical depth τ_S times the fraction of photon absorptions $1^1S \rightarrow 2^1P^o$ that are true absorptions (i.e., do not immediately decay back to 1^1S, but rather visit another level). This fraction is $\Gamma_{\text{other}}/\Gamma$. Thus

$$\tau_{\text{abs}}\phi(\nu) = \frac{\Gamma_{\text{other}}}{\Gamma}\tau_S\frac{\Gamma}{4\pi^2(\nu - \nu_0)^2} = \frac{\tau_S\Gamma_{\text{other}}}{4\pi^2(\nu - \nu_0)^2}. \tag{107}$$

With these results, and neglecting the variation of the blackbody function $e^{((-h\nu)/T)}$ across the line, Eq. (103) simplifies to

$$\frac{\partial f_\nu}{\partial\nu} = \eta_c(f_\nu - e^{-h\nu_0/T}) + \frac{\tau_S\Gamma_{\text{other}}}{4\pi^2(\nu - \nu_0)^2}(f_\nu - f_{584}^{\text{eq}}). \tag{108}$$

The next simplification of this equation occurs if we rescale both the frequency and the phase space density axes as

$$y = 4\pi^2\frac{\nu - \nu_0}{\tau_S\Gamma_{\text{other}}} \tag{109}$$

and

$$\xi(y) = \frac{f_\nu - e^{-h\nu_0/T}}{f_{584}^{\text{eq}} - e^{-h\nu_0/T}}, \tag{110}$$

leading to

$$\frac{d\xi}{dy} = \tau_c\xi + \frac{\xi - 1}{y^2}, \tag{111}$$

where

$$\tau_c \equiv \frac{\tau_S\Gamma_{\text{other}}\eta_c}{4\pi^2}. \tag{112}$$

For blackbody radiation entering the line, we have the boundary condition $\xi(+\infty) = 0$.

This reduces the radiative transfer equation to a single ODE that depends on the single dimensionless parameter τ_c, which is (roughly speaking) the optical depth to H I photoionization within the part of the line that is optically thick to true absorption by He I 584 Å. This parameter is exponentially increasing in the early parts of helium recombination, and becomes of order unity at $z \approx 2100$. It is the parameter that controls which process is a more important sink for resonance line photons: H I continuum opacity ($\tau_c \gg 1$) or escape via redshifting ($\tau_c \ll 1$).

The net rate at which photons are emitted in the He I line in photons per H nucleus per unit time is then obtained by integrating the absorption/emission term in the rate equation,

$$\mathcal{F}^{(584)} = \int\frac{8\pi\nu^2}{n_H c^3}H\nu\tau_{\text{abs}}\phi(\nu)(f_{584}^{\text{eq}} - f_\nu)d\nu. \tag{113}$$

This integral can be converted into an integral over y and f_ν can be replaced with $\xi(y)$; making these substitutions and approximating $\nu \approx \nu_0$ in the prefactors gives

$$\mathcal{F}^{(584)} = \frac{8\pi H\nu_0^3}{n_H c^3}(f_{584}^{\text{eq}} - e^{-h\nu_0/T})\mathcal{E}, \tag{114}$$

where

$$\mathcal{E} = \int\frac{1 - \xi(y)}{y^2}dy \tag{115}$$

depends on the single parameter τ_c. We note that without the factor of $\mathcal{E}$, Eq. (114) would be the standard decay rate formula with escape probability $P_{\text{esc}} = 1/\tau_S$ (appropriate in the high optical depth limit). Thus $\mathcal{E}$ can be thought of as a correction factor associated with the H I continuum opacity.

Unfortunately, the integral in Eq. (115) is not amenable to direct calculation, since the integrand is ill-conditioned near $y \approx 0$. We may obtain an alternative form by taking the integral from $y = -Y$ to $y = +Y$ (we will take the limit $Y \rightarrow \infty$), plugging in Eq. (111) into the integrand, and then realizing that $\xi(+Y) \rightarrow 0$ due to our boundary condition, we find

$$\mathcal{E} = \xi(-Y) + \tau_c\int_{-Y}^{Y}\xi(y)dy. \tag{116}$$

Formally, for $\tau_c > 0$, we have $\xi(-Y) \rightarrow 0$ for $Y \rightarrow \infty$, but numerically one must keep the first term at small τ_c. Equation (111) is stiff and can be solved by the backward Euler method; however even this is too slow to use in a "fast" recombination code. Therefore, we have constructed a fitting function for $\mathcal{E}$, valid to within 0.8% for all positive τ_c:

$$\mathcal{E}(\tau_c) \approx \sqrt{1 + \pi^2\tau_c} + \frac{7.74\tau_c}{1 + 70\tau_c}. \tag{117}$$

Note that this has the correct limiting behavior $\mathcal{E} \rightarrow 1$ for $\tau_c \rightarrow 0$, and for positive τ_c has $\mathcal{E} > 1$, as one would expect. The full function $\mathcal{E}(\tau_c)$ and the approximation of Eq. (117) are shown in Fig. 9.

The use of Eqs. (114) and (117) are sufficient to solve for the behavior of the 584 Å line.

3. The 591 Å line

The intercombination line He I] $2^3P_1^o \rightarrow 1^1S$ at 591 Å can reach optical depths of order unity during helium recombination. Therefore it must be considered carefully. However, because the damping wings are optically thin, it is only the line center that is of interest. Because there is negligible H I continuum opacity in the core during helium recombination, we use the Sobolev approximation for the 591 Å line. If there were no 584 Å line, then we would have
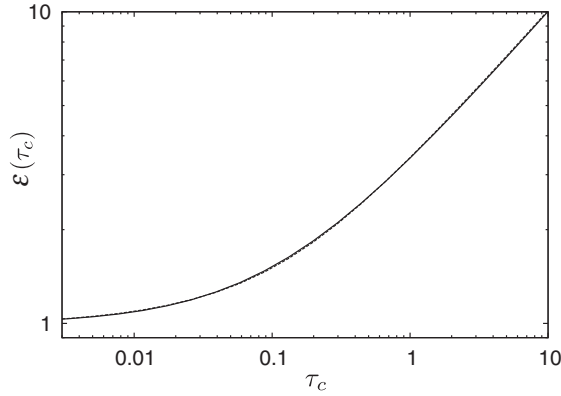
FIG. 9. The function $\mathcal{E}(\tau_c)$ (solid line) and our approximation (dashed line).

FIG. 10. A comparison of the HYREC He II → I recombination history to that computed by the "full physics" code of Switzer & Hirata [33]. The maximum deviation is 0.3%.

blackbody radiation entering the line and could write the net decay rate as

$$\frac{8\pi H\nu_0^3}{n_H c^3}(1 - e^{-\tau_{591}})(f_{591}^{eq} - e^{-(E_{[2^3P^o]}-E_{[1^1S]})/T}), \quad (118)$$

where τ_{591} is the optical depth of the 591 Å line, and $f_{591}^{eq} \equiv x_{[2^3P^o]}/(3x_{[1^1S]})$ is the equilibrium value of the photon occupation number in the line. Based on the ratio of Einstein coefficients [50], we take $\tau_{591} = 1.023 \times 10^{-7}\tau_{584}$.

However, the 591 Å line can also absorb photons that redshifted out of the 584 Å line. We treat this problem by an "on-the-spot" approximation: we assume that for each photon emitted in the 584 Å line, there is a probability that the photon is reabsorbed of

$$P_{reabs} = e^{-\eta_c(E_{[2^1P^o]}-E_{[2^3P^o]})/h}(1 - e^{-\tau_{591}}), \quad (119)$$

where the first factor is the probability that the photon can redshift from 584 Å to 591 Å without being destroyed by H I continuum opacity, and the second factor is the probability that, once it reaches the 591 Å line, the photon is indeed absorbed. Thus the 591 Å line has two effects: first, it contributes to the net formation of the He I ground state in accordance with Eq. (118), and second it reduces the net formation from the 584 Å line by a factor of $1 - P_{reabs}$. Thus we write

$$\mathcal{F}^{(591)} = \frac{8\pi H\nu_0^3}{n_H c^3}(1 - e^{-\tau_{591}})$$
$$\times (f_{591}^{eq} - e^{-(E_{[2^3P^o]}-E_{[1^1S]})/T}) - P_{reabs}\mathcal{F}^{(584)}.$$
$$(120)$$

4. Hydrogen recombination corrections

As a final step, we include in our rate equation the contribution to the ionization fraction due to changes in the hydrogen Saha equilibrium:
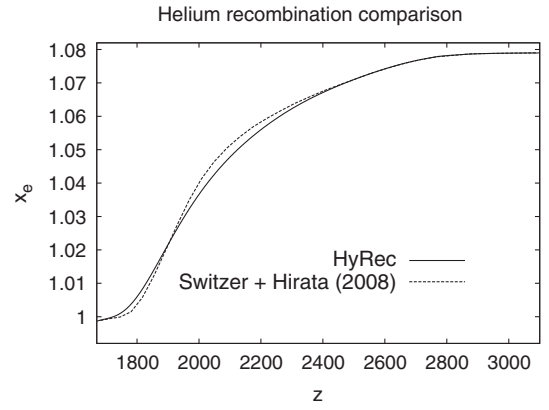
$$\mathcal{F}^{(H)} = -\frac{dx_{HI}}{dt}\bigg|_{Saha}, \quad (121)$$

which is obtained by two-sided finite differencing with $\Delta z = \pm 0.5$. This contribution to $\dot{x}_e$ is necessary for full accuracy at the very last stages of He recombination.

5. Integration details

We may now construct the overall electron fraction evolution, $\mathcal{F} = \mathcal{F}^{(2\gamma)} + \mathcal{F}^{(584)} + \mathcal{F}^{(591)} + \mathcal{F}^{(H)}$. The equation is stiff at early times, which is why we do not turn it on until the free electron fraction given by the post-Saha expansion differs from the Saha equilibrium value by more than 5×10^{-4}. We then use the ODE integrator described in Appendix C.

The treatment of helium is turned off once the abundance of He$^+$ ions is less than 10^{-4} per hydrogen atom, after which we switch to following hydrogen recombination (we checked that hydrogen and helium recombination never overlap and can be followed separately).

D. Comparison to detailed calculations

The ultimate test of the sweeping approximations made in this section is their comparison against more detailed computations of the He I recombination history. We compare against Switzer & Hirata [33] in Fig. 10. The maximum error is 0.3%, which is roughly equal to the stated theoretical uncertainty in the Switzer & Hirata calculation.

VII. CONCLUSIONS

We have presented a complete treatment of primordial hydrogen and helium recombination, including all the effects that have been shown to be important so far. Our computation accounts for the multilevel character of hydrogen and the nonequilibrium of angular momentum

substates, radiative feedbacks, two-photon transitions, and frequency diffusion in Lyα for hydrogen recombination. For helium recombination, we account for HI continuum opacity in the He I $2^1P^o - 1^1S$ line, decays in the $2^3P^o - 1^1S$ intercombination line, and feedback between these lines. We have implemented all these effects in a single recombination code, HYREC, which can compute a recombination history in ~ 2 seconds on a standard laptop for a given set of cosmological parameters. Provided collisional transitions can be neglected (which remains to be established), we estimate the errors of our computation to be a few times 10^{-3} during helium recombination and a few times 10^{-4} during hydrogen recombination, including both numerical errors and errors due to the assumptions and approximations made for physical effects. If collisional transitions are shown to have a significant effect on recombination, our code can be easily updated to account for them with very little loss of computational efficiency.

It has been argued that corrections to the recombination history due to radiative transfer effects are relatively independent of cosmology [52], and that one could therefore compute them once and use the resulting correction function to account for them for any given cosmology. Alternatively, one might run a grid of recombination histories for different cosmologies and construct a fitting function [6,53]. Our point of view here is that the physics of primordial recombination is simple enough, and an exact calculation from first principles is now fast enough that there should be no reason to use fudge factors and approximate correction functions. This is especially relevant if one wishes to extend the standard recombination calculation by introducing "exotic" new physics. We would like to emphasize that the fast computation presented here, using the EMLA method, is very well adapted for the computation of the recombination history, but that the standard MLA approach and fast interpolation methods may still be useful for the computation of the recombination spectrum.

We believe our code is accurate enough (aside from neglecting collisional transitions) and has a sufficiently small runtime to be incorporated in Monte Carlo Markov chains for upcoming CMB data analysis from the *Planck* mission.

ACKNOWLEDGMENTS

APPENDIX A: PROOF OF SOME RELATIONS INVOLVING EFFECTIVE RATES

1. Preliminaries

Here we use the same notation as in Paper I. Capital indices K, L refer to "interior" excited states, and lowercase indices i, i refer to "interface" excited states. We define the rate matrix $\mathbf{M}$ with coefficients:

$$M_{KL} = \delta_{KL}\Gamma_K - (1 - \delta_{KL})R_{K,L}. \qquad (A1)$$

In Paper I, we have shown that the populations of the interface states are given by

$$X_K = \sum_L (\mathbf{M}^{-1})_{LK}\left[n_H x_e^2 \alpha_L + \sum_i x_i R_{i,L} \right]. \qquad (A2)$$

We also showed that the probabilities P_K^i and P_K^e are given by

$$P_K^i = \sum_L (\mathbf{M}^{-1})_{KL} R_{L,i}, \qquad (A3)$$

and

$$P_K^e = \sum_L (\mathbf{M}^{-1})_{KL}\beta_L. \qquad (A4)$$

In Appendix C of Paper I, we showed that $\mathbf{M}$ satisfies the following detailed balance relation:

$$Q_K(\mathbf{M}^{-1})_{KL} = Q_L(\mathbf{M}^{-1})_{LK}, \qquad (A5)$$

where $Q_K = g_K e^{((-E_K)/T_r)}$ is the contribution of individual states to the partition function and g_K is the degeneracy of the state K.

2. Rate of change of the free electron fraction

In this section we derive Eq. (31), which was not derived in Paper I. In the standard MLA formulation, the rate of change of the free electron fraction can be written as

$$\dot{x}_e = -\sum_K [n_H x_e^2 \alpha_K - X_K \beta_K] - \sum_i [n_H x_e^2 \alpha_i - x_i \beta_i]. \qquad (A6)$$

This formula is never used in standard MLA codes, as it requires a summation over a large number of nearly cancelling terms, and MLA codes use instead $\dot{x}_e = -\dot{x}_{1s}$ to compute the rate of change of the free electron fraction. Equation (A6) remains however formally correct. Using Eqs. (A2) and (A4), we rewrite

$$\sum_K X_K \beta_K = \sum_{K,L} \beta_K (\mathbf{M}^{-1})_{LK} \left[n_H x_e^2 \alpha_L + \sum_i x_i R_{i,L} \right]$$

$$= \sum_L P_L^e \left[n_H x_e^2 \alpha_L + \sum_i x_i R_{i,L} \right]$$

$$= \sum_L n_H x_e^2 \alpha_L - n_H x_e^2 \sum_i \sum_L \alpha_L P_L^i$$

$$+ \sum_i x_i \sum_L R_{i,L} P_L^e, \qquad (A7)$$

where in the last equality we have used the complementarity relation $\sum_i P_K^i + P_K^e = 1$. Inserting this result into Eq. (A6), and using the definitions of the effective recombination coefficients and photoionization rates Eqs. (23) and (24), we immediately recover Eq. (31).

3. Proof of Eq. (41)

Consider Eq. (36) with $\tilde{\Gamma}_K \approx \Gamma_K(T_r)$. The formal solution for the $\tilde{P}_K^{1s}$ is given by

$$\tilde{P}_K^{1s} = \sum_L (\mathbf{M}^{-1})_{KL} \tilde{R}_{L,1s}. \qquad (A8)$$

Therefore one may rewrite Eq. (34), for $i = 2s, 2p$:

$$\tilde{\mathcal{R}}_{i,1s} = \tilde{R}_{i,1s} + \sum_K \lambda_{i,K}(T_r) \tilde{R}_{K,1s}, \qquad (A9)$$

where we have defined

$$\lambda_{i,K}(T_r) \equiv \sum_L R_{i,L}(\mathbf{M}^{-1})_{LK} = \sum_L R_{i,L} \frac{Q_K}{Q_L}(\mathbf{M}^{-1})_{KL}$$

$$= \sum_L R_{L,i} \frac{Q_K}{Q_i}(\mathbf{M}^{-1})_{KL} = \frac{g_K}{g_i} e^{((-E_{Ki})/T_r)} P_K^i(T_r), \qquad (A10)$$

where in the second line we used Eq. (A5), in the third line we used the detailed balance relation verified by $R_{i,L}$ and $R_{L,i}$, and in the last line we used the formal solution for P_K^i, Eq. (A3). We therefore obtain Eq. (41).

4. Expression of X_K in terms of x_i, x_e and effective rates

Taking $T_m = T_r$ and using the detailed balance relation $g_e n_H \alpha_L = Q_L \beta_L$, we rewrite Eq. (A2) as follows:

$$X_K = g_e^{-1} x_e^2 \sum_L Q_L (\mathbf{M}^{-1})_{LK} \beta_L + \sum_i x_i \sum_L (\mathbf{M}^{-1})_{LK} \frac{Q_L}{Q_i} R_{L,i}$$

$$= g_e^{-1} x_e^2 Q_K \sum_L (\mathbf{M}^{-1})_{KL} \beta_L + \sum_i x_i \frac{Q_K}{Q_i} \sum_L (\mathbf{M}^{-1})_{KL} R_{L,i}, \qquad (A11)$$

where in the last equality we have used Eq. (A5). Using the formal solutions for the probabilities Eqs. (A3) and (A4), we see that we recover Eq. (42).

APPENDIX B: EXTRAPOLATION OF THE EFFECTIVE RATES TO $n_{max} = \infty$

We have tabulated the effective rates $\mathcal{A}_{2s}(T_m, T_r)$, $\mathcal{A}_{2p}(T_m, T_r)$ and $\mathcal{R}_{2s,2p}(T_r)$, including all excited states up to the principal quantum number n_{max}, for several values of n_{max} up to 600, over the temperature range $0.004 \text{ eV} \leq T_r \leq 0.4 \text{ eV}, \ 0.1 \leq T_m/T_r \leq 1$. This range of temperatures corresponds to $20 < z < 1650$ for a wide range of cosmologies. For every pair (T_m, T_r), we have fitted the effective rates by the following functional form:

$$\mathcal{A}_i(T_m, T_r; n_{max}) = \mathcal{A}_i(T_m, T_r; \infty)\left(1 - \frac{\kappa}{(n_{max})^\gamma}\right), \quad (B1)$$

and similarly for $\mathcal{R}_{2s,2p}$, where κ and γ depend on T_m and T_r as well as on the coefficient being fitted. This allows us to extrapolate the effective rates to $n_{max} \rightarrow \infty$. Of course, this is only a formal extrapolation, as for n larger than a few thousands, the excited states of hydrogen are no more well defined (see Ref. [54] for a discussion). The extrapolated rates are still more accurate than those computed with a finite number of states. The residuals of the fit have a maximum relative amplitude of 5×10^{-4} over the whole range of temperature considered, for $200 \leq n_{max} \leq 600$, and more than an order of magnitude smaller on the restricted range $T_r \gtrsim 0.04 \text{ eV}$, $T_m/T_r \gtrsim 0.8$ which corresponds to $z \gtrsim 200$ (note that neglecting the overlap of the high-lying Lyman lines leads to errors in the effective rates of similar amplitude [15]). For reference, the maximum relative difference between the effective rates computed with $n_{max} = 600$ and their extrapolation at $n_{max} = \infty$ is 0.05 over the whole range of temperature considered, and 0.002 over the restricted range corresponding to $z \gtrsim 200$. We checked that our method recovers the correct case-B recombination coefficient $\alpha_B(T_m) \equiv \sum_{i=2s,2p} \mathcal{A}_i(T_m, T_r = 0; \infty)$. Our extrapolated α_B agrees with the fit of Ref. [55] to better that 0.2% for $T_m > 40 \text{ K}$, which is the accuracy claimed by the authors of Ref. [55].

APPENDIX C: NUMERICAL ODE INTEGRATOR

For the sake of computational efficiency, we use a second order ODE integrator that uses derivatives computed at previous time steps. This allows us to evaluate derivatives only once at each time step. Explicitly, to numerically solve the equation $y'(x) = f(x, y)$, we use evenly spaced steps Δx, and obtain the solution at the $(i + 1)$th step as follows:

$$y_{i+1} = y_i + \Delta y_i, \qquad \Delta y_i = \Delta x [1.25 y_i' - 0.25 y_{i-2}'], \qquad (C1)$$

where $y_i' = f(x_i, y_i)$ is stored at each time step for later use. For the case of interest, we have $x = \ln a$, $y = x_e$ and $f = \dot{x}_e/H$.

APPENDIX D: POST-SAHA EXPANSION AT EARLY PHASES OF HYDROGEN AND HE II → I RECOMBINATIONS

As explained in Appendix D of Paper I, the ODE describing hydrogen recombination is stiff at $z \gtrsim 1500$ and so is the ODE describing He II → I recombination at $z \gtrsim 2800$. We therefore use an expansion around the Saha equilibrium solution:

$$x_e \approx x_e^S + \frac{d(x_e^S)}{dt} \Big/ \frac{\partial \dot{x}_e}{\partial x_e} \Big|_{x_e^S}, \tag{D1}$$

where x_e^S is the Saha equilibrium value of the free electron fraction.

1. Hydrogen recombination

The Saha equilibrium value of the free electron fraction is the solution of the following equation:

$$\frac{(x_e^S)^2}{1 - x_e^S} = s \equiv g_e e^{-E_I/T}, \tag{D2}$$

where g_e was given in Eq. (3) and $T = T_m = T_r$ at early times. The numerator in Eq. (D1) can be obtained analytically by differentiating Eq. (D2):

$$\frac{d(x_e^S)}{dt} = -\frac{H(\frac{E_I}{T} - \frac{3}{2})(x_e^S)^2}{2x_e^S + s}. \tag{D3}$$

For the denominator in Eq. (D1), we numerically differentiate the derivative $\dot{x}_e$ obtained when accounting for two-photon processes and diffusion, using a two-sided finite difference with $\Delta x_e = \pm 0.01(1 - x_e^S)$.

2. He II → I recombination

In that case the free electron fraction in Saha equilibrium is given by $x_e^S = 1 + q$, where q can be obtained from Eq. (97). As in the hydrogen case, differentiation of Eq. (97) gives us an analytic expression for the numerator in Eq. (D1). We numerically differentiate the derivative $\dot{x}_e$ given by Eq. (99) using a two-sided finite difference with $\Delta x_e = \pm 0.01(1 + f_{He} - x_e^S)$.

[1] P. J. E. Peebles, Astrophys. J. **153**, 1 1968).
[2] Y. B. Zeldovich, V. G. Kurt, and R. A. Syunyaev, Zh. Eksp. Teor. Fiz. **55**, 278 (1968); , Sov. Phys. JETP **28**, 146 (1969).
[3] E. Komatsu *et al.*, Astrophys. J. Suppl. Ser. **192**, 18 (2011).
[4] The Planck Collaboration, arXiv:astro-ph/0604069.
[5] W. Hu, D. Scott, N. Sugiyama, and M. White, Phys. Rev. D **52**, 5498 (1995).
[6] A. Lewis, J. Weller, and R. Battye, Mon. Not. R. Astron. Soc. **373**, 561 (2006).
[7] J. A. Rubiño-Martín, J. Chluba, W. A. Fendt, and B. D. Wandelt, Mon. Not. R. Astron. Soc. **403**, 439 (2010).
[8] S. Seager, D. D. Sasselov, and D. Scott, Astrophys. J. Suppl. Ser. **128**, 407 (2000).
[9] S. Seager, D. D. Sasselov, and D. Scott, Astrophys. J. Lett. **523**, L1 (1999).
[10] J. A. Rubiño-Martín, J. Chluba, and R. A. Sunyaev, Mon. Not. R. Astron. Soc. **371**, 1939 (2006).
[11] J. Chluba, J. A. Rubiño-Martín, and R. A. Sunyaev, Mon. Not. R. Astron. Soc. **374**, 1310 (2007).
[12] D. Grin and C. M. Hirata, Phys. Rev. D **81**, 083005 (2010).
[13] J. Chluba, G. M. Vasil, and L. J. Dursi, Mon. Not. R. Astron. Soc. **407**, 599 (2010).
[14] Y. Ali-Haïmoud and C. M. Hirata, Phys. Rev. D **82**, 063521 (2010).
[15] Y. Ali-Haïmoud, D. Grin, and C. M. Hirata, Phys. Rev. D **82**, 123502 (2010).
[16] J. Chluba and R. A. Sunyaev, Astron. Astrophys. **475**, 109 (2007).

[17] E. E. Kholupenko, A. V. Ivanchik, and D. A. Varshalovich, Phys. Rev. D **81**, 083004 (2010).
[18] J. Chluba and R. A. Sunyaev, Astron. Astrophys. **496**, 619 (2009).
[19] S. I. Grachev and V. K. Dubrovich, Astron. Lett. **34**, 439 (2008).
[20] C. M. Hirata and J. Forbes, Phys. Rev. D **80**, 023001 (2009).
[21] J. Chluba and R. A. Sunyaev, Astron. Astrophys. **503**, 345 (2009).
[22] J. Chluba and R. A. Sunyaev, Astron. Astrophys. **446**, 39 (2006).
[23] E. E. Kholupenko and A. V. Ivanchik, Astron. Lett. **32**, 795 (2006).
[24] C. M. Hirata, Phys. Rev. D **78**, 023001 (2008).
[25] V. K. Dubrovich and S. I. Grachev, ''Dynamics of the Primordial Hydrogen and Helium (HeI) Recombination in the Universe'' (unpublished).
[26] J. Chluba and R. A. Sunyaev, Astron. Astrophys. **480**, 629 (2008).
[27] J. Chluba and R. A. Sunyaev, Astron. Astrophys. **512**, A53 (2010).
[28] T. Matsuda, H. Satō, and H. Takeda, Prog. Theor. Phys. **42**, 219 (1969).
[29] T. Matsuda, H. Sato, and H. Takeda, Prog. Theor. Phys. **46**, 416 (1971).
[30] E. E. Kholupenko, A. V. Ivanchik, and D. A. Varshalovich, Mon. Not. R. Astron. Soc. **378**, L39 (2007).
[31] E. R. Switzer and C. M. Hirata, Phys. Rev. D **77**, 083006 (2008).
[32] C. M. Hirata and E. R. Switzer, Phys. Rev. D **77**, 083007 (2008).

[33] E. R. Switzer and C. M. Hirata, Phys. Rev. D **77**, 083008 (2008).

[34] J. A. Rubiño-Martín, J. Chluba, and R. A. Sunyaev, Astron. Astrophys. **485**, 377 (2008).

[35] E. E. Kholupenko, A. V. Ivanchik, and D. A. Varshalovich, Astron. Lett. **34**, 725 (2008).

[36] E. R. Switzer and C. M. Hirata, Phys. Rev. D **72**, 083002 (2005).

[37] J. Chluba, G. M. Vasil, and L. J. Dursi, Mon. Not. R. Astron. Soc. **407**, 599 (2010).

[38] J. Chluba and R. M. Thomas, arXiv:1010.3631.

[39] D. E. Osterbrock and G. J. Ferland, *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei*, edited by D. E. Osterbrock and G. J. Ferland (University Science Books, Sausalito, CA, 2006).

[40] J. Chluba and R. A. Sunyaev, Astron. Astrophys. **458**, L29 (2006).

[41] W. Y. Wong and D. Scott, Mon. Not. R. Astron. Soc. **375**, 1441 (2007).

[42] G. B. Rybicki, Astrophys. J. **647**, 709 (2006).

[43] G. B. Rybicki and I. P. dell'Antonio, Astrophys. J. **427**, 603 (1994).

[44] S. I. Grachev, Astrophysics (Engl. Transl.) **30**, 211 (1989).

[45] J. H. Krolik, Astrophys. J. **338**, 594 (1989).

[46] J. H. Krolik, Astrophys. J. **353**, 21 (1990).

[47] S. I. Grachev and V. K. Dubrovich, Astrophysics (Engl. Transl.) **34**, 124 (1991).

[48] A. Burgess, Mem. R. Astron. Soc. **69**, 1 (1965).

[49] D. C. Morton, Q. X. Wu, and G. W. F. Drake, Can. J. Phys. **84**, 83 (2006).

[50] G. W. F. Drake and D. C. Morton, Astrophys. J. Suppl. Ser. **170**, 251 (2007).

[51] G. W. F. Drake, Phys. Rev. A **34**, 2871 (1986).

[52] J. A. Rubiño-Martín, J. Chluba, W. A. Fendt, and B. D. Wandelt, Mon. Not. R. Astron. Soc. **403**, 439 (2010).

[53] W. A. Fendt, J. Chluba, J. A. Rubiño-Martín, and B. D. Wandelt, Astrophys. J. Suppl. Ser. **181**, 627 (2009).

[54] D. Grin, Ph.D., thesis, California Institute of Technology, 2010).

[55] D. Pequignot, P. Petitjean, and C. Boisson, Astron. Astrophys. **251**, 680 (1991).