Ultrafast effective multilevel atom method for primordial hydrogen recombination

Yacine Ali-Haïmoud and Christopher M. Hirata

California Institute of Technology, Mail Code 350-17, Pasadena, California 91125, USA
(Received 11 June 2010; published 17 September 2010)

Cosmological hydrogen recombination has recently been the subject of renewed attention because of its importance for predicting the power spectrum of cosmic microwave background anisotropies. It has become clear that it is necessary to account for a large number $n \gtrsim 100$ of energy shells of the hydrogen atom, separately following the angular momentum substates in order to obtain sufficiently accurate recombination histories. However, the multilevel atom codes that follow the populations of all these levels are computationally expensive, limiting recent analyses to only a few points in parameter space. In this paper, we present a new method for solving the multilevel atom recombination problem, which splits the problem into a computationally expensive atomic physics component that is independent of the cosmology and an ultrafast cosmological evolution component. The atomic physics component follows the network of bound-bound and bound-free transitions among excited states and computes the resulting effective transition rates for the small set of "interface" states radiatively connected to the ground state. The cosmological evolution component only follows the populations of the interface states. By pretabulating the effective rates, we can reduce the recurring cost of multilevel atom calculations by more than 5 orders of magnitude. The resulting code is fast enough for inclusion in Markov chain Monte Carlo parameter estimation algorithms. It does not yet include the radiative transfer or high-n two-photon processes considered in some recent papers. Further work on analytic treatments for these effects will be required in order to produce a recombination code usable for *Planck* data analysis.

I. INTRODUCTION

The advent of high precision cosmic microwave background (CMB) experiments, such as *Planck* [1], has recently motivated several authors to revisit the theory of cosmological recombination pioneered by Peebles [2] and Zeldovich *et al.* [3] in the 1960s. The free electron fraction as a function of redshift $x_e(z)$ is one of the major theoretical uncertainties in the prediction of the CMB temperature and polarization anisotropy power spectra [4–6]. To obtain a recombination history accurate to the percent level, it is necessary to account for a high number of excited states of hydrogen, up to a principal quantum number $n_{\max} = \mathcal{O}(100)$ [7,8]. The desired subpercent accuracy can only be reached when explicitly resolving the out-of-equilibrium angular momentum substates, which requires the multilevel atom (MLA) codes to follow $N_{\text{level}} = n_{\max}(n_{\max} + 1)/2$ individual states. Moreover, the ordinary differential equations (ODEs) describing the level populations are stiff, requiring the solution of large $N_{\text{level}} \times N_{\text{level}}$ systems of equations at each integration time step. This problem has been solved by several authors [9–11], but each of these codes takes hours to days to run.

Eventually, it is necessary to be able to produce not only accurate but also *fast* recombination histories, to be included in Markov chain Monte Carlo (MCMC) codes for cosmological parameter estimation. The MCMC requires CMB power spectra (and hence recombination histories) to be generated at each proposed point in cosmological parameter space, with a typical chain sampling $\mathcal{O}(10^5)$

points [12]. Furthermore, dozens of MCMCs are often run with different combinations of observational constraints and different parameter spaces. This makes it impractical to include recombination codes that run for more than a few seconds in the MCMC. One solution is to precompute recombination histories $x_e(z|H_0, T_{\text{CMB}}, \Omega_m h^2, \Omega_b h^2, Y_{\text{He}}, N_\nu)$ on a grid of cosmological parameters, and then use elaborate interpolation algorithms to evaluate the recombination history for any cosmology [13], or to construct fitting functions [8,14]. However, such procedures need to be retrained every time additional parameters are added, and are rather unsatisfying regarding their physical significance.

In this work we present a new method of solution for the recombination problem, perfectly equivalent to the standard MLA method, but much more efficient computationally. The basic idea is that the vast majority of the excited hydrogen levels are populated and depopulated only by *optically thin* radiative transitions (bound-bound and bound-free) in a bath of thermal photons; we show that their effect can be "integrated out" leaving only a few functions of the matter and radiation temperatures T_m and T_r (this list would include the free electron density n_e if we incorporated collisions), which can be pretabulated. In an actual call to the recombination code from an MCMC, it is then only necessary to solve an *effective* MLA (hereafter, EMLA) with a smaller number of levels (perhaps only 3: $2s$, $2p$, and $3p$), which eliminates the computationally difficult $N_{\text{level}} \times N_{\text{level}}$ system solution in the traditional MLA. [The idea is similar in spirit to the line-of-sight

integral method for the computation of the CMB power spectrum [15], which eliminated a large number of independent variables from the cosmological perturbation theory system of ODEs (the high-order moments of the radiation field, Θ_ℓ for $\ell \gg 1$) in favor of pretabulated spherical Bessel functions.] Our method achieves a speedup of the recombination calculation by 5 to 6 orders of magnitude.

We note that our method only eliminates the computational complexity associated with the high-n excited states and does not include continuum processes and radiative transfer effects that have been studied by previous authors [16–26]. However, we note that there has been much progress in analytic treatments of these effects [21–23]; ultimately, we expect to improve these analytic treatments and graft them (and an analytic treatment of helium recombination [27–31]) onto the ultrafast EMLA code described herein to yield a recombination code that is accurate enough for *Planck* data analysis.

This paper is organized as follows. In Sec. II we review the general picture of hydrogen recombination, and the bound-bound and bound-free transition rates involved in the calculation. In Sec. III we describe the standard MLA method. We present our new EMLA method in Sec. IV and demonstrate its equivalence with the standard MLA formulation. We describe our numerical implementation and results in Sec. V, and conclude in Sec. VI. Appendix A is dedicated to demonstrating the invertibility of the system defining the EMLA equations. Appendix B proves a complementarity relation between effective transition probabilities. We prove detailed balance relations between effective transition rates in Appendix C. Appendix D exposes the post-saha approximation we use at early times when computing recombination histories.

II. BOUND-BOUND AND BOUND-FREE TRANSITION RATES

The evolution of the free electron fraction is governed by the network of transitions between bound states of hydrogen as well as recombination and photoionization rates. Before giving detailed expressions for these rates, let us first outline the general picture of the process of recombination.

It has long been known that direct recombinations to the ground state are ineffective for recombination [2,3], since the resulting emitted photons are immediately reabsorbed by hydrogen in the ground state, as soon as the neutral fraction is higher than $\sim 10^{-9}$. Electrons and protons can efficiently combine only to form hydrogen in excited states. The minute amount of excited hydrogen at all relevant times during cosmological recombination is not sufficient to distort the blackbody radiation field near the ionization thresholds of the excited states. Recombination to the excited states is therefore a thermal process: it depend on the matter temperature T_m which characterizes

the free electrons and protons velocity distribution, and also on the radiation temperature T_r, since the abundant low-energy thermal photons can cause stimulated recombinations. Photoionization rates from excited states depend only on the radiation temperature since they do not involve free electrons in the initial state.

Transitions between bound excited states may be radiative or collisional. Radiative transition rates are well known and depend only on the radiation temperature characterizing the blackbody radiation field, undistorted in the vicinity of the optically thin lines from the Balmer series and beyond. Collisional transition rates are much less precisely known, but depend only on the matter temperature and the abundance of charged particles causing the transitions (i.e. free electrons and free protons, which, once helium has recombined, have the same abundance $n_e = n_p$ due to charge neutrality).

Finally, some of the excited states can radiatively decay to the ground state. The most obvious route to the ground state is through Lyman transitions from the p states. However, because of the very high optical depth of these transitions, emitted Lyman photons are immediately reabsorbed by hydrogen atoms in their ground state. This "bottleneck" can only be bypassed by the systematic redshifting of photons, which can escape reabsorption once their frequency is far enough below the resonant frequency of the line. The relevant transition rate in this case is the *net* decay rate to the ground state, which is a statistical average accounting for the very small escape probability of Lyman photons. Two-photon transitions are usually much slower than single-photon transitions. However, the rate of two-photon decays from the metastable $2s$ state is comparable to the net decay rate in the highly self-absorbed Lyman transitions, and this process should therefore be included in a recombination calculation [2,3].

We now give explicit expressions for the bound-bound and bound-free rates discussed above. Subscripts nl refer to the bound state of principal quantum number n and angular momentum quantum number l. We denote α_{fs} the fine structure constant, $\mu_e \equiv m_e m_p/(m_e + m_p)$ the reduced mass of the electron-proton system, E_I the ionization energy of hydrogen, and $E_n \equiv -E_I n^{-2}$ the energy of the nth shell. Finally, we denote by $f_{BB}(E, T_r) \equiv (e^{E/T_r} - 1)^{-1}$ the photon occupation number at energy E in the blackbody radiation field at temperature T_r.

A. Recombination to and photoionization from the excited states

The recombination coefficient to the excited state nl, including stimulated recombinations, is denoted $\alpha_{nl}(T_m, T_r)$ (it has units of cm^3 s^{-1}). The photoionization rate per atom in the state nl is denoted $\beta_{nl}(T_r)$. Both can be expressed in terms of the bound-free radial matrix elements $g(n, l, \kappa, l')$ [32]. Defining

$$\gamma_{nl}(\kappa) \equiv \frac{2}{3n^2} \alpha_{\text{fs}}^3 \frac{E_I}{h} (1 + n^2\kappa^2)^3$$
$$\times \sum_{l'=l\pm1} \max(l, l') g(n, l, \kappa, l')^2, \qquad (1)$$

where κ denotes the momentum of the outgoing electron in units of $\hbar/a_0$ (where a_0 is the reduced-mass Bohr radius), the recombination coefficient is given by [32]

$$\alpha_{nl}(T_m, T_r) = \frac{h^3}{(2\pi\mu_e T_m)^{3/2}} \times \int_0^{+\infty} e^{-E_I\kappa^2/T_m} \gamma_{nl}(\kappa)$$
$$\times [1 + f_{\text{BB}}(E_{\kappa n}, T_r)] d(\kappa^2), \qquad (2)$$

where $E_{\kappa n} \equiv E_I(\kappa^2 + n^{-2})$. The photoionization rate depends only on the radiation temperature and can be obtained by detailed balance considerations from the recombination coefficient:

$$\beta_{nl}(T_r) = \frac{(2\pi\mu_e T_r)^{3/2}}{(2l+1)h^3} e^{E_n/T_r} \alpha_{nl}(T_m = T_r, T_r). \qquad (3)$$

B. Transitions between excited states

We denote $R_{nl \to n'l'}$ the transition rate from the excited state nl to the excited state $n'l'$. It has units of $\sec^{-1}$ per atom in the initial state. Transitions among excited states can be either radiative or collisional:

$$R_{nl \to n'l'} = R_{nl \to n'l'}^{\text{rad}}(T_r) + R_{nl \to n'l'}^{\text{coll}}(T_m, n_e), \qquad (4)$$

where $n_e = n_p$ is the abundance of free electrons or free protons. In this paper, we follow exclusively the radiative rates. These are given by

$$R_{nl \to n'l'}^{\text{rad}} = \begin{cases} A_{nl,n'l'}[1 + f_{\text{BB}}(E_{nn'}, T_r)] & E_n > E_{n'} \\ \frac{g_{l'}}{g_l} e^{-E_{n'n}/T_r} R_{n'l' \to nl}^{\text{rad}} & E_n < E_{n'} \end{cases}, \qquad (5)$$

where $E_{nn'} \equiv E_n - E_{n'}$ is the energy difference between the excited levels, $g_l \equiv 2l + 1$ is the degeneracy of the state nl, and $A_{nl,n'l'}$ is the Einstein A coefficient for the $nl \to n'l'$ transition, which may be obtained from the radial matrix element $R_{n'l'}^{nl}$ [33]:

$$A_{nl,n'l'} = \frac{2\pi}{3} \alpha_{\text{fs}}^3 \frac{E_I}{h} \left(\frac{1}{n'^2} - \frac{1}{n^2}\right)^3 \frac{\max(l, l')}{2l+1} |R_{n'l'}^{nl}|^2. \qquad (6)$$

C. Transitions to the ground state

Finally, the ground state population $x_{1s} \approx 1 - x_e$ evolves due to transitions from and into the np and $2s$ states (two-photon transitions from higher energy states are dominated by "1 + 1" photon decays, already accounted for). Photons emitted in the Lyman lines are very likely to be immediately reabsorbed, and the only meaningful quantity for these transitions is the *net* decay rate in the line, which is a statistical average over a large number of atoms, and accounts for the very low escape probability of a photon

emitted in the line. In the Sobolev approximation [34] with optical depth $\tau_{np,1s} \gg 1$, the net decay rate in the $np \to 1s$ transition is

$$\dot{x}_{1s}|_{np} = -\dot{x}_{np}|_{1s} = \frac{A_{np,1s}}{\tau_{np,1s}} (x_{np} - 3x_{1s}f_{np}^+)$$
$$= \frac{8\pi H}{3\lambda_n^3 n_H x_{1s}} (x_{np} - 3x_{1s}f_{np}^+), \qquad (7)$$

where $\lambda_n \equiv hc/E_{n1}$ is the transition wavelength, and f_{np}^+ is the photon occupation number at the blue side of the corresponding Ly-n line. In this paper, we will take $f_{np}^+ = f_{\text{BB}}(E_{n1}, T_r)$, i.e. assume the incoming radiation on the blue side of the line has a blackbody spectrum. This assumption is actually violated due to feedback from higher-frequency Lyman lines (e.g. radiation escaping from Lyβ can redshift into Lyα) [27,35,36]; while our formalism is general enough to incorporate different f_{np}^+, we have not yet implemented this in our code.

The $2s$ state cannot decay to the ground state through a radiatively allowed transition. This decay is, however, possible with a two-photon emission, which, although slow, is comparable in efficiency to the highly self-absorbed Lyman transitions. The simplest expression for the net $2s \to 1s$ two-photon decay rate is

$$\dot{x}_{1s}|_{2s} = -\dot{x}_{2s}|_{1s} = \Lambda_{2s1s}(x_{2s} - x_{1s}e^{-E_2/T_r}), \qquad (8)$$

where $\Lambda_{2s1s} \approx 8.22 \text{ s}^{-1}$ is the total $2s \to 1s$ two-photon decay rate [37].

In each case, we denote the *net* downward rate in the $i \to 1s$ transition, where $i \in \{2s, 2p, 3p, \ldots\}$:

$$\dot{x}_{1s}|_i = -\dot{x}_i|_{1s} = x_i \tilde{R}_{i \to 1s} - x_{1s} \tilde{R}_{1s \to i}, \qquad (9)$$

where the rates $\tilde{R}$ depend on atomic physics, T_r, and the optical depths in the Lyman lines.

Both the Sobolev approximation for the $np \to 1s$ transitions Eq. (7), and the simple expression Eq. (8) for the net $2s \to 1s$ two-photon decay do not account for subtle yet important radiative transfer effects. An accurate recombination calculation should account for time-dependent effects in Lyα [21,24], a suite of two-photon continuum processes [17,18,21,25], and resonant scattering in Lyα [23,26]. These are not included in the present code, and we plan to add them in the future using analytic treatments.

III. THE STANDARD MLA METHOD

Although the standard MLA formulation does not make this distinction, we cast the excited states of hydrogen into two categories. On the one hand, most excited states are not directly radiatively connected to the ground state. We call these states "interior" states and denote X_K the fractional abundance of hydrogen in the interior state $K \in \{3s, 3d, 4s, 4d, 4f, 5s, \ldots\}$. On the other hand, the $2s$ and np states ($n \geq 2$) are directly radiatively connected with the ground state. We call these states "interface" states and

denote x_i the fractional abundance of hydrogen in the interface state $i \in \{2s, 2p, 3p, \ldots\}$.

In the standard MLA formulation, the free electron fraction $x_e(z)$ is evolved by solving the hierarchy of coupled differential equations: for the interior states,

$$\dot{X}_K = x_e^2 n_H \alpha_K + \sum_L X_L R_{L \to K} + \sum_j x_j R_{j \to K}$$
$$- X_K \Big(\beta_K + \sum_L R_{K \to L} + \sum_j R_{K \to j} \Big); \qquad (10)$$

for the interface states,

$$\dot{x}_i = x_e^2 n_H \alpha_i + \sum_L X_L R_{L \to i} + \sum_j x_j R_{j \to i} + x_{1s} \tilde{R}_{1s \to i}$$
$$- x_i \Big(\beta_i + \sum_L R_{i \to L} + \sum_j R_{i \to j} + \tilde{R}_{i \to 1s} \Big); \qquad (11)$$

and for the free electrons and ground state,

$$\dot{x}_e = -\dot{x}_{1s} = x_{1s} \sum_i \tilde{R}_{1s \to i} - \sum_i x_i \tilde{R}_{i \to 1s}. \qquad (12)$$

The radiative rates between excited states are many orders of magnitude larger than the rate at which recombination proceeds, which is of the order of the Hubble rate. Even the relatively small net rates out of the interface states ($\Lambda_{2s,1s}$ and $A_{2p,1s}/\tau_{2p,1s}$) are still more than 12 orders of magnitude larger than the Hubble rate. The populations of the excited states can therefore be obtained to high accuracy in the steady-state approximation (this approximation is ubiquitous in many problems and has long been used in the context of cosmological recombination [2,10,21], where its accuracy has been tested explicitly [11]). Setting $\dot{X}_K$ and $\dot{x}_i$ to zero in Eqs. (10) and (11), we see that the problem amounts to first solve a system of linear algebraic equations for the X_K, x_i, with an inhomogeneous term depending on x_e, and then use the populations x_i in Eq. (12) to evolve the free electron fraction. The solution of the system of Eqs. (10) and (11) needs to be done at *every time step*, since the inhomogeneous term of the equation depends on the ionization history, which explicitly depends on time as well as on the cosmological parameters. Recent work [10,11] has shown that to compute sufficiently accurate recombination histories, one needs to account for excited states up to a principal quantum number $n_{max} \sim 100$, resolving the angular momentum substates. This requires solving an $\mathcal{O}(10^4 \times 10^4)$ system of equations at each time step, which, even with modern computers, is extremely time consuming.

IV. NEW METHOD OF SOLUTION: THE EFFECTIVE MULTILEVEL ATOM

We now give a computationally efficient method of solution for the primordial recombination problem. We factor the effect of the numerous transitions involving interior states in terms of effective transitions into and out of the much smaller number of interface states. Once the rates of these effective transitions are tabulated, the cosmological evolution of the free electron fraction can be obtained from a simple effective few-level atom calculation. We describe the method in Sec. IV A and give the proof of its exact equivalence to the standard MLA method in Sec. IV B. In the subsequent Sec. IV C, we consider which states should be treated as interface states.

A. Motivations and general formulation

We first note that the only quantity of importance for CMB power spectrum calculations is the free electron fraction as a function of redshift, $x_e(z)$. The populations of the excited states are calculated only as an intermediate step—if they are desired (e.g. to calculate Hα scattering features [38]), the populations of the excited states can be obtained by solving Eqs. (10) and (11) once the free electron fraction is known. Furthermore, only the interface states $2s$ and np are directly connected to the ground state and directly appear in the evolution equation for the free electron fraction Eq. (12). All other (interior) excited states are only connected with other excited states or with the continuum, through optically thin radiative transitions (and to a lesser extent through collisions [9]). Interior states are only transitional states: an electron in the interior rapidly transitions through spontaneous and stimulated decays or absorptions caused by the blackbody radiation field (or collisions with free electrons and protons), until it either is photoionized or reaches an interface state. There can be a very large number of transitions before any of these outcomes occurs, but the passage through the interior is always very short compared to the overall recombination time scale, and can be considered as instantaneous (for the same reason that the steady-state approximation is valid in the standard MLA formulation).

Instead of computing the fraction of hydrogen in each interior state K, one can rather evaluate the probabilities that an atom initially in the interior state K ultimately reaches one of the interface states or gets photoionized. Of course, after reaching an interface state, the atom may perfectly transition back to an interior state, or get photoionized. However, we consider the probability of *first* reaching a given interface state before any other one, which is uniquely defined. For an atom in the interior state K, we denote by P_K^i the probability of ultimately reaching the interface state i, and P_K^e the probability of ultimately being photoionized. The probabilities P_K^i must self-consistently account for both direct transitions $K \to i$ and all possible indirect transitions $K \to L \to i$ (with an arbitrary number of intermediate states). Mathematically, this translates to the system of linear equations:

$$P_K^i = \sum_L \frac{R_{K \to L}}{\Gamma_K} P_L^i + \frac{R_{K \to i}}{\Gamma_K}, \qquad (13)$$

where Γ_K is the total width (or inverse lifetime) of the state K:

$$\Gamma_K \equiv \sum_L R_{K\to L} + \sum_j R_{K\to j} + \beta_K. \qquad (14)$$

Similarly, the P_K^e must satisfy the self-consistency relations:

$$P_K^e = \sum_L \frac{R_{K\to L}}{\Gamma_K} P_L^e + \frac{\beta_K}{\Gamma_K}, \qquad (15)$$

We show in Appendix A that these linear systems are invertible and therefore uniquely determine P_K^i and P_K^e. In Appendix B we prove the complementarity relation,

$$\sum_i P_K^i + P_K^e = 1, \qquad (16)$$

which has the simple physical interpretation that an atom in the Kth interior state eventually reaches an interface state or is photoionized with unit probability.

Once these probabilities are known, it is possible to describe the large number of transitions between all the states in a simplified manner, in terms of effective rates into and out of the interface states. To clarify the explanation, we illustrate in Fig. 1 the processes described below.

An electron and a proton can effectively recombine to the interface state i either through a direct recombination (with coefficient α_i) or following a recombination to an interior state K (with coefficient α_K), from which a sequence of interior transitions may ultimately lead to the interface state i with probability P_K^i. The effective recombination coefficient to the interface state i is therefore

$$\mathcal{A}_i \equiv \alpha_i + \sum_K \alpha_K P_K^i. \qquad (17)$$

Conversely, an atom in the interface state i may effectively be ionized either through a direct photoionization (with rate β_i) or after being first excited to an interior state K (with rate $R_{i\to K}$), from which the atom may ultimately be photoionized after a series of interior transitions with probability P_K^e. The effective photoionization rate from the interface state i is therefore
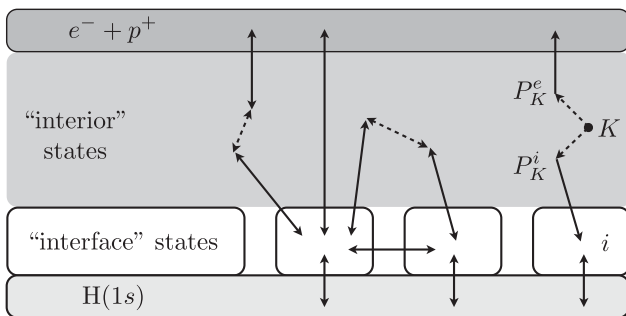

FIG. 1. Schematic representation of the formulation of the recombination problem adopted in this work. Dotted arrows represent possibly numerous fast transitions within the interior.

$$\mathcal{B}_i \equiv \beta_i + \sum_K R_{i\to K} P_K^e. \qquad (18)$$

Finally, atoms can effectively transition from an interface state i to another interface state j, either through a direct transition if it is allowed or after first transitioning through the interior. The effective transfer rate between the ith and jth interface states is therefore

$$\mathcal{R}_{i\to j} \equiv R_{i\to j} + \sum_K R_{i\to K} P_K^j \qquad (j \neq i). \qquad (19)$$

The rate of change of the population of the interface state i is therefore

$$\dot{x}_i = x_e^2 n_H \mathcal{A}_i + \sum_{j\neq i} x_j \mathcal{R}_{j\to i} + x_{1s} \tilde{R}_{1s\to i}$$
$$- x_i \Big(\mathcal{B}_i + \sum_{j\neq i} \mathcal{R}_{i\to j} + \tilde{R}_{i\to 1s} \Big), \qquad (20)$$

where we have included the effective transitions described above, as well as transitions from and to the ground state.

The system of Eqs. (13)–(20) is exactly equivalent to the standard MLA formulation, as we shall show in Sec. IV B below.

Let us now consider the dependences of the effective rates. In the purely radiative case, the probabilities P_K^i and P_K^e depend only on the radiation temperature T_r, since transitions between excited states and photoionizations depend only on the locally thermal radiation field. As a consequence, the effective recombination rates $\mathcal{A}_i(T_m, T_r)$ are only functions of matter and radiation temperatures and the effective photoionization and bound-bound rates $\mathcal{B}_i(T_r)$ and $\mathcal{R}_{i\to j}(T_r)$ are functions of the radiation temperature only. When including collisional transitions, all effective rates become functions of the three variables T_r, T_m, and n_e. In all cases, effective rates can easily be tabulated and interpolated when needed for a recombination calculation.

Intuitively, we would expect that $\mathcal{A}_i$, $\mathcal{B}_i$, and $\mathcal{R}_{i\to j}$ satisfy the detailed balance relations,

$$g_i e^{-E_i/T_r} \mathcal{R}_{i\to j}(T_r) = g_j e^{-E_j/T_r} \mathcal{R}_{j\to i}(T_r) \qquad (21)$$

and

$$g_i e^{-E_i/T_r} \mathcal{B}_i(T_r) = \frac{(2\pi \mu_e T_r)^{3/2}}{h^3} \mathcal{A}_i(T_m = T_r, T_r). \qquad (22)$$

We show in Appendix C that these equations are indeed valid. This means that we only need to tabulate half of the $\mathcal{R}_{i\to j}$ [the other half can be obtained from Eq. (21)] and all the $\mathcal{A}_i$ [the $\mathcal{B}_i$ can be obtained from Eq. (22); in particular, we do not need to solve for the P_K^e].

We note that the probabilities P_K^i, P_K^e are a generalization of the cascade matrix technique introduced by Seaton [39]. Seaton's calculation assumed a vanishing ambient radiation field, so that electrons can only "cascade down" to lower energy states. In the context of the recombination of the

primeval plasma, one cannot ignore the strong thermal radiation field, and electrons rather "cascade up and down," following spontaneous and stimulated decays or photon absorption events. The spirit of our method is, however, identical to Seaton's cascade-capture equations [39], where the "cascading" process is decoupled from the particular process populating the excited states, or from the depopulation of the interface states.

B. Equivalence with the standard MLA method

This section is dedicated to proving the equivalence of the EMLA equations, Eqs. (13)–(20), with the standard MLA equations, Eqs. (10) and (11), in the steady-state limit for the interior states (i.e. where we set $\dot{X}_K \approx 0$). The steady-state approximation does not need to be made for the interface states to demonstrate the equivalence of the two formulations (but we do use it for practical computations since it is valid to very high accuracy).

We denote by N the number of interior states and n_* the number of interface states [we will address in Sec. IV C the issue of which states need to be considered as interface states].

We begin by defining the $N \times N$ rate matrix $\mathbf{M}$ whose elements are

$$M_{KL} \equiv \delta_{KL}\Gamma_K - (1 - \delta_{KL})R_{K \to L}. \qquad (23)$$

We also define the $n_* + 1$ length-N vectors $\mathbf{P}^i$ and $\mathbf{P}^e$ whose elements are the probabilities P_K^i and P_K^e, respectively, and the $n_* + 1$ length-N vectors $\mathbf{R}^i$ and $\mathbf{R}^e$ of components

$$R_K^i \equiv R_{K \to i}, \qquad (24)$$

$$R_K^e \equiv \beta_K. \qquad (25)$$

The defining equations for the probabilities, Eqs. (13) and (15) can be written in matrix form $\mathbf{M}\mathbf{P}^i = \mathbf{R}^i$ and $\mathbf{M}\mathbf{P}^e = \mathbf{R}^e$, respectively (after multiplication by Γ_K). We show in Appendix A that the matrix $\mathbf{M}(T_r)$ is invertible, for any temperature $T_r \geq 0$. The formal solutions for the probabilities are therefore

$$\mathbf{P}^i = \mathbf{M}^{-1}\mathbf{R}^i, \qquad (26)$$

$$\mathbf{P}^e = \mathbf{M}^{-1}\mathbf{R}^e. \qquad (27)$$

We also define the length-N vector $\mathbf{X}$ which contains the populations of the interior states X_K, and the length-N vector $\mathbf{S}$ of components

$$S_K \equiv x_e^2 n_H \alpha_K + \sum_j x_j R_{j \to K}. \qquad (28)$$

A careful look at Eq. (10) in the steady-state approximation ($\dot{X}_K = 0$) shows that it is the matrix equation $\mathbf{M}^T\mathbf{X} = \mathbf{S}$, which has the solution:

$$\mathbf{X} = (\mathbf{M}^T)^{-1}\mathbf{S} = (\mathbf{M}^{-1})^T\mathbf{S}. \qquad (29)$$

Both Eqs. (11) and (20) can be cast in the form

$$\dot{x}_i = x_e^2 n_H \alpha_i + \sum_{j \neq i} x_i R_{j \to i} + x_{1s}\tilde{R}_{1s \to i}$$
$$- x_i\left(\beta_i + \sum_{j \neq i} R_{i \to j} + \tilde{R}_{i \to 1s}\right) + \dot{x}_i|_{\text{interior}}. \qquad (30)$$

The only *a priori* different term is the net transition rate from the interior to the state i, $\dot{x}_i|_{\text{interior}}$. In the standard MLA formulation, Eq. (11), this term is

$$\dot{x}_i|_{\text{interior}}^{\text{(MLA)}} = \sum_K (X_K R_K^i - x_i R_{i \to K}) \qquad (31)$$

$$= \mathbf{X}^T\mathbf{R}^i - x_i\sum_K R_{i \to K}. \qquad (32)$$

With our new formulation, Eq. (20), using the definitions of the effective rates Eqs. (17)–(19), the net transition rate from the interior to the state i is

$$\dot{x}_i|_{\text{interior}}^{\text{(EMLA)}} = \sum_K\left[x_e^2 n_H \alpha_K P_K^i + \sum_{j \neq i} x_j R_{j \to K} P_K^i\right.$$
$$\left. - x_i R_{i \to K}\left(P_K^e + \sum_{j \neq i} P_K^j\right)\right]. \qquad (33)$$

Using the complementarity relation Eq. (16), we rewrite $P_K^e + \sum_{j \neq i} P_K^j = 1 - P_K^i$. We then recognize that the common factor of P_K^i is just the Kth component of the vector $\mathbf{S}$, Eq. (28), so we can rewrite Eq. (33) as

$$\dot{x}_i|_{\text{interior}}^{\text{(EMLA)}} = \mathbf{S}^T\mathbf{P}^i - x_i\sum_K R_{i \to K}. \qquad (34)$$

From the formal solution for the populations of the interior states Eq. (29), we have

$$\mathbf{X}^T\mathbf{R}^i = \mathbf{S}^T\mathbf{M}^{-1}\mathbf{R}^i = \mathbf{S}^T\mathbf{P}^i, \qquad (35)$$

where the second equality is obtained from the formal solution for the probabilities P_K^i, Eq. (26). We therefore see from Eqs. (32) and (34) that

$$\dot{x}_i|_{\text{interior}}^{\text{(MLA)}} = \dot{x}_i|_{\text{interior}}^{\text{(EMLA)}}, \qquad (36)$$

and hence the two formulations are *exactly* equivalent. They only differ by the order in which the bilinear product $\mathbf{S}^T\mathbf{M}^{-1}\mathbf{R}^i$ is evaluated.

C. Choice of interface states

If one naively includes all np states up to $n = n_{\max} = \mathcal{O}(100)$ in the list of interface states, the interpolation of effective rates can become somewhat cumbersome as it involves $\mathcal{O}(10^4)$ functions of one to three variables. However, only the lowest few of these states actually

have significant transition rates to the ground state; indeed, most of the decays to the ground state proceed through either $2s$ (two-photon decay) or $2p$ (Lyα escape), as anticipated in the earliest studies [2,3].

The rate of Lyman line escape is dominated by the lowest few lines. For example, if the relative populations of the excited states were given by the Boltzmann ratios (which is a good approximation until late times), then the net decay rate in the $np \rightarrow 1s$ transition (not accounting for feedback from the next line) would be proportional to

$$\dot{x}_{np \rightarrow 1s} \propto (1 - n^{-2})^3 e^{-E_n/T_r}. \tag{37}$$

This relation would imply that the Lyβ escape rate is <1% of the Lyα escape rate, and the higher-order Lyman lines contribute even less. Our previous computations of the escape rates (e.g. Ref. [21]) agree with this expectation. These considerations imply that for $n \geq 3$, $\dot{x}_{1s}|_{np} \ll \dot{x}_{1s}|_{2p}$ in Eq. (12). Moreover, an atom in the np state with $n \geq 3$ is much more likely to spontaneously decay to $n's$ or $n'd$, with $2 \leq n' < n$, than to emit a Lyman-n photon that successfully escapes the line. This implies that $|\dot{x}_{np}|_{1s}| \ll |\dot{x}_{np}|$ in Eq. (11).

In addition to a very low net decay rate out of the np states for $n \geq 3$, feedback between neighboring lines further suppresses their efficiency as interface states. The few photons that escape the Ly$(n + 1)$ line will be reabsorbed almost certainly in the next lower line, after a redshift interval

$$\Delta z = z_{em} - z_{ab} = (1 + z_{ab})\left(\frac{E_{n+1,1}}{E_{n1}} - 1\right). \tag{38}$$

Feedback between the lowest-lying lines is not instantaneous: $\Delta z/(1 + z_{ab}) = 0.185$ for Ly$\beta \rightarrow$ Lyα feedback, 0.055 for Ly$\gamma \rightarrow$ Lyβ, and 0.024 for Ly$\delta \rightarrow$ Lyγ. However, for higher-order lines, feedback rapidly becomes nearly instantaneous as $\Delta z/(1 + z_{ab}) \sim 2/n^3$. Thus the effect of the higher Lyman lines is even weaker than Eq. (37) would suggest. Recent work [36] has shown that including lines above Lyβ results in a fractional error $|\Delta x_e|/x_e$ of at most $\approx 3 \times 10^{-4}$.

We therefore conclude that very accurate recombination histories can be obtained by only including $2s, 2p, \ldots, n_*p$ as interface states and neglecting higher-order Lyman transitions altogether. We will use $n_* = 3$ in this paper, and investigate the optimal value of this cutoff more quantitatively in future work.

Our formulation in terms of effective transition rates and interface states is therefore much better adapted for a fast recombination calculation that the standard MLA formulation. To compute accurate recombination histories, explicitly accounting for high-n shells of hydrogen, one first needs to tabulate the $\{\mathcal{A}_i\}$ and $\{\mathcal{R}_{i \rightarrow j}\}$ on temperature grids. The computation of the effective rates is the time-consuming part of the calculation; however, since they are independent of the cosmological parameters, this can be done once, and not repeated for each cosmology. The free electron fraction can then be computed very quickly for any given cosmology by solving the $n_* + 1$ Eqs. (20) and (12), interpolating the effective rates from the pre-computed tables. Note that Eq. (20) is a simple $n_* \times n_*$ system of linear algebraic equations in the steady-state approximation.

V. IMPLEMENTATION AND RESULTS

Here we give some details on the implementation of our EMLA code. Section V A describes the computation of the effective rates (the computationally expensive part of the calculation, which needs to be done only once). Section V B describes the implementation of the ultrafast effective few-level atom calculation. We show our recombination histories and compare our results with the existing standard MLA code RECSPARSE [10] in Sec. V C.

A. Computation of the effective rates

We have implemented the calculation of the effective rates in the purely radiative case. Bound-free rates were computed by numerically integrating Eq. (2) using an 11-point Newton-Cotes method, where the radial matrix elements $g(n, l, \kappa, l')$ were obtained using the recursion relation given by Burgess [32]. Einstein A coefficients were computed by using the recursion relations obtained by Hey [40] for the radial matrix elements $R_{n'l'}^{nl}$. Finally, we obtained the probabilities P_K^i using a sparse matrix technique similar to that of Ref. [10] when solving Eq. (13). We accounted explicitly for all excited states up to a principal quantum number n_{max}, resolving angular momentum substates. We tabulated the effective rates $\mathcal{A}_i(T_m, T_r)$ on a grid of 200 log-spaced points in T_r from 0.04 to 0.5 eV and 20 linearly spaced points in T_m/T_r from 0.8 to 1.0, and $\mathcal{R}_{i \rightarrow j}(T_r)$ on the grid of points in T_r. The maximum relative change in the effective rates ($\Delta \ln \mathcal{A}_i$ or $\Delta \ln \mathcal{R}_{i \rightarrow j}$) over the whole range of temperatures considered is 0.051 when comparing $n_{max} = 64$ vs 128, 0.015 when comparing $n_{max} = 128$ vs 250, and 0.005 when comparing $n_{max} = 250$ vs 500.

In the left panel of Fig. 2, we show the total effective recombination coefficient $\mathcal{A}_B(T_m, T_r) \equiv \mathcal{A}_{2s}(T_m, T_r) + \mathcal{A}_{2p}(T_m, T_r)$ computed for $n_* = 2$ (i.e. with interfaces states $2s$ and $2p$ only, neglecting all Lyman transitions above Lyman-α), normalized to the case-B recombination coefficient $\alpha_B(T_m)$. Note that $\alpha_B(T_m)$ is just $\mathcal{A}_B(T_m, T_r = 0|n_{max} = \infty)$ with our notation; indeed, for $T_r = 0$, $\beta_K = 0$ and therefore $P_K^e = 0$ for all K so $\sum_i P_K^i = 1$ and hence $\sum_i \mathcal{A}_i = \sum_i \alpha_i + \sum_K \alpha_K = \sum_{nl} \alpha_{nl}$, where the last sum is over all excited states. We can see that as the radiation temperature increases (i.e. as the redshift increases in Fig. 2), the convergence with n_{max} becomes faster. This is to be expected, since for higher T_r, highly excited hydrogen is more easily photoionized, i.e. P_{nl}^e
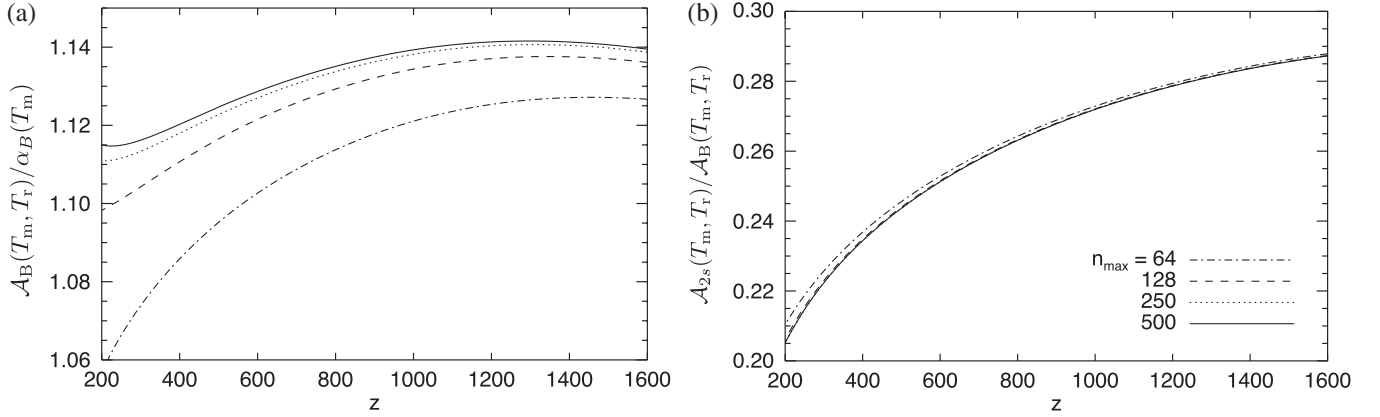
(a)

(b)

FIG. 2. Left panel: "Exact fudge factor" as a function of redshift $\mathcal{A}_B(T_m, T_r)/\alpha_B(T_m)$, for several values of n_{max}, using $T_m(z)$ computed by RECSPARSE for cosmological parameters as in Ref. [10]. We use the fit of Ref. [41] for the case-B recombination coefficient $\alpha_B(T_m)$. For comparison, the code RECFAST uses a constant fudge factor $F = 1.14$ to mimic the effect of high-n shells. Right panel: Fraction of the effective recombinations to the $n = 2$ shell that lead to atomic hydrogen in the $2s$ state. In both cases the effective rates were computed for $n_* = 2$, i.e. with interface states $2s$ and $2p$ only, neglecting escape from the Lyman $\beta, \gamma, \ldots$ lines.

becomes closer to unity. In that case adding more shells to the calculation does not matter so much because recombinations to the highest shells are very inefficient, due to the high probability of a subsequent photoionization.

In the right panel of Fig. 2, we show the ratio $\mathcal{A}_{2s}(T_m, T_r)/\mathcal{A}_B(T_m, T_r)$, which is the fraction of recombinations to the $n = 2$ shell that are to the $2s$ level. This fraction is in general different from the intuitive value of $1/4$, and its exact value depends on temperature.

B. Ultrafast EMLA code

In order to actually compute the recombination history, we require an evolution equation for the free electron fraction,

$$\dot{x}_e(x_e, n_H, H, T_m, T_r), \tag{39}$$

and in some cases a similar equation for $\dot{T}_m$. For concreteness, we implement the case of 3 interface states $i \in \{2s, 2p, 3p\}$ ($n_* = 3$).

To compute $\dot{x}_e$, we first obtain the downward $\mathcal{R}_{i \to j}(T_r)$ from our table via cubic polynomial (4-point) interpolation and $\mathcal{A}_i(T_m, T_r)$ via bicubic interpolation (two-dimensional in $\ln(T_r)$ and T_m/T_r using 4×4 points). The upward $\mathcal{R}_{j \to i}(T_r)$ are obtained using Eq. (21), and the effective photoionization rates $\mathcal{B}_i(T_r)$ are obtained using Eq. (22). We then solve for the $\{x_i\}$ using Eq. (20), and finally obtain $\dot{x}_e$ using Eq. (12).

The matter temperature is determined by the Compton evolution equation,

$$\dot{T}_m = -2HT_m + \frac{8\sigma_T a_r T_r^4 x_e (T_r - T_m)}{3(1 + f_{He} + x_e)m_e c}, \tag{40}$$

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

the electron mass, and c is the speed of light. At high redshift, one may use the steady-state solution (see Appendix A of Ref. [21]),

$$T_m \approx T_{m,ss} = T_r \left[1 + \frac{3(1 + f_{He} + x_e)m_e cH}{8\sigma_T a_r T_r^4 x_e} \right]^{-1}. \tag{41}$$

At the highest redshifts, the ODE describing hydrogen recombination is stiff; therefore for $z > 1570$ we follow the recombination history using perturbation theory around the Saha approximation, as described in Appendix D. At $500 < z < 1570$ we use Eq. (41) to set the matter temperature, and a fourth-order Runge-Kutta integration algorithm (RK4) to follow the single ODE for $x_e(z)$; and at $z < 500$ we use RK4 to follow the two ODEs for $x_e(z)$ and $T_m(z)$ simultaneously. The integration step size is $\Delta z = -1.0$ (negative since we go from high to low redshifts).

C. Results and code comparison

We have tabulated the effective rates for $n_{max} = 16, 32, 64, 128, 250$, and 500. It is, in principle, possible to compute the effective rates for an arbitrarily high n_{max}, but it is not meaningful to do so as long as collisional transitions are not properly accounted for. The recurring computation time of our ultrafast EMLA code is 0.08 seconds on a MacBook laptop computer with a 2.1 GHz processor, independently of n_{max}. Our recombination histories are shown in Fig. 3. We compared our results with the existing standard MLA code RECSPARSE for $n_{max} = 16, 32, 64, 128$, and 250. As can be seen in Fig. 4, the two codes agree to better than 8×10^{-5} across the range $200 < z < 1600$, despite having different methods for accounting for the excited states, and independent implementations for matrix elements and ODE integration.
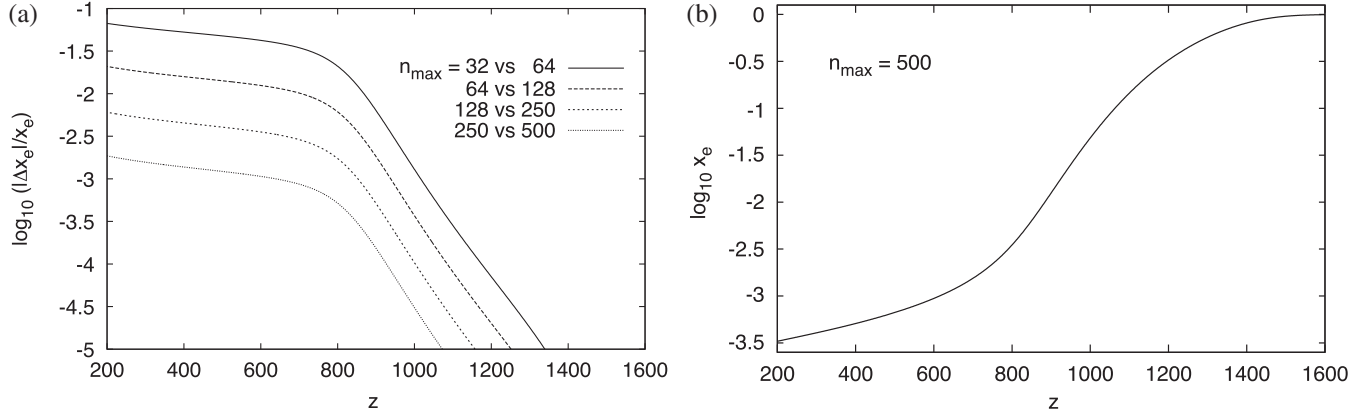
FIG. 3.   Left panel: Relative differences between recombination histories computed with successively more accurate effective rates. Right panel: Recombination history for effective rates computed with $n_{max} = 500$, i.e. accounting explicitly for 125 250 states of the hydrogen atom.
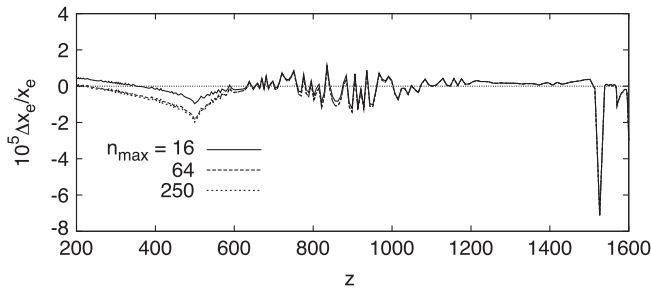

FIG. 4.   A comparison of our ultrafast code to RECSPARSE [10], for different values of n_{max}. The vertical axis is the fractional difference in free electron abundance rescaled by 10^5 (positive indicating that RECSPARSE gives a larger x_e). We see that the maximum fractional deviation is $< 8 \times 10^{-5}$. The feature around $z = 1540$ is due to a time step change in RECSPARSE.

VI. CONCLUSIONS AND FUTURE DIRECTIONS

We have shown that the computation of primordial hydrogen recombination can be factored into two independent calculations. On the one hand, most excited states are not directly radiatively connected to the ground state, and undergo transitions caused by the thermal bath of black-body photons at the relevant frequencies, as well as the thermal electrons and protons. One can account for these numerous transitions with effective transition rates into and out of the interface states which are connected to the ground state. The computationally intensive aspect of a recombination calculation, in fact, resides in the evaluation of these effective rates, which are functions of matter and radiation temperature only. This calculation being independent of cosmological parameters, it can be done prior to any recombination calculation, once and for all. A simple effective few-level atom can then be evolved for any set of cosmological parameters, without any need for "fudge factors" or approximations.

This work does *not* present a final recombination code satisfying the accuracy requirements for future CMB experiments. First, collisional transitions were not included. They may be particularly important for the high-n states. The effective rates computed here are therefore only approximating the correct rates in the limit of zero density. Our formalism is general and collisions can be included as soon as accurate rates are available (the main change would be that the interpolation tables would require $\ln(n_e)$ as an additional independent variable). Second, we have not included important radiative transfer effects, such as feedback between low-lying Lyman lines [35,36], two-photon decays from $n \geq 3$ [16,19–21,25], resonant scattering in Lyα [22,23,26], or overlap of the high-lying Lyman lines (work in preparation). To preserve the computational efficiency of our method, fast analytic approximations have to be developed to include these effects, which will be the subject of future work.

APPENDIX A: INVERTIBILITY OF THE SYSTEM DEFINING THE P_K^i, P_K^e

In this section we show that the matrix $\mathbf{M}(T_r)$ defined in Eq. (23) is nonsingular, for any value of the radiation temperature $T_r \geq 0$.

Let us consider the eigenvalue equation $\mathbf{Mb} = \mathbf{0}$ and select a particular K_1 such that $|b_{K_1}| \geq |b_L|$ for all L. The eigenvalue equation implies

$$0 = |M_{K_1K_1}b_{K_1} + \sum_{L \neq K} M_{K_1L}b_L|$$

$$\geq M_{K_1K_1}|b_{K_1}| - \sum_{L \neq K_1} |M_{K_1L}||b_L|, \qquad (A1)$$

where we have used the inverse triangle inequality. The matrix $\mathbf{M}$ is diagonally dominant, i.e.

$$\forall \ K, \qquad M_{KK} = \Gamma_K \geq \sum_{L \neq K} R_{K \to L} = \sum_{L \neq K} |M_{KL}|. \quad (A2)$$

Using the inequality (A2) for $K = K_1$ in Eq. (A1), we obtain

$$0 \geq \sum_{L \neq K_1} |M_{K_1L}|(|b_{K_1}| - |b_L|). \qquad (A3)$$

For any interior state K_1, there always exists a sequence of transitions that ultimately leads to some interface state i, $K_1 \to K_2 \to \cdots \to K_n \to i$, for any temperature $T_r \geq 0$ (i.e. there are no "dead end" interior states). In particular, $|M_{K_1K_2}| = R_{K_1 \to K_2} > 0$. For Eq. (A3) to hold, it is therefore necessary that $|b_{K_1}| = |b_{K_2}|$. Repeating the above reasoning recursively leads to $|b_{K_1}| = |b_{K_2}| = \cdots = |b_{K_n}|$.

For the last interior state of this sequence, K_n, the inequality (A2) is strict since $R_{K_n \to i} > 0$. The eigenvalue equation projected on K_n leads to Eq. (A1) for K_n:

$$0 \geq M_{K_nK_n}|b_{K_n}| - \sum_{L \neq K_n} |M_{K_nL}||b_L|. \qquad (A4)$$

If $\mathbf{b} \neq \mathbf{0}$, then $|b_{K_n}| > 0$ and the strict inequality (A2) for $K = K_n$ used in Eq. (A4) implies the contradictory result

$$0 > \sum_{L \neq K_n} |M_{K_nL}|(|b_{K_n}| - |b_L|) \geq 0. \qquad (A5)$$

As a consequence, $\mathbf{Mb} = \mathbf{0}$ implies that $\mathbf{b} = \mathbf{0}$ necessarily. This proves that $\mathbf{M}(T_r)$ is nonsingular, for any $T_r \geq 0$.

APPENDIX B: PROOF OF THE COMPLEMENTARITY RELATION $\sum_i P_K^i + P_K^e = 1$

We define the length-N vector $\mathbf{V} \equiv (1, 1, \ldots, 1)^T$, and note that

$$(\mathbf{MV})_K = \sum_L M_{KL} = \sum_j R_{K \to j} + \beta_K \qquad (B1)$$

(the $R_{K \to L}$ terms cancel). In matrix form, this reads:

$$\mathbf{MV} = \sum_j \mathbf{R}^j + \mathbf{R}^e = \mathbf{M}\left[\sum_j \mathbf{P}^j + \mathbf{P}^e\right]. \qquad (B2)$$

The matrix $\mathbf{M}$ being invertible, this implies $\sum_j \mathbf{P}^j + \mathbf{P}^e = \mathbf{V}$, which once projected on each component K is just the complementarity relation Eq. (16).

APPENDIX C: DETAILED BALANCE RELATIONS

This appendix is dedicated to proving the detailed balance relations for $\mathcal{R}_{i \to j}$ and $\mathcal{B}_i$. Defining the contributions of individual states to the partition function,

$$Q_K \equiv g_K e^{-E_K/T_r} \qquad (C1)$$

and similarly for Q_i, we make use of the standard principle of detailed balance for rates connecting individual states,

$$Q_K R_{K \to L} = Q_L R_{L \to K}, \qquad (C2)$$

and similarly $Q_K R_{K \to i} = Q_i R_{i \to K}$.

We begin by defining the $N \times N$ nonsingular diagonal matrix $\mathbf{F}$ that is proportional to the equilibrium abundances,

$$F_{KL} \equiv Q_K \delta_{KL}. \qquad (C3)$$

Then Eq. (C2) combined with the definition Eq. (23) implies that $\mathbf{FM}$ is *symmetric*. It therefore follows that its matrix inverse $\mathbf{M}^{-1}\mathbf{F}^{-1}$ is symmetric, and hence that

$$\frac{(\mathbf{M}^{-1})_{KL}}{Q_L} = \frac{(\mathbf{M}^{-1})_{LK}}{Q_K}. \qquad (C4)$$

The transition rate, Eq. (19), can be expanded using Eq. (26) as

$$\mathcal{R}_{i \to j} = R_{i \to j} + \sum_{K,L} (\mathbf{M}^{-1})_{KL} R_{i \to K} R_{L \to j}. \qquad (C5)$$

We then see that

$$Q_i \mathcal{R}_{i \to j} = Q_i R_{i \to j} + \sum_{K,L} Q_i (\mathbf{M}^{-1})_{KL} R_{i \to K} R_{L \to j}$$

$$= Q_i R_{i \to j} + \sum_{K,L} Q_K (\mathbf{M}^{-1})_{KL} R_{K \to i} R_{L \to j}$$

$$= Q_j R_{j \to i} + \sum_{K,L} Q_L (\mathbf{M}^{-1})_{LK} R_{K \to i} R_{L \to j}$$

$$= Q_j R_{j \to i} + \sum_{K,L} Q_j (\mathbf{M}^{-1})_{LK} R_{K \to i} R_{j \to L}$$

$$= Q_j \mathcal{R}_{j \to i}, \qquad (C6)$$

where we have used Eq. (C2) twice and in the third equality used Eq. (C4). This proves Eq. (21).

We may also relate the effective recombination and photoionization rates. To do so, we consider the case of $T_m = T_r$ and define

$$q \equiv \left(\frac{2\pi \mu_e T_r}{h^2}\right)^{3/2}. \qquad (C7)$$

Then Eq. (3) can be written as $Q_{nl}\beta_{nl} = q\alpha_{nl}$. Using Eq. (27), we see that

$$
\begin{aligned}
Q_i \mathcal{B}_i &= Q_i \beta_i + \sum_{K,L} Q_i R_{i \to K} (\mathbf{M}^{-1})_{KL} \beta_L \\
&= q\alpha_i + \sum_{K,L} Q_K R_{K \to i} (\mathbf{M}^{-1})_{KL} \beta_L \\
&= q\alpha_i + \sum_{K,L} Q_L R_{K \to i} (\mathbf{M}^{-1})_{LK} \beta_L \\
&= q\alpha_i + \sum_{K,L} q R_{K \to i} (\mathbf{M}^{-1})_{LK} \alpha_L = q\mathcal{A}_i, \quad \text{(C8)}
\end{aligned}
$$

where in the last equality we have used Eq. (17) with the P_K^i determined by Eq. (26). This proves Eq. (22).

APPENDIX D: IMPLEMENTATION OF POST-SAHA CORRECTION AT EARLY TIMES

At the highest redshifts, the ODE describing hydrogen recombination is stiff, and we follow the recombination history using perturbation theory around the Saha approximation, which we describe here. The idea is that the actual ionization fraction $x_e(z)$ is slightly greater than the Saha equation would predict because there are more recombinations than ionizations ($\dot{x}_e < 0$) and a slight deviation from thermodynamic equilibrium is required in order to drive this imbalance. We may thus take an ODE for the recombination history (for simplicity we use the Peebles ODE [2] with an updated recombination coefficient α_B [41]), and Taylor expand it around the Saha ionization fraction:

$$
\dot{x}_e^{\mathrm{P}}(x_e, z) = D_1(x_e - x_e^{\mathrm{Saha}}) + \mathcal{O}(x_e - x_e^{\mathrm{Saha}})^2. \quad \text{(D1)}
$$

Here the superscript "P" denotes the Peebles ODE, and the zeroth-order coefficient in the Taylor series vanishes since thermal equilibrium considerations imply $\dot{x}_e^{\mathrm{P}}(x_e^{\mathrm{Saha}}, z) = 0$. The coefficient D_1 may be obtained by numerical differentiation; we use a two-sided finite difference with $\Delta x_e = \pm 0.01 x_e^{\mathrm{Saha}}(1 - x_e^{\mathrm{Saha}})$. Then we obtain a post-Saha corrected solution by setting the left-hand side of Eq. (D1) with $\dot{x}_e^{\mathrm{Saha}}(z)$ (again obtained by a two-sided finite difference with $\Delta z = \pm 1$):

$$
x_e^{\mathrm{corr}}(z) = x_e^{\mathrm{Saha}}(z) + \frac{\dot{x}_e^{\mathrm{Saha}}(z)}{D_1}. \quad \text{(D2)}
$$

At the transition redshift $z_t = 1570$, we use $x_e^{\mathrm{corr}}(z_t)$ as an initial condition for the integration with our full ODE.

[1] Planck Collaboration, arXiv:astro-ph/0604069.
[2] P. J. E. Peebles, Astrophys. J. **153**, 1 (1968).
[3] Y. B. Zel'dovich, V. G. Kurt, and R. A. Sunyaev, Zh. Eksp. Teor. Fiz. **55**, 278 (1968) [Sov. Phys. JETP **28**, 146 (1969)].
[4] W. Hu, D. Scott, N. Sugiyama, and M. White, Phys. Rev. D **52**, 5498 (1995).
[5] A. Lewis, J. Weller, and R. Battye, Mon. Not. R. Astron. Soc. **373**, 561 (2006).
[6] J. A. Rubiño-Martín, J. Chluba, W. A. Fendt, and B. D. Wandelt, Mon. Not. R. Astron. Soc. **403**, 439 (2010).
[7] S. Seager, D. D. Sasselov, and D. Scott, Astrophys. J. Suppl. Ser., **128**, 407 (2000).
[8] S. Seager, D. D. Sasselov, and D. Scott, Astrophys. J. Lett. **523**, L1 (1999).
[9] J. Chluba, J. A. Rubiño-Martín, and R. A. Sunyaev, Mon. Not. R. Astron. Soc. **374**, 1310 (2007).
[10] D. Grin and C. M. Hirata, Phys. Rev. D **81**, 083005 (2010).
[11] J. Chluba, G. M. Vasil, and L. J. Dursi, arXiv:1003.4928.
[12] D. Larson, J. Dunkley, G. Hinshaw, E. Komatsu, M. R. Nolta, C. L. Bennett, B. Gold, M. Halpern, R. S. Hill, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, L. Page, K. M. Smith, D. N. Spergel, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright, arXiv:1001.4635.
[13] W. A. Fendt, J. Chluba, J. A. Rubiño-Martín, and B. D. Wandelt, Astrophys. J. Suppl. Ser. **181**, 627 (2009).
[14] W. Y. Wong, A. Moss, and D. Scott, Mon. Not. R. Astron. Soc. **386**, 1023 (2008).
[15] U. Seljak and M. Zaldarriaga, Astrophys. J. **469**, 437 (1996).
[16] V. K. Dubrovich and S. I. Grachev, Astron. Lett. **31**, 359 (2005).
[17] J. Chluba and R. A. Sunyaev, Astron. Astrophys. **446**, 39 (2006).
[18] E. E. Kholupenko and A. V. Ivanchik, Astron. Lett. **32**, 795 (2006).
[19] W. Y. Wong and D. Scott, Mon. Not. R. Astron. Soc. **375**, 1441 (2007).
[20] J. Chluba and R. A. Sunyaev, Astron. Astrophys. **480**, 629 (2008).
[21] C. M. Hirata, Phys. Rev. D **78**, 023001 (2008).
[22] S. I. Grachev and V. K. Dubrovich, Astron. Lett. **34**, 439 (2008).
[23] C. M. Hirata and J. Forbes, Phys. Rev. D **80**, 023001 (2009).
[24] J. Chluba and R. A. Sunyaev, Astron. Astrophys. **496**, 619 (2009).
[25] J. Chluba and R. A. Sunyaev, Astron. Astrophys. **512**, A53 (2010).
[26] J. Chluba and R. A. Sunyaev, Astron. Astrophys. **503**, 345 (2009).
[27] E. R. Switzer and C. M. Hirata, Phys. Rev. D **77**, 083006 (2008).
[28] C. M. Hirata and E. R. Switzer, Phys. Rev. D **77**, 083007 (2008).
[29] E. R. Switzer and C. M. Hirata, Phys. Rev. D **77**, 083008 (2008).
[30] E. E. Kholupenko, A. V. Ivanchik, and D. A. Varshalovich, Mon. Not. R. Astron. Soc. **378**, L39 (2007).

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

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

[33] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Academic Press, New York, 1957).

[34] V. V. Sobolev (Harvard University Press, Cambridge, MA, 1960).

[35] J. Chluba and R. A. Sunyaev, Astron. Astrophys. **475**, 109 (2007).

[36] E. E. Kholupenko, A. V. Ivanchik, and D. A. Varshalovich, Phys. Rev. D **81**, 083004 (2010).

[37] S. P. Goldman, Phys. Rev. A **40**, 1185 (1989).

[38] C. Herńandez-Monteagudo, J. A. Rubiño-Martín, and R. A. Sunyaev, Mon. Not. R. Astron. Soc. **380**, 1656 (2007).

[39] M. J. Seaton, Mon. Not. R. Astron. Soc. **119**, 90 (1959).

[40] J. D. Hey, J. Phys. B **39**, 2641 (2006).

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