HOW EXACTLY DID THE UNIVERSE BECOME NEUTRAL?

SARA SEAGER[1] AND DIMITAR D. SASSELOV
Astronomy Department, Harvard University, 60 Garden Street, Cambridge, MA 02138

AND

DOUGLAS SCOTT
Department of Physics and Astronomy, University of British Columbia, Vancouver, BC V6T 1Z1, Canada
Received 1998 November 30; accepted 2000 January 25

ABSTRACT

We present a refined treatment of H, He I, and He II recombination in the early universe. The difference from previous calculations is that we use multilevel atoms and evolve the population of each level with redshift by including all bound-bound and bound-free transitions. In this framework we follow several hundred atomic energy levels for H, He I, and He II combined. The main improvements of this method over previous recombination calculations are (1) allowing excited atomic level populations to depart from an equilibrium distribution, (2) replacing the total recombination coefficient with recombination to and photoionization from each level calculated directly at each redshift step, and (3) correct treatment of the He I atom, including the triplet and singlet states.

We find that $x_e (\equiv n_e/n_H)$ is approximately 10% smaller at redshifts $\lesssim 800$ than in previous calculations, as a result of the nonequilibrium of the excited states of H that is caused by the strong but cool radiation field at those redshifts. In addition, we find that He I recombination is delayed compared with previous calculations and occurs only just before H recombination. These changes in turn can affect the predicted power spectrum of microwave anisotropies at the few percent level. Other improvements, such as including molecular and ionic species of H, including complete heating and cooling terms for the evolution of the matter temperature, including collisional rates, and including feedback of the secondary spectral distortions on the radiation field, produce negligible change to the ionization fraction. The lower x_e at low z found in this work affects the abundances of H molecular and ionic species by 10%–25%. However, this difference is probably not larger than other uncertainties in the reaction rates.

Subject headings: atomic processes — cosmic microwave background — cosmology: theory — early universe

1. INTRODUCTION

The photons that Penzias & Wilson (1965) detected coming from all directions with a temperature of about 3 K have traveled freely since their last Thomson scattering, when the universe became cool enough for the ions and electrons to form neutral atoms. During this recombination epoch, the opacity dropped precipitously, matter and radiation were decoupled, and the anisotropies of the cosmic microwave background (CMB) radiation were essentially frozen in. These anisotropies of the CMB have now been detected on a range of scales (e.g., White, Scott, & Silk 1994; Smoot & Scott 1998), and developments in the field have been very rapid. Recently, two post-*COBE* missions, the *Microwave Anisotropy Probe* (*MAP*) and the *Planck* satellite, were approved with the major science goal of determining the shape of the power spectrum of anisotropies with experimental precision at a level similar to current theoretical predictions.

Detailed understanding of the recombination process is crucial for modeling the power spectrum of CMB anisotropies. Since the seminal work of the late 1960s (Peebles 1968; Zeldovich et al. 1968), several refinements have been introduced by, for example, Matsuda, Sato, & Takeda (1971), Zabotin & Nasel'skii (1982), Lyubarsky & Sunyaev (1983), Jones & Wyse (1985), Krolik (1990), and others, but in fact little has changed (a fairly comprehensive overview of earlier work on recombination is to be found in § IIIC of Hu

et al. 1995, hereafter HSSW95). More recently, refinements have been made independently in the radiative transfer to calculate secondary spectral distortions (Dell'Antonio & Rybicki 1993; Rybicki & Dell'Antonio 1994) and in the chemistry (Stancil, Lepp, & Dalgarno 1996b). These improvements may have noticeable effects (at the 1% level) on the calculated shapes of the power spectrum of anisotropies. Given the potential to measure important cosmological parameters with *MAP* and *Planck* (e.g., Jungman et al. 1996; Bond, Efstathiou, & Tegmark 1997; Zaldarriaga, Spergel, & Seljak 1997; Eisenstein, Hu, & Tegmark 1998; Bond & Efstathiou 1998), it is of great interest to make a complete and detailed calculation of the process of recombination. Our view is that this is in principle such a simple process that it should be so well understood that it could never affect the parameter estimation endeavor.

Our motivation is to carry out a "modern" calculation of the cosmic recombination process. The physics is well understood, and so it is surprising that cosmologists have not moved much beyond the solution of a single ordinary differential equation, as introduced in the late 1960s. With today's computing power, there is no need to make the sweeping approximations that were implemented 30 years ago. We therefore attempt to calculate to as great an extent as possible the *full* recombination problem. The other difference compared with three decades ago is that we are now concerned with high-precision calculations because of the imminent prospect of high-fidelity data.

It is our intention to present here a coupled treatment of the nonequilibrium radiative transfer and the detailed

[1] Now at School of Natural Sciences, Institute for Advanced Study, Olden Lane, Princeton, NJ, 08540.

chemistry. The present investigation was motivated by indications (HSSW95) that multilevel nonequilibrium effects in H and He, as well as in some molecular species, may have measurable effects on the power spectrum of CMB anisotropies by affecting the low-z and high-z tails of the visibility function $e^{-\tau}d\tau/dz$ (where τ is the optical depth).

To that effect, this paper presents a study of the recombination era by evolving neutral and ionized species of H and He and molecular species of H simultaneously with the matter temperature. We believe our work represents the most accurate picture to date of how exactly the universe as a whole became neutral.

2. BASIC THEORY

2.1. *The Cosmological Picture*

We will assume that we live in a homogeneous, expanding universe within the context of the canonical hot big bang paradigm. The general picture is that at some sufficiently early time the universe can be regarded as an expanding plasma of hydrogen plus some helium, with around 10^9 photons per baryon and perhaps some nonbaryonic matter. As it expanded and cooled there came a time when the protons were able to keep hold of the electrons and the universe became neutral. This is the period of cosmic recombination.

In cosmology it is standard to use redshift z as a time coordinate, so that high redshift represents earlier times. Explicitly, $a(t) = 1/(1 + z)$ is the scale factor of the universe, normalized to be unity today, and with the relationship between scale factor and time depending on the particular cosmological model. It is sufficient to consider only hydrogen and helium recombination, since the other elements exist in minute amounts. The relevant range of redshift is then $\lesssim 10,000$, during which the densities are typical of those that astrophysicists deal with every day, and the temperatures are low enough that there are no relativistic effects.

2.2. *The Radiation Field*

In describing the radiation field and its interaction with matter, we must use a specific form of the radiative transfer equation. It should describe how radiation is absorbed, emitted, and scattered as it passes through matter in a medium that is homogeneous, isotropic, infinite, and expanding. The basic time-dependent form of the equation of transfer is

$$\frac{1}{c}\frac{\partial I(r, \hat{n}, v, t)}{\partial t} + \frac{\partial I(r, \hat{n}, v, t)}{\partial l} = j(r, \hat{n}, v, t)$$
$$- \kappa(r, \hat{n}, v, t)I(r, \hat{n}, v, t) . \quad (1)$$

Here the symbols are as follows: $I(r, \hat{n}, v, t)$ is the specific intensity of radiation at position r, traveling in direction $\hat{n}$ (the unit direction vector), with frequency v at a time t (in units of ergs s^{-1} cm^{-2} Hz^{-1} sr^{-1}); l is the path length along the ray (and is a coordinate-independent path length); j is the emissivity, which is calculated by summing products of upper excitation state populations and transition probabilities over all relevant processes that can release a photon at frequency v, including electron scattering; and κ is the absorption coefficient, which is the product of an atomic absorption cross section and the number density of absorbers summed over all states that can interact with photons of frequency v.

In the homogeneous and isotropic medium of the early universe, we can integrate equation (1) over all solid angles ω (i.e., integrating over the unit direction vector $\hat{n}$):

$$\frac{4\pi}{c}\frac{\partial J(r, v, t)}{\partial t} + \nabla \cdot F(r, v, t) = -\oint [j(r, \hat{n}, v, t)$$
$$- \kappa(r, \hat{n}, v, t)I(r, \hat{n}, v, t)]d\omega . \quad (2)$$

Here $J(v, t)$ is the mean intensity, the zeroth-order moment of the specific intensity over all angles (in units of ergs s^{-1} cm^{-2} Hz^{-1}); F is the flux of radiation, which is the net rate of radiant energy flow across an arbitrarily oriented surface per unit time and frequency interval; and c is the speed of light. If the radiation field is isotropic, there is a ray-by-ray cancellation in the net energy transport across a surface, and the net flux is zero. Also, because of the isotropy of the radiation field and the medium being static, we can drop the dependence upon angle of j and κ in equation (2). With the definition of J, $\oint I d\omega = 4\pi J$, this simplifies equation (2) to

$$\frac{1}{c}\frac{\partial J(v, t)}{\partial t} = j(v, t) - \kappa(v, t)J(v, t) . \quad (3)$$

The above equation is for a static medium. An isotropically expanding medium would reduce the number density of photons as a result of the expanding volume and reduce their frequencies as a result of redshifting. The term resulting from the density change will be simply a $3\dot{a}(t)/a(t)$ factor, while the redshifting term will involve the frequency derivative of $J(v, t)$ and hence a $v\dot{a}(t)/a(t)$ factor.

Then the equation for the evolution of the radiation field as affected by the expansion and the sources and sinks of radiation becomes

$$\frac{dJ(v, t)}{dt} = \frac{\partial J(v, t)}{\partial t} - vH(t)\frac{\partial J(v, t)}{\partial v}$$
$$= -3H(t)J(v, t) + c[j(v, t) - \kappa(v, t)J(v, t)] , \quad (4)$$

where $H(t) \equiv \dot{a}/a$.

This equation is in its most general form and difficult to solve; fortunately, we can make two significant simplifications because the primary spectral distortions[2] are of negligible intensity (Dell'Antonio & Rybicki 1993) and the quasi-static solution for spectral-line profiles is valid (Rybicki & Dell'Antonio 1994). The first simplification is that, for the purposes of this paper (in which we do not study secondary spectral distortions), we set $J(v, t) = B(v, t)$, the Planck function, which is observed to approximate $J(v)$ to at least 1 part in 10^4 (Fixsen et al. 1996). Thus, we eliminate explicit frequency integration from the simultaneous integration of all equations (§ 2.6). The validity of this assumption is shown in § 3.5, where we follow the dominant secondary distortions of H Lyα and H two-photon transition by including their feedback on the recombination process and find that the secondary spectral distortions from the other Lyman lines and He are not strong enough to feed back on the recombination process.

The second simplification is in the treatment of the evolution of the resonance lines (Lyα, etc.), which must still be treated explicitly; because of cosmological redshifting they cause $J(v, t) \neq B(v, t)$ in the lines. These we call the primary distortions. We use escape probability methods for moving (expanding) media (§§ 2.3.3 and 3.1). This simplification is

---

[2] Not to be confused here with the power spectrum of spatial anisotropies.

not an approximation but is an exact treatment—a simple solution to the multilevel radiative transfer problem afforded by the physics of the expanding early universe.

Not only does using $B(v, t)$ with the escape probability method instead of $J(v, t)$ simplify the calculation and reduce computing time enormously, but the effects from following the actual radiation field will be small compared to the main improvements of our recombination calculation, which are the level-by-level treatment of H, He I, and He II, calculating recombination directly, and the correct treatment of He I triplet and singlet states.

2.3. The Rate Equations

The species we evolve in the expanding universe are H I, H II, He I, He II, He III, e^-, H^-, H_2, and H_2^+. The chemistry of the early universe involves the reactions of association and dissociation among these species, facilitated by interactions with the radiation field, $J(v, t)$. The rate equations for an atomic system with N energy levels can be described as

$$a(t)^{-3} \frac{d[n_i(t)a(t)^3]}{dt} = [n_e(t)n_c(t)P_{ci} - n_i(t)P_{ic}]$$
$$+ \sum_{j=1}^{N} [n_j(t)P_{ji} - n_i(t)P_{ij}] , \quad (5)$$

where the P_{ij} are the rate coefficients between bound levels i and j, and the P_{ic} are the rate coefficients between bound levels and the continuum c; $P_{ij} = R_{ij} + n_e C_{ij}$ and $P_{ic} = R_{ic} + n_e C_{ic}$, where R refers to radiative rates and C to collisional rates. Here the values of n are physical (as opposed to comoving) number densities: n_i refers to the number density of the ith excited atomic state, n_e to the number density of electrons, and n_c to the number density of a continuum particle such as a proton, He II, or He III; $a(t)$ is the cosmological scale factor. The rate equations for molecules take a slightly different form because their formation and destruction depend on the rate coefficients for the reactions discussed in § 2.3.4, and molecular bound states are not included.

2.3.1. Photoionization and Photorecombination

By calculating photorecombination rates R_{ci} directly to each level for multilevel H, He I, and He II atoms, we avoid the problem of finding an accurate recombination coefficient, the choice of which has a large effect on the power spectra (HSSW95).

Photoionization rates are calculated by integrals of the incident radiation field $J(v, t)$ and the bound-free cross section $\sigma_{ic}(v)$. The photoionization rate in s^{-1} is

$$R_{ic} = 4\pi \int_{v_0}^{\infty} \frac{\sigma_{ic}(v)}{h_P v} J(v, t)dv . \quad (6)$$

Here i refers to the ith excited state and c refers to the continuum; v_0 is the threshold frequency for ionization from the ith excited state. The radiation field $J(v, t)$ depends on frequency v and time t. With n_i as the number density of the ith excited state, the number of photoionizations per unit volume per unit time (hereafter photoionization rate) is $n_i R_{ic}$.

By using the principle of detailed balance in the case of local thermodynamic equilibrium (LTE), the radiative recombination rate can be calculated from the photoionization rate. Then, as described below, the photorecombination rate can be generalized to the non-LTE case by scaling

the LTE populations with the actual populations and substituting the actual radiation field for the LTE radiation field. In LTE, the radiation field $J(v, t)$ is the Planck function $B(v, t)$. $B(v, t)$ is a function of time t during recombination because $T_R = 2.728[1 + z(t)]$. We will call the LTE temperature T ($T = T_R = T_M$ at early times, where T_R is the radiation temperature and T_M the matter temperature). To emphasize the Planck function's dependence on temperature, we will use $B(v, T)$, where T is a function of time.

By detailed balance in LTE we have

$$(n_e n_c R_{ci})^{LTE} = (n_i R_{ic})^{LTE} .$$

Radiative recombination includes spontaneous and stimulated recombination, so we must rewrite the above equation as

$$(n_e n_c R_{ci})^{LTE} = (n_e n_c R_{ci}^{spon})^{LTE} + (n_e n_c R_{ci}^{stim})^{LTE}$$
$$= [(n_i R_{ic})^{LTE} - (n_i R_{ic}^{stim})^{LTE}]$$
$$+ (n_i R_{ic}^{stim})^{LTE} . \quad (7)$$

Using the definition of R_{ic} in equation (6),

$$(n_e n_c R_{ci})^{LTE} = 4\pi n_i^{LTE} \int_{v_0}^{\infty} \frac{\sigma_{ic}(v)}{h_P v}$$
$$\times B(v, T)(1 - e^{-h_P v/k_B T})dv$$
$$+ 4\pi n_i^{LTE} \int_{v_0}^{\infty} \frac{\sigma_{ic}(v)}{h_P v} B(v, T)e^{-h_P v/k_B T} dv . \quad (8)$$

The first term on the right-hand side is the spontaneous recombination rate, and the second term on the right-hand side is the stimulated recombination rate. Here h_P is Planck's constant and k_B is Boltzmann's constant. The factor $(1 - e^{-h_P v/k_B T})$ is the correction for stimulated recombination (see Mihalas 1978, § 4.3, for a derivation of this factor). Stimulated recombination can be treated either as negative ionization or as positive recombination; the physics is the same (see Seager & Sasselov 2000 for some subtleties). With the LTE expression for recombination (eq. [8]), it is easy to generalize to the non-LTE case, considering spontaneous and stimulated recombination separately. Because the matter temperature T_M and the radiation temperature T_R differ at low z, it is important to understand how recombination depends on each of these separately.

Spontaneous recombination involves a free electron, but its calculation requires no knowledge of the local radiation field because the photon energy is derived from the electron's kinetic energy. In other words, whether or not LTE is valid, the LTE spontaneous recombination rate holds per ion, as long as the velocity distribution is Maxwellian. The local Planck function (as representing the Maxwell distribution) depends on T_M because the Maxwell distribution describes a collisional process. Furthermore, since the Maxwellian distribution depends on T_M, so does the spontaneous rate. To get the non-LTE rate, we only have to rescale the LTE ion density to the actual ion density:

$$n_e n_c R_{ci}^{spon} = 4\pi \frac{n_e n_c}{(n_e n_c)^{LTE}} n_i^{LTE} \int_{v_0}^{\infty} \frac{\sigma_{ic}(v)}{h_P v}$$
$$\times B(v, T_M)(1 - e^{-h_P v/k_B T_M})dv \quad (9)$$
$$= 4\pi n_e n_c \left(\frac{n_i}{n_e n_c}\right)^{LTE} \int_{v_0}^{\infty} \frac{\sigma_{ic}(v)}{h_P v} \frac{2h_P v^3}{c^2}$$
$$\times e^{-h_P v/k_B T_M} dv . \quad (10)$$

To generalize the stimulated recombination rate from the LTE rate to the non-LTE rate, we rescale the LTE ion density to the actual ion density and replace the LTE radiation field by the actual radiation field $J(v, t)$ because that is what is "stimulating" the recombination. The correction for stimulated recombination depends on T_M because the recombination process is collisional; the term always remains in the LTE form because equation (8) was derived from detailed balance, so we have

$$n_e n_c R_{ci}^{\text{stim}} = 4\pi \frac{n_e n_c}{(n_e n_c)^{\text{LTE}}} n_i^{\text{LTE}} \int_{v_0}^{\infty} \frac{\sigma_{ic}(v)}{h_P v} J(v, t) e^{-h_P v/k_B T_M} dv .$$

(11)

Therefore, the total non-LTE recombination rate ($R_{ci}^{\text{spon}} + R_{ci}^{\text{stim}}$) is

$$n_e n_c R_{ci} = n_e n_c \left(\frac{n_i}{n_e n_c}\right)^{\text{LTE}} 4\pi \int_{v_0}^{\infty} \frac{\sigma_{ic}(v)}{h_P v}$$
$$\times \left[\frac{2 h_P v^3}{c^2} + J(v, t)\right] e^{-h_P v/k_B T_M} dv . \quad (12)$$

The LTE population ratios $(n_i/n_e n_c)^{\text{LTE}}$ depend only on T_M through the Saha relation:

$$\left(\frac{n_i}{n_e n_c}\right)^{\text{LTE}} = \left(\frac{h^2}{2\pi m_e k_B T_M}\right)^{3/2} \frac{g_i}{2 g_c} e^{E_i/k_B T_M} . \quad (13)$$

Here m_e is the electron mass, the atomic parameter g is the degeneracy of the energy level, and E_i is the ionization energy of level i. In the recombination calculation presented in this paper, we use the Planck function $B(v, T_R)$ instead of the radiation field $J(v, T)$ as described earlier.

In the early universe Case B recombination is used. This excludes recombinations to the ground state and considers the Lyman lines to be optically thick. An implied assumption necessary to compute the photoionization rate is that the excited states ($n \geq 2$) are in equilibrium with the radiation. Our approach is more general than Case B because we do not consider the Lyman lines to be optically thick and do not assume equilibrium among the excited states. For more details on the validity of Case B recombination, see § 3.2.5. To get the total recombination coefficient, we sum over captures to all excited levels above the ground state.

To summarize, the form of the total photoionization rate is

$$\sum_{i>1}^{N} n_i R_{ic} = \sum_{i>1}^{N} n_i 4\pi \int_{v_0}^{\infty} \frac{\sigma_{ic}(v)}{h_P v} B(v, T_R) dv , \quad (14)$$

and the total recombination rate is

$$\sum_{i>1}^{N} n_e n_c R_{ci} = n_e n_c \sum_{i>1}^{N} \left(\frac{n_i}{n_e n_c}\right)^{\text{LTE}} 4\pi \int_{v_0}^{\infty} \frac{\sigma_{ic}(v)}{h_P v}$$
$$\times \left(\frac{2 h_P v^3}{c^2} + B(v, T_R)\right) e^{-h_P v/k_B T_M} dv . \quad (15)$$

2.3.2. Comparison with the "STANDARD" Recombination Calculation of Hydrogen

The "standard" recombination calculation refers to the calculation widely used today and first derived by Peebles (1968, 1993), Zeldovich, Kurt, & Sunyaev (1968), and Zeldovich & Novikov (1983), updated with the most recent

parameters and recombination coefficient (HSSW95) (see also § 3.1).

For a 300-level H atom in our new recombination calculation, equations (14) and (15) include 300 integrals at each redshift step. The standard recombination calculation does not go through this time-consuming task but avoids it entirely by using a precalculated recombination coefficient that is a single expression dependent on T_M only. The recombination coefficient to each excited state i is defined by

$$\alpha_i(T_M) = R_{ci}^{\text{spon}} . \quad (16)$$

Here α is a function of T_M because spontaneous recombination is a collisional process, as described previously. The total Case B recombination coefficient (α_B) is obtained from

$$\alpha_B(T_M) = \sum_{i>1}^{N} \alpha_i(T_M) . \quad (17)$$

We will refer to $\alpha_B(T_M)$ as the "precalculated recombination coefficient" because the recombination to each atomic level i and the summation over i are precalculated for LTE conditions (see Hummer 1994 for an example of how these recombination coefficients are calculated). Some more elaborate derivations of $\alpha(T_M) = f(T_M, n)$ have also been tried (e.g., Boschan & Biltzinger 1998).

The standard recombination calculation uses a photoionization coefficient $\beta_B(T_M)$, which is derived from detailed balance using the recombination rate

$$(n_i \beta_i)^{\text{LTE}} = (n_e n_p \alpha_i)^{\text{LTE}} . \quad (18)$$

To get the non-LTE rate, one uses the actual populations n_i,

$$n_i \beta_i(T_M) = n_i \left(\frac{n_e n_p}{n_i}\right)^{\text{LTE}} \alpha_i(T_M) , \quad (19)$$

or, with the Saha relation (eq. [13]),

$$n_i \beta_i(T_M) = n_i \left(\frac{2\pi m_e k_B T_M}{h_P^2}\right)^{3/2} \frac{2}{g_i} e^{-E_i/k_B T_M} \alpha_i(T_M) . \quad (20)$$

Constants and variables are as described before. To get the total photoionization rate, β_i is summed over all excited levels. Because the "standard" calculation avoids use of all levels i explicitly, the n_i are assumed to be in equilibrium with the radiation and thus can be related to the first excited state number density n_{2s} by the Boltzmann relation,

$$n_i = n_{2s} \frac{g_i}{g_{2s}} e^{-(E_2 - E_i)/k_B T_M} . \quad (21)$$

With this relation, the total photoionization rate is

$$\sum_{i>1}^{N} n_i \beta_i = n_{2s} \alpha_B e^{-E_{2s}/k_B T_M} \left(\frac{2\pi m_e k_B T_M}{h_P^2}\right)^{3/2} \equiv n_{2s} \beta_B . \quad (22)$$

In this expression for the total photoionization rate, the excited states are populated according to a Boltzmann distribution. T_M is used instead of T_R because the Saha and Boltzmann equilibrium used in the derivation are collisional processes. The expression says nothing about the excited levels being in equilibrium with the continuum because the actual values of n_e, n_1, and n_{2s} are used, and the n_i are proportional to n_{2s}.

To summarize, the standard calculation uses a single expression for each of the total recombination rate and the total photoionization rate that is dependent on T_M only.

The total photoionization rate is

$$\sum_{i>1}^{N} n_i R_{ic} = n_{2s} \beta_B(T_M)$$

$$= n_{2s} \alpha_B(T_M) e^{-E_{2s}/k_B T_M} (2\pi m_e k_B T_M)^{3/2}/h_P^3 , \quad (23)$$

and the total recombination rate is

$$\sum_{i>1}^{N} n_e n_p R_{ci} = n_e n_p \alpha_B(T_M) . \quad (24)$$

Comparing the right-hand side of equation (23) to our level-by-level total photoionization rate (eq. [14]), the main improvement in our method over the standard one is clear: we use the actual excited level populations n_i, *assuming no equilibrium distribution among them*. In this way we can test the validity of the equilibrium assumption. Far less important is that the standard recombination treatment cannot distinguish between T_R and T_M, even though photoionization and stimulated recombination are functions of T_R, while spontaneous recombination is a function of T_M, as shown in equations (14) and (15). The nonequilibrium of excited states is important at the 10% level in the residual ionization fraction for $z \lesssim 800$, while using T_R in photoionization and photoexcitation is only important at the few percent level for $z \lesssim 300$ (for typical cosmological models). Note that although the precalculated recombination coefficient includes spontaneous recombination only, stimulated recombination (as a function of T_M) is still included as negative photoionization via detailed balance (see eq. [7]).

2.3.3. *Photoexcitation*

In the expanding universe, redshifting of the photons must be taken into account (see eq. [4]). Line photons emitted at one position may be redshifted out of interaction frequency (redshifted more than the width of the line) by the time they reach another position in the flow. We use the Sobolev escape probability to account for this, a method which was first used for the expanding universe by Dell'Antonio and Rybicki (1993). The Sobolev escape probability (Sobolev 1946), also sometimes called the large velocity gradient approximation, is not an approximation but is an exact, simple solution to the multilevel radiative transfer in the case of a large velocity gradient. It is this solution that allows the explicit inclusion of the line distortions to the radiation field—without it our detailed approach to the recombination problem would be intractable. We will call the net bound-bound rate for each line transition ΔR_{ji}, where j is the upper level and i is the lower level:

$$\Delta R_{ji} = p_{ij}\{n_j[A_{ji} + B_{ji} B(v_{ij}, t)] - n_i B_{ij} B(v_{ij}, t)\} . \quad (25)$$

Here the terms A_{ji}, B_{ji}, and B_{ij} are the Einstein coefficients; the escape probability p_{ij} is the probability that photons associated with this transition will "escape" without being further scattered or absorbed. If $p_{ij} = 1$, the photons produced in the line transition escape to infinity—they contribute no distortion to the radiation field. If $p_{ij} = 0$, no photons escape to infinity; all of them get reabsorbed, and the line is optically thick. This is the case of primary distortions to the radiation field, and the Planck function cannot be used for the line radiation. In general, $p_{ij} \ll 1$ for the Lyman lines and $p_{ij} = 1$ for all other line transitions. With this method we have described the redshifting of photons

through the resonance lines and found a simple solution to the radiative transfer problem for all bound-bound transitions. The rest of this section is devoted to deriving p_{ij}.

For the case of no cosmological redshifting, the radiative rates per cubic centimeter for transitions between excited states of an atom are

$$n_i R_{ij} = n_i B_{ij} \overline{J} \quad (26)$$

and

$$n_j R_{ji} = n_j A_{ji} + n_j B_{ji} \overline{J} , \quad (27)$$

where

$$\overline{J} = \int_0^{\infty} J(v, t)\phi(v)dv \quad (28)$$

and $\phi(v)$ is the line profile function with its area normalized by

$$\int_0^{\infty} \phi(v)dv = 1 . \quad (29)$$

The line profile $\phi(v)$ is taken to be a Voigt function that includes natural and Doppler broadening. In principle, equation (28) is the correct approach. In practice we take $\phi(v)$ as a delta function and use $J(v, t)$ instead of $\overline{J}$. The smooth radiation field is essentially constant over the width of the line, and so the line shape is not important; we get the same results using $\overline{J}$ or $J(v, t)$.

The Sobolev escape probability considers the distance over which the expansion of the medium induces a velocity difference equal to the thermal velocity (for the case of a Doppler width): $L = v_{th}/|v'|$, where v_{th} is the thermal velocity width and v' the velocity gradient. The theory is valid when this distance L is much smaller than typical scales of macroscopic variation of other quantities.

We follow Rybicki (1984) in the derivation of the Sobolev escape probability. The general definition of escape probability is given by the exponential extinction law,

$$p_{ij} = \exp\left[-\tau(v_{ij})\right] , \quad (30)$$

where v_{ij} is the frequency for a given line transition, and $\tau(v_{ij})$ is the monochromatic optical depth forward along a ray from a given point to the boundary of the medium. Here $\tau(v_{ij})$ is defined by

$$d\tau(v_{ij}) = -\tilde{k}\phi(v_{ij})dl , \quad (31)$$

where $\tilde{k}$ is the integrated line absorption coefficient, so that the monochromatic absorption coefficient or opacity is $\kappa = \tilde{k}\phi(v_{ij})$, and l is the distance along the ray from the emission point ($l = 0$). Rewriting the optical depth for a line profile function (which has units of inverse frequency) of the dimensionless frequency variable $x = (v - v_{ij})/\Delta$, with Δ the width of the line in Doppler units and v_{ij} the central line frequency, we have

$$d\tau(v_{ij}) = -\frac{\tilde{k}}{\Delta} \phi(x)dl . \quad (32)$$

Here κ is the absorption coefficient (defined in eq. [1]), with $\tilde{k}$ just dividing out the line profile function, and

$$\tilde{k} = \frac{h_P v}{4\pi} (n_i B_{ij} - n_j B_{ji}) . \quad (33)$$

Using the Einstein relations $g_i B_{ij} = g_j B_{ji}$ and $A_{ji} = (2h\nu^3/c^2)B_{ji}$, the absorption coefficient can also be written as

$$\tilde{k} = \frac{A_{ji}\lambda_{ij}^2}{8\pi}\left(n_i\frac{g_j}{g_i} - n_j\right). \qquad (34)$$

Note that distances along a ray l correspond to shifts in frequency x. This is due to the Doppler effect induced by the velocity gradient and is the essence of the Sobolev escape probability approach. For example, the Lyα photons cannot be reabsorbed in the Lyα line if they redshift out of the frequency interaction range. This case will happen at some frequency x or at some distance l from the photon emission point, where, because of the expansion, the photons have redshifted out of the frequency interaction range. For an expanding medium with a constant velocity gradient $v' = dv/dl$, the escape probability along a ray is then

$$p_{ij} = \exp\left[-\frac{\tilde{k}}{\Delta}\int_0^\infty \phi(x - l/L)dl\right]$$
$$\equiv \exp\left[-\tau_S\int_{-\infty}^x \phi(\tilde{x})d\tilde{x}\right]. \qquad (35)$$

The velocity field has in effect introduced an intrinsic escape mechanism for photons; beyond the interaction limit with a given atomic transition, the photons can no longer be absorbed by the material, even if it is of infinite extent, but escape freely to infinity (Mihalas 1978). Here the Sobolev optical thickness along the ray is defined by

$$\tau_S \equiv \frac{\tilde{k}}{\Delta}L, \qquad (36)$$

where L is the Sobolev length

$$L = v_{th}/|v'| = \sqrt{\frac{3k_B T_M}{m_{atom}}}\Big/|v'| \qquad (37)$$

and Δ is the width of the line, which in the case of Doppler broadening is

$$\Delta = \frac{\nu_0}{c}\sqrt{\frac{3k_B T_M}{m_{atom}}}. \qquad (38)$$

With these definitions, equation (36) becomes

$$\tau_S = \frac{\lambda_{ij}\tilde{k}}{|v'|}. \qquad (39)$$

In the expanding universe, the velocity gradient v' is given by the Hubble expansion rate $H(z)$, and using the above definition for $\tilde{k}$,

$$\tau_S = \frac{A_{ji}\lambda_{ij}^3[n_i(g_j/g_i) - n_j]}{8\pi H(z)}. \qquad (40)$$

To find the Sobolev escape probability for the ray, we average over the initial frequencies x, using the line profile function $\phi(x)$ from equation (29),

$$p_{ij} = \int_{-\infty}^\infty dx\, \phi(x)\exp\left[-\tau_S\int_{-\infty}^x \phi(\tilde{x})d\tilde{x}\right]$$
$$= \int_0^1 d\zeta\, \exp\left(-\tau_S\zeta\right),$$

i.e.,

$$p_{ij} = \frac{1 - \exp(-\tau_S)}{\tau_S}. \qquad (41)$$

Note that this expression is independent of the line profile shape $\phi(x)$. The escape probability p_{ij} is defined as a frequency average at a single point. Finally, we must average over angle, but in the case of the isotropically expanding universe, the angle-averaged Sobolev escape probability takes the same form as p_{ij} above. (For further details on the Sobolev escape probability, see Rybicki 1984 or Mihalas 1978, § 14.2.)

How does the Sobolev optical depth relate to the usual meaning of optical depth? The optical depth for a specific line at a specific redshift point is equivalent to the Sobolev optical depth,

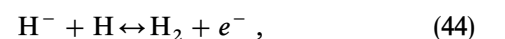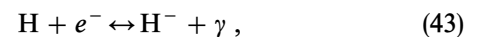$$d\tau(v_{ij}) = -\tau_S\,\phi(x)\frac{dl}{L}. \qquad (42)$$

If no other line or continuum photons are redshifted into that frequency range before or after the redshift point, then the optical depth at a given frequency today (i.e., summed over all redshift points) will be equivalent to the Sobolev optical depth at that past point. Generally, behaviors in frequency and space are interchangeable in a medium with a velocity gradient.

In order to derive an expression for the bound-bound rate equations, we must consider the mean radiation field $\bar{J}$ in the line. For the case of spectral distortions, $\bar{J}$ does not equal the Planck function at the line frequency. We use the core saturation method (Rybicki 1984) to get $\bar{J}$ using p_{ij}; from this we get the net rate of de-excitations in that transition ($j \to i$), given in equation (25). In general, only the Lyman lines of H and He II and the He I n^1p–1^1s lines have $p_{ij} < 1$. With this solution we have accomplished two things: (1) described the redshifting of photons through the resonance lines, and (2) found a simple solution to the radiative transfer problem for all bound-bound lines.
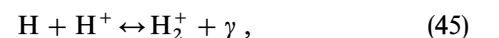
Peculiar velocities during the recombination era may cause line broadening of the same order of magnitude as thermal broadening over certain scales (A. Loeb 1998, private communication). Because we use $J(v) = B(v, T)$, the radiation field is essentially constant over the width of the line, and so the line shape is not important. Similarly, the peculiar velocities will not affect the Sobolev escape probability because it is independent of line shape (eq. [41]). If peculiar velocities were angle dependent, there would be an effect on the escape probability, which is an angle-averaged function. Only for computing spectral distortions to the CMB, where the line shape is important, should line broadening from peculiar velocities be included.
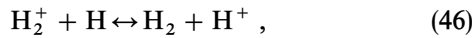
2.3.4. Chemistry

Hydrogen molecular chemistry has been included because it may affect the residual electron densities at low redshift ($z < 200$). During the recombination epoch, the H_2 formation reactions include the H^- processes

$$H + e^- \leftrightarrow H^- + \gamma, \qquad (43)$$

$$H^- + H \leftrightarrow H_2 + e^-, \qquad (44)$$

and the H_2^+ processes

$$H + H^+ \leftrightarrow H_2^+ + \gamma, \qquad (45)$$

$$H_2^+ + H \leftrightarrow H_2 + H^+ , \tag{46}$$

together with

$$H + H_2 \leftrightarrow H + H + H \tag{47}$$

and

$$H_2 + e^- \leftrightarrow H + H + e^- \tag{48}$$
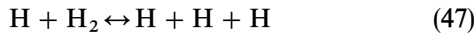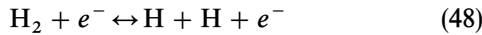
(see Lepp & Shull 1984). The direct three-body process for H_2 formation is significant only at much higher densities. The most recent rate coefficients and cross sections are listed with their references in the Appendix (see also Cen 1992; Puy et al. 1993; Tegmark et al. 1997; Abel et al. 1997; Galli & Palla 1998).

We have not included the molecular chemistry of Li, He, or D. In general, molecular chemistry only becomes important at values of $z < 200$. Recent detailed analyses of H, D, and He chemistry (Stancil et al. 1996a) and of Li and H chemistry (Stancil, Lepp, & Dalgarno 1996b; Stancil & Dalgarno 1997) presented improved relative abundances of all atomic, ionic, and molecular species. The values are certainly too small to have any significant effect on the CMB power spectrum.

There are two reasons for this. The visibility function is $e^{-\tau} d\tau/dz$, where τ is the optical depth. The main component of the optical depth is Thomson scattering by free electrons. Because the populations of Li, LiH, HeH^+, HD, and the other species are so small relative to the electron density, they do not affect the contributions from Thomson scattering. Second, these atomic species and their molecules themselves make no contribution to the visibility function because they have no strong opacities. HD has only a weak dipole moment. And while LiH has a very strong dipole moment, its opacity during this epoch is expected to be negligible because of its tiny ($<10^{-18}$) fractional abundance (Stancil et al. 1996b). Similarly, the fractional abundance of H_2D^+ ($<10^{-22}$) is too small to have an effect on the CMB spectrum (Stancil et al. 1996b). (For more details, see Palla et al. 1995 and Galli & Palla 1998.) An interesting additional point is that, because of the smaller energy gap between $n = 2$ and the continuum in it, Li I actually recombines at a slightly lower redshift than hydrogen does (see, e.g., Galli & Palla 1998). Of course this has no significant cosmological effects.

We have also excluded atomic D from the calculation. D, like H, has an atomic opacity much lower than Thomson scattering for the recombination era conditions. D parallels H in its reactions with electrons and protons and recombines in the same way and at the same time as H (see Stancil et al. 1996b). Although the abundance of D is small ([D/H] $\simeq 10^{-5}$), its Lyman photons are still trapped because they are shared with hydrogen. This is seen, for example, by the ratio of the isotopic shift of D Lyα to the width of the H Lyα line, on the order of 10^{-2}. Therefore, by excluding D we expect no change in the ionization fraction and hence none in the visibility function.

While the nonhydrogen chemistry is still extremely important for cooling and triggering the collapse of primordial gas clouds, it is not relevant for CMB power spectrum observations at the level measurable by *MAP* and *Planck*.

2.4. *Expansion of the Universe*

The differential equations in time for the number densities and matter temperature must be converted to differential equations in redshift by multiplying by a factor of dz/dt. The redshift z is related to time by the expression

$$\frac{dz}{dt} = -(1 + z)H(z) , \tag{49}$$

with scale factor

$$a(t) = \frac{1}{1 + z} . \tag{50}$$

Here $H(z) = \dot{a}/a$ is the Hubble factor

$$H(z)^2 = H_0^2 \left[\frac{\Omega_0}{1 + z_{eq}} (1 + z)^4 + \Omega_0 (1 + z)^3 \right.$$
$$\left. + \Omega_K (1 + z)^2 + \Omega_\Lambda \right], \tag{51}$$

where Ω_0 is the density contribution, Ω_K is the curvature contribution, and Ω_Λ is the contribution associated with the cosmological constant, with $\Omega_0 + \Omega_R + \Omega_K + \Omega_\Lambda = 1$ and $\Omega_R = \Omega_0/(1 + z_{eq})$. Here z_{eq} is the redshift of matter-radiation equality,

$$1 + z_{eq} = \Omega_0 \frac{3(H_0 c)^2}{8\pi G(1 + f_\nu)U} , \tag{52}$$

where f_ν is the neutrino contribution to the energy density in relativistic species ($f_\nu \simeq 0.68$ for three massless neutrino types), G the gravitational constant, U the photon energy density, and H_0 the Hubble constant today, which will be written as $100 \, h$ km s^{-1} Mpc^{-1}. Since we are interested in redshifts $z \sim z_{eq}$, it is crucial to include the radiation contribution explicitly (see HSSW95).

2.5. *Matter Temperature*

The important processes that are considered in following the matter temperature are Compton cooling, adiabatic cooling, and bremsstrahlung cooling. Less important but also included are photoionization heating, photorecombination cooling, radiative and collisional line cooling, collisional ionization cooling, and collisional recombination cooling. Note that throughout the relevant time period, collisions and Coulomb scattering hold all the matter species at very nearly the same temperature. "Matter" here means protons (and other nuclei), plus electrons, plus neutral atoms; dark matter is assumed to be decoupled.

Compton cooling is a major source of energy transfer between electrons and photons. It is described by the rate of transfer of energy per unit volume between photons and free electrons when the electrons are near thermal equilibrium with the photons:

$$\frac{dE_{e,\gamma}}{dt} = \frac{4\sigma_T U n_e k_B}{m_e c} (T_R - T_M) , \tag{53}$$

or

$$\frac{dT_M}{dt} = \frac{8}{3} \frac{\sigma_T U n_e}{m_e c n_{tot}} (T_R - T_M) \tag{54}$$

(Weymann 1965), where $E_{e,\gamma}$ is the electron energy density, k_B, m_e, and c are constants as before, σ_T is the Thomson scattering cross section, T_R is the radiation temperature, and T_M is the electron or matter temperature. To get from equation (53) to equation (54) we use the energy of all particles; collisions among all particles keep them at the same tem-

perature. Here n_{tot} represents the total number density of particles, which includes all of the species mentioned in § 2.3, while U represents the radiation energy density (integrated over all frequencies) in units of ergs cm^{-3}:

$$U = \int_0^\infty u(v)dv , \qquad (55)$$

where $u(v, t) = 4\pi J(v, t)/c$. In thermal equilibrium the radiation field has a frequency distribution given by the Planck function, $J(v, t) = B(v, T_R)$, and thus in thermal equilibrium the energy density is

$$u(v, t) = \frac{4\pi}{c} B_v(T_R) , \qquad (56)$$

and the total energy density U is given by Stefan's law

$$U = \frac{8\pi h_P}{c^3} \int_0^\infty (e^{h_P v/k_B T_R} - 1)^{-1} v^3 \, dv = a_R T_R^4 . \qquad (57)$$

The spectrum of the CMB remains close to blackbody because the heat capacity of the radiation is very much larger than that of the matter (Peebles 1993), i.e., there are vastly more photons than baryons.

Adiabatic cooling as a result of the expansion of the universe is described by

$$\frac{dT_M}{dt} = -2H(t)T_M \qquad (58)$$

since $\gamma = 5/3$ for an ideal gas implies $T_M \propto (1 + z)^2$. The following cooling and heating processes are often represented by approximate expressions. We used the exact forms, with the exception of bremsstrahlung cooling and the negligible collisional cooling: (1) bremsstrahlung, or free-free cooling:

$$\Lambda_{\text{brem}} = \frac{2^5 \pi e^6 Z^2}{3^{3/2} h_P m_e c^3} \left(\frac{2\pi k_B T}{m_e}\right)^{1/2}$$
$$\times g_{\text{ff}} n_e(n_p + n_{\text{He II}} + 4n_{\text{He III}}) , \qquad (59)$$

where g_{ff} is the free-free Gaunt factor (Seaton 1960), n_p the number density of protons, $n_{\text{He II}}$ and $n_{\text{He III}}$ the number density of singly and doubly ionized helium, respectively, and other symbols are as previously described; (2) photoionization heating:

$$\Pi_{p-i} = \sum_{i=1}^N n_i 4\pi \int_{v_0}^\infty \frac{\sigma_{ic}(v)}{h_P v} B(v, T_R)h_P(v - v_0)dv ; \qquad (60)$$

(3) photorecombination cooling:

$$\Lambda_{p-r} = \sum_{i=1}^N n_e n_c \left(\frac{n_i}{n_e n_c}\right)^{\text{LTE}} 4\pi \int_{v_i}^\infty \frac{\sigma_i(v)}{h_P v}$$
$$\times \left[\frac{2h_P v^3}{c^2} + B(v, T_R)\right] e^{-h_P v/k_B T_M} h_P(v - v_0)dv ; \qquad (61)$$

(4) line cooling:

$$\Lambda_{\text{line}} = h_P v_0 \Delta R_{ji} ; \qquad (62)$$

(5) collisional ionization cooling:

$$\Lambda_{c-i} = h_P v_0 C_{ic} ; \qquad (63)$$

and (6) collisional recombination heating:

$$\Lambda_{c-r} = h_P v_0 C_{ci} . \qquad (64)$$

Here v_0 is the frequency at the ionization edge. We used approximations for collisional ionization and recombination cooling because these collisional processes are essentially negligible during the recombination era. C_{ic} and C_{ci} are the collisional ionization and recombination rates, respectively, computed as in, e.g., Mihalas (1978), § 5.4.

Thus, with

$$\frac{dT_M}{dz} = \frac{dt}{dz} \frac{dT_M}{dt} , \qquad (65)$$

the total rate of change of matter temperature with respect to redshift becomes

$$(1 + z)\frac{dT_M}{dz} = \frac{8\sigma_T U}{3H(z)m_e c} \frac{n_e}{n_e + n_H + n_{\text{He}}}$$
$$\times (T_M - T_R) + 2T_M + \frac{2}{3k_B n_{\text{tot}}}$$
$$\times \frac{1}{H(z)} (\Lambda_{\text{brem}} - \Pi_{p-i} + \Lambda_{p-r}$$
$$+ \Lambda_{c-i} + \Lambda_{c-r} + \Lambda_{\text{line}}) . \qquad (66)$$

Here n_{He} is the total number density of helium, and the denominator $n_e + n_H + n_{\text{He}}$ (n_{tot} from eq. [54]) takes into account the fact that the energy is shared among all the available matter particles. All the terms except adiabatic cooling in equation (66) involve matter energy conversion into photons. In particular, Compton and bremsstrahlung cooling are the most important, and they can be thought of as keeping T_M very close to T_R until their timescales become long compared with the Hubble time, and thereafter the matter cools as $T_M \propto (1 + z)^2$. Previous recombination calculations only included Compton and adiabatic cooling; however, the additional terms add improvements only at the $10^{-3}\%$ level in the ionization fraction. The reason for the negligible improvement is that it makes little difference which mechanism keeps T_M close to T_R early on, and adiabatic cooling still becomes important at the same time.

2.6. Summary of Equations

The system of equations to be simultaneously integrated in redshift is

$$(1 + z)\frac{dn_i(z)}{dz} = -\frac{1}{H(z)}$$
$$\times \left\{[n_e(z)n_c(z)P_{ci} - n_i(z)P_{ic}] + \sum_{j=1}^N \Delta R_{ji}\right\} + 3n_i(z) , \qquad (67)$$

$$(1 + z)\frac{dT_M}{dz} = \frac{8\sigma_T U}{3H(z)m_e c} \frac{n_e}{n_e + n_H + n_{\text{He}}}$$
$$\times (T_M - T_R) + 2T_M + \frac{2}{3k_B n_{\text{tot}}}$$
$$\times \frac{1}{H(z)} (\Lambda_{\text{brem}} - \Pi_{p-i} + \Lambda_{p-r}$$
$$+ \Lambda_{c-i} + \Lambda_{c-r} + \Lambda_{\text{line}}) , \qquad (68)$$

and

$$(1 + z)\frac{dJ(v, z)}{dz} = 3J(v, z) - \frac{c}{H(z)} [j(v, z) - \kappa(v, z)J(v, z)] . \qquad (69)$$

For $J(v, z) = B(v, z) = B(v, T_R)$ (see § 2.2), equation (69) can be omitted because the expansion of the universe preserves the thermal spectrum of noninteracting radiation, and we can use the Sobolev escape probability method for the primary spectral line distortions. The system of coupled equations (67) that we use contains up to 609 separate equations, 300 for H (one for each of a maximum of 300 levels we considered), 200 for He I, 100 for He II, 1 for He III, 1 for electrons, 1 for protons, and 1 for each of the five molecular or ionic H species. This system of equations, along with equation (68), is extremely stiff, that is, the dependent variables are changing on very different timescales. We used the Bader-Deuflhard semi-implicit numerical integration scheme, which is described in Press et al. (1992). To test the numerical integration we checked at each time step that the total charge and total number of particles are conserved to one part in 10^7.

3. RESULTS AND DISCUSSION

By an "effective three-level" H atom we mean a hydrogen atom that includes the ground state, first excited state, and continuum. In an effective three-level atom, the energy levels between $n = 2$ and the continuum are accounted for by a recombination coefficient that includes recombinations to those levels. This should be distinguished from an *actual* three-level atom, which would completely neglect all levels above $n = 2$ and would be a hopeless approximation. Good accuracy is obtained by considering an n-level atom, where n is large enough. In practice, we find that a 300-level atom is more than adequate. We do not explicitly include angular momentum states l, whose effect we expect to be negligible. In contrast to the effective three-level H atom, the 300-level H atom has no recombination coefficient with "extra" levels. The "standard" recombination calculation refers to the calculation with the effective three-level atom that is widely used today and first derived by Peebles (1968) and Zeldovich et al. (1968), updated with the most recent parameters and recombination coefficient (HSSW95).

The primordial He abundance was taken to be $Y_P = 0.24$ by mass (Schramm & Turner 1998). The present-day CMB temperature T_0 was taken to be 2.728 K, the central value determined by the FIRAS experiment (Fixsen et al. 1996).

3.1. The " Effective Three-Level" Hydrogen Atom

For comparison with the standard recombination calculation that only includes hydrogen (see Peebles 1968, 1993; Scott 1988), we reduce our chemical reaction network to an effective three-level atom, i.e., a two-level hydrogen atom plus continuum. The higher atomic energy levels are included by way of the recombination coefficient, which can effectively include recombination to hundreds of levels. The following reactions are included: $H_{n=2, l=2s} + \gamma \leftrightarrow e^- + H^+$, $H_{n=1} + \gamma \leftrightarrow H_{n=2, l=2p}$, and $H_{n=1} + 2\gamma \leftrightarrow H_{n=2, l=2s}$.

As described in Peebles (1993), we omit the recombinations and photoionizations to the ground state because any recombination directly to the ground state will emit a photon with energy greater than 13.6 eV, where there are few blackbody photons, and this will immediately reionize a neighboring H atom. We include the two-photon rate from the $2s$ state with the rate $\Lambda_{2s-1s} = 8.22458$ s^{-1} (Goldman 1989). The most accurate total Case B recombination coeffi-

cient is by Hummer (1994) and is fitted by the function

$$\alpha_B = 10^{-19} \frac{at^b}{1 + ct^d} \text{ m}^3 \text{ s}^{-1} , \quad (70)$$

where $a = 4.309$, $b = -0.6166$, $c = 0.6703$, $d = 0.5300$, and $t = T_M/10^4$ K (Péquignot et al. 1991; see also Verner & Ferland 1996).

Consideration of detailed balance in the effective three-level atom leads to a single ordinary differential equation for the ionization fraction:

$$\frac{dx_e}{dz} =$$

$$\frac{[x_e^2 n_H \alpha_B - \beta_B(1 - x_e)e^{-h_P v_{2s}/k_B T_M}][1 + K\Lambda_{2s-1s} n_H(1 - x_e)]}{H(z)(1 + z)[1 + K\Lambda_{2s-1s} n_H(1 - x_e) + K\beta_B n_H(1 - x_e)]} \quad (71)$$

(see, e.g., Peebles 1968; extra terms included in Jones & Wyse 1985, for example, are negligible). Here x_e is the residual ionization fraction, that is, the number of electrons compared to the total number of hydrogen nuclei (n_H). Here the Case B recombination coefficient $\alpha_B = \alpha_B(T_M)$, the total photoionization rate $\beta_B = \alpha_B(2\pi m_e k_B T_M/h_P^2)^{3/2}$ $\exp(-E_{2s}/k_B T_M)$ as described in § 2.3.2, v_{2s} is the frequency of the $2s$ level from the ground state, and the redshifting rate $K \equiv \lambda_\alpha^3/[8\pi H(z)]$, where λ_α is the Lyα rest wavelength. Note that T_M is used in equation (71) and in β_B because the temperature terms come from detailed balance derivations that use Boltzmann and Saha equilibrium distributions, which are collisional descriptions. In the past, this equation has been solved [for $x_e(z)$] simultaneously with a form of equation (66) containing only adiabatic and Compton cooling. We refer to the approach of equation (71) as the "standard calculation."

For the comparison test with the standard recombination calculation, we also use an effective recombination coefficient, but we use three equations to describe the three reactions listed above. That is, we simplified equation (67) to three equations, one for the ground-state population (n_1), one for the first excited state population (n_2), and one for the electrons (for H recombination $n_e = n_p$):

$$(1 + z)\frac{dn_1(z)}{dz} = -\frac{1}{H(z)}(\Delta R_{2p-1s} + \Delta R_{2s-1s}) + 3n_1 , \quad (72)$$

$$(1 + z)\frac{dn_2(z)}{dz} = -\frac{1}{H(z)}[n_e(z)n_p(z)\alpha_B - n_{2s}(z)\beta_B$$
$$- \Delta R_{2p-1s} - \Delta R_{2s-1s}] + 3n_2 , \quad (73)$$

$$(1 + z)\frac{dn_e(z)}{dz} = -\frac{1}{H(z)}[n_{2s}(z)\beta_B - n_e(z)n_p(z)\alpha_B] + 3n_e . \quad (74)$$

The remaining physical difference between our effective three-level atom approach and that of the standard calculation is the treatment of the redshifting of H Lyα photons (included in the ΔR_{2p-1s} terms). In our calculation the redshifting is accounted for by the Sobolev escape probability (see § 2.3.3). Following Peebles (1968, 1993), the standard calculation accounts for the redshifting by approximating the intensity distribution as a step and in effect takes the ratio of the redshifting of the photons through the line to

the expansion scale that produces the same amount of red-shifting. It can be shown that Peebles's step method considered as an escape probability scales as $1/\tau_S$, where τ_S is the Sobolev optical depth. For high Sobolev optical depth, which holds for H Lyα during recombination for any cosmological model, the Sobolev escape probability also scales as $1/\tau_S$:

$$\lim_{\tau_S \gg 1} p_{ij} = \lim_{\tau_S \gg 1} \frac{1}{\tau_S}(1 - e^{-\tau_S}) = \frac{1}{\tau_S}. \tag{75}$$

Therefore, the two approximations are equivalent for Lyα, although we would expect differences for lines with $\tau_S \lesssim 1$, where $p \to 1$. Because we treat recombination in the same way as Peebles (1968, 1993), no individual treatment of other lines is permitted, and therefore there are no other differences between the two calculations for this simple case. Note that with Peebles's step method to compute ΔR_{2p-1s} and the assumption that $n_1 = n_H - n_p$, the above equations will reduce to the single ordinary differential equation (eq. [71]).

The results from our effective three-level recombination calculation are shown in Figure 1, plotted along with values from a separate code as used in HSSW95, which represents the standard recombination calculation updated with the most recent parameters. The resulting ionization fractions are equal, which shows that our new approach gives exactly the standard result when reduced to an effective three-level atom. Two other results are plotted for comparison, namely, values of x_e taken from Peebles (1968) and Jones & Wyse (1985). Their differences can be largely accounted for by the use of an inaccurate recombination coefficient with $\alpha_B(T_M) \propto T_M^{-1/2}$.

As an aside, we note the behavior for $z \lesssim 50$ in our curve and the HSSW95 one. This is caused by inaccuracy in the recombination coefficient for very low temperatures. The downturn is entirely artificial and could be removed by using an expression for $\alpha_B(T)$ that is more physical at small temperatures. The results of our detailed calculations are not believable at these redshifts either since accurate modeling becomes increasingly difficult as a result of numerical precision as T approaches zero. But in any case the optical 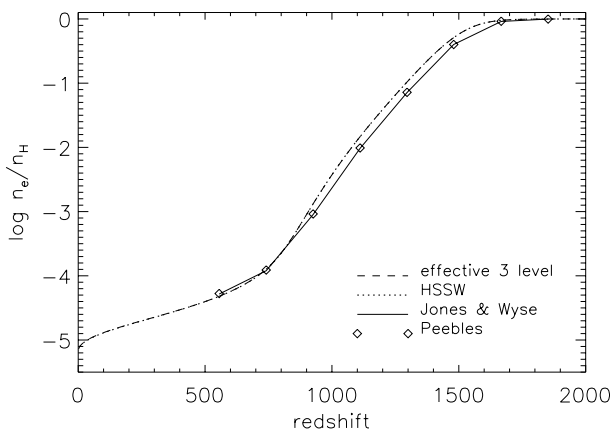depth back to such redshifts is negligible, and the real universe is reionized at a similar epoch (between $z = 5$ and 50 certainly).

3.2. Multilevel Hydrogen Atom

The purpose of a multilevel hydrogen atom is to improve the recombination calculation by following the population of each atomic energy level with redshift and by including all bound-bound and bound-free transitions. This includes recombination to and photoionization from all levels *directly* as a function of time, in place of a parameterized recombination and photoionization coefficient. The individual treatment of all levels in a coupled manner allows for the development of departures from equilibrium among the states with time and feedback on the rate of recombination. Since the accuracy of the recombination coefficient is probably the single most important effect in obtaining accurate power spectra (HSSW95), it makes sense to follow the level populations as accurately as possible.

In the multilevel H atom recombination calculation, we do not consider individual l states (with the exception of $2s$ and $2p$) but assume that the l sublevels have populations proportional to $2l + 1$. The l sublevels only deviate from this distribution in extreme nonequilibrium conditions (such as planetary nebulae). In their H recombination calculation, Dell'Antonio & Rybicki (1993) looked for such l-level deviations for $n \leq 10$ and found none. For $n > 10$, the l states are even less likely to differ from an equilibrium distribution because the energy gaps between the l sublevels are increasingly smaller as n increases.

3.2.1. Results from a Multilevel H Atom

Figure 2 shows the ionization fraction x_e from recombination of a two-, 10-, 50-, 100-, and 300-level H atom, compared with the standard effective three-level results. The fraction x_e converges for the highest n-level atom calculations. The effective three-level atom actually includes about 800 energy levels via the recombination coefficient (e.g., Hummer 1994).

Figure 2 shows that the more levels that are included in the hydrogen atom, the lower the residual x_e. The simple explanation is that the probability for electron capture increases with more energy levels per atom. Once captured, the electron can cascade downward before being reionized. Together this means that adding more higher energy levels per atom increases the rate of recombination. Eventually x_e converges as the atom becomes complete in terms of electron energy levels, i.e., when there is no gap between the highest energy level and the continuum (see Fig. 3). Ultimately the uppermost levels will have gaps to the continuum that are smaller than the thermal broadening of those levels, and so energy levels higher than about $n = 300$ do not need to be considered, except perhaps at the very lowest redshifts.

For other reasons entirely, our complete (300-level) H atom recombination calculation gives an x_e lower than that of the effective three-level atom calculation. The faster production of hydrogen atoms is due to nonequilibrium processes in the excited states of H, made obvious by our new, level-by-level treatment of recombination. The details are described in § 3.2.2 below.

3.2.2. Faster H Recombination in Our Level-by-Level Recombination Calculation

The lower x_e in our calculation compared to the standard calculation is caused by the strong but cool radiation field. Specifically, both a faster downward cascade rate and a
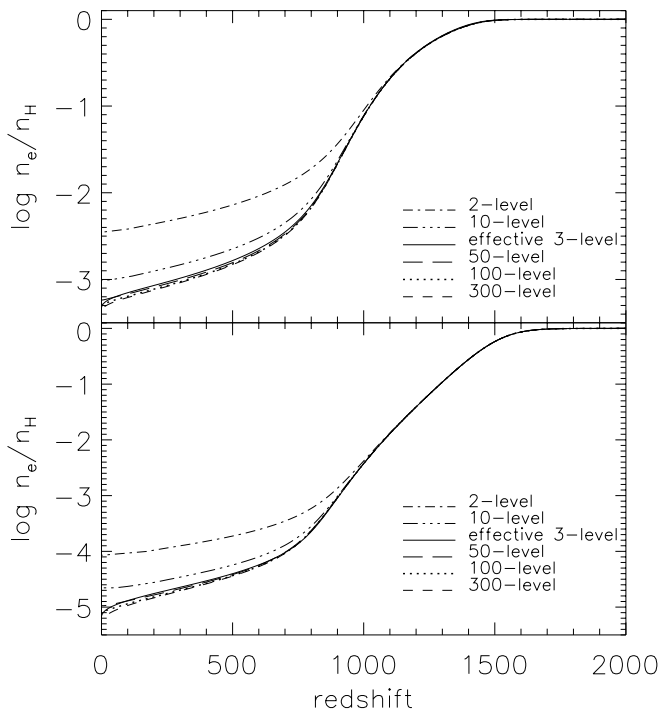

FIG. 1.—Comparison of effective three-level hydrogen recombination for the parameters $\Omega_{tot} = 1.0$, $\Omega_B = 1.0$, $h = 1.0$. Note that the Jones & Wyse (1985) and Peebles (1968) curves overlap, as does our curve with the HSSW95 one.

FIG. 2.—Multilevel hydrogen recombination for the standard CDM parameters $\Omega_{tot} = 1.0$, $\Omega_B = 0.05$, $h = 0.5$ (*top*), and for $\Omega_{tot} = 1.0$, $\Omega_B = 1.0$, $h = 1.0$ (*bottom*), both with $Y_P = 0.24$, $T_0 = 2.728$ K. The "effective three-level" calculation is essentially the same as in HSSW95 and uses a recombination coefficient that attempts to account for the net effect of all relevant levels. We find that we require a model that considers close to 300 levels for full accuracy. Note also that although we plot all the way to $z = 0$, we know that the universe becomes reionized at $z > 5$ and that our calculations (as a result of numerical precision at low T) become unreliable for $z \lesssim 50$ in the upper two curves and for $z \lesssim 20$ in the other curves.

lower total photoionization rate contribute to a faster net recombination rate.

By following the population of each atomic energy level with redshift, we relax the assumption used in the standard calculation that the excited states are in equilibrium. In addition, we calculate all bound-bound rates that control equilibrium among the bound states. In the standard calculation, equilibrium among the excited states $n \geq 2$ is assumed, meaning that the net bound-bound rates are zero. Figure 4 shows that the net bound-bound rates are actually different from zero at $z \lesssim 1000$. The reason for this is that at low temperatures, the strong but cool radiation field means that high-energy transitions are rare as a result of few high-energy photons. More specifically, photoexcitation and stimulated photode-excitation for high-energy transitions become rare (e.g., 70–10, 50–4, etc.). In this case, spontaneous de-excitation dominates, causing a faster downward cascade to the $n = 2$ state. In addition, the faster downward cascade rate is faster than the photoionization rate from the upper state, and one might view this as radiative decay stealing some of the depopulation "flux" from photoionization. Both the faster downward cascade and the lower photoionization rate contribute to the faster net recombination rate.

The cool radiation field is strong, so photoexcitations and photode-excitations are rapid among nearby energy levels (e.g., 70–65, etc.; see Fig. 4). What we see in Figure 4 is that with time, after $z \lesssim 1000$, the $n = 70$ energy level becomes progressively decoupled from the distant lower energy levels ($n = 2, 3, \ldots 20, \ldots$) but remains tightly coupled to its nearby "neighbors" ($n = 60, 65$, etc.). This explains the departures from an equilibrium Boltzmann distribution (in the excited states) as seen in the shape of the curves in Figure 5.

Figure 5 illustrates the nonequilibrium of the excited states by showing the ratio of populations of the excited states compared to a Boltzmann equilibrium distribution with respect to $n = 2$. We find that the upper levels of the hydrogen atom are not in thermal equilibrium with the radiation, i.e., the excited levels are not populated according to a Boltzmann distribution. The excited states are in fact overpopulated relative to a Boltzmann distribution. This is not a surprise for the population of the $n = 2$ state, which is strongly overpopulated compared to the $n = 1$ ground state, and so should be all $n \geq 2$ states because all Lyman
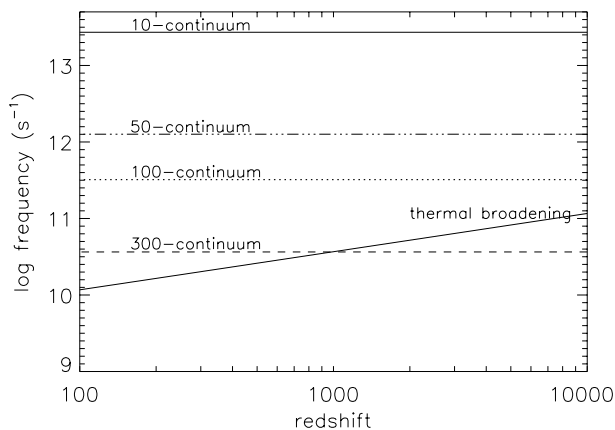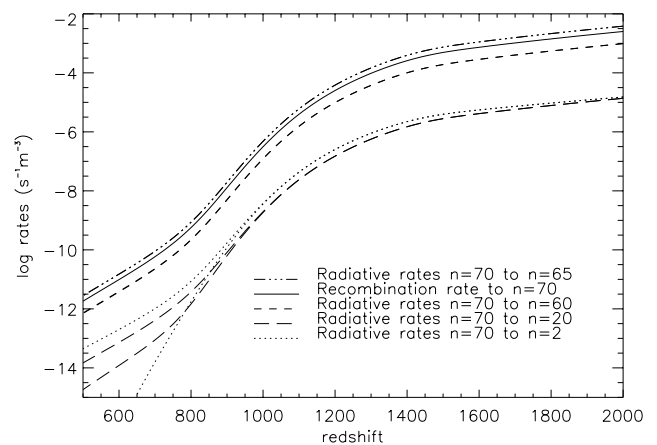

FIG. 3.—Energy separations between various hydrogen atomic levels and the continuum. The solid curve shows the energy width of the same energy levels as a result of thermal broadening. Thermal broadening was calculated using $v(2k_B T_M/m_H c^2)^{1/2}$. The frequency of the highest atomic energy level ($n = 300$, with an energy from the ground state of 109676.547 cm^{-1}; the continuum energy level is 109677.766 cm^{-1}) was used for v, but a thermal broadening value for any atomic energy level would overlap on this graph.

FIG. 4.—How the bound-bound rates of large energy separation (e.g., $n = 70$–20) go out of equilibrium at low T, illustrated using an upper level of $n = 70$ for definiteness. The case of equilibrium corresponds to net bound-bound rates of zero. Each upward and downward bound-bound rate for a given transition is represented by the same curve; nonequilibrium occurs where a single curve separates into two as redshift decreases. The rates shown are for the sCDM model.

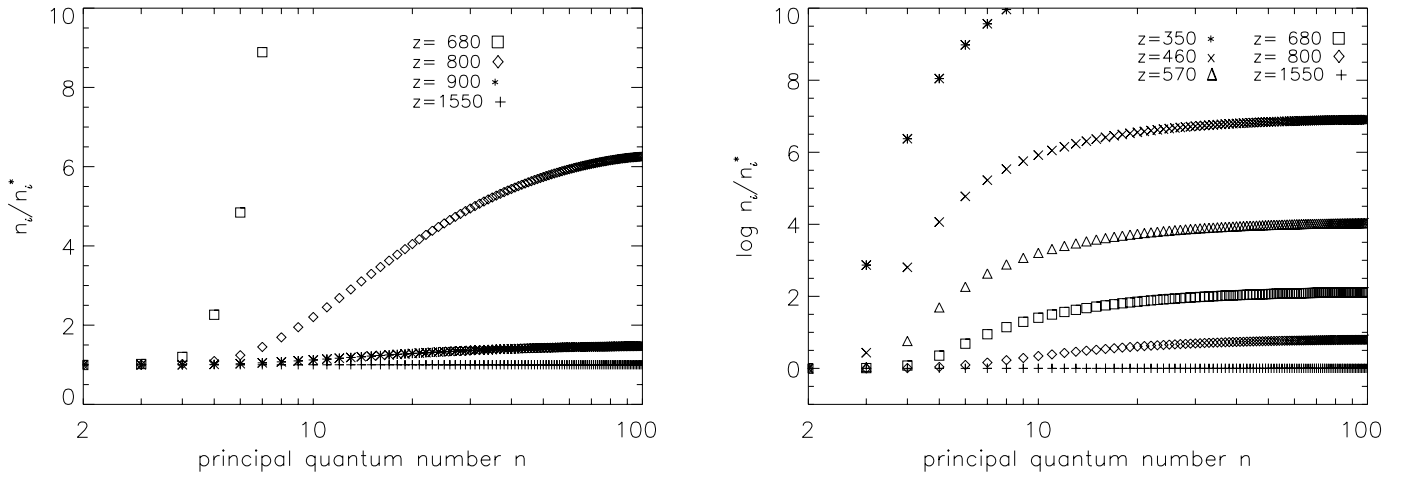FIG. 5.—Ratio of the actual number densities of the excited states (n_i) and number densities for a Boltzmann distribution of excited states (n_i^*) at different redshifts for recombination within the standard CDM model. See § 3.2.2 for details. We give two different plots with linear and logarithmic y-axes and different redshift values to show a wider range of out-of-equilibrium conditions. The ratio approaches a constant for high n at a given z because the high-energy (Rydberg) states are all very close in energy and thus have similar behavior (i.e., remain coupled to the radiation field and each other).

lines remain optically thick during recombination. What *is* surprising is that all excited states develop a further overpopulation *with respect to* $n = 2$ and each other. Note that this is not a population inversion. The recombination rate to a given high level is faster than the downward cascade rate, and this causes a "bottleneck" creating the overpopulation. Figure 5 shows that all states are in equilibrium at high redshifts, with the highest states going out of equilibrium first, followed by lower and lower states as the redshift decreases. The factor by which the excited states are overpopulated approaches a constant at high n for a given redshift, with this factor increasing as z decreases. The ratio is constant because the high-energy level Rydberg states have very similar energy levels to each other, with a relatively large energy separation from the $n = 2$ state (i.e., the exponential term in eq. [21] dominates over the g_i ratios, and the exponential term is similar for all of the Rydberg states). Figure 5 also shows an enormous ratio at low redshift ($z < 500$) for number densities of the actual excited states to the number densities of a Boltzmann distribution of excited states, on the order of 10^6. At such a low redshift, there are almost no electrons in the excited states ($\sim 10^{-20}$ cm^{-3}), and so unlike the case of higher redshifts, the ratio is only an illustration of the strong departure from an equilibrium distribution; the actual populations are very low in any case.

In comparison with the standard equilibrium capture-cascade calculation for α_B, the unusual situation described above (caused by the strong but cool radiation field) leads to higher effective recombination rates for the majority of excited states without increasing photoionization proportionally. This results in a higher net rate of production of neutral hydrogen atoms, i.e., a lower x_e.

3.2.3. *Accurate Recombination versus Recombination Coefficient*

To demonstrate why the nonequilibrium in the excited states of H affects the recombination rate, we must consider the difference in our new treatment of recombination compared to the standard treatment. An important new benefit of our level-by-level calculation lies in replacing the recombination coefficient with a direct calculation of recombination to and photoionization from each level at each

redshift step. In other words, we calculate the recombination rate and the photoionization rate using individual level populations and parameters of the excited states i,

$$\sum_{i=1}^{N} n_e n_p R_{ci} = n_e n_p \sum_{i=1}^{N} \left(\frac{n_i}{n_e n_p}\right)^{\text{LTE}} 4\pi \int_{v_0}^{\infty} \frac{\sigma_i(v)}{h_{\text{P}} v}$$
$$\times \left[\frac{2h_{\text{P}} v^3}{c^2} + B(v, T_R)\right] e^{-h_{\text{P}} v/k_{\text{B}} T_M} dv , \quad (76)$$

and similarly for the photoionization rate,

$$\sum_{i=1}^{N} n_i R_{ic} = \sum_{i=1}^{N} n_i 4\pi \int_{v_0}^{\infty} \frac{\sigma_{ic}(v)}{h_{\text{P}} v} B(v, T_R) dv . \quad (77)$$

For the standard recombination calculation, the recombination rate is

$$\sum_{i=1}^{N} n_e n_p R_{ci} = n_e n_p \alpha_B(T_M) , \quad (78)$$

and the photoionization rate is

$$\Sigma_i n_i R_{ic} \equiv n_{2s} \beta_B(T_M)$$
$$= n_{2s} \alpha_B(T_M) e^{-E_{2s}/k_{\text{B}} T_M} (2\pi m_e k_{\text{B}} T_M)^{3/2}/h_{\text{P}}^3 . \quad (79)$$

In this last equation, the excited state populations are hidden by the Boltzmann relation with n_{2s} (see § 2.3.2). The important point here is that our method allows redistribution of the H level populations over all 300 levels at each redshift step, which feeds back on the recombination process via equation (77) and leads to the lower x_e shown in Figure 2. This redistribution of the level populations is not possible in the standard calculation's equation (79), which only considers the populations n_e, n_1, and n_{2s} and considers the excited level populations $n > 2s$ to be proportional to n_{2s} in an equilibrium distribution.

A small improvement in our new recombination treatment over the standard treatment is in our distinguishing the various temperature dependencies of recombination. Photoionization and stimulated recombination are radiative, so they should depend on T_R. Spontaneous recombination is collisional and depends on T_M (see § 2.3.1). In the standard calculation, the radiative nature of recombination
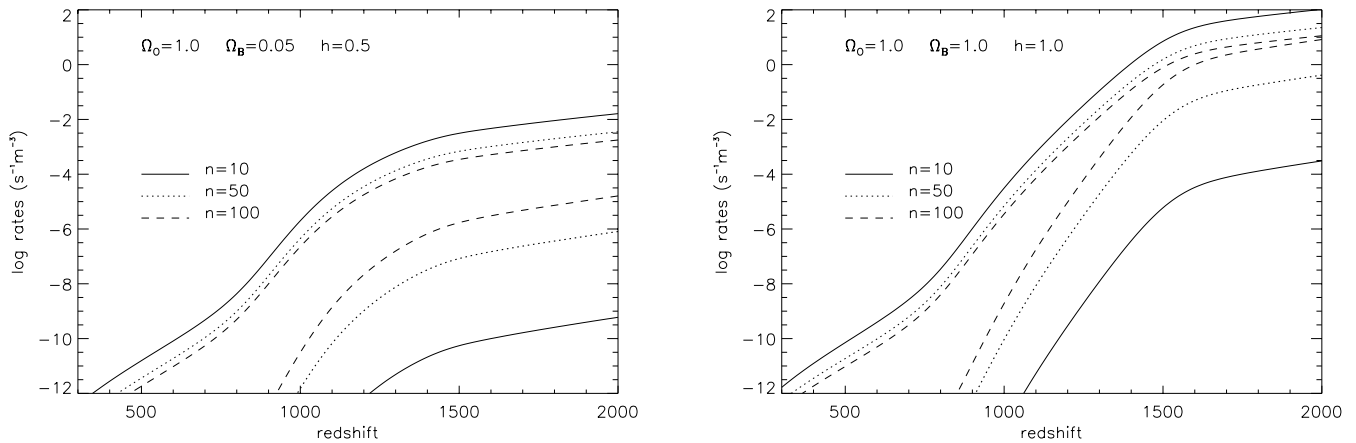
FIG. 6.—Comparison of ionization rates. The upper curves are photoionization rates, the lower curves are collisional ionization rates. The left panel shows a model with the standard CDM parameters, while the right panel shows an extreme baryon cosmology.

and photoionization is overlooked because both the recombination coefficient and the photoionization coefficient are a function of T_M only (eqs. [78] and [79]). Although adiabatic cooling (eq. [58]) does not dominate until quite low redshifts ($z \lesssim 100$), it still contributes partially to matter cooling throughout recombination. The resulting difference between T_M and T_R in the net recombination rate affects x_e at the few percent level at $z \lesssim 300$ for the popular cosmologies and has an even larger effect for high-Ω_B models.

3.2.4. Collisions

The standard recombination calculation omits collisional excitation and ionization because at the relevant temperatures and densities they are negligible for a three-level hydrogen atom (Matsuda et al. 1971). We have found that the collisional processes are also not important for the higher levels, even though those electrons are bound with little energy. In high-Ω_B models, collisional ionization and collisional recombination rates for the highest energy levels are of the same order of magnitude as the photoionization and recombination rates, though not greater than them (see Fig. 6).

3.2.5. Departures from Case B

Case B recombination excludes recombination to the ground state and considers the Lyman lines to be optically thick (i.e., photons associated with all permitted radiative transitions to $n = 1$ are assumed to be instantly reabsorbed). An implied assumption necessary to compute the photoionization rate is that the excited states ($n \geq 2$) are in equilibrium with the radiation. Unlike the standard recombination calculation, our method allows departures from Case B because the Lyman lines are treated by the Sobolev escape probability method, which is valid for any optical thickness, and our method allows departures from equilibrium of the excited states (§ 3.2.1). We find that the excited states depart from equilibrium at redshifts $\lesssim 800$, so Case B does not hold then. However, our calculations show that for hydrogen all Lyman lines are indeed optically thick during all of hydrogen recombination, so Case B holds for H recombination above redshifts $\simeq 800$.

Figure 7 shows that the Lyman lines are not optically thick at earlier times, e.g., during helium recombination, where we find some optically thin H Lyman lines. The Sobolev escape probability treats this consistently, which is

necessary because we evolve H, He I, and He II simultaneously.

3.2.6. Other Recent Studies

The previous study that was closest in approach to our own was that of Dell'Antonio & Rybicki (1993), who calculated recombination for a 10-level hydrogen atom in order to estimate the spectral distortions to the CMB blackbody radiation spectrum. Ten levels are insufficient to calculate recombination accurately because the higher energy levels of the atom are completely ignored (see Fig. 3). However, the accuracy of the ionization fraction (x_e) was sufficient to determine the magnitude of the spectral distortions. Their recombination model treated individual levels but used a recombination coefficient to each level of the form $T_M^{-1/2}$. Because the form of the recombination coefficient dominates the H recombination process, our models are not equivalent, and so there is little use in comparing the results.

More recently, Boschan & Biltzinger (1998) derived a new parameterized recombination coefficient to solve the recombination equation of the standard calculation and to generate spectral distortions in the CMB. Their calculation
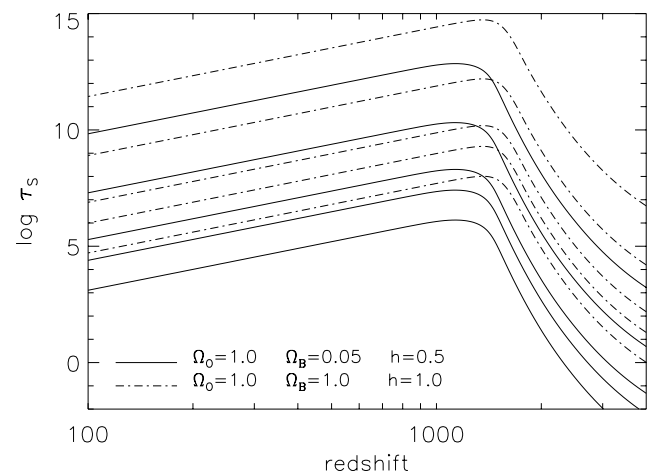

FIG. 7.—Sobolev optical depth (τ_S) for the H Lyman lines for two different cosmological models. From upper to lower the curves represent the optical depth in the Lyman transitions from $n = 2$ (i.e., Lyα), 10, 50, 100, and 270. The curves show that all the Lyman transitions are optically thick during H recombination ($z \lesssim 2000$) but some are optically thin ($\tau_S < 1$) earlier, during He recombination ($z \gtrsim 2000$).

differs from ours in that their recombination coefficient is precalculated. Hence, it is not an interactive part of the calculation and does not allow the advantages that our calculation does, mainly the feedback of the nonequilibrium in the excited states on the net recombination rate. While they include pressure broadening for a cutoff in the partition function, they neglect thermal broadening. A more serious problem is their method of inclusion of stimulated recombination, as originally suggested by Sasaki & Takahara (1993), who included stimulated recombination as positive recombination instead of negative ionization. The physics (as described in § 2.3.1) and our computational results are the same regardless of whether stimulated recombination is treated as positive recombination or negative ionization. However, this may not be the case computationally for the standard calculation, if it is not treated with care. We defer a full discussion of these matters to a separate paper (Seager & Sasselov, in preparation).

We have also investigated how we can approximate our calculations, so that other researchers can obtain approximately accurate results without the need to follow 300 levels in a hydrogen atom. Because the net effect of our new H calculation is a faster recombination (a lower freeze-out ionization fraction), our results can be reproduced by artificially speeding up recombination in the standard calculation. Further details are described in Seager, Sasselov, & Scott (1999).

3.3. Helium

We compute helium and hydrogen recombination simultaneously. The recombination of He III into He II and He II into He I is calculated in much the same way as hydrogen, with recombination, photoionization, redshifting of the n^1p–1^1s lines (in H these are the Lyman lines), inclusion of the 2^1s–1^1s two-photon rates, collisional excitation, collisional de-excitation, collisional ionization, and collisional recombination, as described in §§ 2.3.1–2.3.3. The multilevel He I atom includes the first four angular momentum states up to the level $n = 22$, above which only the principal quantum number energy levels and transitions are used. Figure 8 (which shows the levels up to $n = 4$ only) indicates how much more complicated the He I atom is compared with H or the hydrogenic He II. Our multilevel He II atom includes the first four angular momentum states up to the level $n = 4$, above which only the principal quantum number energy levels and transitions are used. Photoionizations from any He I excited state are allowed only into the ground state of He II because there are few photons energetic enough (>40 eV) to do more than that. Two electron transitions in He I are negligible at recombination era temperatures.

Cosmological helium recombination was discussed explicitly in Matsuda et al. (1969, 1971, hereafter MST), Sato, Matsuda, & Takeda (1971), and to a lesser extent in Lyubarsky & Sunyaev (1983), while several other papers give results but no details (e.g., Lepp & Shull 1984; Fahr & Loch 1991; Galli & Palla 1998). The main improvement in our calculation over previous treatments of helium is that we use a multilevel He II atom, a multilevel He I atom with triplets and singlets treated correctly, and evolve the population of each energy level with redshift by including all bound-bound and bound-free transitions. This is not possible for the standard recombination calculation method (eq. [71]) extended to He I, using an effective three-level He I
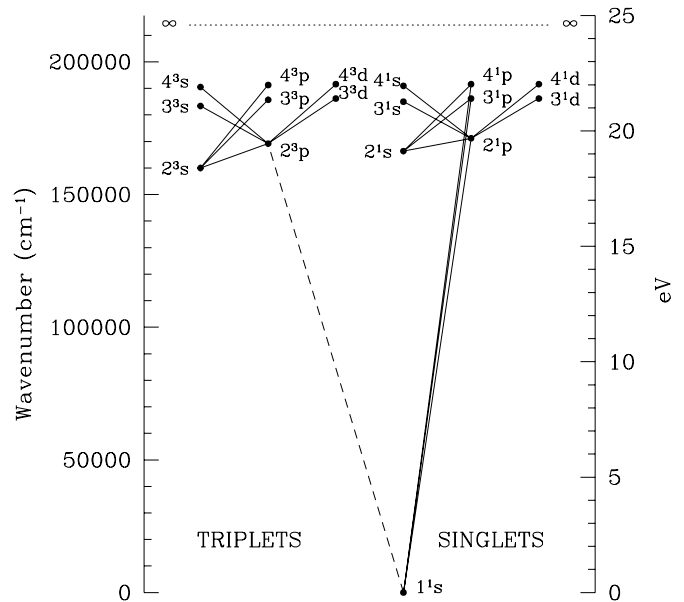

Fig. 8.—Grotrian diagram for He I, showing the states with $n \leq 4$ and the continuum. In practice, our model atom explicitly contains the first four angular momentum states up to $n = 20$ and 120 principal quantum number energy levels beyond.

atom with only a singlet ground state, singlet first excited state, and continuum.

3.3.1. Results from He I Recombination

Figure 9 shows the ionization fraction x_e through He II, He I, and H recombination, plotted against the standard H calculation that includes He II and He I recombination via the Saha equation. For completeness we give the helium
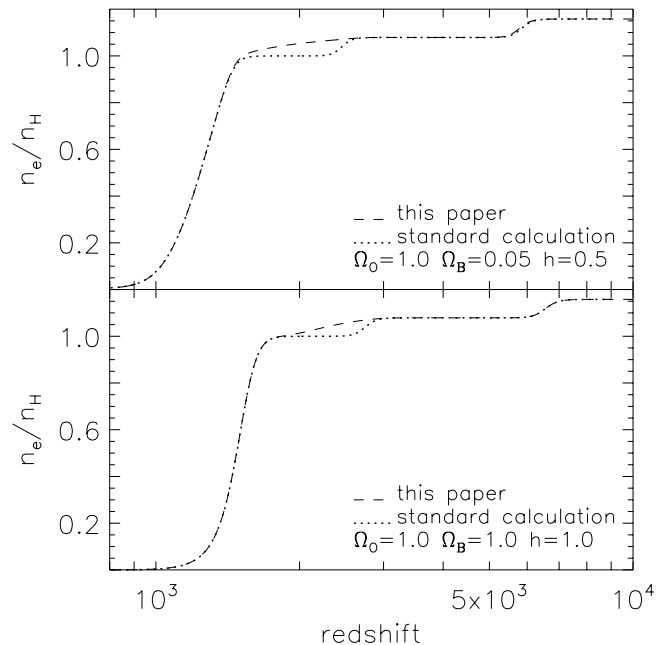

Fig. 9.—Helium and hydrogen recombination for two cosmological models with $Y_P = 0.24$ and $T_0 = 2.728$ K. The first step from right to left is recombination of He III to He II, the second step is He II to He I, and the third step is H recombination.

Saha equations here: for He I ↔ He II,

$$\frac{(x_e - 1)x_e}{1 + f_{He} - x_e} = 4 \frac{(2\pi m_e k_B T)^{3/2}}{h_P^3 n_H} e^{-\chi_{He\,I}/k_B T} , \qquad (80)$$

and for He II ↔ He III,

$$\frac{(x_e - 1 - f_{He})x_e}{1 + 2f_{He} - x_e} = \frac{(2\pi m_e k_B T)^{3/2}}{h_P^3 n_H} e^{-\chi_{He\,II}/k_B T} . \qquad (81)$$

Here the χ's are ionization potentials, n_H is the total number density of hydrogen, f_{He} is the total number fraction of helium to hydrogen $f_{He} = n_{He}/n_H = Y_P/4(1 - Y_P)$, and our definition of $x_e \equiv n_e/n_H$ results in the complicated-looking left-hand sides. The extra factor of 4 on the right-hand side for He I ↔ He II arises from the statistical weight factors.

While our improved x_e agrees fairly closely with Saha recombination for He II (see § 3.3.4), the difference in x_e from Saha recombination during He I recombination is dramatic. Our new detailed treatment of He I shows He I recombination finishing just after the start of H recombination (see Fig. 9), i.e., significantly delayed compared with the Saha equilibrium case. This is different from the earlier calculations (e.g., MST), in which He I recombination is finished well before H recombination begins. In this previous case, He I recombination still affected the CMB anisotropy power spectrum on small angular scales because the diffusion damping length grows continuously and is sensitive to the full thermal history (HSSW95). In our new case, particularly for our low Ω_B models, He I recombination is still finishing at the very beginning of H recombination, which further affects the power spectrum at large angular scales (see § 3.7). We show a "blowup" of the two helium recombination epochs in Figure 10.

3.3.2. Physics of He I Recombination

The physics of He I recombination can be summarized as follows. There are three major aspects to it: (1) the He I has excited states that are able to retain charge; but (2) being very close to the continuum, the highly excited states are easily photoionized by the radiation field at $z \simeq 3000$; then (3) we have a standard hydrogenic-like Case B recombi-

nation, which is unaffected by neutral H removing He I 2^1p-1^1s (resonance line) photons.

The He I atom has a metastable, i.e., very slow, set of states—the triplets (e.g., n^3p-n^1s). Therefore, overall the excited states of He I can naturally retain more charge than a simple hydrogenic system under Boltzmann equilibrium. The situation would resemble what we found for H recombination with the enhanced populations of the higher states and would lead to faster reduction of x_e. However, the high excited states of He I are much more strongly "packed" toward the continuum compared to those of H; the energy difference between the $3p$ levels and the continuum is 1.6 eV for He I versus 1.5 eV for H, compared to 24.6 eV versus 13.6 eV for the ground-state continuum energy difference. This is enough to depopulate the triplets (whose "ground state" is $n = 2^3s$), given the much higher radiation temperature during He I recombination. Left on its own under these circumstances, He I would recombine much like the standard Case B effective three-level H atom, i.e., slower than Saha recombination. There is one possible obstacle—it is the existence of some neutral H, which could "steal" He I resonance line photons, invalidate the effective Case B, and make it a Saha recombination instead. However, our detailed calculation shows that neutral H during He I recombination is not able to accomplish that, and the process is *not* described by Saha equilibrium.

Figure 11 shows that Saha equilibrium recombination is invalid for He I, by comparing two competing absorption processes of the He I 2^1p-1^1s (in H this is Lyα) photons: (1) photoexcitation by the He I 1^1s-2^1p transition and (2) photoionization of the ground state of H. The third curve shows the Saha equilibrium rate. Figure 11 clearly shows that process (2) is negligible (in contradiction to the discussion in HSSW95). The destruction rate of the He I 2^1p-1^1s photons by the cosmological redshift and by the 2^1p-1^1s two-photon rate is much smaller than the He I 1^1s-2^1p photoexcitation rate. To be doubly sure that absorption of these photons by hydrogen is negligible, we explicitly included the relevant rate in our models and found no discernible effects. In order for He I recombination
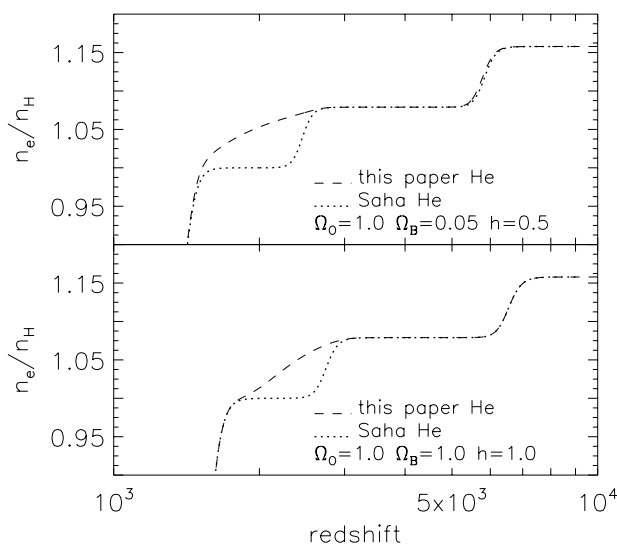
FIG. 10.—Details of helium recombination for the standard CDM cosmology (*top*) and the high-Ω_B cosmology (*bottom*). The dashed lines show our new results, and the dotted lines show the results assuming Saha equilibrium.
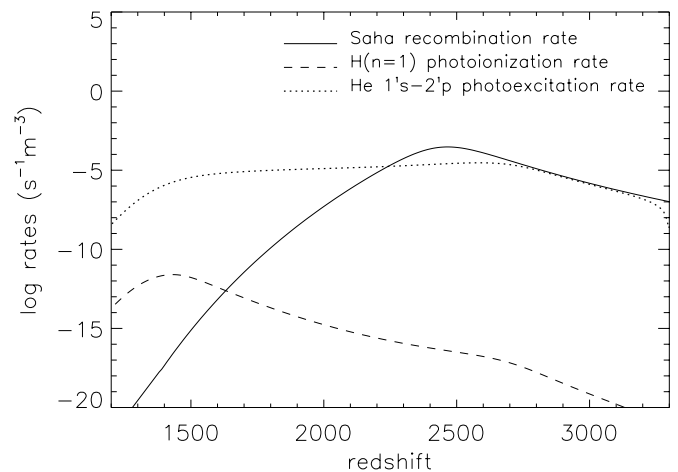
FIG. 11.—This figure shows why the Saha equilibrium recombination rate (R_{Saha}) for He I is not valid. Comparing the dotted and dashed lines, the photoexcitation rate (i.e., photoabsorption rate) for He 1^1s-2^1p (R_{He}) is orders of magnitude greater than the photoionization rate for H (R_H) from the same He I 2^1p-1^1s photon pool; there is no possibility for H to "steal" the photons to speed up He I recombination. For Saha recombination to be valid, $R_H > R_{He}$, as well as $R_H \geq R_{Saha}$. For the sCDM model shown here, He I recombination begins around $z = 3000$.

to be approximated by Saha equilibrium, one of the three processes described above would have to be faster than or equal to the Saha equilibrium rate, which we do not find to be the case.

The recombination of He I is slow for the same reasons that H recombination is, namely, because of the optically thick n^1p-1^1s transitions, which slow cascades to the ground state, and the exclusion of recombinations to the ground state. In other words, He I follows a Case B recombination. Because the "bottleneck" at $n = 2$ controls recombination, it is not surprising that He I and H recombination occur at a similar redshift; the ionization energy of $n = 2$ is similar in both. He I recombination is slower than H recombination because of its different atomic structure. The excited states of He I are more tightly packed, and the 2^1p-1^1s energy difference is greater than that of H. The strong radiation field keeps the ratio of photoionization rate/downward cascade rate higher than in the H case, resulting in a slower recombination.

We find that the strong radiation field also causes the triplet states to be virtually unpopulated. The lack of electrons in triplet states is easily understood by considering the blackbody radiation spectrum. At He I recombination ($z \simeq 3000$), the blackbody radiation peak is around 2 eV, so there are around 11 orders of magnitude more photons that can ionize the lowest triplet state 2^3s (4.8 eV) than the singlet ground state (24.6 eV) since both are on the steeply decreasing Wien tail. It is interesting to note that in planetary nebulae, where the young, hot ionizing star produces most of its energy in the UV, the opposite occurs: the He I atoms have few electrons in the singlet states; instead, most of them are in the triplet states.

There is one more possible method to speed up He I recombination, and that is collisional rates between the triplets and singlet states. If fast enough, the collisional rates would provide another channel to keep hold of captured electrons (by pumping them into the triplet states faster than they can be reionized). The triplets are 3 times as populated as the singlets as a result of the statistical weight factors. By forcing the collisional rates to be greater than the recombination rates and the bound-bound radiative rates, we find an extremely fast He I recombination, approximated by the Saha equilibrium. Essentially, we force electrons from the singlets into the triplets faster than they can cascade downward and faster than they can be photoionized out of the triplets. In reality, the collisions are negligible, a few orders of magnitude less than the radiative rates. It is important to note that apart from collisions, the singlet and triplet states are *only* connected via the n^3p-n^1s transitions, which are orders of magnitude slower than the 2^3s-1^1s rate. We note here that MST stated that the collisional rates were high enough to cause equilibrium between the triplet and singlet states. One must be careful to compare all relevant rates, and we keep all of them in our code. We find that the allowed radiative rates (e.g., photoexcitation and photode-excitation) are greater than the collisional rates. Therefore, the allowed radiative rates control the excited states' population distribution, *not* the collisional rates. In other words, electrons in the singlet states are jumping between bound singlet states faster than the collisional rates can send them into the triplet states.

3.3.3. *Effective Three-Level Calculation for* He I

We note here that MST used an effective three-level He I singlet atom and calculated He I recombination in the same way as the standard H calculation (eq. [71]) with the appropriate He I parameters. When we follow their treatment, we get essentially the same result as our multilevel He I calculation. We are not sure why MST obtained such a fast He I recombination.

As with hydrogen, we have also investigated what is required to achieve an accurate solution for helium, without modeling the full suite of atomic processes. We have found that the use of the "effective three-level" equations for helium (as described in MST), together with an appropriate recombination coefficient for singlets only (eq. [82]), results in a very accurate treatment of $x_e(z)$ during the time of helium recombination. In detail, it is necessary to follow hydrogen and helium recombination simultaneously, increasing the number of differential equations to solve. However, little accuracy is in fact lost by treating them independently since recombination is governed by dramatic changes in timescales through Boltzmann factors and the like and is affected little by small changes in the number of free electrons at a given time. Further details are discussed in Seager et al. (1999).

Although our model does not explicitly use a recombination coefficient, it does allow us to calculate one easily. To aid other researchers, it is worth presenting a fit for the singlet-only Case B recombination coefficient for He I (including recombinations to all states except the ground state) from the data in Hummer & Storey (1998). Hummer & Storey (1998) compute photoionization cross sections that are more accurate than the ones we use (Hofsaess 1979) but are not publicly available. Following the functional forms used in the fits of Verner & Ferland (1996), we find

$$\alpha_{\mathrm{He}} = a \left\{ \left(\frac{T_M}{T_2} \right)^{1/2} \left[1 + \left(\frac{T_M}{T_2} \right)^{1/2} \right]^{1-b} \right.$$
$$\left. \times \left[1 + \left(\frac{T_M}{T_1} \right)^{1/2} \right]^{1+b} \right\}^{-1} \mathrm{m^3 \ s^{-1}}, \quad (82)$$

with $a = 10^{-16.744}$, $b = 0.711$, $T_1 = 10^{5.114}$ K, and T_2 fixed arbitrarily at 3 K. This fit is good to less than 0.1% over the relevant temperature range (4000–10,000 K), and still fairly accurate over a much wider range of temperatures.

3.3.4. *He II Recombination*

He II recombination occurs too early to affect the power spectrum of CMB anisotropies. For completeness, we mention it briefly here. He II recombination is fast because of the very fast two-photon rate. For most cosmologies the two-photon rate is faster than the net recombination rate, meaning that as fast as electrons are captured from the continuum they can cascade down to the ground state. Because of this, there is essentially no "bottleneck" at the $n = 2$ level. In high-baryon models, He II recombination can be approximated using the Saha recombination. As shown in Figure 9, He II recombination is slightly slower than the Saha recombination for low-baryon models.

3.3.5. *What Controls Recombination?*

H recombination is largely controlled by the 2s–1s two-photon rate, which, except for low-baryon cases, is much faster than the H Lyα rate. The net recombination rate, net 2s–1s rate, and net Lyα rate are compared for different cosmologies in Figure 12. Figures 13 and 14 show the same rate comparison for He I and He II. The three figures all have the same scale on the x- and y-axes for easy comparison. He I recombination is controlled by the 2^1p-1^1s rate
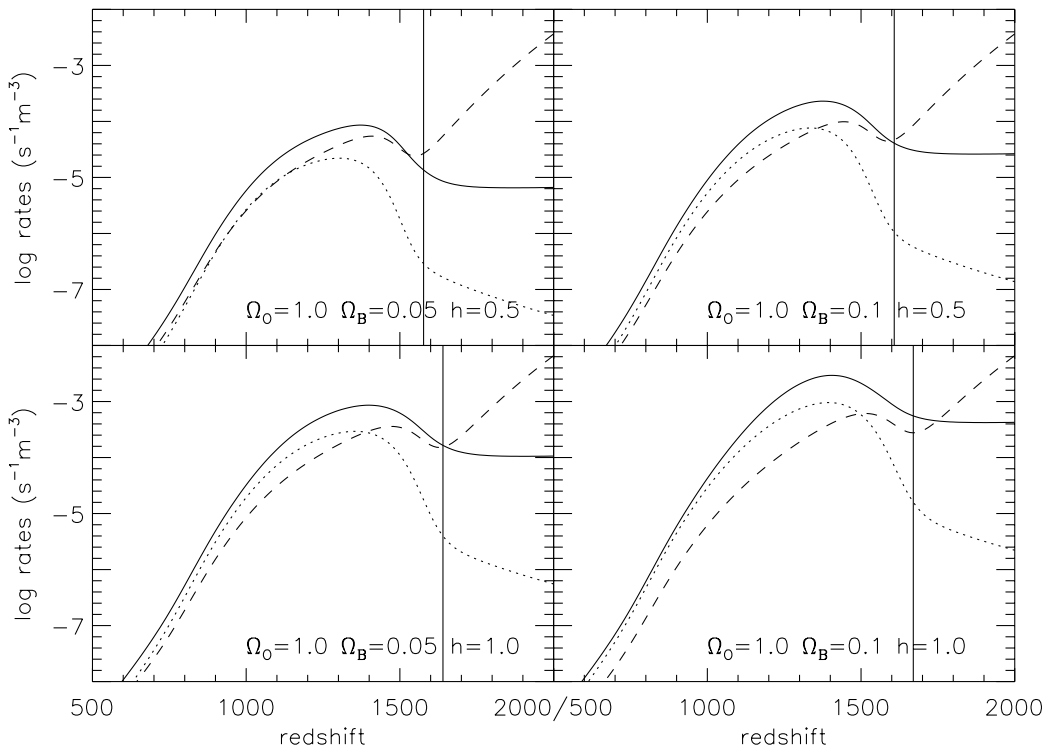
FIG. 12.—What controls H recombination? The net 2p–1s rate (*dashed*) compared to the net 2s–1s two-photon rate (*dotted*) and the net recombination rate (*solid*) for four different cosmologies. Except for low-Ω_B and low-h models (e.g., the sCDM model), the 2s–1s rate dominates. The solid vertical line represents where 5% of the atoms have recombined.

rather than the 2^1s–1^1s rate, as previously stated (e.g., MST). Figure 13 (for He I) also illustrates the slow net recombination rate, which is the primary factor in the slow Case B He I recombination. Figure 14 also illustrates that

He II in the high-Ω_B and h models has a 2s–1s rate faster than the net recombination rate, meaning that there is no slowdown of recombination as a result of $n = 2$, and the Saha equilibrium approximation is valid (see § 3.3.4).
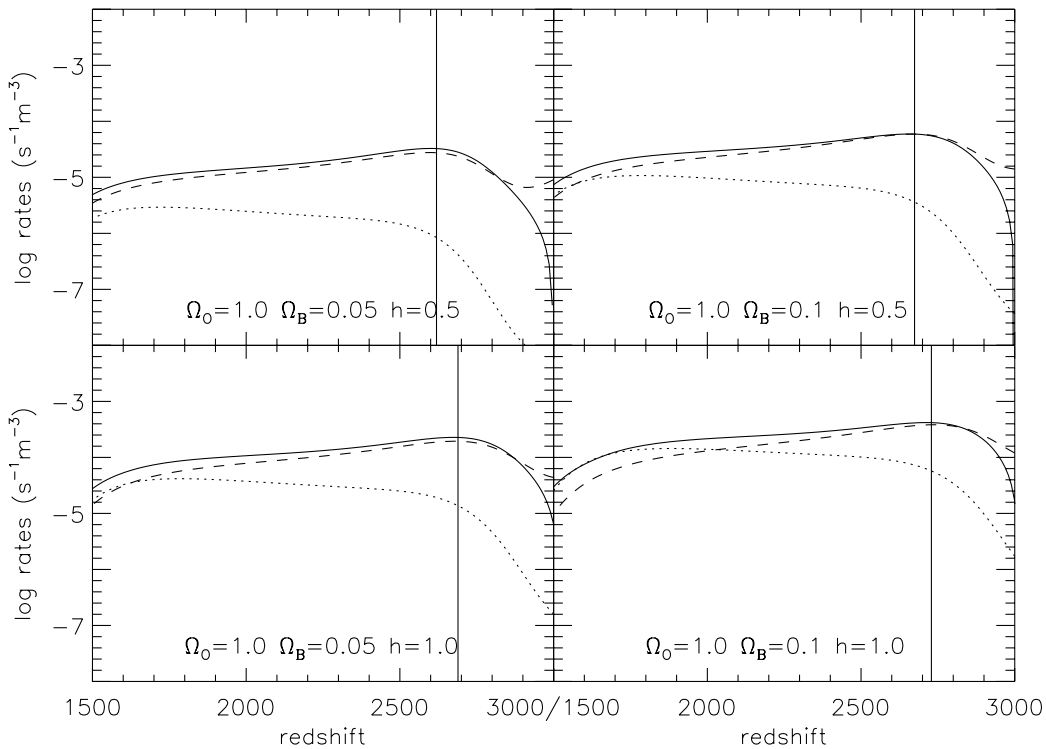

FIG. 13.—What controls He I recombination? The net 2^1p–1^1s rate (*dashed* lines) compared to the net 2^1s–1^1s two-photon rate (*dotted* lines) and the net recombination rate (*solid* lines) for four different cosmologies. Except for high-Ω_B and high-h models, the 2^1p–1^1s rate dominates, in contrast to H. The solid vertical line represents where 5% of the atoms have recombined.
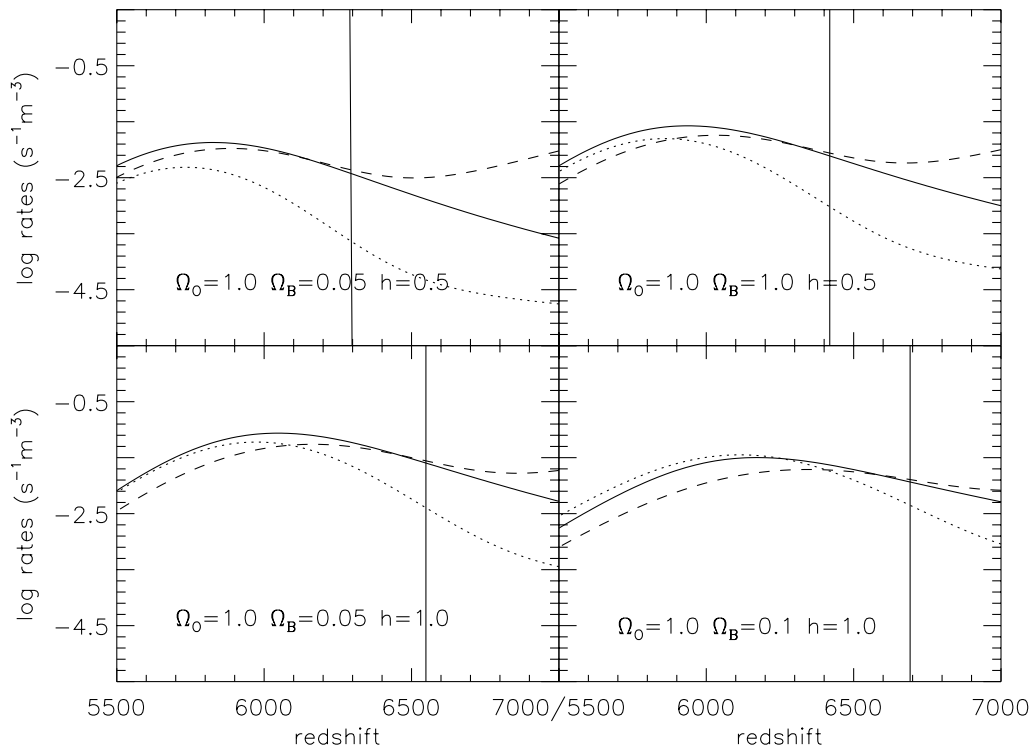
FIG. 14.—What controls He II recombination? The net $2p$–$1s$ rate (*dashed* lines) compared to the net $2s$–$1s$ two-photon rate (*dotted* lines) and the net recombination rate (*solid* lines) for four different cosmologies. Except for low-Ω_B and low-h models, the $2s$–$1s$ rate dominates during recombination, and the $2p$–$1s$ at the start of recombination. The solid vertical line represents where 5% of the atoms have recombined.

The rates change with cosmological model. Physically, this is because all of the rates are very sensitive to the baryon density. The 2^1p–1^1s rates are further affected by the Hubble factor because the Sobolev approximation (eqs. [41] and [40]) depends on the velocity gradient. Whether most of the atoms in the universe recombined via a $2p$–$1s$ or a $2s$–$1s$ two-photon transition depends on the precise values of the cosmological parameters. A confident answer to that question is still not known, given today's parameter uncertainties.

3.4. Atomic Data and Estimate of Uncertainties

Our approach in this work has been to include all relevant degrees of freedom of the recombining matter in a consistent and coupled manner. This requires special attention to the quality of the atomic data used. The challenge lies in building a consistent model for *all* energy levels and transitions, not just for the low-lying ones, which are often better known experimentally and theoretically.

3.4.1. H *and* He II

Hydrogen and the hydrogenic ion of helium have exactly known rate coefficients for radiative processes from precise quantum mechanical calculations (uncertainties below 1%). We use exact values for the bound-bound radiative transitions and for radiative recombination (see Appendix), as in, e.g., Hummer (1994). For more details, see Hummer & Storey (1987), but also Brocklehurst (1970) and Johnson (1972). In particular, the rate of radiative recombination to level n of a hydrogenic ion can be evaluated from the photoionization (bound-free) cross section for level n, $\sigma_{nc}(v)$, with the standard assumption of detailed balance (see § 2.3.1). For hydrogenic bound-free cross sections, we follow, in essence, Seaton's work (Seaton 1959), with its asymptotic expansion for the Gaunt factor (see Brocklehurst 1970).

Note that the weak dependence of the Gaunt factor on wavelength has a noticeable effect in our final recombination rate calculation. Given our application, we do not require the resolution of resonances, as achieved for a few transitions by the Opacity Project (TOPbase; Cunto et al. 1993). Like Hummer (1994), we work with the n-levels assuming that the l sublevels have populations proportional to $2l + 1$. The resulting uncertainties for hydrogenic radiative rates at low temperatures ($T \leq 10^5$ K) certainly do not exceed the 1% level.

Collisional rate coefficients cannot be calculated exactly. So, compared to the hydrogenic radiative rate coefficients, the situation for the bound-bound collisional rates and collisional ionization is poor, with errors typically about 6% and as high as 20% in some cases (Percival & Richards 1978). A number of methods are used to evaluate electron-impact excitation cross sections of hydrogen-like ions (Fisher et al. 1997). These most recent values compare well to the older sources (Johnson 1972; Percival & Richards 1978). The helium ion, He II, is hydrogenic and was treated accordingly. We basically followed Hummer & Storey (1987) and Hummer (1994) in building the model atom. For our application, collisional processes are negligible, so that the large uncertainties that still persist for the collisional rates have no impact on our results.

3.4.2. He I

Helium, in its neutral state, poses a challenge for building a multilevel atomic model of high precision. Unlike atomic hydrogen, no exact solutions to the Schrödinger equation are available for helium. However, very high precision approximations are now available (Drake 1993, 1994), which we have used. These approximations are essentially exact for all practical purposes. The largest relativistic cor-

rection comes from singlet-triplet mixing between states with the same n, L, and J but is still small. Transition rates were calculated following the recent comprehensive He I model built by Smits (1996) and some values in Theodosiou (1987). The source of our photoionization cross sections was TOPbase (Cunto et al. 1993) and Hofsaess (1979) for small n; above $n = 10$, we used scaled hydrogenic values. New detailed calculations (Hummer & Storey 1998) show that the He I photoionization cross sections become strictly hydrogenic at about $n > 20$. The uncertainties in the He I radiative rates are at the 5% level and below.

The situation with the collisional rates for He I is predictably much worse than for He I radiative rates, with good R-matrix calculations existing only for $n \leq 5$ (Sawey & Berrington 1993). The collisional rates at large n are a crucial ingredient in determining the amount of singlet-triplet mixing, but fortunately collisions are not very important for the low-density conditions in the early universe, so the large uncertainty in these rates does not effect our calculation. The Born approximation, which assumes proportionality to the radiative transition rates, is used (see Smits 1996) to calculate the collisional cross sections for large n.

For the 2^1s–1^1s two-photon rate for He I, we used the value $\Lambda_{\text{HeI}} = 51.3 \text{ s}^{-1}$ (Drake et al. 1969), which differs from a previously used value (Dalgarno 1966) by $\sim 10\%$. An uncertainty even of this magnitude would still make little difference in the final results. For the He II $2s$–$1s$ two-photon rate, we used the value $\Lambda_{\text{HeI}} = 526.5 \text{ s}^{-1}$ (for hydrogenic ions this is essentially Z^6 times the value for H) from Lipeles et al. (1965). Dielectronic recombination for He I is not at all important during He I recombination. While dielectronic recombination dominates at temperatures above 6×10^4 K, for the range of temperatures relevant here it is at least 10 orders of magnitude below the radiative recombination rate (using the fit referred to in Abel et al. 1997).

3.4.3. *Combined Error from Atomic Data*

We have gathered together the uncertainties in the atomic data in order to estimate the resulting uncertainty in our derivation of x_e. The atomic data with the dominant effect on our calculation are the set of bound-free cross sections for all H and He I levels—not so much any individual values, but the overall consistency of the sets (which are taken from different sources). The differences between our model atom and Hummer's (1994) reflect the uncertainty in the atomic data. To test the effect on our hydrogenic results, we compared the x_e results of an effective three-level atom using Hummer's (1994) recombination coefficient with the results using a recombination coefficient calculated with our own model H atom. We find maximum differences of 1% at $z = 300$, which corresponds to measurable effects on CMB anisotropies of much less than 1%.

The error in x_e arising from He I is more difficult to calculate. We estimate it to be considerably less than 1% because the low-level ($n \leq 4$) bound-bound and bound-free radiative rates dominate He I recombination, and, as described above, those data are accurate.

3.5. *Secondary Distortions in the Radiation Field*

In our recombination calculation we follow "secondary" distortions in the radiation field that could affect the recombination process at a later time. The secondary distortions are caused by the primary distortions that are frozen into

the radiation field. At a later time they are redshifted into interaction frequency with other atomic transitions. Explicitly, we follow (1) H Lyα photons and (2) H $2s$–$1s$ photons. By the time of H recombination, these photons have been redshifted into an energy range where they could photoionize H ($n = 2$). In addition, we follow (3) He I 2^1p–1^1s and (4) He I 2^1s–1^1s. By the time of H recombination, these photons have been redshifted into an energy range that could photoionize H ($n = 1$). And finally, we also follow (5) He II Lyα photons and (6) He II $2s$–$1s$ photons. By the time of H recombination, these photons have been redshifted into an energy range that could photoionize H ($n = 1$). These He II photons bypass He I because the photons have not been redshifted into a suitable energy range for interaction.

Here we only attempt to investigate the maximum effects of secondary spectral distortions. To that end, we do not include additional distortions that are smaller. For example, Lyman lines other than (1), (3), and (5), whose distortions are smaller than Lyα, will produce a comparably smaller feedback on photoionization. The He I singlet recombination photons could theoretically photoionize He I triplet states, but, as previously discussed, there are virtually no electrons in the triplet states, so this process is also negligible. Another possible effect is due to the similar energy levels of H and He II: $\Delta E_{\text{HeII}} = 4\Delta E_{\text{H}}$. For example, the transition from He II ($n = 4$) to ($n = 2$) produces the same frequency photons as the transition from H ($n = 2$) to ($n = 1$). These transitions are theoretically competing for photons, and this effect can be important for other astrophysical situations (e.g., planetary nebulae) where H and He II simultaneously exist. However, any such effect is negligible for primeval recombination because during He II recombination the amount of neutral H is very small ([H/He II] $< 10^{-8}$), and during H recombination there is almost no He II ([He II/H] $< 10^{-10}$).

Because we are only investigating maximum effects, we assume the photons were emitted at line center and are redshifted undisturbed until their interaction with H ($n = 1$) or H ($n = 2$), as described above. We also assume two photons at half the energy for the $2s$–$1s$ transitions, compared to the $2p$–$1s$ transitions. The distorting photons emitted at a time z_{em} are absorbed at a later time z, where

$$z = z_{\text{em}} \nu_{\text{edge}}/\nu_{\text{em}} . \tag{83}$$

Here ν_{edge} is the photoionization edge frequency where the photons are being absorbed, and ν_{em} is the photon's frequency at emission. The distortions are calculated as

$$
\begin{aligned}
J(\nu, z) = {} & h\nu(z)cp_{ij}(z_{\text{em}}) \\
& \times \left[(n_j(z_{\text{em}})\{A_{ji} + B_{ji} B[\nu_{\text{em}}, T_R(z_{\text{em}})]\} \right. \\
& \left. - n_i(z_{\text{em}})B_{ij} B[\nu_{\text{em}}, T_R(z_{\text{em}})]) \right] ,
\end{aligned}
\tag{84}
$$

where $B[\nu_{\text{em}}, T_R(z_{\text{em}})]$ is the Planck function at the time of emission; A_{ji}, B_{ji}, and B_{ij} are the Einstein coefficients; p_{ij} is the Sobolev escape probability for the line; and the other variables are as described previously.

The distortions (1) and (2) were previously discussed by Rybicki & Dell'Antonio (1993). They pointed out that the effect from (1) should be small because the Lyα distortion must be redshifted by at least a factor of 3 to have any effect. This means that the Lyα photons produced at $z \lesssim 2500$ will only affect the Balmer continuum at $z \lesssim 800$ when the recombination process (and any possibility of photoionization) is almost entirely over.

We find that including the distortions (1)–(6) improves x_e during H recombination at less than the 0.01% level. This difference is far too small to make a significant change in the power spectrum, and it is negligible compared to the major improvements in this paper, which are the level-by-level treatment of H, He I, and He II, allowing departures of the excited state populations from an equilibrium distribution, calculating recombination directly to each excited state, and the correct treatment of He I triplet and singlet states. However, the removal of these distorting photons by photoionization must be taken into account when calculating spectral distortions to the CMB blackbody, which we plan to study in a later paper.

3.6. Chemistry

Including the detailed hydrogen chemistry (see § 2.3.4) marginally affects the fractional abundances of protons and electrons at low z. However, the correction is of the order 10^{-2} x_e at $z < 150$. This change in the electron density would change the Thomson scattering optical depth by $\sim 10^{-5}$, too little to make a difference in the CMB power spectrum.

On the other hand, as shown in Figure 15, the different $x_e(z)$ that we find will lead to fractional changes of similar size in molecular abundances at low z since H_2, for example, is formed via H^-, which is affected by the residual free electron density. The delay in He I recombination compared to previous studies causes a similar delay in formation of He molecules (P. Stancil 1998, private communication). However, with the exception of He_2^+, no changes are greater than those caused by the residual free electron density at freeze-out. Since molecules can be important for the cooling of primordial gas clouds and the formation of the first objects in the universe, the precise determination of molecular abundances is an important issue (e.g., Lepp & Shull 1984; Tegmark et al. 1997; Abel et al. 1997; Galli & Palla 1998). However, the roughly 10%–20% change in the abundance of some chemical species is probably less than other uncertainties in the reaction rates (A. Dalgarno 1998, private communication). With this in mind, we suspect no drastic implications for theories of structure formation.

3.7. Power Spectrum

Even relatively small differences in the recombination history of the universe can have potentially measurable effects on the CMB anisotropies, and so we might expect our two main changes (one in H and one in He) to be noticeable in the power spectrum. As a first example, Figure 16 compares the difference in the anisotropy power spectrum derived from our new $x_e(z)$ to that derived from the standard recombination x_e (essentially identical to that described in HSSW95), for hydrogen recombination only. Here the C_l are squares of the amplitudes in a spherical harmonic decomposition of anisotropies on the sky (the azimuthal index m depends on the choice of axis and so is irrelevant for an isotropic universe). They represent the power and angular scale of the CMB anisotropies by describing the rms temperatures at fixed angular separations averaged over the whole sky (see, e.g., White, Scott, & Silk 1994). These C_l depend on the ionization fraction x_e through the precise shape of the thickness of the photon last-scattering surface (i.e., the visibility function). Since the detailed shape of the power spectrum may allow determination of fundamental cosmological parameters, the significance of the change in x_e is evident. To determine the effect of the change in x_e, we have used the code CMBFAST written and made available by Seljak & Zaldarriaga (1996), with a slight modification to allow for the input of an arbitrary recombination history.

The dominant physical effect arising from the new H calculation comes from the change in x_e at low z. A process seldom mentioned in discussions of CMB anisotropy physics (which is otherwise quite comprehensive; e.g., Hu, Silk, & Sugiyama 1997) is that the low-z tail of the visibility function results in partial erasure of the anisotropies produced at $z \sim 1000$. The optical depth in Thomson scattering back to, for example, $z = 800$ $[\tau = c\sigma_T \int n_e (dt/dz) dz]$ can be several percent. This partial rescattering of the photons leads to partial erasure of the C_l by an amount $e^{-2\tau}$. Let us look at the standard cold dark matter (CDM) calculation first (Fig. 16a). Our change in the optical depth back to $z \simeq 800$ (see Fig. 2) is around 1% less than that obtained using the standard calculation, and so we find that the anisotropies suffer less partial erasure by about 2%. There is no effect on angular scales larger than the horizon at the scattering epoch (here redshifts of several hundred), so that all multipoles are affected except for the lowest hundred or so l's. Hence, this effect is largely a change in the overall normalization of the power spectrum, with some additional differences at low l that will be masked by the "cosmic
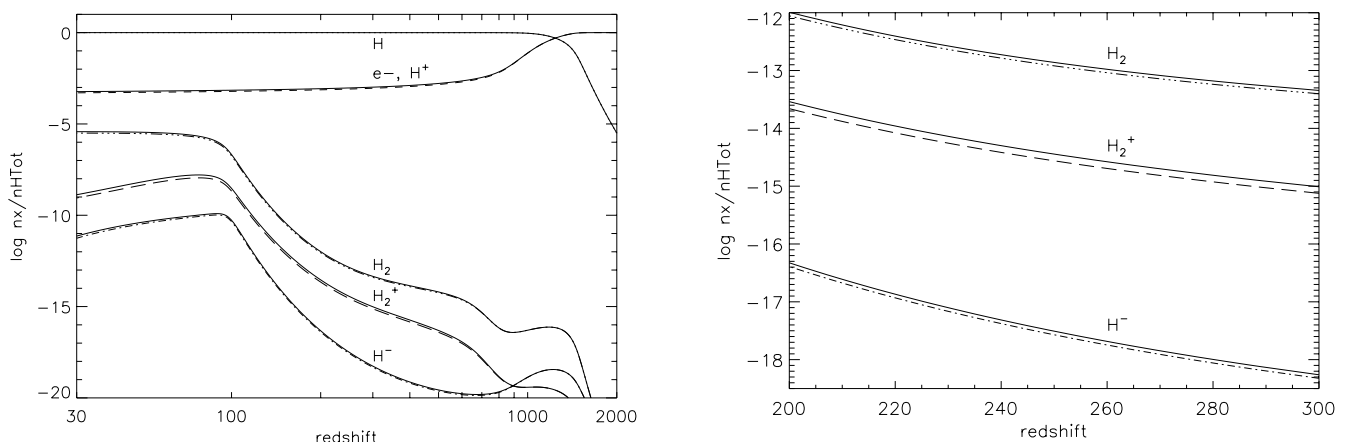

FIG. 15.—The effect of the improved treatment of recombination on H chemistry. Shown is the standard CDM model. Solid lines are values from the standard calculation, dashed and dotted lines from our improved results.
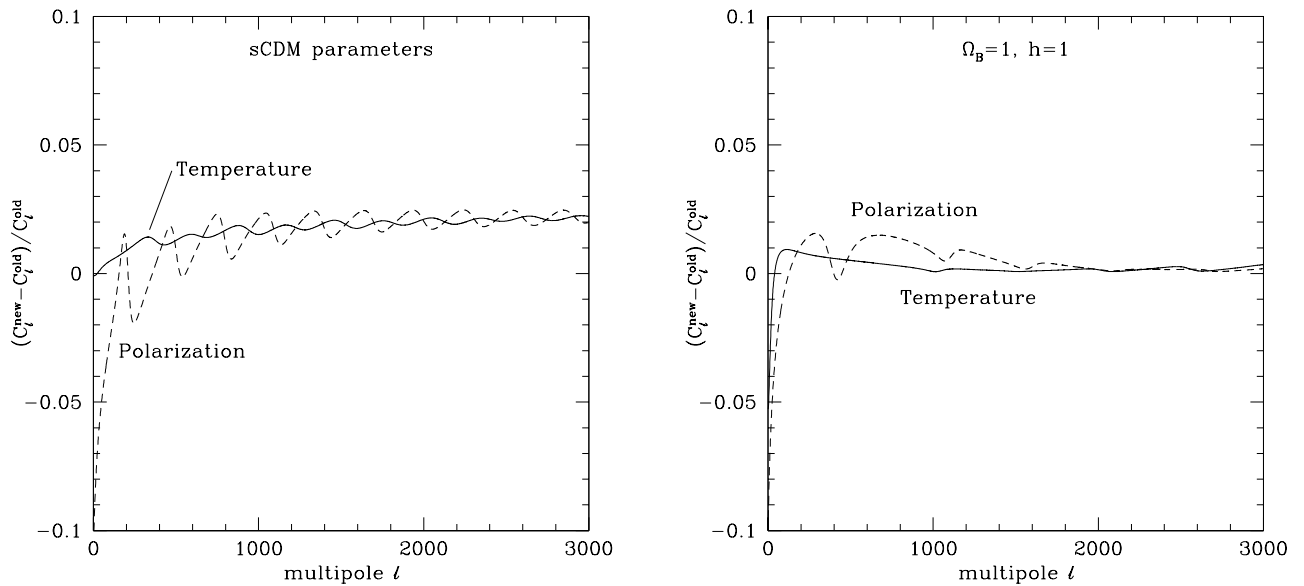
FIG. 16.—Differences in CMB power spectra arising from the improved treatment of hydrogen for (a) the standard CDM parameters $\Omega_{\rm tot} = 1.0$, $\Omega_B = 0.05$, $h = 0.5$, $Y_P = 0.24$, $T_0 = 2.728$ K, and (b) an extreme baryon model with $\Omega_B = 1$, $h = 1.0$ (for which there is relatively little effect). The fractional difference plotted is between our new hydrogen recombination calculation and the standard hydrogen recombination calculation (e.g., in HSSW95), with the sense of $C_l^{\rm new} - C_l^{\rm old}$, and with the two calculations normalized to have the same amplitude for the initial conditions. The solid lines are for temperature, while the dashed lines are for the (" E ") mode polarization power spectrum.

variance." In addition, there are smaller effects as a result of changes in the *generation* of anisotropies in the low-z tail, giving small changes in the acoustic peaks, which can be seen as wiggles in the figure. Since the partial erasing effect is essentially unchanged in the case of the $\Omega_B = h = 1.0$ model, these otherwise subdominant effects are more obvious in Figure 16b.

Differences in the power spectra are rather small in absolute terms, so Figure 16 plots the relative difference. We have shown this for our two chosen models, one being standard cold dark matter (a), which we will refer to as sCDM, and the other being an extreme baryon-only model (b). These models are meant to be representative only, and changes in cosmological parameters will result in curves that differ in detail. We describe how to calculate an approximately correct recombination history for arbitrary models in a separate paper (Seager et al. 1999). Since the main effect is similar to an overall amplitude change, we normalized our CMB power spectra to have the same large-scale matter power spectrum, which is equivalent to normalizing to the same amplitude for the initial conditions. The amplitude of the effect of our new H calculation clearly depends on the cosmology. For the high-Ω_B and h case, the freeze-out value of x_e is much smaller (around 10^{-5}), and since the fractional change in x_e is similarly $\sim 10\%$, the absolute change in ionization fraction is much lower than for the sCDM model. The integrated optical depth is directly proportional to $\Omega_B h \Delta x_e$, which is small, despite the increase in Ω_B and h. Hence, we see a much smaller increase from our hydrogen improvement in Figure 16b. The normalization change is rather difficult to see since it is masked by relatively small changes around the power spectrum peaks, giving wiggles in the difference spectrum.

The dashed lines in Figure 16 show the effect on the power spectrum for CMB polarization. In standard models polarization is typically at the level of a few percent of the anisotropy signal and so will be difficult to measure in detail

(see Hu & White 1997b for a discussion of CMB polarization). We show the results here to indicate that there are further observational consequences of our improved recombination calculation (explicitly, we have plotted the " E " mode of polarization; see, e.g., Seljak 1997). The effect of our improvements on the polarization can be understood similarly through the visibility function. Since the polarization power spectrum tends to have sharper acoustic peaks, the wiggles in the difference spectrum are more pronounced than for the temperature anisotropies. Note that the large relative differences at low l are actually very small in absolute terms since the polarization signal is so small there. The polarization-temperature correlation power spectrum and the " B " mode of polarization (for models with gravity waves) could also be plotted, but little extra insight is gained, and so we avoid this for the sake of clarity.

The other major difference we find compared with previous treatments is in the delayed recombination of He I. In Figure 17, we show the effect of our new He I calculation, again as a fractional change in the CMB anisotropy power spectrum versus multipole l. The change in the recombination of He I affects the density of free electrons just before hydrogen recombination, which in turn affects the diffusion of the photons and baryons and hence the damping scale for the acoustic oscillations that give rise to the peaks in the power spectrum. The phases of the acoustic oscillations will also be affected somewhat, which shows up in the wiggles in the difference spectrum. For CDM-like models, the main effect is the change in the damping scale, since we now think there are more free electrons at $z \sim 1500$–2000. The resulting change in the C_l is essentially the same as assuming the wrong angular scale for the damping of the anisotropies (see Hu & White 1997a), which is the same physical effect that HSSW95 found in arguing for the need to include He I recombination *at all* for obtaining percent accuracy in the C_l. The effect of this improved He I on the
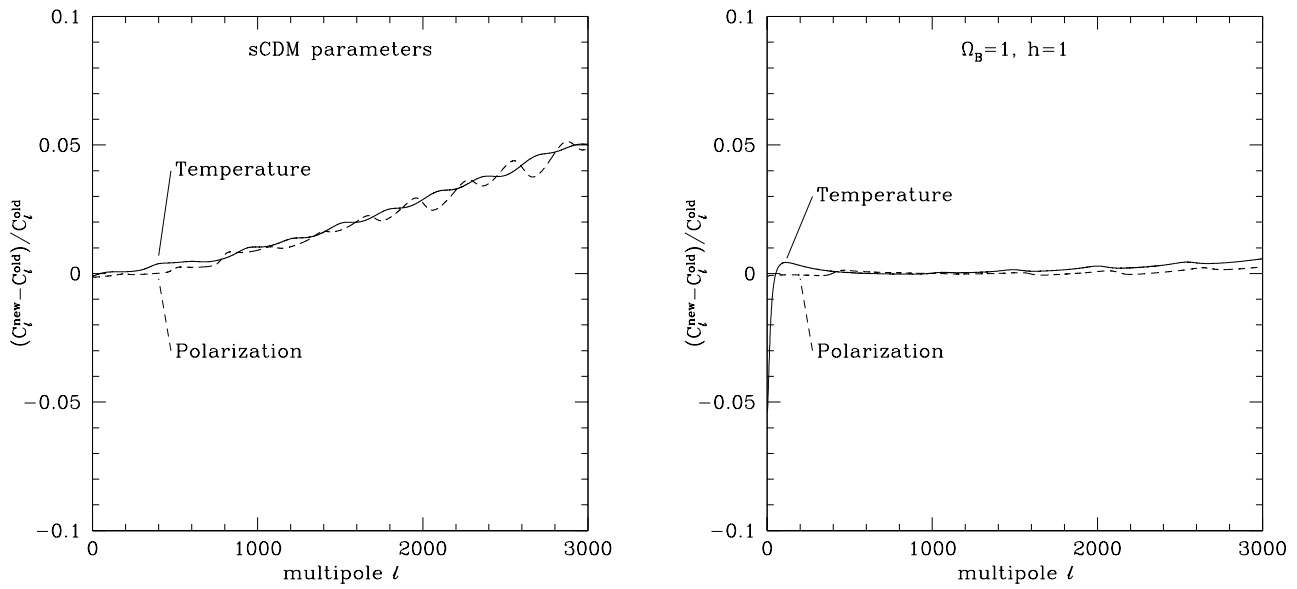
FIG. 17.—Effect of the improved treatment of helium on the CMB power spectra for (a) the standard CDM model, and (b) the extreme baryonic model (for which there is essentially no change). The fractional difference plotted is between our new helium recombination calculation and the assumption that helium follows Saha equilibrium (as in HSSW95), with the same "effective three-level" hydrogen recombination used in both cases. Again the sense is $C_l^{new} - C_l^{old}$, solid lines are temperature, and dashed lines are polarization.

power spectrum will depend on the background cosmology through the baryon density ($\propto \Omega_B h^2$) and the horizon size at last scattering through $\Omega_0 h^2$; hence there is no simple fitting formula, and it is necessary to calculate the effect on the anisotropy damping tail for each cosmological model considered.

There are really two parts to the He I effect. First of all, the extra x_e makes the tight coupling regime tighter, so that the photon mean free path is shorter and the length scale for diffusion is smaller. Second, the effective damping scale comes from an average over the visibility function, so an increase in the high-z tail also leads to a smaller damping scale. The CMB anisotropies can be thought of as a series of acoustic peaks multiplied by a roughly exponential damping envelope, with the characteristic multipole of the

cutoff being determined by the damping length scale. As a result of this smaller damping scale, the high-l part of the power spectrum is less suppressed, and so we see an increase in Figure 17a toward high l. For the $\Omega_B = h = 1$ model (Fig. 17b), we see only a very small effect at the highest l's. This is easily understood by examining Figure 10, where we see that He I recombination is pushed back to higher redshifts than for the sCDM case and also that H recombination happens earlier, which, together with the higher Ω_B and h, shifts the peak of the visibility function to lower redshifts relative to the recombination curve. Hence, the high-z tail of the visibility function is much less affected in this case, and our improved He I calculation has essentially negligible effect. However, for less extreme models we find that the He I effect is always at least marginally significant.
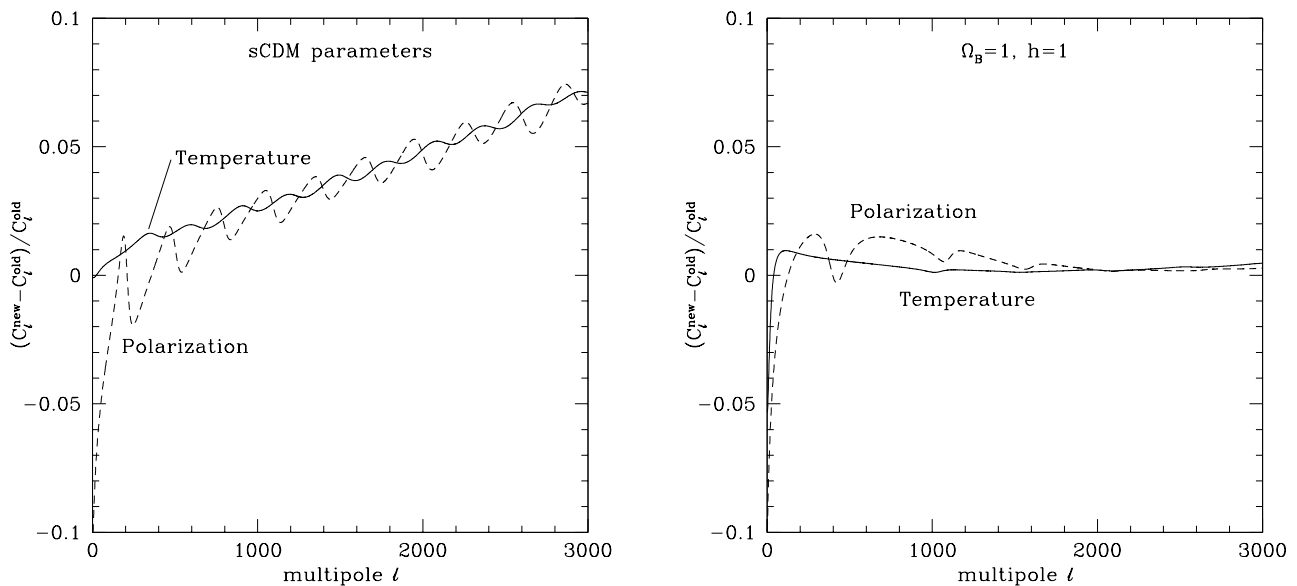

FIG. 18.—The total effect of our improvements on the CMB power spectra for (a) the standard CDM model and (b) an extreme baryonic model. These plots are essentially the sum of the separate effects of hydrogen and helium.

Taking the two main effects together, for the sCDM model, we find that they have essentially the same sign, so that the total effect of our new calculation is more dramatic (see Fig. 18a). Both effects lead to a slight increase in the anisotropies, particularly at small angular scales (high l). Although the exact details will depend on the underlying cosmological model, we see the same general trend for most of the parameters that we have considered. The change can be higher than 5% for the smallest scales, although it should be remembered that the absolute amplitude of the anisotropies at such scales is actually quite low. Nevertheless, changes of this size are well above the level that is relevant for future determinations of the power spectrum. As a rough measure, the cosmic variance at $l \sim 1000$ is about 3%. Hence, a 1% change over a range of say 1000 multipoles is something like a 10 σ effect for a cosmic variance limited experiment. What we can see is that although the effects are far from astonishing, they are at a level that is potentially measurable. Hence our improvements are significant in terms of using future CMB data sets to infer the values of cosmological parameters; if not properly taken into account, these subtle effects in the atomic physics of hydrogen and helium might introduce biases in the determination of fundamental parameters.

3.8. *Spectral Distortions to the CMB*

With the model described in this paper we plan to calculate spectral distortions to the blackbody radiation of the CMB today (see also Dubrovich 1975; Lyubarsky & Sunyaev 1983; Fahr & Loch 1991; Dell'Antonio & Rybicki 1993; Burdyuzha & Chekmezov 1994; Dubrovich & Stolyarov 1995, 1997; Boschan & Biltzinger 1998). The main emission from Lyα and the two-photon process will be in the far-infrared part of the spectrum. Transitions among the very high energy levels, which have very small energy separations, may produce spectral distortions in the radio. Although far weaker than distortions by the lower Lyman lines, they will be in a spectral region less contaminated by background sources.

Detecting such distortions will not be easy since they are generally swamped by Galactic infrared or radio emission and other foregrounds. Our new calculation does not yield any vast improvement in the prospects for detection. However, confirmation of the presence of these recombination lines would be a definitive piece of supporting evidence for the whole big bang paradigm. Moreover, detailed measurement of the lines, if ever possible to carry out, would be a direct diagnostic of the recombination process. For these reasons, we will present spectral results elsewhere.

4. CONCLUSIONS

One point we would like to stress is that our detailed calculation agrees very well with the results of the effective three-level atom. This underscores the tremendous achievement of Peebles, Zeldovich, and colleagues in so fully understanding cosmic recombination 30 years ago. However, the great goal of modern cosmology is to deter-mine the cosmological parameters to an unprecedented level of precision, and in order to do so it is now necessary to understand very basic things, like recombination, much more accurately.

We have shown that improvements upon previous recombination calculations result in a roughly 10% change in x_e at low redshift for most cosmological models, plus a substantial delay in He I recombination, resulting in a few percent change in the CMB power spectrum at small angular scales. Specifically, the low-redshift difference in x_e is due to the H excited states' departure from an equilibrium distribution. This, in turn, comes from the level-by-level treatment of a 300-level H atom, which includes all bound-bound radiative rates, and which allows feedback of the disequilibrium of the excited states on the recombination process. The large improvement in x_e during He I recombination comes from the correct treatment of the atomic levels, including triplet and singlet states. While it was already understood that He I recombination would affect the power spectrum at high multipoles (HSSW95), our improved He I recombination affects even the start of H recombination for traditional low-Ω_B models. There is thus a substantially bigger change in the C_l, reaching to larger angular scales.

Careful use of T_M rather than T_R can also have noticeable consequences, as to a lesser extent can the treatment of Lyα redshifting using the Sobolev escape probability. Our other new contributions to the recombination calculation produce negligible differences in x_e. Collisional excitation and ionization for H, He I, and He II are of little importance. Inclusion of additional cooling and heating terms in the evolution of T_M also produce little change in x_e. The largest secondary spectral distortions do not feed back on the recombination process to a level greater than 0.01% in x_e. Finally, the H chemistry occurs too low in redshift to make any noticeable difference in the CMB power spectrum.

Although we have tried to be careful to consider every process we can think of, it is certainly possible that other subtle effects remain to be uncovered. We hope that we do not have to wait another 30 years for the next piece of substantial progress in understanding how the universe became neutral.

APPENDIX

For reaction rates and cross sections see Table 1. Where cross sections are listed instead of reaction rates, we calculate the reaction rate through the integrals for photoionization (or photodissociation) and recombination, as described in § 2.3.1.

SEAGER, SASSELOV, & SCOTT

TABLE 1

REACTION RATES AND CROSS SECTIONS

Reaction	Rate Coefficient (cm^3 s^{-1})	Reference
$H^- + H \rightarrow H_2 + e^-$	1.30×10^{-9}	1
$H_2 + e^- \rightarrow H^- + H$	$2.70 \times 10^{-8} T_M^{-3/2} \exp(-43000/T_M)$	2
$H_2^+ + H \rightarrow H_2 + H^+$	6.40×10^{-10}	3
$H_2 + H^+ \rightarrow H_2^+ + H$	$2.4 \times 10^{-9} \exp(-21,200/T_M)$	4
$H + H_2 \rightarrow H + H + H$	$1.0 \times 10^{-10} \exp(-52,000/T_M)$	5
$H_2 + e^- \rightarrow H + H + e^-$	$2.0 \times 10^{-9} (T_M/300)^{0.5} \exp(-116,300/T_M)$	6
$H_2^+ + \gamma \leftrightarrow H + H^+$	See expression in reference.	4
$H^- + \gamma \leftrightarrow H + e^-$	See table in reference.	7
$H + \gamma \leftrightarrow H^+ + e^-$	See § 3.4	
$He + \gamma \leftrightarrow He^+ + e^-$	See § 3.4	
$He^+ + \gamma \leftrightarrow He^{++} + e^-$	See § 3.4	

NOTE.—The last five entries refer to cross sections.

REFERENCES.—(1) de Jong 1972, (2) Hirasawa 1969, (3) Karpas et al. 1979, (4) Shapiro & Kang 1987, (5) Dove & Mandy 1986, (6) Gredel & Dalgarno 1995, (7) Wishart 1979; Broad & Reinhardt 1976.

REFERENCES

Abel, T., Anninos, P., Zhang, Y., & Norman, M. L. 1997, NewA, 2, 181
Bond, J. R., & Efstathiou, G. 1998, MNRAS, 304, 75
Bond, J. R., Efstathiou, G., & Tegmark, M. 1997, MNRAS, 291, L33
Boschan, P., & Biltzinger, P. 1998, A&A, 336, 1
Broad, J. T., & Weinhardt, W. P. 1976, Phys. Rev., A14, 2159
Brocklehurst, M. 1970, MNRAS, 148, 417
Burdyuzha, V. V., & Chekmezov, A. N. 1994, AZh, 71, 341
Cen, R. 1992, ApJS, 78, 341
Cunto, W., Mendoza, C., Ochsenbein, F., & Zeippen, C. J. 1993, A&A, 275, L5
Dalgarno, A. 1966, MNRAS, 131, 311
de Jong, T. 1972, A&A, 20, 263
Dell'Antonio, I. P., & Rybicki, G. B. 1993, in ASP Conf. Ser. 51, Observational Cosmology, ed. G. Chincarini et al. (San Francisco: ASP), 548
Dove, J. E., & Mandy, M. E. 1986, ApJ, 311, L93
Drake, G. F. W. 1993, in Long Range Casimir Forces: Theory and Recent Experiments in Atomic Systems, ed. F. S. Levin & D. Micha (New York: Plenum), 107
———. 1994, Adv. At. Mol. Phys., 32, 93
Drake, G. F. W., Victor, G. A., & Dalgarno, A. 1969, Phys. Rev., 180, 25
Dubrovich, V. K. 1975, Soviet Astron. Lett., 1, 196
Dubrovich, V. K., & Stolyarov, V. A. 1995, A&A, 302, 635
———. 1997, Astrophys. Lett., 23, 565
Eisenstein, D. J., Hu, W., & Tegmark, M. 1998, ApJ, 518, 2
Fahr, H. J., & Loch, R. 1991, A&A, 246, 1
Fisher, V. I., Ralchenko, Y. V., Bernshtam, V. A., Goldgirsh, A., Maron, Y., Vainshtein, L. A., Bray, I., & Golten, H. 1997, Phys. Rev. A, 55, 329
Fixsen, D. J., Cheng, E. S., Gales, J. M., Mather, J. C., Shafer, R. A., & Wright, E. L. 1996, ApJ, 473, 576
Galli, D., & Palla, F. 1998, A&A, 335, 403
Goldman, S. P. 1989, Phys. Rev. A, 40, 1185
Gredel, R., & Dalgarno, A. 1995, ApJ, 446, 852
Hirasawa, T. 1969, Prog. Theor. Phys., 42, 3, 523
Hofsaess, D. 1979, At. Data Nucl. Data Tables, 24, 285
Hu, W., Scott, D., Sugiyama, N., & White, M. 1995, Phys. Rev. D, 52, 5498 (HSSW95)
Hu, W., Silk, J., & Sugiyama, N. 1997, Nature, 386, 37
Hu, W., & White, M. 1997a, ApJ, 479, 568
———. 1997b, NewA, 2, 323
Hummer, D. G. 1994, MNRAS, 268, 109
Hummer, D. G., & Storey, P. J. 1987, MNRAS, 224, 801
———. 1998, MNRAS, 297, 1073
Johnson, L. C. 1972, ApJ, 174, 227
Jones, B. J. T., & Wyse, R. F. G. 1985, A&A, 149, 144
Jungman, G., Kamionkowski, M., Kosowsky, A. G., & Spergel, D. N. 1996, Phys. Rev. D, 54, 1332
Karpas, A., Anicich, V., & Huntress, W. T. 1979, J. Chem. Phys., 70, 6, 2877
Krolik, J. H. 1990, ApJ, 353, 21
Lepp, S., & Shull, J. M. 1984, ApJ, 280, 465
Lipeles, M., Novick, R., & Tolk, N. 1965, Phys. Rev. Lett., 15, 690
Lyubarsky, Y. E., & Sunyaev, R. A. 1983, A&A, 123, 171
Matsuda, T., Sato, H., & Takeda, H. 1969, Prog. Theor. Phys., 42, 219
———. 1971, Prog. Theor. Phys., 46, 416
Mihalas, D. M. 1978, in Stellar Atmospheres (San Francisco: Freeman), § 4.3, 5.4, 14.2

Palla, F., Galli, D., & Silk, J. 1995, ApJ, 451, 44
Peebles, P. J. E. 1968, ApJ, 153, 1
———. 1993, Principles of Physical Cosmology (Princeton: Princeton Univ. Press), chap. 6
Penzias, A. A., & Wilson, R. W. 1965, ApJ, 142, 419
Péquignot, D., Petitjean, P., & Boisson, C. 1991, A&A, 251, 680
Percival, I. C., & Richards, D. 1978, MNRAS, 183, 329
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes in C (2d ed.; Cambridge: Cambridge Univ. Press)
Puy, D., Alecian, G., Le Bourlot, J., Léorat, J., & Pineau des Forêts, G. 1993, A&A, 267, 337
Rybicki, G. B. 1984, in Methods in Radiative Transfer, ed. W. Kalkofen (Cambridge: Cambridge Univ. Press), chap. 3
Rybicki, G. B., & Dell'Antonio, I. P. 1993, ApJ, 427, 603
Sasaki, S., & Takahara, F. 1993, PASJ, 45, 655
Sato, H., Matsuda, T., & Takeda, H. 1971, Prog. Theor. Phys. Suppl., 49, 11
Sawey, P. M. J., & Berrington, K. A. 1993, At. Data Nucl. Data Tables, 55, 81
Schramm, D. N., & Turner, M. S. 1998, Rev. Mod. Phys., 70, 318
Scott, D. 1988, Ph.D. thesis, Cambridge Univ.
Seager, S., & Sasselov, D. D. 2000, in preparation
Seager, S., Sasselov, D. D., & Scott, D. 1999, ApJ, 523, L1
Seaton, M. J. 1959, MNRAS, 119, 81
———. 1960, Rep. Prog. Phys., 23, 313
Seljak, U. 1997, ApJ, 482, 6
Seljak, U., & Zaldarriaga, M. 1996, ApJ, 463, 1
Shapiro, P. R., & Kang, H. 1987, ApJ, 318, 32
Smits, D. P. 1996, MNRAS, 278, 683
Smoot, G. F., & Scott, D. 1998, in Caso C., et al., Eur. Phys. J., C3, 1, Rev. Part. Phys.
Sobolev, V. V. 1946, Moving Atmospheres of Stars (Leningrad: Leningrad State Univ. [in Russian]); English transl., 1960 (Cambridge: Harvard Univ. Press)
Stancil, P. C., & Dalgarno, A. 1997, ApJ, 479, 543
Stancil, P. C., Dalgarno, A., Zygelman, B., & Lepp, S. 1996a, BAAS, 188, 1209
Stancil, P. C., Lepp, S., & Dalgarno, A. 1996b, ApJ, 458, 401
Tegmark, M., Silk, J., Rees, M. J., Blanchard, A., Abel, T., & Palla, F. 1997, ApJ, 474, 1
Theodosiou, C. E. 1987, At. Data Nucl. Data Tables, 36, 97
Verner, D. A., & Ferland, G. J. 1996, ApJS, 103, 467
Weymann, R. 1965, Phys. Fluids, 8, 2112
White, M., Scott, D., & Silk, J. 1994, ARA&A, 32, 319
Wishart, A. W. 1979, MNRAS, 187, 59
Zabotin, N. A., & Nasel'skii, P. D. 1982, Soviet Astron., 26, 272
Zaldarriaga, M., Spergel, D. N., & Seljak, U. 1997, ApJ, 488, 1
Zeldovich, Ya. B., Kurt, V. G., & Sunyaev, R. A. 1968, Zh. Eksp. Teor. Fiz., 55, 278; English transl., 1969, Soviet Phys.—JETP Lett., 28, 146
Zeldovich, Ya. B., & Novikov, I. D. 1983, Relativistic Astrophysics, Vol. 2, The Structure and Evolution of the Universe (Chicago: Univ. Chicago Press), § 8.2