Mon. Not. R. Astron. Soc. **367,** 259–274 (2006) doi:10.1111/j.1365-2966.2005.09949.x

Wouthuysen–Field coupling strength and application to high-redshift 21-cm radiation

Christopher M. Hirata★
Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA

Accepted 2005 November 30. Received 2005 November 26; in original form 2005 July 5

ABSTRACT

The first ultraviolet sources in the universe are expected to have coupled the H I spin temperature to the gas kinetic temperature via scattering in the Lyα resonance (the 'Wouthuysen–Field effect'). By establishing an H I spin temperature different from the temperature of the cosmic microwave background, the Wouthuysen–Field effect should allow observations of H I during the reionization epoch in the redshifted 21-cm hyperfine line. This paper investigates four mechanisms that can affect the strength of the Wouthuysen–Field effect that were not previously considered. (1) Photons redshifting into the H I Lyman resonances may excite an H atom and result in a radiative cascade terminating in two-photon $2s_{1/2} \rightarrow 1s_{1/2}$ emission, rather than always degrading to Lyα as usually assumed. (2) The fine structure of the Lyα resonance alters the photon frequency distribution and leads to a suppression of the scattering rate. (3) The spin-flip scatterings change the frequency of the photon and cause the photon spectrum to relax not to the kinetic temperature of the gas but to a temperature between the kinetic and spin temperatures, effectively reducing the strength of the Wouthuysen–Field coupling. (4) Near line centre, a photon can change its frequency by several times the line width in a single scattering event, thus potentially invalidating the usual calculation of the Lyα spectral distortion based on the diffusion approximation. It is shown that (1) suppresses the Wouthuysen–Field coupling strength by a factor of up to $\sim$2, while (2) and (3) are important only at low kinetic temperatures. Effect (4) has a $\leqslant$3 per cent effect for kinetic temperatures $T_{\rm k} \geqslant 2\,{\rm K}$. In particular, if the pre-reionization intergalactic medium was efficiently heated by X-rays, only effect (1) is important. Fitting formulae for the Wouthuysen–Field coupling strength are provided for the range of $T_{\rm k} \geqslant 2\,{\rm K}$ and Gunn–Peterson optical depth $10^5 < \tau_{\rm GP} < 10^7$ so that all of these effects can be easily incorporated into 21-cm codes.

Key words: radiative transfer – intergalactic medium.

1 INTRODUCTION

The cosmic reionization is one of the unexplored frontiers of astrophysics. Currently, we have only a few limited observational constraints on the nature of the intergalactic medium (IGM) during this era and the objects that must have formed during it. The major constraints on reionization currently come from the H I Lyα absorption at a wavelength of $\lambda_{{\rm Ly}\alpha} = 1216\,$Å and from the polarization of the cosmic microwave background (CMB). In particular, observations of complete Lyα absorption at $z \sim 6$ in quasar spectra have pinpointed this epoch as the end of reionization (Becker et al. 2001; Fan et al. 2002), whereas the CMB polarization data from the *Wilkinson Microwave Anisotropy Probe* (*WMAP*) suggest a significant ionization at higher redshifts, e.g. for instantaneous reionization *WMAP* finds reionization at $z = 20^{+10}_{-9}$ (Bennett et al. 2003; Kogut et al. 2003).

While the Lyα and *WMAP* polarization data are currently our best source of information about the early ionization history of the IGM and the ionizing sources responsible for reionization, these techniques leave several fundamental questions unanswered. The Lyα absorption saturates at relatively low neutral fraction $x_{\rm H\,I} \ll 1$ and cannot probe the bulk of the reionization epoch. The CMB polarization does probe the bulk of the reionization epoch, but on the large angular scales of interest, cosmic variance limits the precision with which information can be extracted (Hu & Holder 2003) and polarized foregrounds may prove to be a further limitation. The large-scale polarization also only probes the mean ionization of the universe, and has coarse redshift information. CMB anisotropies on small scales are sensitive to patchy reionization, but these come with no redshift information, so their interpretation could be difficult (Doré, Hennawi & Spergel 2004).

★E-mail: chirata@sns.ias.edu

One promising source of information about the reionization history that overcomes both of these problems is the hyperfine 21.1-cm line of H I. For most of the reionization era, H I is present in significant quantities. Moreover, radio interferometry may make 21-cm inhomogeneities observable across a range of angular scales, and because the 21-cm radiation is a spectral line, frequency information immediately gives the redshift. Thus the 21-cm line has attracted much interest as a probe of the high-redshift IGM (Hogan & Rees 1979; Madau, Meiksin & Rees 1997; Ciardi & Madau 2003; Iliev et al. 2003; Wyithe & Loeb 2004; Zaldarriaga, Furlanetto & Hernquist 2004). Several experiments are currently being performed or planned to observe the high-redshift 21-cm signal, including the Primeval Structure Telescope (Peterson, Pen & Wu 2004), the Low-Frequency Array,[1] the Mileura Wide-Field Array,[2] and the Square Kilometre Array.[3]

The 21-cm line is sensitive to several properties of the IGM including its density, neutral fraction $x_{\rm H\,I}$, and spin temperature $T_{\rm s}$. Before the first ultraviolet (UV) sources turn on, the spin temperature is determined by a competition between the tendency of radiative transitions to bring $T_{\rm s}$ into equilibrium with the CMB at T_γ and the tendency of atomic collisions to bring $T_{\rm s}$ into equilibrium with the gas kinetic temperature $T_{\rm k}$. Loeb & Zaldarriaga (2004) have shown that at redshifts $z \sim 30$ the radiative transitions dominate over collisions in regions of the universe near the mean density. Collisions still dominate in the highest-density regions of the universe such as minihaloes, which can be hotter than the CMB and thus appear in emission (Iliev et al. 2002; Ahn et al. 2005; Kuhlen, Madau & Montgomery 2005). Therefore at these redshifts, the 21-cm signal should consist of very weak absorption from most of the volume, plus emission from the high-density regions.

However, once the first galaxies form, UV radiation is released into the IGM. This radiation can Raman-scatter through the Lyα resonances and convert hydrogen atoms between the two hyperfine levels $F = 0$ and 1. The photons within the Lyα resonance region can exchange energy with H I atoms via the Doppler shift, hence they are expected to come to Boltzmann equilibrium with the gas kinetic temperature, and so the Raman scattering should tend to bring $T_{\rm s}$ into equilibrium with $T_{\rm k}$. This process is known as the Wouthuysen–Field effect, after Wouthuysen (1952) and Field (1958); this effect, together with the CMB and collisions, controls the H I spin temperature during reionization. Once the Wouthuysen–Field effect turns on, one should observe a strong absorption signal at $21(1 + z)$ cm if $T_{\rm k} < T_\gamma$ as expected if the IGM has expanded adiabatically since thermal decoupling from the CMB at $z \sim 200$, or an emission signal if the neutral IGM has been heated efficiently by X-rays (Madau et al. 1997).

The main purpose of this paper is to investigate in more detail the physics of the Wouthuysen–Field effect as applied to the high-redshift IGM. Much progress in this direction has recently been made due to the work of Chen & Miralda-Escudé (2004) and Barkana & Loeb (2005b), who have, respectively, investigated the mean Wouthuysen–Field coupling rate and the perturbations caused by the fluctuating density of galaxies. However, there are several physical effects that were neglected in these papers, but are investigated here. First, it is usually assumed that any UV photon emitted in the band between the Lyman edge at 912 Å and Lyα at 1216 Å, will redshift into a Lyman-series resonance and be degraded to Lyα via a radiative cascade. However, some radiative cascades in H I terminate in the two-photon transition from $2s_{1/2}$ to $1s_{1/2}$, and these produce no Lyα. It is shown that all photons emitted between Lyβ (1026 Å) and Lyγ (973 Å), and most photons between Lyγ and the Lyman edge, are 'lost' in this way. This reduces the Wouthuysen–Field coupling rate since the latter is determined by the flux of Lyα photons. Here, it is shown that the reduction can be as much as a factor of $\sim$2 for hard source spectra.

Secondly, the photon spectrum in the vicinity of Lyα and the associated spin-flip rate are considered in detail, taking into consideration the fine and hyperfine structures of Lyα, the frequency dependence of the spin-flip probability (which was previously assumed to be a constant 4/27), the $\Delta\nu = \pm 1.4$ GHz change of frequency of photons during spin-flip scatterings. Additionally, the validity of treating the Lyα spectral feature via the Fokker–Planck equation (i.e. as a diffusive process) is investigated. These corrections are only important at low kinetic temperatures, since at high $T_{\rm k}$ the smearing of the Lyα line profile by the Doppler effect during repeated scatterings overwhelms the 11-GHz $2p_{1/2}$–$2p_{3/2}$ fine structure splitting and the even smaller hyperfine splitting. For example, they suppress the Wouthuysen–Field effect by $\sim$10 per cent at $T_{\rm k} = 5$ K and $\sim$1 per cent at $T_{\rm k} = 50$ K. Chen & Miralda-Escudé (2004) argue that X-rays from supernovae or X-ray binaries are likely to have heated the IGM to high temperatures ($T_{\rm k} \gg T_\gamma$) well before the end of reionization; if this did indeed happen, then the fine and hyperfine structure effects considered here are completely negligible.

One could ask whether it is worth investigating effects such as two-photon decay or fine and hyperfine structures when there are larger sources of uncertainty in predicting the 21-cm signal during the early stages of reionization, in particular whether or not H_2 cooling is active in low-mass haloes, the star formation efficiency, the initial mass function and the X-ray luminosities of early galaxies. Of course, answering these questions is a major motivation for 21-cm observations. This paper takes the perspective that one can only address these questions if the theoretically tractable parts of the problem, such as the Wouthuysen–Field coupling strength, have been solved. Otherwise, degeneracies exist in the data that cannot be broken, e.g. one could change the emitted UV spectra of the stars and also change how the Lyα production probability $P_{n\rm p}$ depends on quantum level n. Also, one cannot establish that an effect such as the fine structure of Lyα is negligible until it has been calculated.

The results of this paper mostly affect the 21-cm signal during a narrow redshift range near the beginning of reionization. This is because at earlier times there were no UV photons, so the Wouthuysen–Field effect is unimportant, and at later times there were so many UV photons that the Wouthuysen–Field effect is the only important mechanism determining the spin temperature, so that $T_{\rm s} = T_{\rm k}$ regardless of the details. The transition region, in which UV photons compete with the CMB for control of the spin temperature, may have been brief but it is a gold mine of information on early galaxies. For example, Barkana & Loeb (2005b) have suggested that the fluctuations in the UV radiation could be detectable, providing information about the clustering of the first stars.

This paper is organized as follows. Section 2 explains the formalism used to predict the 21-cm signal and defines relevant notation. The main results of the paper are in Section 3, including the calculation of the probabilities for Lyα emission and two-photon decay, the new calculation of the Lyα line profile and a fitting formula for the Wouthuysen–Field coupling efficiency. Section 4 illustrates how the changes in the physics affect the 21-cm signal in two toy models of reionization. Conclusions are given in Section 5.

[1] http://www.lofar.org/
[2] http://web.haystack.mit.edu/arrays/MWA/index.html
[3] http://www.skatelescope.org/

2 HIGH-REDSHIFT H I 21-cm RADIATION

This section reviews the basic theory of the 21-cm radiation from the pre-reionization IGM. More details can be found in the references.

The brightness temperature of the 21-cm signal is determined by the spin temperature T_s of the H I according to the relation (e.g. Zaldarriaga et al. 2004)

$$T_b = \frac{3c^3\hbar A_{10} n_H x_{HI}}{16\nu_{10}^2 k_B (1+z)^2 (d\nu_\parallel/dr_\parallel)} \left(1 - \frac{T_\gamma}{T_s}\right), \quad (1)$$

where $d\nu_\parallel/dr_\parallel$ is the physical velocity gradient at redshift z; A_{10} is the intrinsic width of the $F = 1$ hyperfine level; $\nu_{10} = 1.42\,\mathrm{GHz}$ is the H I hyperfine transition frequency; n_H is the proper number density of hydrogen nuclei; x_{HI} is the fraction of hydrogen that is neutral; $T_\gamma = 2.73(1+z)\,\mathrm{K}$ is the CMB temperature; and T_s is the H I spin temperature. In the linear regime, the velocity field is related to the matter field by (Bharadwaj & Ali 2004; Barkana & Loeb 2005a)

$$\frac{d\nu_\parallel}{dr_\parallel} = \frac{H(z)}{1+z}\left(1 + f\partial_\chi \nabla^{-2}\delta\right). \quad (2)$$

Here, $f = \mathrm{d}\left[\ln(D/a)\right]/\mathrm{d}\ln a$ depends on the growth factor and $f \approx 1$ in the matter-dominated era (a good approximation at the redshift of reionization), and χ is the comoving radial distance. Plugging in numbers from the currently favoured cosmology gives

$$T_b \approx 28\,\mathrm{mK}\left(\frac{1+z}{10}\right)^{1/2}\left(1 - \frac{T_\gamma}{T_s}\right)\left(1 + f\partial_\chi\nabla^{-2}\delta\right)^{-1}. \quad (3)$$

Here χ is comoving radial distance. In the second line the homogeneous universe and peculiar velocity terms have been separated out from each other.

The spin temperature is determined by three effects: the radiative coupling to the CMB, the Wouthuysen–Field and collisional coupling to the gas kinetic temperature T_k. These effects compete to determine the fraction y of hydrogen atoms in the $F = 1$ excited hyperfine level. This fraction is related to the spin temperature via

$$\frac{y}{1-y} = 3e^{-T_\star/T_s} \quad \rightarrow \quad y = \frac{3}{3 + e^{T_\star/T_s}}, \quad (4)$$

where $T_\star = h\nu_{10}/k_B = 68.2\,\mathrm{mK}$. Sometimes we will write the populations of the excited and ground hyperfine levels $y_1 = y$ and $y_0 = 1 - y$ for simplicity. At $T_s \gg T_\star$, one may use the approximation

$$y \approx \frac{3}{4} - \frac{3T_\star}{16T_s}. \quad (5)$$

The evolution of y can be broken into its CMB, Wouthuysen–Field and collisional terms,

$$\dot{y} = \dot{y}_\gamma + \dot{y}_\alpha + \dot{y}_c. \quad (6)$$

The radiative term is given by

$$\dot{y}_\gamma = -\frac{4A_{10}T_\gamma}{T_\star}\left(y - \frac{3}{4} + \frac{3T_\star}{16T_\gamma}\right), \quad (7)$$

where T_γ is the photon temperature. The factor of $4T_\gamma/T_\star$ in front accounts for the acceleration of the radiative transition via stimulated emission and absorption (which contributes a factor of 3 since the $F = 0$ state can be excited to any of the three $F = 1$ states), which dominate over spontaneous emission for $T_\gamma \gg T_\star$. The collisional term is

$$\dot{y}_c = -\frac{4A_{10}T_\gamma}{T_\star}x_c\left(y - \frac{3}{4} + \frac{3T_\star}{16T_k}\right), \quad (8)$$

where

$$x_c = \frac{\kappa_{10} n_H T_\star}{A_{10} T_\gamma} \quad (9)$$

and κ_{10} is the collisional rate coefficient (Zygelman 2005).[4] The Wouthuysen–Field rate is

$$\dot{y}_\alpha = -\frac{4A_{10}T_\gamma}{T_\star}x_\alpha\left(y - \frac{3}{4} + \frac{3T_\star}{16T_k}\right). \quad (10)$$

This is given by

$$x_\alpha = \frac{8\pi\lambda_{Ly\alpha}^2\gamma T_\star}{9A_{10}T_\gamma}S_\alpha J_\alpha, \quad (11)$$

where $\gamma = 50\,\mathrm{MHz}$ is the half width at half-maximum (HWHM) of the Lyα resonance. Here, J_α is the flux of Lyα photons (in $\mathrm{cm^{-2}\,s^{-1}\,Hz^{-1}\,sr^{-1}}$), and S_α is a factor of order unity that accounts for spectral distortions. Chen & Miralda-Escudé (2004) provide values for S_α that are typically of order unity. In this paper, the values of S_α are revised downward slightly after accounting for several new processes that affect the colour temperature and spectral profile of the Lyα feature.

The final spin temperature is the steady-state solution to equation (6),

$$1 - \frac{T_\gamma}{T_s} = \frac{x_\alpha + x_c}{1 + x_\alpha + x_c}\left(1 - \frac{T_\gamma}{T_k}\right). \quad (12)$$

3 WOUTHUYSEN–FIELD COUPLING EFFICIENCY

Lyα photons are produced in neutral regions of the universe in one of two ways: either photons can be cosmologically redshifted into the Lyα resonance, or they can be emitted as part of the radiative cascade to the H I ground state following capture of a higher-order Lyman-series photon. Once produced, Lyα photons couple the spin temperature of H I to the gas kinetic temperature via the Wouthuysen–Field mechanism until they are redshifted out of the resonance. This section computes the Wouthuysen–Field coupling rate as a function of the radiation field entering each of the Lyman lines. Section 3.1 computes the probability that a photon entering a Lyman-series resonance cause a radiative cascade in the excited H I atom that terminates with a two-photon decay from the $2s_{1/2}$ level and produces no Lyα. Decay of an H I atom from 2s involves a competition between the two-photon process and collisions that transfer the atom to 2p; only the latter yields Lyα photons (Spitzer & Greenstein 1951; Seaton 1955a). The usual assumption is that the Lyα-producing channels dominate; however in the IGM the opposite is true: collisions are negligible (see Appendix A). Section 3.2 investigates the effect of fine and hyperfine structure and frequency changes during spin-flip events using the Fokker–Planck equation. Fitting formulae for these results are presented in Section 3.3. Section 3.4 tests the assumptions of the Fokker–Planck equation by comparing its predictions to Monte Carlo simulations that are computationally intensive but do not make any approximations to the frequency redistribution matrix. There it is shown that the Fokker–Planck equation reproduces the Wouthuysen–Field coupling strength predicted by the simulations to within $\leqslant 3$ per cent at $T_k \geqslant 2\,\mathrm{K}$.

3.1 Lyα production efficiency

H I in the IGM is normally found in its ground configuration, 1s. If a photon is emitted into the IGM at energies between the Lyα

[4] This is roughly given by $\kappa_{10} \approx 4\tilde{\kappa}_{10}/3$, where $\tilde{\kappa}_{10}$ is the coefficient tabulated by Allison & Dalgarno (1969).

resonance at 10.2 eV and the Lyman edge at 13.6 eV, it redshifts cosmologically until it reaches one of the Lyman-series resonances. Because the Lyman lines in the neutral IGM are optically thick, the photon will be absorbed and one H I atom is boosted into the np configuration ($n \geqslant 2$). The excited state is unstable and decays through a radiative cascade. Ultimately the cascade ends in one of three possibilities: (i) an Lyα photon is emitted from the 2p configuration, leaving the H I atom in the ground configuration; (ii) the H I atom reaches the metastable 2s configuration or (iii) the H I atom decays directly from n'p (with $2 < n' < n$) to 1s, emitting a higher-order (Lyβ, Lyγ, etc.) photon. In case (iii), the emitted photon immediately re-excites an H I atom to the n'p configuration; the process of absorption and re-emission ultimately terminates in either (i) or (ii). In case (i), the original photon is downgraded to Lyα. Appendix A shows that in case (ii) the atom in the 2s configuration decays almost always via two-photon emission. The latter process, of course, produces no Lyα. Thus the Lyα photon production rate depends on the branching fractions for cases (i) and (ii), which are evaluated next.

[There is so much H I in the early universe that some of the electric quadrupole lines $1s_{1/2} \rightarrow nd_{3/2,5/2}$ are optically thick. Since some of these lines have slightly higher energy than the electric dipole lines $1s_{1/2} \rightarrow np_{1/2,3/2}$ due to fine structure, one might worry that a photon will redshift into the quadrupole resonance first and excite a hydrogen atom to the nd rather than np configuration. However a simple calculation shows that for $n \geqslant 3$, the fine structure splitting $29n^{-3}$ GHz between $np_{3/2}$ and $nd_{5/2}$ levels is less than the $1.0(1 - n^{-2})T^{1/2}\,\mathrm{K}^{-1/2}$ GHz Doppler width of the Lyman line for the temperatures $T \geqslant 2$ K expected in the IGM. The splitting between $np_{3/2}$ and $nd_{3/2}$ is even less, as it is due to Lamb shifts and hyperfine splitting. Thus for the purposes of photon absorption, the $np_{3/2}$ and $nd_{3/2,5/2}$ levels are degenerate and absorption occurs in the stronger electric dipole line.]

Let us define P_{nl} to be the probability for an H I atom in the nl configuration to decay ultimately via Lyα emission. The 2s $\rightarrow$ 1s two-photon emission probability is then $1 - P_{nl}$. The probabilities can be determined iteratively via the usual equation

$$P_{nl} = \frac{\sum_{n'=2}^{n-1} \sum_{l'=0}^{n'-1} A_{nl \rightarrow n'l'} P_{n'l'}}{\sum_{n'=2}^{n-1} \sum_{l'=0}^{n'-1} A_{nl \rightarrow n'l'}}, \tag{13}$$

where the $A_{nl \rightarrow n'l'}$ are the decay rate coefficients (in e.g. s^{-1}) to the specified states. The n p $\rightarrow$ 1s rate is removed from the sum since it results in a Lyman-series photon that immediately re-excites a hydrogen atom to np, and decays from non-p states to 1s are forbidden. Since an H I atom in the 2s configuration always undergoes two-photon emission, whereas an atom in 2p undergoes Lyα emission, equation (13) can be initialized with $P_{2s} = 0$ and $P_{2p} = 1$. The resulting probabilities for producing Lyα photons are shown in Fig. 1 and Table 1. Note in particular that all photons that redshift into Lyα end up in the Lyα resonance ($P_{2p} = 1$), whereas none of the photons that redshift into Lyβ do ($P_{3p} = 0$) because the 3p configuration always decays to 1s or 2s on account of electric dipole selection rules. Photons entering higher-order Lyman resonances can go either way ($0 < P_{np} < 1$).

3.2 Scattering rate

The efficiency of Wouthuysen–Field coupling is determined by the Lyα spin-flip rate x_α and the degree to which the photon spectrum in the vicinity of Lyα has relaxed to the gas kinetic temperature T_k. It is generally believed that relaxation of the colour temperature to T_k is complete if the optical depth through the Lyα resonance (i.e. the Gunn–Peterson depth τ_GP) is high enough; Deguchi & Watson (1985) showed that if Lyα can be treated as a single line, this relaxation occurs for $\tau_\mathrm{GP} \geqslant 10^5$, which holds at all redshifts prior to reionization. The usual computation also assumes that each Lyα scattering by an H I atom in the $1s_{1/2}(F = 1)$ level has a 4/27 probability of transferring the atom to $1s_{1/2}(F = 0)$, as computed by Field (1958, 1959).

The 4/27 probability was derived assuming that the $J(\nu)$ is constant across the Lyα multiplet. While this is appropriate in the context of ISM studies where $k_\mathrm{B}T_\mathrm{k}/h$ is much greater than the width of the Lyα spectral feature (Field 1959), the pre-reionization IGM may have been cold, with the minimum temperature determined by the onset of X-ray heating (Chen & Miralda-Escudé 2004). In this case, the use of frequency-averaged cross-sections, as in Field (1958), is no longer valid and one must treat the line profile in detail. This is done in Appendix B, where the Lyα line profile is broken into the parts $\phi_{F_\mathrm{i}F_\mathrm{f}}(\Delta\nu)$ that give the rate of scattering from initial total spin $F_\mathrm{i} \in \{0, 1\}$ to final $F_\mathrm{f} \in \{0, 1\}$. Also, the Wouthuysen–Field coupling implies some transfer of energy between the Lyα photons and the hydrogen spins, hence the colour temperature relaxes not to T_k but to some value intermediate between T_k and T_s. This effect reduces x_α because the Wouthuysen–Field energy transfer rate

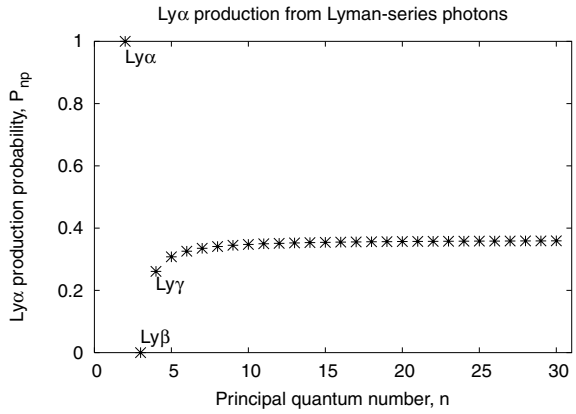

Figure 1. The probabilities P_{np} of producing a Lyα photon following excitation of H I to the np configuration. For example, if a photon redshifts into the Lyγ resonance (1s $\rightarrow$ 4 p), there is a probability $P_{4p} = 0.26$ that the photon is degraded to Lyα and a probability $1 - P_{4p} = 0.74$ that it is lost to two-photon emission and never contributes to the Wouthuysen–Field coupling.

Table 1. The probabilities P_{np} of producing a Lyα photon following excitation of H I to the np configuration.

n	P_{np}	n	P_{np}	n	P_{np}
		11	0.3496	21	0.3572
2	1.0000	12	0.3512	22	0.3575
3	0.0000	13	0.3524	23	0.3578
4	0.2609	14	0.3535	24	0.3580
5	0.3078	15	0.3543	25	0.3582
6	0.3259	16	0.3550	26	0.3584
7	0.3353	17	0.3556	27	0.3586
8	0.3410	18	0.3561	28	0.3587
9	0.3448	19	0.3565	29	0.3589
10	0.3476	20	0.3569	30	0.3590

contains the temperature difference between T_c and T_s, instead of between T_k and T_s.

This section introduces this new physics to obtain the Lyα spectral distortion and to compute x_α. It is based on the treatment of the photon spectrum using the Fokker–Planck equation, which assumes that the change in frequency $\delta\nu$ in a single scattering event is small in comparison to the frequency scale over which the photon intensity or the scattering coefficients change. These assumptions are not strictly valid, and for this reason Section 3.4 will be devoted to testing them.

A distinction is made between 'continuum' photons that cosmologically redshift into Lyα, and 'injected' photons that are produced as part of a radiative cascade. We find that in terms of the Wouthuysen–Field coupling, there is very little difference between these, in accordance with the results of Chen & Miralda-Escudé (2004).

The kinetic temperature range considered will be $T_k \geqslant 2\,\mathrm{K}$, which occurs if the universe cools adiabatically until $z = 9$. In practice, lower temperatures were probably not reached: the universe may have been partially or fully reionized by $z = 9$, and even inefficient heating sources such as Lyα heating could have kept the universe warmer than 2 K throughout reionization (e.g. Chen & Miralda-Escudé 2004).

3.2.1 The Lyα spectral distortion

The steady-state Fokker–Planck equation used by Chen & Miralda-Escudé (2004) is easily modified to include the full (non-Voigt) line profile. In the vicinity of the Lyα resonance, the equation can be written as

$$\frac{\partial}{\partial\nu}\left(-\mathcal{A}J + D\frac{\partial J}{\partial\nu}\right) + C\psi(\nu) = \dot{J}(\nu) = 0, \tag{14}$$

where $\mathcal{A}$ is the frequency drift (in $\mathrm{Hz\,s^{-1}}$), D is the frequency diffusivity (in $\mathrm{Hz^2\,s^{-1}}$), C is the photon source term and ψ is the frequency distribution with which photons are injected. The drift and diffusivity can be decomposed as

$$\mathcal{A} = \mathcal{A}_H + \mathcal{A}_k + \mathcal{A}_s \tag{15}$$

and

$$D = D_k + D_s + D_{int}. \tag{16}$$

This includes terms due to Hubble expansion (subscript H), kinetic/Doppler coupling (k) and spin coupling (s). The diffusion term contains an 'interference' contribution if the diffusion due to kinetic coupling is correlated with that due to spin coupling; it is shown later in this section that D_{int} can be neglected. (Hubble expansion causes a drift in the frequency, but no diffusion.) The Hubble expansion term is trivial, $\mathcal{A}_H = -H\nu_{Ly\alpha}$.

The kinetic and spin diffusion terms can be worked out from the usual Fokker–Planck rules, which state that for any process X,

$$\mathcal{A}_X = \Gamma_{scat}\langle\delta\nu_X\rangle \tag{17}$$

and

$$D_X = \frac{1}{2}\Gamma_{scat}\left\langle\delta\nu_X^2\right\rangle, \tag{18}$$

where Γ_{scat} is the scattering rate (in $\mathrm{s^{-1}}$) and $\langle\delta\nu_X\rangle$ and $\langle\delta\nu_X^2\rangle$ are the mean change in frequency and mean square change in frequency during a scattering. The (spin-averaged) scattering rate is

$$\Gamma_{scat} = \frac{3}{2}\lambda_{Ly\alpha}^2\gamma n_H x_{H1} c\bar{\phi}(\nu), \tag{19}$$

where

$$\bar{\phi}(\nu) = \frac{1}{4}(\phi_{00} + \phi_{01}) + \frac{3}{4}(\phi_{10} + \phi_{11}) \tag{20}$$

is the spin-averaged cross-section appropriate for $T_s \gg T_\star$; cf. equation (B17). As usual with Fokker–Planck equations, the drift and diffusion terms obey an Einstein relation

$$\mathcal{A}_X = -\frac{h}{k_B T_X} D_X, \tag{21}$$

where T_X is the temperature of the reservoir with which the photon exchanges energy during process X. Here X is either k (Doppler shift, for which T_k appears in equation 21) or s (spin coupling with T_s). Physically, the kinetic drift term $\mathcal{A}_k$ corresponds to the loss of photon energy due to atomic recoil. The spin drift term $\mathcal{A}_s$ corresponds to the loss of photon energy due to having more atoms in the $F = 0$ than $F = 1$ level, so that if the photon spectrum were flat ($\mathrm{d}J/\mathrm{d}\nu = 0$) the photons would on average lose more energy in spin-flip excitations than they gain in de-excitations.

The kinetic diffusion has been worked out by Rybicki & Dell'Antonio (1994),[5] with the result that $\langle\delta\nu_k^2\rangle = 2\sigma_\nu^2$ and hence

$$D_k = \frac{3}{2}\lambda_{Ly\alpha}^2\gamma n_H x_{H1} c\sigma_\nu^2\bar{\phi}(\nu), \tag{22}$$

where σ_ν is the 1σ Doppler width. [In the Fokker–Planck approximation, and for an isotropic situation, the angular dependence of the cross-section enters into equation (22) only through the combination $\langle 1 - \boldsymbol{n}\cdot\boldsymbol{n}'\rangle$ where $\boldsymbol{n}$ and $\boldsymbol{n}'$ are the incoming and outgoing photon directions; see equations (A15) and (A16) of Rybicki & Dell'Antonio (1994). So long as only the electric dipole transitions are involved in scattering, the probability for scattering into direction $\boldsymbol{n}'$ is the same as that into $-\boldsymbol{n}'$, and the angular dependence requires no modification to equation (22).] Equation (21) then gives $\mathcal{A}_k$.

Chen & Miralda-Escudé (2004) included in their Fokker–Planck equation only the Hubble drift, kinetic drift and kinetic diffusivity (in their equation 13, the Hubble drift is the γ_S term and the kinetic drift is the η term). However, the hyperfine splitting of the ground state allows a photon to change its frequency during scattering by $\pm\nu_{10}$, even in the centre-of-mass frame. This results in spin contributions to the drift and diffusivity. Spin diffusivity results only from those Lyα scattering events that change the total spin state of the atom; in the limit $T_s \gg T_\star$,

$$D_s = \frac{3}{4}\lambda_{Ly\alpha}^2\gamma n_H x_{H1} c\nu_{10}^2\left(\frac{1}{4}\phi_{01} + \frac{3}{4}\phi_{10}\right). \tag{23}$$

Equation (21) can then be used to obtain $\mathcal{A}_s$.

Finally, there is the interference diffusivity D_{int} in equation (16). This term comes from the fact that one cannot exactly separate $\langle\delta\nu^2\rangle$ into kinetic and spin parts, $\langle\delta\nu_k^2\rangle + \langle\delta\nu_s^2\rangle$, and is equal to

$$D_{int} = \Gamma_{scat}\langle\delta\nu_k\delta\nu_s\rangle. \tag{24}$$

In our particular case, the deviation of $\delta\nu_k$ from its mean value $\langle\delta\nu_k\rangle$ is proportional to $\boldsymbol{n}\cdot\boldsymbol{n}'$. However as argued above, the probabilities of scattering the photon in directions $\boldsymbol{n}'$ and $-\boldsymbol{n}'$ are equal; the same argument holds for the conditional probabilities for fixed final spin F_f. Therefore $\delta\nu_k$ is uncorrelated with $\delta\nu_s$, and

[5] Rybicki & Dell'Antonio (1994) work in terms of the variable x, which is related to the detuning by $\Delta\nu = \sqrt{2}\sigma_\nu x$. In this paper, including equation (22), I have converted to $\Delta\nu$.

$D_{\rm int} = \Gamma_{\rm scat}\langle\delta\nu_{\rm k}\rangle\langle\delta\nu_{\rm s}\rangle$. Combining with equations (15), (16) and (21) shows that

$$\frac{D_{\rm int}}{\sqrt{D_{\rm k}D_{\rm s}}} = \frac{h\left\langle\delta\nu_{\rm k}^2\right\rangle^{1/2}}{k_{\rm B}T_{\rm k}}\frac{h\left\langle\delta\nu_{\rm s}^2\right\rangle^{1/2}}{k_{\rm B}T_{\rm s}}. \quad (25)$$

It is readily verified that $h\langle\delta\nu_{\rm X}^2\rangle^{1/2} \ll k_{\rm B}T_{\rm X}$ for X = k, s so long as $T_{\rm k} \gg (h\nu_{{\rm Ly}\alpha})^2/k_{\rm B}m_{\rm p}\,c^2 = 1.3\,{\rm mK}$ and $T_{\rm s} \gg T_\star$, respectively. Both of these conditions are easily satisfied in the IGM, and so $D_{\rm int} \ll \sqrt{D_{\rm k}D_{\rm s}}$. Since it is strictly true that $\sqrt{D_{\rm k}D_{\rm s}} \leqslant (D_{\rm k}+D_{\rm s})/2$, $D_{\rm int}$ can be dropped.[6]

Equation (14) can be solved by the method of Chen & Miralda-Escudé (2004), which consists of first reducing it to first order,

$$-\mathcal{A}(\nu)J(\nu) + D(\nu)\frac{\partial J(\nu)}{\partial\nu} = -\mathcal{A}(-\infty)J(-\infty) - C\int_{-\infty}^{\nu}\psi(\nu')\,{\rm d}\nu', \quad (26)$$

and then applying an ordinary differential equation (ODE) solver starting from $\nu = -\infty$ and working upward in frequency.[7] Here $J(-\infty) = J_\alpha$ is the total flux at Lyα (including both continuum and injected photons) and C is determined only by the injected photons. One can determine C as follows: substituting $\nu = +\infty$ into equation (26), and recalling that ψ integrates to unity, one finds

$$-\mathcal{A}(+\infty)J(+\infty) = -\mathcal{A}(-\infty)J(-\infty) - C, \quad (27)$$

implying

$$C = -\mathcal{A}(-\infty)[J(+\infty) - J(-\infty)] = H\nu_{{\rm Ly}\alpha}J_\alpha({\rm inj}). \quad (28)$$

3.2.2 *Effect on spin temperature*

Once a solution to equation (14) is obtained, one can go back and estimate the Wouthuysen–Field effect on the spin temperature. The rate per atom $\Gamma_{\alpha 10}$ for converting $F = 1$ hydrogen atoms to $F = 0$ is

$$\Gamma_{10} = 4\pi\int J(\nu)\sigma(1\rightarrow 0;\nu)\,{\rm d}\nu = 6\pi\lambda_{{\rm Ly}\alpha}^2\gamma J_\alpha\int\frac{J(\nu)}{J_\alpha}\phi_{10}(\nu)\,{\rm d}\nu, \quad (29)$$

and a similar rate holds for $F = 0 \rightarrow 1$ conversions. If y is the fraction of hydrogen atoms in the excited hyperfine level $F = 1$, then the Wouthuysen–Field contribution to $\dot{y}$ is

$$\dot{y}_\alpha = (1-y)\Gamma_{01} - y\Gamma_{10} = (\Gamma_{01}+\Gamma_{10})(y - y_{\alpha,{\rm ss}}), \quad (30)$$

where

$$y_{\alpha,{\rm ss}} = \frac{\Gamma_{01}}{\Gamma_{01}+\Gamma_{10}} \quad (31)$$

is the steady-state occupation fraction of the excited level if the Wouthuysen–Field effect were the only effect operating and if the Lyα spectral shape were fixed. For the special case where the photon spectrum is thermal across the Lyα line with colour temperature $T_{\rm c}$, $J(\nu) \propto \exp(-h\nu/k_{\rm B}T_{\rm c})$, one would have $y_{\alpha,{\rm ss}} = 3/4 - 3T_\star/16T_{\rm c}$. In reality, the spectrum in the vicinity of the Lyα resonance is non-thermal, and the effective colour temperature $-(h/k_{\rm B})\,{\rm d}\ln J/{\rm d}\nu$ is between $T_{\rm k}$ and $T_{\rm s}$. However, $y_{\alpha,{\rm ss}}$ as defined by equation (31) still exists. One can therefore define an effective colour temperature $T_{\rm c}^{\rm eff}$ by

$$e^{-T_\star/T_{\rm c}^{\rm eff}} \equiv \frac{y_{\alpha,{\rm ss}}}{3(1-y_{\alpha,{\rm ss}})} \quad\rightarrow\quad y_{\alpha,{\rm ss}} \approx \frac{3}{4} - \frac{3T_\star}{16T_{\rm c}^{\rm eff}}. \quad (32)$$

Then comparison of equation (30) with equation (10) yields

$$x_\alpha = \frac{(\Gamma_{01}+\Gamma_{10})T_\star}{4A_{10}T_\gamma}\frac{\left(T_{\rm c}^{\rm eff}\right)^{-1} - T_{\rm s}^{-1}}{T_{\rm k}^{-1} - T_{\rm s}^{-1}}. \quad (33)$$

This is the equation used to determine x_α. Note that x_α depends on all three temperatures $T_{\rm k}$, $T_{\rm s}$ and T_γ, both explicitly and through the dependence on Γ_{01}, Γ_{10} and $T_{\rm c}^{\rm eff}$. The explicit dependence on the radiation temperature can be eliminated by using equation (11) to write

$$S_\alpha = \frac{9(\Gamma_{01}+\Gamma_{10})}{32\pi\lambda_{{\rm Ly}\alpha}^2\gamma J_\alpha}\frac{\left(T_{\rm c}^{\rm eff}\right)^{-1} - T_{\rm s}^{-1}}{T_{\rm k}^{-1} - T_{\rm s}^{-1}}. \quad (34)$$

The value of S_α thus depends only on $T_{\rm k}$, $T_{\rm s}$, the injection profile $\psi(\nu)$, H, and $n_{\rm H}x_{\rm H\,I}$. It does not depend on J_α because of the linearity of equation (14). Furthermore, if one multiplies both H and $n_{\rm H}\,x_{\rm H\,I}$ by some scaling factor β while holding J_α fixed, then $\mathcal{A}$, D and C are all multiplied by β, hence the solution to equation (14) and the value of S_α are unchanged. Therefore, S_α can really be written purely as a function of $T_{\rm k}$, $T_{\rm s}$, $\psi(\nu)$ and the Gunn–Peterson depth (Gunn & Peterson 1965)

$$\tau_{\rm GP} = \frac{3n_{\rm H}x_{\rm H\,I}\lambda_{{\rm Ly}\alpha}^3\gamma}{2H}, \quad (35)$$

which differs from the ratio $n_{\rm H}\,x_{\rm H\,I}/H$ only by fundamental constants.

As an example, Fig. 2 shows the values of S_α for the particular case of $T_{\rm s} = 57\,{\rm K}$ and $\tau_{\rm GP} = 2\times10^6$, which are reasonable for redshifts $z \approx 20$ in late-reionization scenarios where the Wouthuysen–Field coupling is still weak (i.e. $x_\alpha \ll 1$). The figure shows both the 'old' calculation, which neglected fine structure and spin diffusivity, and assumed $T_{\rm c}^{\rm eff} = T_{\rm k}$, and the 'new' calculation which includes fine structure and spin diffusivity and accounts for incomplete relaxation of the photon spectrum ($T_{\rm c}^{\rm eff} \neq T_{\rm k}$). The feature in the 'all' curve at $T_{\rm k} = 57\,{\rm K}$ represents the fact that the denominator in equation (34) has a singularity when $T_{\rm k} = T_{\rm s}$, since even in this case, the Hubble expansion term in the Fokker–Planck equation implies that $T_{\rm c}^{\rm eff}$ is not exactly equal to $T_{\rm k}$. Because this feature corresponds only to a small change in $T_{\rm c}^{\rm eff}$ it has no important physical consequences, rather it is just an annoying feature of the variable S_α. In Section 3.3, I introduce a modified variable $\tilde{S}_\alpha$ that avoids any singular behaviour.

3.3 Practical calculation

The scattering function S_α is convenient conceptually, however in actual computation the presence of $T_{\rm k}^{-1} - T_{\rm s}^{-1}$ in the denominator is problematic. This problem is solved by splitting S_α into two parts,

$$S_\alpha = \tilde{S}_\alpha\frac{\left(T_{\rm c}^{\rm eff}\right)^{-1} - T_{\rm s}^{-1}}{T_{\rm k}^{-1} - T_{\rm s}^{-1}}, \quad (36)$$

[6] It is a good thing that $D_{\rm int}$ can be neglected, since if we had $h\langle\delta\nu_{\rm X}^2\rangle^{1/2} \geqslant k_{\rm B}T_{\rm X}$ then the typical change in frequency in a single scattering would be comparable to $k_{\rm B}T_{\rm X}/h$, i.e. to the width of the spectral features. In this case the assumptions of the Fokker–Planck equation would not be valid.

[7] The initial condition is technically given by $J(-\infty) = J_\alpha$. In practice, any errors in the initial condition of equation (26) are damped as $\propto \exp\int[\mathcal{A}/D]\,{\rm d}\nu$. Since $\mathcal{A} \leqslant \mathcal{A}_{\rm H} < 0$ and the diffusivity $D \rightarrow 0$ far from resonance, all solutions of equation (26) rapidly converge to the physical solution if the integration is initiated at sufficiently negative $\Delta\nu$.

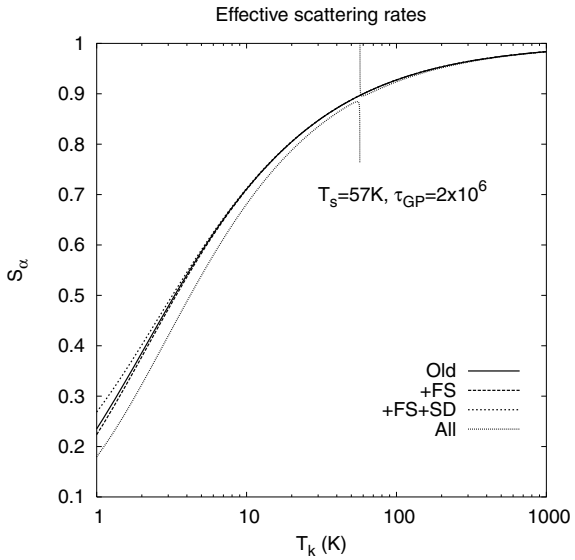

Figure 2. The Wouthuysen–Field effective coupling S_α. The solid curve shows the old calculation, which treats the Lyα resonant cross-section as a single Voigt profile and assumes $T_c^{\rm eff} = T_k$. The long-dashed curve ('+FS') shows what happens when one includes the fine and hyperfine structure of the Lyα line profile using equation (B18). The short-dashed curve ('+FS+SD') also includes the spin diffusivity, i.e. the change of frequency of a photon when it scatters a hydrogen atom and flips the spin state. Finally, the dotted curve ('all') represents the full calculation and includes the correct colour temperature $T_c^{\rm eff}$ instead of assuming that it is completely relaxed to the kinetic temperature T_k. Note that all of the effects are most important at low T_k. (See the text for a discussion of the singularity in the 'all' curve at $T_k = 57$ K.)

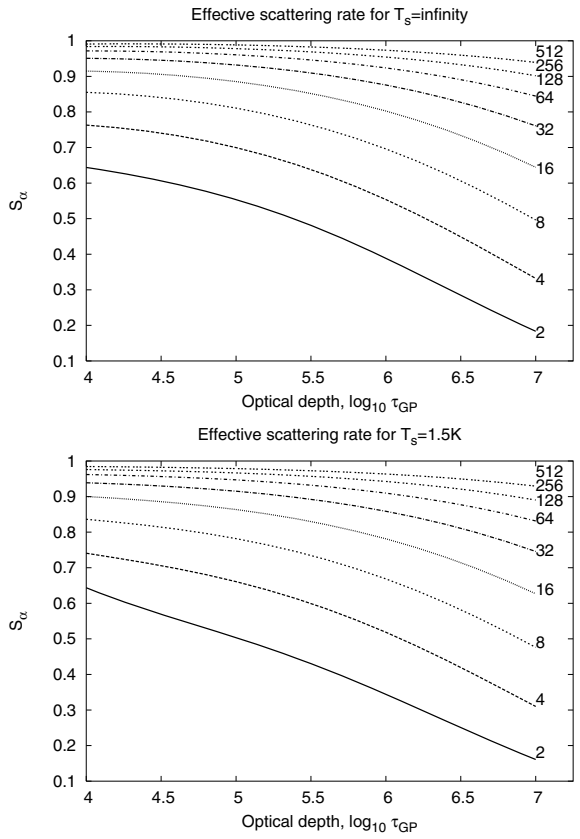

Figure 3. The values of $\tilde{S}_\alpha$ for $T_s = \infty$ (top panel) and 2 K (bottom panel) for continuum photons. The curves are drawn for 11 values of T_k spaced logarithmically from 1 to 10^4 K with intervals of $10^{0.4}$. The labels shown are the values of log $_{10}$ T_k, with T_k in K.

where

$$\tilde{S}_\alpha = \frac{9(\Gamma_{01}+\Gamma_{10})}{32\pi\lambda_{{\rm Ly}\alpha}^2\gamma J_\alpha} = \frac{27}{16}\int \frac{J(\nu)}{J_\alpha}[\phi_{10}(\nu)+\phi_{10}(\nu)]\,{\rm d}\nu. \tag{37}$$

Here $\tilde{S}_\alpha$ and $T_c^{\rm eff}$ are functions of $\tau_{\rm GP}$, T_s and T_k, and the second equality uses equation (29). One can define a modified Lyα coupling parameter $\tilde{x}_\alpha$ by

$$\tilde{x}_\alpha = \frac{8\pi\lambda_{{\rm Ly}\alpha}^2\gamma T_\star}{9A_{10}T_\gamma}\tilde{S}_\alpha J_\alpha. \tag{38}$$

The overall spin temperature is then given by

$$T_s^{-1} = \frac{T_\gamma^{-1} + \tilde{x}_\alpha\left(T_c^{\rm eff}\right)^{-1} + x_c T_k^{-1}}{1+\tilde{x}_\alpha + x_c}. \tag{39}$$

Note that since $\tilde{S}_\alpha$ and $T_c^{\rm eff}$ are functions of T_s as well as T_k and $\tau_{\rm GP}$, equation (39) is an implicit equation for the spin temperature. The dependence is however weak, so a simple and robust way to find T_s for given T_γ, T_k, J_α, and $\tau_{\rm GP}$ is to iteratively compute $T_c^{\rm eff}$ and $\tilde{S}_\alpha$ for some value of T_s, and then update T_s using equation (39). Initializing the iteration with $T_s({\rm init}) = T_\gamma$ results in convergence to better than 1 per cent after less than five iterations for reasonable values of T_k ($T_k > 1$ K).

The functions $\tilde{S}_\alpha$ and $T_c^{\rm eff}$ cannot be computed in closed analytic form, and can be expensive to evaluate numerically as they require solution of an ODE. Therefore the simplest method to obtain them is to first compute values on a grid of points in $(\tau_{\rm GP}, T_s, T_k)$, and then build a fitting formula. The following formula for $\tilde{S}_\alpha$ reproduces our numerical results to within 1 per cent in the range $T_k \geqslant 2$ K, $T_s \geqslant 2$ K and $10^5 \leqslant \tau_{\rm GP} \leqslant 10^7$ for continuum photons:

$$\begin{aligned}\tilde{S}_\alpha = {} & \left(1 - 0.063\,1789T_k^{-1} + 0.115\,995T_k^{-2}\right.\\ & \left. - 0.401\,403T_s^{-1}T_k^{-1} + 0.336\,463T_s^{-1}T_k^{-2}\right)\\ & \times (1 + 2.983\,94\xi + 1.535\,83\xi^2 + 3.852\,89\xi^3)^{-1},\end{aligned} \tag{40}$$

where

$$\xi = \left(10^{-7}\tau_{\rm GP}\right)^{1/3}T_k^{-2/3} \tag{41}$$

and the temperatures T_k and T_s are in K. A simpler formula holds for $T_c^{\rm eff}$ over the same range,

$$\left(T_c^{\rm eff}\right)^{-1} = T_k^{-1} + 0.405\,535T_k^{-1}\left(T_s^{-1} - T_k^{-1}\right), \tag{42}$$

where again T_k is in K. This reproduces the $(T_c^{\rm eff})^{-1}$ values from the Fokker–Planck equation to 1 per cent.

The numerically computed (i.e. not from the fitting formula) function $\tilde{S}_\alpha$ is shown in Fig. 3.

For the injected photons, it is found that equation (40) reproduces the Fokker–Planck results for $\tilde{S}_\alpha$ to better than 3 per cent. Equation (42) reproduces the colour temperature $T_c^{\rm eff}$ to better than 4 per cent at $T_k < 10^3$ K. At higher temperatures $10^3 < T_k < 10^4$ K, the error increases to 12 per cent with the fitting formula underestimating the colour temperature. This is because for the very high temperatures the photon spectrum is essentially flat, with the slope $(T_c^{\rm eff})^{-1}$ being very close to zero. The absolute error in $(T_c^{\rm eff})^{-1}$ at $T_k > 10^3$ K is never greater than 3.7×10^{-5} K^{-1}, which is <1 per cent of T_γ^{-1} at all redshifts of interest.

3.4 Monte Carlo simulations

In Section 3.2, we solved for the Lyα spectral distortion assuming the Fokker–Planck equation to be valid. This equation rests on several assumptions whose validity must be considered and tested. The simplest way to test the assumptions is to use a Monte Carlo simulation, which is done in this section.

Chen & Miralda-Escudé (2004) argued that the Fokker–Planck equation is valid whenever the scale in frequency over which the photon spectrum varies is much greater than σ_v. This is true in the damping tails of the Lyα resonances, but not in the Doppler cores since one must also drop the derivatives of the line profiles $\phi_{F_\mathrm{i}F_\mathrm{f}}$, as was done in deriving equation (A15) of Rybicki & Dell'Antonio (1994). The Lyα Doppler core extends out to $\geqslant 3.3\sigma_v$ at $T_\mathrm{k} \geqslant 2\,\mathrm{K}$, so within this region the Fokker–Planck equation does not reproduce the frequency redistribution matrix. Of course, for the case considered by Chen & Miralda-Escudé (2004) the portion of the spectrum near line centre is in thermal equilibrium with the atoms, with colour temperature T_k. If equilibrium applies, the correct solution is obtained regardless of the frequency redistribution matrix. In this paper, however, we have introduced spin diffusivity for which one may have $T_\mathrm{s} \neq T_\mathrm{k}$, and thermal equilibrium does not apply. At high kinetic temperatures, this is irrelevant because the change in frequency $\sim\nu_{10}$ due to spin-flip events is negligible compared to the change $\sim\sigma_v$ due to the Doppler effect, and the photons equilibrate at colour temperature T_k. But at low kinetic temperature if $T_\mathrm{s} \neq T_\mathrm{k}$ no such equilibrium occurs, the exact form of the frequency redistribution matrix matters, and the validity of the Fokker–Planck equation must be verified.

The results from the Fokker–Planck equation can be checked in two basic ways: one could construct the integro-differential equations for $J(\nu)$ and solve them, or one could do a Monte Carlo simulation in which the distribution $J(\nu)$ is sampled rather than explicitly represented as a function. In our case, the inclusion of fine/hyperfine structure and spin-flip (Raman) scattering makes the redistribution matrix much more complicated than the 'R_II' form of Deguchi & Watson (1985) or Rybicki & Dell'Antonio (1994), so the Monte Carlo method is used here.

3.4.1 Methodology

The basic procedure for the Monte Carlo simulation is:

(1) Start a photon at some starting frequency $\nu = \nu_\mathrm{start}$.

(2) Determine the optical depth $\delta\tau$ through which the photon travels before it scatters by selecting it from an exponential distribution: $P(\delta\tau)\,\mathrm{d}\delta\tau = \mathrm{e}^{-\delta\tau}\,\mathrm{d}\delta\tau$.

(3) Determine the photon's frequency $\nu^{(1)}$ when it scatters by solving the equation,

$$\delta\tau = \tau_\mathrm{GP} \int_{\nu^{(1)}}^{\nu} \sum_{F_\mathrm{i},F_\mathrm{f}} y_{F_\mathrm{i}} \phi_{F_\mathrm{i}F_\mathrm{f}}(\nu')\,\mathrm{d}\nu'. \tag{43}$$

The Gunn–Peterson depth τ_GP normalizes the total optical depth. The Doppler-convolved line profiles $\phi_{F_\mathrm{i}F_\mathrm{f}}$ appear in equation (43). If the optical depth $\delta\tau$ is not reached by the time the integration reaches a terminating frequency $\nu^{(1)} = \nu_\mathrm{term}$, the simulation is stopped.

(4) When the photon scatters off an H atom, choose the initial and final spin states of the H atom. The probability for $F_\mathrm{i} \rightarrow F_\mathrm{f}$ scattering is $y_{F_\mathrm{i}}\phi_{F_\mathrm{i}F_\mathrm{f}} / \sum_{F_\mathrm{i}'F_\mathrm{f}'} y_{F_\mathrm{i}'}\phi_{F_\mathrm{i}'F_\mathrm{f}'}$.

(5) Once the initial and final spin states are selected, one must obtain the velocity $\boldsymbol{v}$ of the atom that does the scattering. It is most convenient to express this velocity in frequency units, $\boldsymbol{u} = \nu_{\mathrm{Ly}\alpha}\,\boldsymbol{v}/c$. The component parallel to the initial direction of propagation of the photon is denoted by $u_\parallel$. Its probability distribution is

$$P(u_\parallel)\,\mathrm{d}u_\parallel = \frac{\mathrm{e}^{-u_\parallel^2/2\sigma_v^2}\phi^\mathrm{u}_{F_\mathrm{i}F_\mathrm{f}}\left[\nu^{(1)} - u_\parallel\right]\,\mathrm{d}u_\parallel}{\int_{-\infty}^{\infty}\mathrm{e}^{-u_\parallel'^2/2\sigma_v^2}\phi^\mathrm{u}_{F_\mathrm{i}F_\mathrm{f}}\left[\nu^{(1)} - u_\parallel'\right]\,\mathrm{d}u_\parallel'}, \tag{44}$$

where the superscript u denotes the unconvolved line profile. The perpendicular component in the plane of scattering (i.e. containing the initial and final directions of the photon) is $u_\perp$ and has a Gaussian probability distribution with zero mean and variance σ_v^2. For a Maxwellian velocity distribution, $u_\parallel$ and $u_\perp$ are independent.

(6) Obtain the scattering angle χ, i.e. the angle between the incoming and outgoing photon directions. This is obtained via

$$P(\chi)\,\mathrm{d}\chi = \frac{1}{2}\left[1 + 5\varpi_2\left(\frac{3}{2}\cos^2\chi - \frac{1}{2}\right)\right]\sin\chi\,\mathrm{d}\chi \tag{45}$$

within the range $0 \leqslant \chi \leqslant \pi$; cf. equation (B12). The phase function ϖ_2 is evaluated using equation (B22) at the frequency in the frame of the atom, $\nu^{(1)} - u_\parallel'$, instead of $\nu^{(1)}$.

(7) The post-scattering frequency of the photon is determined by conservation of energy. Specifically, the atom picks up a recoil velocity $\delta\boldsymbol{v}$ with components $\delta v_\parallel = (1 - \cos\chi)h\nu_{\mathrm{Ly}\alpha}/m_\mathrm{p}\,c$ and $\delta v_\perp = (\sin\chi)h\nu_{\mathrm{Ly}\alpha}/m_\mathrm{p}c$. Its kinetic energy then changes by $m_\mathrm{p}\boldsymbol{v}\cdot\delta\boldsymbol{v} + (1/2)m_\mathrm{p}|\delta\boldsymbol{v}|^2$. The atom also changes its hyperfine energy by $(F_\mathrm{f} - F_\mathrm{i})h\nu_{10}$. Putting these results together implies a post-scattering frequency

$$\nu^{(2)} = \nu^{(1)} - (F_\mathrm{f} - F_\mathrm{i})\nu_{10} - (u_\parallel + \eta)(1 - \cos\chi) - u_\perp\sin\chi, \tag{46}$$

where $\eta = h\nu_{\mathrm{Ly}\alpha}^2/m_\mathrm{p}c^2$.

(8) Replace $\nu := \nu^{(2)}$ and return to step #2.

The Monte Carlo method is straightforward in concept; the major non-trivial aspect is the construction of random numbers. The distribution of $\delta\tau$ in step #2 and that of $u_\perp$ in step #5 are exponential and Gaussian, respectively, and are computed using the 'expdev' and 'gasdev' functions from the Numerical Recipes in C (Press et al. 1992). The distribution of χ in step #6 is also straightforward: the variable $\mu = \cos\chi$ is in the range $-1 \leqslant \mu \leqslant +1$, and a simple rejection method with a constant comparison function (e.g. section 7.3 of Press et al. 1992) works very efficiently. The challenge is the distribution of $u_\parallel$ in step #5 because in most cases the distribution is polymodal with $P(u_\parallel)$ sometimes varying by several orders of magnitude between the very narrow resonance peaks. This algorithm is presented in Appendix C.

The starting and terminating frequencies also require some work. There are two requirements on these. First, one does not want to miss the spin-flip scattering events that can occur in the Lyα damping wings; and secondly, one does not want to artificially terminate photons that reach ν_term that in reality would be scattered back to line centre. The first issue can be addressed by considering the number of spin-flip scatterings that occur in the damping wings. Using equation (B18), we can find the integrated spin-flip cross-section in the far damping wings. For example, for $F = 0 \rightarrow 1$ scattering,

$$\int_{\nu_\mathrm{A}+750\,\mathrm{GHz}}^{\infty}\phi_{01}^u(\nu')\,\mathrm{d}\nu' \approx \int_{-\infty}^{\nu_\mathrm{A}-750\,\mathrm{GHz}}\phi_{01}^u(\nu')\,\mathrm{d}\nu' \approx 3\times10^{-10}; \tag{47}$$

the corresponding value for $1 \rightarrow 0$ scattering is 1×10^{-10}. Thus the fraction of the spin-flip events that occur more than 750 GHz from resonance can be neglected. The Doppler smearing does not change this conclusion since at the temperatures of interest, $\sigma_v \ll$ 750 GHz and hence the spin-flip cross-sections more than 750 GHz

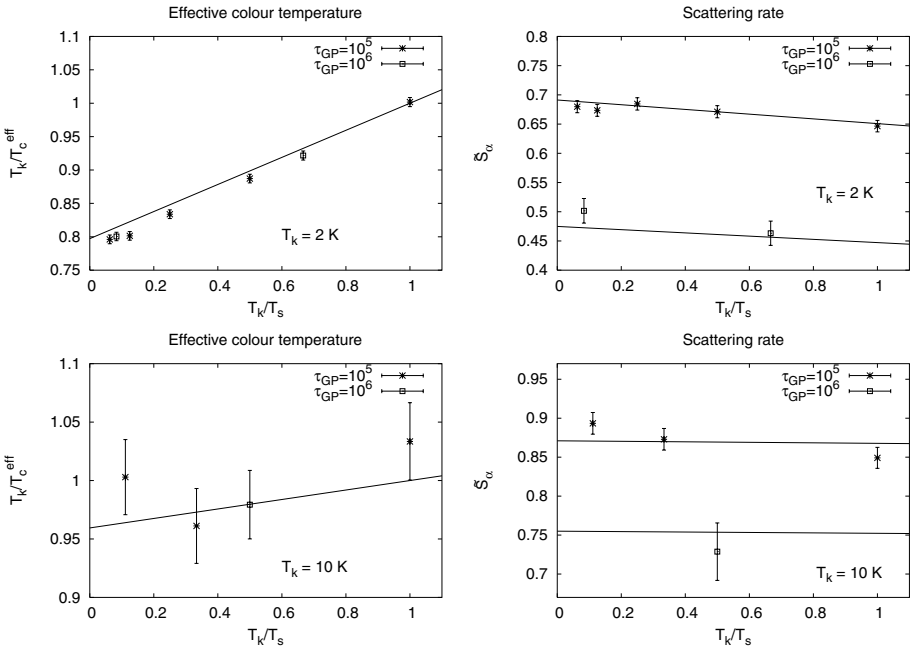

Figure 4. The colour temperatures and spin-flip scattering rates for several values of T_k and τ_{GP}. The points with error bars come from Monte Carlo simulations at the specified values of T_k/T_s and τ_{GP}. The curves in the left-hand panel come from equation (40), and those in the right-hand panel come from equation (42) using $\tau_{GP} = 10^5$ (upper curve) and 10^6 (lower curve). The $\tau_{GP} = 10^5$ points were obtained by simulating 4096 Monte Carlo photons, while the $\tau_{GP} = 10^6$ points were obtained with 512. All of these simulations are for continuum photons, with the photons being started well to the blue side of Lyα at $\nu_{start} = \nu_{Ly\alpha} + 750$ GHz.

from resonance are not significantly affected by the Doppler effect.[8] We thus use $\nu_{start} = \nu_A + 750$ GHz.

We next consider the possibility of a photon reaching $\nu_{term} = \nu_A - 1$ THz and scattering back to line centre. A simple way of evaluating how important this is to go to the Fokker–Planck equation (which is valid in the damping tails) and injecting photons at the frequency ν_{term} instead of at line centre. Even in the worst case used in the Monte Carlo simulations below ($T_k = 10$ K, $T_s = \infty$ and $\tau_{GP} = 10^6$), this gives $\tilde{S}_\alpha = 1.0 \times 10^{-12}$, which implies that photons that pass through ν_{term} and then scatter contribute this amount to the scattering rate. Since this is negligible, we conclude that for the parameters simulated, $\nu_{term} = \nu_A - 1$ THz is an acceptable terminating frequency.

Once the Monte Carlo simulation has been run, one can construct the quantities $\tilde{S}_\alpha$ and T_c^{eff} as follows. Suppose that during the course of the simulation, one observes $N_{F_iF_f}$ of the $F_i \to F_f$ scattering events. The rate per unit volume (i.e. in cm^{-3} s^{-1}) at which photons are redshifting into the Lyα resonance is

$$\dot{n}_\gamma = 4\pi H \nu_{Ly\alpha} c J_\alpha, \tag{48}$$

where the factor of $4\pi c$ converts the 'per unit area per unit time per unit solid angle' in the definition of J_α into 'per unit volume', and $H\nu_{Ly\alpha}$ is the rate at which the frequency of the photon is changing. The rate of $F_i \to F_f$ scattering events per neutral atom in the F_i level is then

$$\Gamma_{F_iF_f} = \frac{\dot{n}_\gamma \langle N_{F_iF_f}\rangle}{n_H x_{H1} y_{F_i}} \tag{49}$$

(this has units of s^{-1}). Comparison to equation (29), and use of equation (35) to express the Hubble rate and the number densities in terms of τ_{GP}, yields the expression

$$\int \frac{J(\nu)}{J_\alpha}\phi_{F_iF_f}(\nu)\,d\nu = \frac{\langle N_{F_iF_f}\rangle}{\tau_{GP} y_{F_i}}. \tag{50}$$

Equation (50) allows us to obtain $\tilde{S}_\alpha$ by plugging the results into equation (37). One may also obtain the colour temperature by plugging the rates (equation 49) into equations (31) and (32); the result is

$$e^{-T_\star/T_c^{eff}} = \frac{y_0\langle N_{10}\rangle}{3y_1\langle N_{01}\rangle}. \tag{51}$$

Error estimates on $\tilde{S}_\alpha$ and T_c^{eff} may be computed by taking the covariance matrix of N_{10} and N_{01}, obtained from the dispersion among many Monte Carlo simulations, and propagating these to $\tilde{S}_\alpha$ and T_c^{eff} using the usual Jacobian rules.

3.4.2 Results

The Monte Carlo simulations must be used to verify the Fokker–Planck estimates of (i) the colour temperature T_c^{eff} and (ii) the spin-flip rate $\tilde{S}_\alpha$ that describes how rapidly the spins relax to the colour temperature. Results for both of these are shown in Fig. 4 for $T_k = 2$ and 10 K, and at $\tau_{GP} = 10^5$ and 10^6. The agreement with the fitting formulae (equations 40 and 42) is at the $\leqslant 3$ per cent level. It is especially remarkable that the fitting formulae perform very well at reproducing the correct dependence of the colour temperature on T_s at low T_k, since the non-equilibrium effects on the spectral distortion must be taken into account and the slope of the spectrum

[8] The values of the order of 10^{-10} are several orders of magnitude less than that one would calculate for a Lorentzian profile. This is because the spin-flip process can proceed through either the $2p_{1/2}(F = 1)$ or $2p_{3/2}(F = 1)$ hyperfine excited levels. The amplitudes through each of the excited levels add coherently, and they undergo destructive interference when one is far from resonance.

in the Doppler cores of the resonances (where the validity of the Fokker–Planck equation is most questionable) is important.

4 SIMPLE MODEL FOR SPIN TEMPERATURE EVOLUTION

This section presents a simple model for the evolution of T_s as a function of redshift. The purpose of this model is to illustrate how much of a difference the improvements in the physics of the Wouthuysen–Field effect can make in the final result; it is not claimed that they necessarily represent the real universe. Only the mean brightness temperature perturbation T_b in the 21-cm line is calculated here for simplicity. While foreground synchrotron radiation probably precludes a direct measurement of the mean signal T_b (Shaver et al. 1999; Oh & Mack 2003), it is still possible that it could be determined indirectly using redshift space distortions. Specifically, on linear scales the $l = 4$ moment of the power spectrum of 21-cm fluctuations, denoted $P_{\mu^4}(k)$ by Barkana & Loeb (2005a), is simply related to the mean temperature and matter power spectrum via $P_{\mu^4}(k) = T_{b^2} P_\delta(k)$. If linear scales can be observed during the early stages of reionization and the cosmological parameters are known well enough to estimate $P_\delta(k)$, then it may become possible to obtain T_b.

A model for T_s requires a model for the evolution of the CMB temperature, the Lyα flux and the gas kinetic temperature. Of these, finding the CMB temperature is the easiest. It is

$$T_\gamma = T_{\gamma 0}(1+z), \tag{52}$$

where $T_{\gamma 0} = 2.725\,\mathrm{K}$. The Ly$\alpha$ flux is given by

$$J_\alpha = \frac{(1+z)^2}{4\pi} \sum_{n=2}^{\infty} P_{n\mathrm{p}} \int_z^{z_{\max}} \frac{c}{H(z')} \epsilon(\nu_n', z')\,\mathrm{d}z'; \tag{53}$$

see Barkana & Loeb (2005b). The UV source term is $\epsilon(\nu_n', z')$, which is the number of photons emitted per unit comoving volume per unit proper time per unit frequency at redshift z' and frequency ν_n' (see below). The factor of $P_{n\mathrm{p}}$ has been added to account for the fact that not all photons in the 912–1216 Å band degrade to Lyα. The nth term in the sum is the contribution from photons emitted between the $1\mathrm{s} \rightarrow n\,\mathrm{p}$ and $1\mathrm{s} \rightarrow (n+1)\mathrm{p}$ Lyman transitions, which ultimately redshift and excite $1\mathrm{s} \rightarrow n\,\mathrm{p}$; as such, the emitted photon frequency is $\nu_n' = \nu_{1\mathrm{s}\rightarrow n\mathrm{p}}\,(1+z')/(1+z)$ and the maximum redshift from which this photon could have been received is

$$1 + z_{\max} = \frac{\nu_{1\mathrm{s}\rightarrow(n+1)\mathrm{p}}}{\nu_{1\mathrm{s}\rightarrow n\mathrm{p}}}(1+z) = \frac{1-(n+1)^{-2}}{1-n^{-2}}(1+z). \tag{54}$$

The source emissivity is modelled following Barkana & Loeb (2005b) by the equation

$$\epsilon(\nu, z) = \epsilon_b(\nu) \frac{\Omega_b}{m_p \Omega_m} \frac{\mathrm{d}}{\mathrm{d}t} \int f_\star(M,t) M n(M,t)\,\mathrm{d}M, \tag{55}$$

where M is the relevant halo mass, $n(M, t)$ is the comoving number density of haloes at proper time t per unit mass, $f_\star(M, t)$ is the fraction of the baryons that have turned into stars, and $\epsilon_b(\nu)$ is the number of photons emitted per baryon by the stars. This equation assumes that the lifetimes of the UV-emitting stars are short compared to the Hubble time, so that the UV emissivity tracks the instantaneous star formation rate. This is reasonable since most radiation at 912–1216 Å is emitted by the most massive stars with lifetimes of $<10^7$ yr, whereas the Hubble time during reionization is $>10^8$ yr. The halo mass function of Sheth & Tormen (1999) is used.

We consider the temperature evolution of the IGM due to cosmological expansion and X-ray heating. The real universe has inhomogeneities that alter the spin temperature evolution via changes in the kinetic temperature (in shocks, by adiabatic expansion or compression during structure formation, or from inhomogeneous X-ray sources), and by enhancing the collisional coupling in the denser regions. The main effect is to increase the 21-cm emissivity of haloes and filaments (Iliev et al. 2002; Ahn et al. 2005; Kuhlen et al. 2005) and so including them in the model would make the computed signal more positive (or less negative). For example, Ahn et al. (2005) find in a simulation with no X-ray or UV sources that these effects increase the mean signal by $+1$ mK at $z = 18$ and $+5$ mK at $z = 10$. We have not simulated the effect of inhomogeneities in the presence of UV radiation, but they could be larger than that found by Ahn et al. (2005) because the Wouthuysen–Field coupling will make most of the diffuse, unshocked IGM 'visible' and hence the importance of temperature fluctuations in the unshocked phase will be increased. Subject to these caveats, our temperature evolution equation is thus

$$(1+z)\frac{\mathrm{d}T_k}{\mathrm{d}z} = 2T_k - \frac{2\mu m_p \Gamma_X}{3\rho_{b0} k_B H(z)}, \tag{56}$$

assuming a monatomic gas (Chen & Miralda-Escudé 2004) with mean atomic weight of $\mu = 1.22$, as appropriate for a hydrogen–helium mixture with helium mass fraction 0.24.

The X-ray heating Γ_X (in e.g. erg per physical second per comoving cm^3) is

$$\Gamma_X = f_\Gamma f_{Xe} E_X \frac{\Omega_b}{m_p \Omega_m} \frac{\mathrm{d}}{\mathrm{d}t} \int f_\star(M,t) M n(M,t)\,\mathrm{d}M, \tag{57}$$

where f_Γ is the fraction of X-ray energy that goes into heating the IGM, f_{Xe} is the fraction of X-ray photons that escape from an early star cluster or galaxy, and E_X is the energy emitted in X-rays per baryon that forms stars. There are many sources that contribute to E_X, e.g. stars, supernovae, X-ray binaries and quasars, and both the total X-ray emission and the relative contributions from different sources are very uncertain (Glover & Brand 2003). Also, equation (57) assumes that the X-ray heating tracks the star formation rate, which may not be true particularly if quasars contribute significantly to the X-ray emission. Equation (56) has the solution

$$T_k(z) = \left(\frac{1+z}{1+z_0}\right)^2 T_k(z_0) + \frac{2\mu m_p}{3\rho_{b0} k_B} \int_z^{z_0} \frac{\Gamma_X(z')}{H(z')} \frac{(1+z)^2}{(1+z')^3}\,\mathrm{d}z', \tag{58}$$

where z_0 is an arbitrary starting redshift, which can be any time after the thermal decoupling of the gas from the CMB but prior heating is important. We use $z_0 = 50$ and initialize the temperature using RECFAST (Seager, Sasselov & Scott 1999).

An example of this model is shown in Fig. 5. Here, it is assumed that stars form only in haloes with virial temperature $T_{\mathrm{vir}} > 10^4\,\mathrm{K}$ that can cool via atomic transitions. The star formation efficiency is taken as $f_* = 2.5 \times 10^{-4}$ in these haloes, which causes the Lyα coupling to turn on ($x_\alpha = 1$) at $z \approx 21$. Their $\epsilon_b(\nu)$ is assumed to be a blackbody of temperature 10^5 K, as appropriate for massive Population III stars with $M \geqslant 300\,\mathrm{M}_\odot$ (Bromm, Kudritzki & Loeb 2001); the blackbody is normalized to a total energy of 7.1 MeV per H nucleus or 5.4 MeV per baryon, appropriate for complete hydrogen burning to ^{4}He. (Most of the energy of the star is released during the hydrogen-burning stage.) This model contains no X-ray emission. If one assumes that 0.5 per cent of the energy of the stars emerges from early galaxies in the form of X-rays that can heat the IGM (corresponding to $f_{Xe} E_X = 27\,\mathrm{keV}$), and that the

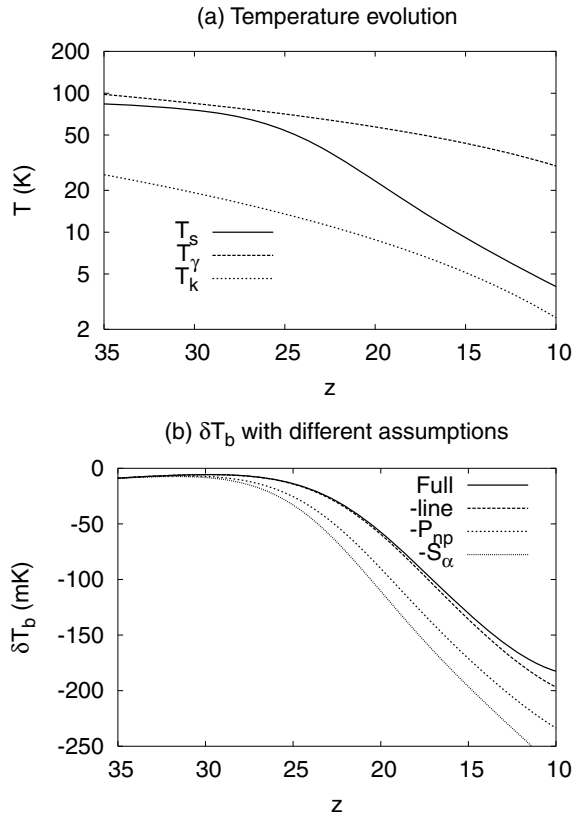

Figure 5. (a) Spin temperature evolution assuming Population III stars forming with efficiency $f_* = 2.5 \times 10^{-4}$ in haloes that can cool via atomic transitions. All sources of heating, including X-rays, have been neglected in this model. (b) The effect of changing the physics of the Wouthuysen–Field effect. The solid curve shows the full calculation for the mean brightness temperature T_b. The long-dashed curve shows the calculation removing the spin diffusivity and fine structure corrections. The short-dashed curve also assumes $P_{np} = 1$ instead of the correct values; this is the curve that would be calculated using the most recent models prior to this paper. The dotted curve makes the further simplification that $S_\alpha = 1$, as was done by Madau et al. (1997).

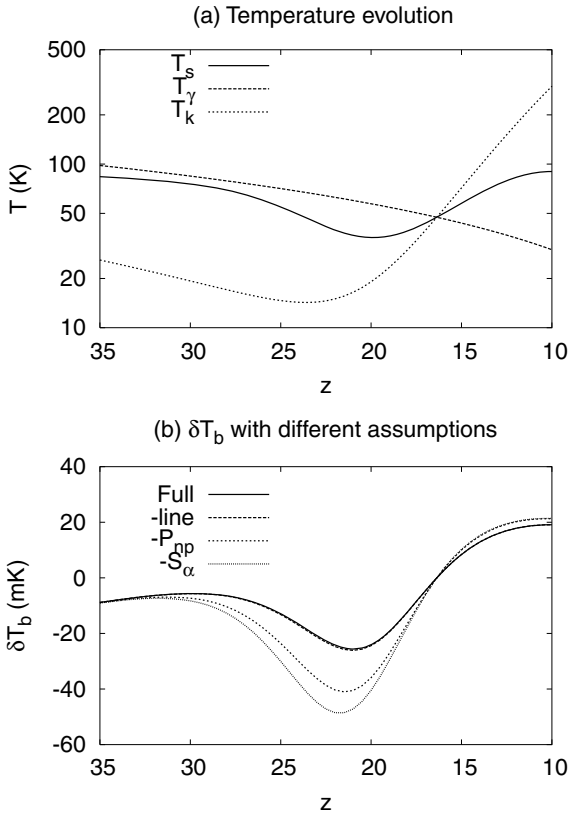

Figure 6. Same as Fig. 5, except with X-rays. Here, it is assumed that the X-rays escaping from early galaxies carry 0.5 per cent of the stellar energy output, corresponding to $f_{\rm Xe} E_{\rm X} = 27\,{\rm keV}$.

heating efficiency is $f_\Gamma = 0.14$ (Shull & van Steenberg 1985; Chen & Miralda-Escudé 2004), then one obtains the model in Fig. 6.[9] In both cases, the best-fitting six-parameter cosmology of Seljak et al. (2005) was used.

In both the examples with and without X-ray emission, a calculation neglecting the Lyα spectral distortion (e.g. Madau et al. 1997) can overestimate the 21-cm signal by as much as a factor of $\sim$2.4, as shown by the dotted curves. Incorporating the simplified model of the Lyα spectral distortion using the Voigt profile (e.g. Chen & Miralda-Escudé 2004) reduces the error to a factor of $\leqslant$1.9, as shown by the short-dashed curves. Most of the remaining error is due to the two-photon decays (included in the long-dashed curves). The inclusion of Lyα fine structure and spin diffusivity (solid line) makes a $<$10 per cent difference in the model with no X-rays and even less in the model with X-rays. Thus, it is seen that the two-photon correction P_{np} can have a large effect on the 21-cm signal.

[9] This amount of X-ray emission corresponds to $\alpha_{\rm X} = 0.028$ in the notation of Chen & Miralda-Escudé (2004).

5 CONCLUSIONS

The H I spin temperature of the IGM is determined by a balance of interaction with the CMB in the 21-cm line, atomic collisions and the Wouthuysen–Field effect. The last of these depends on both the emission rate of UV photons and on the coupling coefficient $P_{np} S_\alpha$. In this paper, I have evaluated the coupling coefficient including several new physical processes, and found that it is lower than previously computed. The most important correction is the inclusion of two-photon decay, $P_{np} < 1$. Fine and hyperfine structure effects and spin diffusivity are small except at low temperatures. The Fokker–Planck equation is found to provide an accurate description of the Wouthuysen–Field effect at the several per cent level even at the lowest temperatures that could reasonably be encountered in the IGM. Fitting formulae for the scattering rate $\tilde{S}_\alpha$ (equation 40) and colour temperature $T_{\rm c}^{\rm eff}$ (equation 42) have been provided, along with bounds on their errors.

The corrections described here pertain to the strength of the Wouthuysen–Field effect and are important only during the era when $x_{\rm c} < x_\alpha \leqslant {\rm O}(1)$. Early on ($z > 30$ in the models of Section 4), the Wouthuysen–Field effect is negligible. Later on ($z < 15$ in the models of Section 4), the Wouthuysen–Field effect becomes saturated in the sense that $x_\alpha \gg 1$ and $T_{\rm s} \approx T_{\rm k}$; in this case changes in the coupling strength have no impact on the observable temperature fluctuations. The changes described here, particularly P_{np}, can however have a very large effect at intermediate redshifts (here $15 < z < 30$) particularly where $x_\alpha \sim 1$. This is the range of redshifts at which Barkana & Loeb (2005b) have suggested that the fluctuations in the Lyα background could be observable, providing

information about early galaxies such as their bias (and hence their halo mass). These authors found that photons redshifting into the higher-order Lyman transitions Lyn ($n \gg 1$) dominate the Lyα fluctuations at $k \geqslant 0.1\,h\,\mathrm{Mpc}^{-1}$; since $P_{np} = 0.36$ for these photons, the power spectrum of these small-scale Lyα fluctuations will be correspondingly reduced. For this application in particular, the inclusion of the two-photon decay mechanism will be valuable in extracting maximal information from 21-cm observations.

ACKNOWLEDGMENTS

I wish to thank P. J. E. Peebles, Uroš Seljak and Kris Sigurdson for useful comments, and X. Chen and J. Miralda-Escudé for discussion about the high-temperature behaviour of the spectral distortion. I acknowledge support from NASA grant NGT5-50383.

REFERENCES

Ahn K., Shapiro P. R., Alvarez M. A., Iliev I. T., Martel H., Ryu D., 2005, New Astron. Rev., in press (astro-ph/0509651)
Allison A., Dalgarno A., 1969, ApJ, 158, 423
Barkana R., Loeb A., 2005a, ApJ, 624, L65
Barkana R., Loeb A., 2005b, ApJ, 626, 1
Becker R. et al., 2001, AJ, 122, 2850
Bennett C. et al., 2003, ApJS, 148, 1
Berestetskii V. B., Lifshitz E. M., Pitaevskii V. B., 1971, Relativistic Quantum Theory, Part 1. Pergamon Press, Ltd., Oxford
Bharadwaj S., Ali S., 2004, MNRAS, 352, 142
Braskén M., Kyrölä E., 1998, A&A, 332, 732
Breit G., Teller E., 1940, ApJ, 91, 215
Bromm V., Kudritzki R., Loeb A., 2001, ApJ, 552, 464
Chen X., Miralda-Escudé J., 2004, ApJ, 602, 1
Ciardi B., Madau P., 2003, ApJ, 596, 1
Deguchi S., Watson W. D., 1985, ApJ, 290, 578
Domke H., Hubeny I., 1988, ApJ, 334, 527
Doré O., Hennawi J. F., Spergel D. N., 2004, ApJ, 606, 46
Edmonds A., 1960, Angular Momentum in Quantum Mechanics. Princeton Univ. Press, Princeton, NJ
Fan X., Narayanan V., Strauss M., White R., Becker R., Pentericci L., Rix H., 2002, AJ, 123, 1247
Field G., 1958, Proc. IREE Aust., 46, 240
Field G., 1959, ApJ, 129, 536
Glover S. C. O., Brand P. W. J. L., 2003, MNRAS, 340, 210
Goeppert-Mayer M., 1931, Ann. d. Phys., 9, 273
Gunn J. E., Peterson B. A., 1965, ApJ, 142, 1633
Hogan C., Rees M., 1979, MNRAS, 188, 791
Hu W., Holder G., 2003, Phys. Rev. D, 68, 023001
Iliev I. T., Shapiro P. R., Ferrara A., Martel H., 2002, ApJ, 572, L123
Iliev I. T., Scannapieco E., Martel H., Shapiro P. R., 2003, MNRAS, 341, 81
Kogut A. et al., 2003, ApJS, 148, 161
Kuhlen M., Madau P., Montgomery R., 2005, ApJL, in press (astro-ph/0510814)
Lee J. S., 1977, ApJ, 218, 857
Lee J. S., 1982, ApJ, 255, 303
Loeb A., Zaldarriaga M., 2004, Phys. Rev. Lett., 92, 211301
Madau P., Meiksin A., Rees M., 1997, ApJ, 475, 429
Oh S. P., Mack K. J., 2003, MNRAS, 346, 871
Omont A., Smith E. W., Cooper J., 1972, ApJ, 175, 185
Peterson J. B., Pen U., Wu X., 2004, Mod. Phys. Let. A, 19, 1001
Press W., Teukolsky S., Vetterling W., Flannery B., 1992, Numerical Recipes in C. The Art of Scientific Computing. Cambridge Univ. Press, Cambridge
Rybicki G., Dell'Antonio I., 1994, ApJ, 427, 603
Seager S., Sasselov D., Scott D., 1999, ApJ, 523, L1
Seaton M., 1955a, MNRAS, 115, 279
Seaton M., 1955b, Proc. Phys. Soc. A, 68, 457
Seljak U. et al., 2005, Phys. Rev. D, 71, 103515
Shapiro J., Breit G., 1959, Phys. Rev., 113, 179
Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A&A, 345, 380
Sheth R., Tormen G., 1999, MNRAS, 308, 119
Shull J. M., van Steenberg M. E., 1985, ApJ, 298, 268
Spitzer L., Greenstein J., 1951, ApJ, 114, 407
Wouthuysen S., 1952, AJ, 57, 31
Wyithe S. J., Loeb A., 2004, ApJ, 610, 117
Zaldarriaga M., Furlanetto S., Hernquist L., 2004, ApJ, 608, 622
Zygelman B., 2005, ApJ, 622, 1356

APPENDIX A: THE FATE OF HYDROGEN IN THE 2s CONFIGURATION

When excited hydrogen atoms decay to the 2s configuration, there are no further one-photon transitions allowed to the ground configuration. In vacuum, the 2s configuration decays via two-photon emission with a rate $\Lambda = 8.2\,\mathrm{s}^{-1}$ (Goeppert-Mayer 1931; Breit & Teller 1940; Shapiro & Breit 1959). The purpose of this appendix is to show that under conditions encountered in the high-redshift IGM, the two-photon process is faster than competing processes, namely collisions and interactions with the CMB.

The $2s_{1/2}$ level can be depopulated by either collisions with neutral atoms or with charged particles. Only crude approximations to these are needed since they will be shown to be negligible. This is convenient, since there are no published rate coefficients for the low temperatures required here. For the collisions with neutral H or He, the rate of de-excitation of $2s_{1/2}$ is $\Gamma_Q \sim n\sigma_Q v$, where n is the number density of H or He, σ_Q is the quenching cross-section, and v is the typical velocity $v \sim 1.3 \times 10^4 T_k^{1/2}\,\mathrm{cm\,s}^{-1}$ (with T_k in K). In order for the collisional de-excitation rate to be 1 per cent of the two-photon rate at $z = 75$, one needs $\sigma_Q \sim 10^{-4}\,T_k^{-1/2}\,\mathrm{cm}^2$, i.e. many orders of magnitude larger than the cross-sections for collision of neutral atoms; thus the atomic collisions contribute negligibly to the depopulation of $H(2s_{1/2})$. Of course in the standard cosmological model there are no Lyα photons at $z = 75$; the result that collisions with neutrals are negligible is even stronger at lower redshifts that are more reasonable for the Wouthuysen–Field effect.

Cross-sections for $H(2s_{1/2})$ with charged particles (e^- or p^+) can be much larger than that for neutral atoms, particularly at low temperature, because the long-range electric field of the passing charged particle can produce a Stark effect that mixes 2s and 2p; once the H atom reaches 2p, it decays quickly by Lyα emission. The rate coefficients W (in e.g. $\mathrm{cm}^3\,\mathrm{s}^{-1}$) scale roughly as $T_k^{-1/2}$ and are dominated by collision with protons (Seaton 1955b). Extrapolating the rate coefficients from Seaton (1955b) at $T_k = 10^4$ K down, and assuming no heating of the IGM so that $T_k = 0.022(1+z)^2$ K, one finds a rate coefficient of $W = 0.36(1+z)^{-1}\,\mathrm{cm}^3\,\mathrm{s}^{-1}$. This is an upper limit because the actual scaling is shallower than $W \propto T_k^{-1/2}$ at low T_k, and because any heating of the IGM increases T_k. The rate of charged particle collisional de-excitation is then

$$\Gamma_{\mathrm{charge}} \sim n_H x_e W = 7 \times 10^{-4} \left(\frac{1+z}{100}\right)^2 x_e\,\mathrm{s}^{-1}, \qquad \text{(A1)}$$

which is much less than the two-photon rate $\Lambda = 8.2\,\mathrm{s}^{-1}$ at all relevant redshifts since the electron-to-hydrogen nucleus ratio x_e is always less than 1.16 (and much less before reionization).

The CMB can depopulate the H I $2s_{1/2}$ level via stimulated emission at the Lamb shift frequency $\nu_{1/2} = 1.06\,\mathrm{GHz}$ to the $2p_{1/2}$ level,

or via radiative excitation to $2p_{3/2}$ at $\nu_{3/2} = 11\,\mathrm{GHz}$. H I atoms in these levels decay by Lyα emission. The rates for these are given by the usual formula

$$\Gamma(2s_{1/2} \rightarrow 2p_j) = \frac{32\pi^3\alpha\nu_j^3}{3c^2}\overline{\sum |\boldsymbol{r}_{2p_j,2s_{1/2}}|^2}\left(\frac{k_B T_\gamma}{h\nu_j}\right), \tag{A2}$$

where the bar and summation indicates that the squares of the dipole matrix elements $\boldsymbol{r}_{2p_j,2_{1/2}}$ are averaged over values of the magnetic quantum number in the $2s_{1/2}$ level and summed over the $2p_j$ level, and the last factor is the number of photons per state. (This is much greater than 1, so spontaneous emission and quantum corrections to the Rayleigh–Jeans formula can be neglected.) The dipole matrix elements $\overline{\sum |\boldsymbol{r}_{2p_j,2s_{1/2}}|^2}$ are $9a_0^2$ for $j = 1/2$ and $18a_0^2$ for $j = 3/2$, where a_0 is the Bohr radius. Substituting into equation (A2) and using $T_\gamma = 2.73(1+z)\,\mathrm{K}$ yields

$$\Gamma(2s_{1/2} \rightarrow 2p_{1/2}) = 4.4 \times 10^{-8}(1+z)\,\mathrm{s}^{-1} \quad \text{and}$$
$$\Gamma(2s_{1/2} \rightarrow 2p_{3/2}) = 9.3 \times 10^{-6}(1+z)\,\mathrm{s}^{-1}. \tag{A3}$$

These rates are negligible compared with the two-photon rate $\Lambda = 8.2\,\mathrm{s}^{-1}$ at all relevant redshifts.

Most of the CMB photons during the reionization era have much higher energies than 1.06 or 11 GHz (for comparison, $k_B T_\gamma/h = 570\,\mathrm{GHz}$ at $1 + z = 10$). These photons can cause non-resonant Raman scattering, $2s_{1/2} \rightarrow 1s_{1/2}$, that puts the hydrogen atom in the ground state and results in the emission of a photon with frequency just above the Lyα frequency. This photon immediately redshifts into the Lyα doublet and can participate in the Wouthuysen–Field effect. The relevant frequencies are all much greater than the fine structure splitting, so at least for a rough estimate one can ignore electron spin in the calculation of the Raman scattering rate. The Raman scattering cross-section is (e.g. Berestetskii, Lifshitz & Pitaevskii 1971, equation 61.8)

$$\begin{aligned}\sigma(\mathrm{i} \rightarrow \mathrm{f}) &= \frac{128\pi^5 e^4 \nu\nu'^3}{9h^2c^4} \\ &\quad \times \sum_{\alpha,\beta}\left|\sum_{m_l=-1}^{1}\frac{\langle 1s|r^\alpha|2p, m_l\rangle\langle 2p, m_l|r^\beta|2s\rangle}{\Delta\nu}\right|^2 \\ &= \frac{128\pi^5 e^4 \nu\nu'^3}{27h^2c^4}\left|\frac{\langle 1s||\boldsymbol{r}||2p\rangle\langle 2p||\boldsymbol{r}||2s\rangle}{\Delta\nu}\right|^2,\end{aligned} \tag{A4}$$

where ν is the incoming frequency, $\nu' = \nu_{\mathrm{Ly}\alpha} + \nu$ is the outgoing frequency, and $\Delta\nu$ is the detuning from the intermediate (2p) state, i.e.

$$h\Delta\nu_n = h\nu + E_{2s} - E_{2p}. \tag{A5}$$

This includes only the 2p intermediate state since the total energy of the atom and photon is only slightly above the $n = 2$ energy level, hence its denominator $\Delta\nu$ is the largest. For the same reason the terms in the Raman matrix element where the outgoing photon is emitted before the incoming photon is absorbed have been dropped. One also has $\nu' \approx \nu_{\mathrm{Ly}\alpha}$, and because of the 2s–2p degeneracy $\Delta\nu_{2p} \approx \nu$. Putting this together and using the hydrogenic matrix elements gives

$$\sigma(\mathrm{i} \rightarrow \mathrm{f}) = \frac{419\,4304\pi^5 e^4 \nu_{\mathrm{Ly}\alpha}^3 a_0^4}{19683h^2c^4\nu}. \tag{A6}$$

The total Raman scattering rate (per atom in the $2s_{1/2}$ level) is given by integration of the cross-section over the blackbody curve:

$$\begin{aligned}\Gamma_{\mathrm{Raman}} &= \int \frac{8\pi\nu^2}{c^2(\mathrm{e}^{h\nu/k_B T_\gamma} - 1)}\sigma(\mathrm{i} \rightarrow \mathrm{f})\,\mathrm{d}\nu \\ &= \frac{167\,772\,16\pi^8 e^4 \nu_{\mathrm{Ly}\alpha}^3 a_0^4 k_B^2}{590\,49h^4c^6}T_\gamma^2 \\ &= \frac{64\pi^2\alpha^6 k_B^2 T_\gamma^2}{81h^2\nu_{\mathrm{Ly}\alpha}} = 1.7 \times 10^{-7}(1+z)^2\,\mathrm{s}^{-1},\end{aligned} \tag{A7}$$

where we have used $a_0 = 3e^2/(8h\nu_{\mathrm{Ly}\alpha})$ and $\alpha = 2\pi e^2/(hc)$. Once again, this rate is negligible compared to $\Lambda = 8.2\,\mathrm{s}^{-1}$ at the redshifts of interest for the Wouthuysen–Field effect.

APPENDIX B: Lyα CROSS-SECTION

In order to compute the Wouthuysen–Field coefficient x_α, it is necessary to know the cross-sections for resonant Rayleigh and Raman scattering between the two hyperfine levels $1s_{1/2}$ ($F = 0, 1$). There are four cross-sections $F \rightarrow F'$, where $F, F' \in \{0, 1\}$, which depend on the photon frequency ν. Similar computations can be found in Domke & Hubeny (1988) and Braskén & Kyrölä (1998), but this appendix includes both the hyperfine structure and the detailed frequency dependence.

The cross-sections can be determined from the reduced dipole matrix elements between the $1s_{1/2}(F)$ and $2p_j(F')$ hyperfine levels. The electron position operator $\boldsymbol{r}$ has reduced matrix element given by the hydrogenic form

$$\langle 2p||\boldsymbol{r}||1s\rangle = \frac{128\sqrt{6}}{243}a_0. \tag{B1}$$

Since the $\boldsymbol{r}$ operator acts only on the positional degrees of freedom of the electron, without regard to electronic or nuclear spin, the hyperfine matrix elements can be obtained entirely from group theory. Applying equation (7.1.7) of Edmonds (1960) twice, and using the fact that H I has electronic spin $S = 1/2$ and (for ^{1}H) nuclear spin $I = 1/2$,

$$\langle n'l'_{j'}(F')||\boldsymbol{r}||nl_j(F)\rangle = \mathcal{I}\langle n'l'||\boldsymbol{r}||nl\rangle, \tag{B2}$$

where the coefficient $\mathcal{I}$ is given by the $6j$ symbols,

$$\begin{aligned}\mathcal{I} &= (-1)^{l'+j+j'+F+1} \\ &\quad \times [(2j+1)(2j'+1)(2F+1)(2F'+1)]^{1/2} \\ &\quad \times \begin{Bmatrix} l' & j' & 1/2 \\ j & l & 1 \end{Bmatrix}\begin{Bmatrix} j' & F' & 1/2 \\ F & j & 1 \end{Bmatrix}.\end{aligned} \tag{B3}$$

Values of $\mathcal{I}$ are shown in Table B1.

Table B1. The six hyperfine components of the Lyα lines. The frequency offset shown is relative to the lowest-frequency line, i.e. it is $\nu - \nu_A$. The values of $\mathcal{I}$ are used in equation (B2).

Line	Lower level	Upper level	I	Frequency offset (GHz)
A	$1s_{1/2}$ ($F = 1$)	$2p_{1/2}$ ($F = 0$)	$+\sqrt{1/3}$	0.000
B	$1s_{1/2}$ ($F = 1$)	$2p_{1/2}$ ($F = 1$)	$-\sqrt{2/3}$	0.059
C	$1s_{1/2}$ ($F = 0$)	$2p_{1/2}$ ($F = 1$)	$-\sqrt{1/3}$	1.479
D	$1s_{1/2}$ ($F = 1$)	$2p_{3/2}$ ($F = 1$)	$-\sqrt{1/3}$	10.945
E	$1s_{1/2}$ ($F = 1$)	$2p_{3/2}$ ($F = 2$)	$+\sqrt{5/3}$	10.968
F	$1s_{1/2}$ ($F = 0$)	$2p_{3/2}$ ($F = 1$)	$+\sqrt{2/3}$	12.365

The matrix element for resonant electric dipole scattering with incoming photon energy $h\nu$ is

$$(c^{\mu\nu})_{\rm fi} = e^2 \sum_a \frac{\langle f|\hat{r}^\mu|a\rangle\langle a|\hat{r}^\nu|i\rangle}{E_a - E_{\rm i} - h\nu - {\rm i}h\Gamma_a/4\pi}, \tag{B4}$$

where Γ_a is the width of the intermediate state a and the width of the initial and final states is neglected. In order to obtain the differential cross-section for $F_{\rm i} \rightarrow F_{\rm f}$ scattering by randomly oriented atoms, one must obtain the spin-K irreducible parts of the scattering tensor,

$$\bar{G}^{(K)}_{F_{\rm i}\rightarrow F_{\rm f}} = \frac{\Pi^{(K)}_{\mu\nu\alpha\beta}}{2F_{\rm i}+1} \sum_{M_{\rm f}=-F_{\rm f}}^{F_{\rm f}} \sum_{M_{\rm i}=-F_{\rm i}}^{F_{\rm i}} (c^{\mu\nu})_{\rm fi} (c^{\alpha\beta})^*_{\rm fi}, \tag{B5}$$

where $\Pi^{(K)}_{\mu\nu\alpha\beta}$ is the projection matrix that selects the spin-K ($K = 0, 1, 2$) part of an arbitrary second-rank tensor $X_{\alpha\beta}$. This decomposition of second-rank tensors is complete, so that

$$\sum_{K=0}^{2} \Pi^{(K)}_{\mu\nu\alpha\beta} = g_{\mu\alpha} g_{\nu\beta}, \tag{B6}$$

where $g_{\mu\nu}$ is the metric tensor, equal to $\delta_{\mu\nu}$ in the usual Cartesian coordinate basis. The most convenient basis for these calculations, however, is not the Cartesian basis but the polar basis (Edmonds 1960) in which the coordinates r^μ are related to Cartesian X, Y and Z via

$$r^{\pm 1} = \mp\frac{1}{\sqrt{2}}(X \pm {\rm i}Y) \quad \text{and} \quad r^0 = Z; \tag{B7}$$

the metric tensor is $g_{\mu\nu} = (-1)^\mu \delta_{\mu,-\nu}$. In this basis the powerful spherical tensor methods can be used. The projection matrix is then

$$\begin{aligned}\Pi^{(K)}_{\mu\nu\alpha\beta} &= \sum_{M_K,\gamma,\delta} g_{\mu\gamma} g_{\nu\delta} \langle 1\gamma; 1\delta | K M_K\rangle \\ &\quad \times \langle K M_K | 1\alpha; 1\beta\rangle \\ &= (2K+1) \sum_{M_K=-K}^{K} (-1)^{M_K} \begin{pmatrix} 1 & 1 & K \\ \mu & \nu & -M_K \end{pmatrix} \\ &\quad \times \begin{pmatrix} 1 & 1 & K \\ \alpha & \beta & M_K \end{pmatrix},\end{aligned} \tag{B8}$$

where in the first line Clebsch–Gordon coefficients have been used to emphasize the nature of $\Pi^{(K)}$ as a projection matrix, and in the second line these have been converted to $3j$ symbols.

Writing equation (B5) in terms of reduced matrix elements, and substituting equation (B8), one obtains

$$\begin{aligned}\bar{G}^{(K)}_{F_{\rm i}\rightarrow F_{\rm f}} &= \frac{2K+1}{2F_{\rm i}+1}\frac{e^4}{h^2} \sum_{M_{\rm f},M_{\rm i},M_K,\mu,\nu,\alpha,\beta,a,b} \\ &\quad \frac{\langle f||\boldsymbol{r}||a\rangle\langle a||\boldsymbol{r}||i\rangle\langle i||\boldsymbol{r}||b\rangle\langle b||\boldsymbol{r}||f\rangle}{(\Delta\nu_{ai} + {\rm i}\Gamma_a/4\pi)(\Delta\nu_{bi} - {\rm i}\Gamma_{\rm b}/4\pi)}(-1)^{M_K} \\ &\quad \times \begin{pmatrix} 1 & 1 & K \\ \mu & \nu & -M_K \end{pmatrix} \begin{pmatrix} 1 & 1 & K \\ \alpha & \beta & M_K \end{pmatrix} \\ &\quad \times (-1)^{F_{\rm f}-M_{\rm f}} \begin{pmatrix} F_{\rm f} & 1 & F_a \\ -M_{\rm f} & \mu & M_a \end{pmatrix} \\ &\quad \times (-1)^{F_a-M_a} \begin{pmatrix} F_a & 1 & F_{\rm i} \\ -M_a & \nu & M_{\rm i} \end{pmatrix} \\ &\quad \times (-1)^{F_{\rm i}-M_{\rm i}} \begin{pmatrix} F_{\rm i} & 1 & F_{\rm b} \\ -M_{\rm i} & \alpha & M_{\rm b} \end{pmatrix} \\ &\quad \times (-1)^{F_{\rm b}-M_{\rm b}} \begin{pmatrix} F_{\rm b} & 1 & F_{\rm f} \\ -M_{\rm b} & \beta & M_{\rm f} \end{pmatrix}.\end{aligned} \tag{B9}$$

The complicated sums of $3j$ symbols can be reduced by applying the reduction formula (equation 6.2.8 of Edmonds 1960) twice and then using $3j$ symbol orthogonality. This eliminates all summation over magnetic quantum numbers:

$$\begin{aligned}\bar{G}^{(K)}_{F_{\rm i}\rightarrow F_{\rm f}} &= \frac{2K+1}{2F_{\rm i}+1}\frac{e^4}{h^2} \sum_{a,b} (-1)^{F_a-F_{\rm b}} \\ &\quad \times \frac{\langle f||\boldsymbol{r}||a\rangle\langle a||\boldsymbol{r}||i\rangle\langle i||\boldsymbol{r}||b\rangle\langle b||\boldsymbol{r}||f\rangle}{(\Delta\nu_{ai} + {\rm i}\Gamma_a/4\pi)(\Delta\nu_{bi} - {\rm i}\Gamma_{\rm b}/4\pi)} \\ &\quad \times \begin{Bmatrix} F_{\rm f} & F_{\rm i} & K \\ 1 & 1 & F_a \end{Bmatrix} \begin{Bmatrix} F_{\rm f} & F_{\rm i} & K \\ 1 & 1 & F_{\rm b} \end{Bmatrix}.\end{aligned} \tag{B10}$$

Similar expressions are given by Omont, Smith & Cooper (1972) and Domke & Hubeny (1988) for the case where there is a single (possibly degenerate) intermediate level.

The cross-section is given by equations (61.7) and (61.9) of Berestetskii et al. (1971).[10] Noting that the phase space factors involving the frequency can be evaluated at $\nu_{{\rm Ly}\alpha}$ with negligible error yields a total cross-section

$$\sigma_{F_{\rm i}\rightarrow F_{\rm f}} = \frac{128\pi^5\nu_{{\rm Ly}\alpha}^4}{9c^4}\left[\bar{G}^{(0)} + \bar{G}^{(1)} + \bar{G}^{(2)}\right]. \tag{B11}$$

The angular dependence is given by

$$\frac{{\rm d}\sigma_{F_{\rm i}\rightarrow F_{\rm f}}}{{\rm d}\Omega} = \frac{\sigma_{F_{\rm i}\rightarrow F_{\rm f}}}{4\pi}[1 + 5\varpi_{2;F_{\rm i}\rightarrow F_{\rm f}} P_2(\cos\theta)], \tag{B12}$$

where P_2 is a Legendre polynomial and the phase function is

$$\varpi_{2;F_{\rm i}\rightarrow F_{\rm f}} = \frac{(1/10)\bar{G}^{(0)} - (1/20)\bar{G}^{(1)} + (1/100)\bar{G}^{(2)}}{\bar{G}^{(0)} + \bar{G}^{(1)} + \bar{G}^{(2)}}. \tag{B13}$$

Repeated scattering of Lyα photons eliminates any polarization so the polarization dependence is not needed.

The scattering cross-sections can then be determined in terms of the detunings for the six hyperfine transitions of Lyα, shown in Table B1, and their HWHM values,

$$\gamma = \frac{\Gamma_{\rm 2p}}{4\pi} = \frac{16\pi^3 e^2 \nu_{{\rm Ly}\alpha}^3}{9hc^3}\,|\langle {\rm 2p}||\boldsymbol{r}||{\rm 1s}\rangle|^2 = 50\,{\rm MHz} \tag{B14}$$

(γ is the same for all 2p levels on account of the sum rules). In writing the cross-sections, it is convenient to define the normalized Lorentzian profiles

$$\phi_{\rm AA} = \frac{\gamma}{\pi\left(\Delta\nu_{\rm A}^2 + \gamma^2\right)} \tag{B15}$$

and the interference profiles

$$\phi_{\rm AB} = \frac{\gamma(\Delta\nu_{\rm A}\Delta\nu_{\rm B} + \gamma^2)}{\pi\left(\Delta\nu_{\rm A}^2 + \gamma^2\right)\left(\Delta\nu_{\rm B}^2 + \gamma^2\right)}. \tag{B16}$$

(Similar definitions are used for profiles of other lines.) The cross-sections in the rest frame of the atom are

$$\sigma(F_{\rm i} \rightarrow F_{\rm f}) = \frac{3}{8\pi}\lambda_{{\rm Ly}\alpha}^2 \Gamma_{\rm 2p} \phi^{\rm u}_{F_{\rm i}F_{\rm f}}, \tag{B17}$$

[10] Berestetskii et al. (1971) denote our $\bar{G}^{(0)}$, $\bar{G}^{(1)}$ and $\bar{G}^{(2)}$ by $3G^0$, G^a and G^s, respectively. This can be seen by noting that their equation (60.10) is precisely the projection $\Pi^{(K)}$, $K = 0, 1, 2$, but in the Cartesian instead of the polar basis.

where

$$\phi^{\rm u}_{00} = \frac{1}{9}\phi_{\rm CC} + \frac{4}{9}\phi_{\rm FF} + \frac{4}{9}\phi_{\rm CF},$$

$$\phi^{\rm u}_{11} = \frac{1}{9}\phi_{\rm AA} + \frac{4}{27}\phi_{\rm BB} + \frac{1}{27}\phi_{\rm DD} + \frac{5}{9}\phi_{\rm EE} + \frac{4}{27}\phi_{\rm BD},$$

$$\phi^{\rm u}_{01} = \frac{2}{9}\phi_{\rm CC} + \frac{2}{9}\phi_{\rm FF} - \frac{4}{9}\phi_{\rm CF} \quad \text{and}$$

$$\phi^{\rm u}_{10} = \frac{2}{27}\phi_{\rm BB} + \frac{2}{27}\phi_{\rm DD} - \frac{4}{27}\phi_{\rm BD}. \tag{B18}$$

(The superscript u indicates that these profiles are unconvolved and do not include thermal broadening.) These satisfy the line profile normalization conditions

$$\sum_{F_{\rm f}=0}^{1}\int_{-\infty}^{\infty}\phi^{\rm u}_{F_{\rm i}F_{\rm f}}(\Delta\nu)\,{\rm d}\Delta\nu = 1. \tag{B19}$$

In gas with a finite temperature, all profiles must be convolved with a Gaussian of 1σ width

$$\sigma_\nu = \sqrt{\frac{k_{\rm B}T}{m_{\rm p}c^2}}\,\nu_{{\rm Ly}\alpha}, \tag{B20}$$

that is,

$$\phi_{F_{\rm i}F_{\rm f}}(\nu) = \int_{-\infty}^{\infty}\phi^{u}_{F_{\rm i}F_{\rm f}}(\nu')\frac{1}{\sqrt{2\pi}\sigma_\nu}{\rm e}^{-(\nu-\nu')^2/2\sigma_\nu^2}\,{\rm d}\nu'. \tag{B21}$$

The phase factors $\varpi_{2;F_{\rm i}\rightarrow F_{\rm f}}$ are

$$\varpi_{2;0\rightarrow0} = \frac{1}{10},$$

$$\varpi_{2;1\rightarrow1} = \frac{1}{40}(4\phi_{\rm BB} + \phi_{\rm DD} + 21\phi_{\rm EE} + 24\phi_{\rm AE} + 4\phi_{\rm BD} + 36\phi_{\rm BE} + 18\phi_{\rm DE}) \times (\phi_{\rm AA} + 4\phi_{\rm BB} + \phi_{\rm DD} + 15\phi_{\rm EE} + 4\phi_{\rm BD})^{-1},$$

$$\varpi_{2;0\rightarrow1} = -\frac{1}{20} \quad \text{and}$$

$$\varpi_{2;1\rightarrow0} = -\frac{1}{20}. \tag{B22}$$

Note that $\varpi_{2;1\rightarrow1}$ is frequency dependent because there are several resonances with different symmetries that contribute to it. This frequency must of course be evaluated in the frame of the atom rather than the frame at rest with respect to the bulk gas.

APPENDIX C: RANDOM-VELOCITY GENERATOR

This appendix presents an algorithm for generating random variables $u_\parallel$ from the distribution of equation (44). This distribution is an appropriately normalized version of a Gaussian (the Maxwellian velocity distribution of the H atoms) times a resonance line profile. In this case the resonance line profile is complicated and has up to four separate resonances, including interference terms. There are existing algorithms (Lee 1977, 1982) for the case where the resonance line profile is Lorentzian in the frame of the atom, and the algorithm given here draws on many of the same concepts. Our version of the algorithm is not highly optimized and there are places where it could be sped up significantly at the expense of additional complexity, but its speed is adequate for our purposes. In particular, the code is fast within 1–$2\sigma_\nu$ of the Doppler cores of the H I $1{\rm s}_{1/2}$–$2{\rm p}_{1/2}$ and $1{\rm s}_{1/2}$–$2{\rm p}_{3/2}$ lines, and since nearly all scatterings occur in these regions there is little to be gained by speeding up the code at other frequencies.

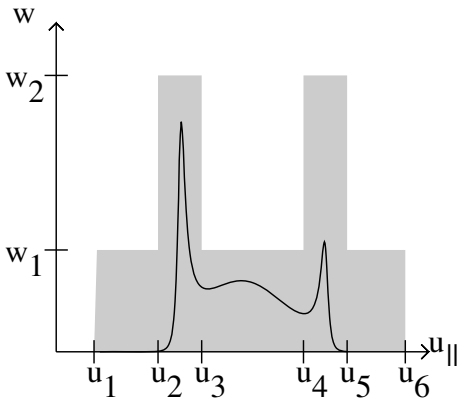

Figure C1. The region chosen for the rejection algorithm. The shaded region corresponds to the region within which $(u_\parallel, w)$ pairs are chosen, and the solid line indicates the upper boundary of the acceptance region. The scale in the figure is schematic only.

The distribution here is generated by first restricting to $|u_\parallel| \leqslant 7\sigma_\nu$, which introduces negligible error since only a fraction $\sim 1.3\times10^{-12}$ of the H atoms have higher velocities $|u_\parallel|$ than this. We then use a rejection method with a piecewise constant comparison function. Specifically, we begin by defining the region in the $(u_\parallel, w)$-plane shown in Fig. C1. The boundaries $\{u_j\}_{j=1}^6$ are chosen to enclose both the Gaussian (Maxwell) distribution and the Lyα resonances. They are

$$u_1 = -7\sigma_\nu,$$

$$u_2 = \nu^{(1)} - \nu_{\rm E} - 5\gamma,$$

$$u_3 = \nu^{(1)} - \nu_{\rm D} + 5\gamma,$$

$$u_4 = \nu^{(1)} - \nu_{\rm B} - 5\gamma,$$

$$u_5 = \nu^{(1)} - \nu_{\rm A} + 5\gamma \quad \text{and}$$

$$u_6 = 7\sigma_\nu \tag{C1}$$

if $F_{\rm i} = 1$. If $F_{\rm i} = 0$ one substitutes $\nu_{\rm A}, \nu_{\rm B} \rightarrow \nu_{\rm C}$ and $\nu_{\rm D}, \nu_{\rm E} \rightarrow \nu_{\rm F}$, since these are the resonances that can be excited from $F_{\rm i} = 0$.

The upper limits are chosen as follows. The non-resonant upper limit is

$$w_1 = \max_{j=1}^{6}\phi^{u}_{F_{\rm i}F_{\rm f}}(\nu^{(1)} - u_j); \tag{C2}$$

this is an upper limit to ${\rm e}^{-u_\parallel^2/2\sigma_\nu^2}\phi^{u}_{F_{\rm i}F_{\rm f}}(\nu^{(1)}-u)$ in the entire region of interest excluding the resonance regions $[u_2, u_3]$ and $[u_4, u_5]$, since ${\rm e}^{-u_\parallel^2/2\sigma_\nu^2}$ never exceeds 1 and $\phi^{u}_{F_{\rm i}F_{\rm f}(\nu(1)-u)}$ has local maxima only at the resonance peaks. The resonant upper limit is

$$w_2 = \mathcal{R}\exp\left(-\frac{1}{2\sigma_\nu^2}|u_\parallel|^2_{\rm min}\right), \tag{C3}$$

where $|u_\parallel|_{\rm min}$ is the minimum value of $u_\parallel$ in the resonant regions $[u_2, u_3]$ and $[u_4, u_5]$. The amplitude $\mathcal{R}$ can be any number greater than the maximum of $\phi^u_{F_{\rm i}F_{\rm f}(\nu')}$; this guarantees that w_2 is an upper limit to $e^{-u_\parallel^2/2\sigma_v^2}\phi^u_{F_{\rm i}F_{\rm f}}(\nu^{(1)}-u)$ within the resonant regions. Here we choose $\mathcal{R}$ to be $0.156/\gamma$, $0.078/\gamma$, $0.026/\gamma$ and $0.207/\gamma$ for $0\rightarrow 0$, $0\rightarrow 1$, $1\rightarrow 0$ and $1\rightarrow 1$ scatterings, respectively.

Once the point $(u_\parallel, w)$ has been chosen, we accept it if

$$w < e^{-u_\parallel^2/2\sigma_v^2}\phi^u_{F_{\rm i}F_{\rm f}}(\nu^{(1)}-u); \tag{C4}$$

if this is not the case, we generate a new point.

This paper has been typeset from a TEX/LATEX file prepared by the author.