



Prospects of Detecting Nonthermal Protons in Solar Flares via Lyman Line Spectroscopy: Revisiting the Orrall–Zirker Effect

Graham S. Kerr^{1,2}, Joel C. Allred³, Adam F. Kowalski^{4,5}, Ryan O. Milligan^{2,6}, Hugh S. Hudson^{7,8},
Natalia Zambrana Prado^{2,3}, Therese A. Kucera³, and Jeffrey W. Brosius^{2,3}

¹ NASA Goddard Space Flight Center, Heliophysics Science Division, Code 671, 8800 Greenbelt Road, Greenbelt, MD 20771, USA; graham.s.kerr@nasa.gov, kerrg@cua.edu

² Department of Physics, Catholic University of America, 620 Michigan Avenue, Northeast, Washington, DC 20064, USA

³ NASA Goddard Space Flight Center, Heliophysics Science Division, Code 671, 8800 Greenbelt Road, Greenbelt, MD 20771, USA

⁴ National Solar Observatory, University of Colorado Boulder, 3665 Discovery Drive, Boulder CO 80303, USA

⁵ Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder 2000 Colorado Ave, CO 80305, USA

⁶ Queen's University Belfast, University Rd, Belfast BT7 1NN, UK

⁷ SUPA School of Physics & Astronomy, University of Glasgow, Glasgow, G12 8QQ, UK

⁸ Space Sciences Laboratory, UC Berkeley, CA 94720, USA

Received 2023 January 29; revised 2023 February 3; accepted 2023 February 3; published 2023 March 22

Abstract

Solar flares are efficient particle accelerators, with a substantial fraction of the energy released manifesting as nonthermal particles. While the role that nonthermal electrons play in transporting flare energy is well studied, the properties and importance of nonthermal protons are rather less well understood. This is in large part due to the paucity of diagnostics, particularly at the lower-energy (deka-keV) range of nonthermal proton distributions in flares. One means to identify the presence of deka-keV protons is by an effect originally described by Orrall & Zirker. In the Orrall–Zirker effect, nonthermal protons interact with ambient neutral hydrogen, and via charge exchange produce a population of energetic neutral atoms (ENAs) in the chromosphere. These ENAs subsequently produce an extremely redshifted photon in the red wings of hydrogen spectral lines. We revisit predictions of the strength of this effect using modern interaction cross sections, and numerical models capable of self-consistently simulating the flaring nonequilibrium ionization stratification, and the nonthermal proton distribution (and, crucially, their feedback on each other). We synthesize both the thermal and nonthermal emission from Ly α and Ly β , the most promising lines that may exhibit a detectable signal. These new predictions are weaker and more transient than prior estimates, but the effects should be detectable in fortuitous circumstances. We degrade the Ly β emission to the resolution of the Spectral Imaging of the Coronal Environment (SPICE) instrument on board Solar Orbiter, demonstrating that though likely difficult, it should be possible to detect the presence of nonthermal protons in flares observed by SPICE.

Unified Astronomy Thesaurus concepts: Solar flares (1496); Solar flare spectra (1982); Solar chromosphere (1479); Radiative transfer (1335); Radiative processes (2055); Charge transfer (2218); Nonthermal radiation sources (1119); Radiative transfer simulations (1967); Hydrodynamical simulations (767); Solar ultraviolet emission (1533)

1. Introduction

Solar flares and other forms of solar magnetic activity often generate high-energy particles, with energies far in excess of the highest mean energies ($kT \approx 1$ keV) for any collisionally relaxed solar plasma population (e.g., Vilmer 2012). The bremsstrahlung continuum allows us to observe electrons above $E \sim 10$ keV by the techniques of X-ray astronomy, and to characterize them spatially, spectrally, and temporally (reviews in the solar flare context include Holman et al. 2011; Kontar et al. 2011). This knowledge underpins the development of “thick target” beam models (e.g., Brown 1971; Hudson 1972; Emslie 1978), which have had extensive theoretical development. The particularly close spatial and temporal association between deka-keV hard X-ray sources and flare ribbons and footpoints observed in the UV/optical/infrared (e.g., Fletcher et al. 2011) supports the idea that accelerated electrons dominate the transport and dissipation of

flare energy during the impulsive phase (Kane & Anderson 1970) of an event.

In some flares we also observe strong emissions from high-energy ions—protons and, to a lesser extent, α particles and other heavier ions (e.g., Shih et al. 2009). In such cases they may well carry energy comparable to nonthermal electrons (Emslie et al. 2012). By excluding nonthermal ions in flare models, we are potentially ignoring up to half of the flare energy transported through the Sun’s atmosphere, or at least an energetically important constituent.

Unfortunately, characterizing their properties (or even their presence in chromospheric flare footpoints) is more challenging as nonthermal ions do not have such convenient signatures as nonthermal electrons do. Chupp et al. (1981) initially discovered solar γ -ray line emission in the form of the 2.223 MeV line produced by deuterium formation ($n + p \rightarrow {}^2\text{H}$). This implied the existence of free neutrons as a part of the flare process, and these could only have come from nuclear reactions: hence, the acceleration of high-energy ions (as already well known from solar cosmic rays, now known as solar energetic particles, SEPs). These first observations also showed the presence of the positron-annihilation line at 511 keV and of inelastic-scattering (spallation) lines, such as the

4.43 MeV line of ^{12}C , as well. These processes had all been anticipated theoretically, based on the prior knowledge of SEPs, but the γ -radiation gave us hints about the interactions of such particles in the lower solar atmosphere and their possible role in flare dynamics. Limitations of detector sensitivity generally make it almost impossible to use γ rays to detect ions from thin-target processes in the corona. The positron-annihilation and deuterium-formation lines result from secondary processes; the initial spallation reactions produce unstable nuclei that emit positrons and neutrons, respectively, and these must slow down collisionally before they interact. The interpretation of these delayed emission lines therefore involves additional model assumptions (e.g., Ramaty et al. 1975).

The γ -ray observations (for ions) have severe limitations in comparison with the hard X-ray bremsstrahlung (for electrons), however. In addition to the major technical problems of astronomy in this photon energy range, the inelastic-scattering lines alone provide almost no spectral information. Thus our remote-sensing knowledge of nonthermal ions actually in the solar flare volume basically only really reflects the proton population at about 10 MeV, the typical energy threshold for these reactions. Given coronal values of $kT_e \approx 100$ eV, or at most a few keV in flares, this leaves a huge gap in our knowledge. Extrapolating down to 1 MeV/nucleon, the energy of the flare ion population already rivals that of the nonthermal electrons in some cases at least (e.g., Ramaty et al. 1995), suggesting orders-of-magnitude uncertainty in total particle energy if one were to consider (still highly nonthermal) particles at 10–100 keV nucleon $^{-1}$.

In spite of the challenges of working in the few-MeV range of photon energies, the Reuven Ramaty High-Energy Spectroscopic Imager (RHESSI; Lin et al. 2002) succeeded in obtaining useful images of solar flares in a small number of events (Hurford et al. 2003, 2006). These results localized high-energy ions, with their sources consistent with the flare loops (Lin 2011), confirming the conclusion drawn from the rapid time variability previously observed. The 2.223 MeV line proved to come from flare footpoint regions, as expected theoretically, though with centroids curiously offset from the 200–300 keV X-ray (i.e., electron) sources by some $[14\text{--}20]'' \pm 5''$ in the three events studied. The RHESSI data also strongly suggested that the acceleration mechanism for the flare ions had a close relationship with that observed via the relatively well-observed X-ray bremsstrahlung above about 300 keV (Shih et al. 2009). This key observation, identifying high-energy particles within closed magnetic field regions (the flare loops), makes the study of ion beams interesting, many decades after this was proposed on the basis of early observations of SEPs (Švestka 1970; Najita & Orrall 1970; Simnett 1995).

The likely existence of ion beams in flares provides a crucial motivation for the work presented in this paper. Here we study the Orrall–Zirker effect (hereinafter OZ), as proposed by Orrall & Zirker (1976). Briefly, successful observations of OZ, via the Ly α line wings, would permit remote sensing of the potentially dominant low-energy ions in the energy range $\sim 0.01\text{--}1$ MeV for the first time. Figure 1 sketches the OZ process: precipitating ions enter the partially ionized lower atmosphere, pick up electrons from neutral H atoms by atomic charge exchange reactions, and then spontaneously radiate a bound–bound photon in flight. This photon would be Doppler shifted by some amount depending on the original energy (velocity) of the proton. In the simplest picture this would enhance the red wing of Ly α or Ly β because of the

essentially downward motion of the beam particle. At exact beam alignment (the observer on the beam axis) 10–100 keV protons would appear as displacements of 5.6–17.7 Å in the red wing of Ly α , well beyond likely flaring line widths of a few Å; we do not yet have enough direct observation of this line to be more precise about this. We note that stellar Ly α spectra obtained by the Hubble Space Telescope (Wood et al. 2005) show flare-star Ly α line profiles no more than a few Å wide in quiescent conditions.

Follow on studies by Canfield & Chang (1985), Fang et al. (1995), and Brosius & Woodgate (1999) confirmed and expanded upon the initial theoretical work of Orrall & Zirker (1976). Fang et al. (1995) noted that the predictions of the strength of nonthermal emission from models that assumed a homogeneous, and low, ionization fraction, were overestimated. However, even though they used a more realistic ionization stratification, they used static semiempirical flare atmospheres and did not model self-consistently the feedback between the nonthermal protons and the atmospheric stratification. Canfield & Chang (1985) and Brosius & Woodgate (1999), investigated the effect that different proton beam distributions had on the appearance of the nonthermal feature, indicating diagnostic potential. Brosius & Woodgate (1999) looked at the temporal evolution of the nonthermal emission, resulting from the atmospheric evolution, using a more sophisticated treatment of the nonthermal proton thermalization than employed by the other studies. They focused on weaker energy fluxes than Canfield & Chang (1985), and noted an initially strong source, that quickly diminished but then increased somewhat. However, they did not calculate the ambient emission⁹ in their flare simulations, to compare against the strength of the nonthermal emission.

The only tantalizing signature of the OZ effect from a bonafide redshifted feature of the OZ effect was reported by Woodgate et al. (1992), who used data from the Goddard High Resolution spectrograph (GHRS) on the Hubble Space Telescope (HST). Those observations of Ly α from a moderately strong stellar flare that occurred on the dMe star AU Mic revealed a short ($t \sim 3$ s) but clear enhancement to redward of the line core relative to the blueward side. Unfortunately, follow-up studies of flares on AU Mic had null detections of nonthermal Ly α emission (Robinson et al. 1993, 2001; Feinstein et al. 2022). As regards solar flares, we have lacked routine high-quality observations of the Lyman lines in flares, and those that we do possess have not revealed the OZ effect. Searching for the equivalent feature from charge exchange between a nonthermal α -particle beam producing nonthermal He II 304 Å emission (Peter et al. 1990), Brosius (2001) found no evidence using Solar and Heliospheric Observatory (SOHO)/Coronal Diagnostics Spectrometer (CDS) data (which observed this line in second order). Similarly, Hudson et al. (2012) were unable to detect nonthermal He II 304 Å emission in large γ -ray events using Solar Dynamics Observatory (SDO)/EUV Variability Experiment (EVE) observations. Outwith the solar flare context, a similar process results in proton aurorae on Mars, where charge exchange between the solar wind and the Martian exosphere results in enhancements to Ly α (see the recent observations and discussions of this effect by Hughes et al. 2019).

⁹ When we refer to “ambient” emission throughout, we mean the intensity in the absence of nonthermal emission following charge exchange; that is the Lyman line enhancements following flare heating in the original RADYN+FP simulations.

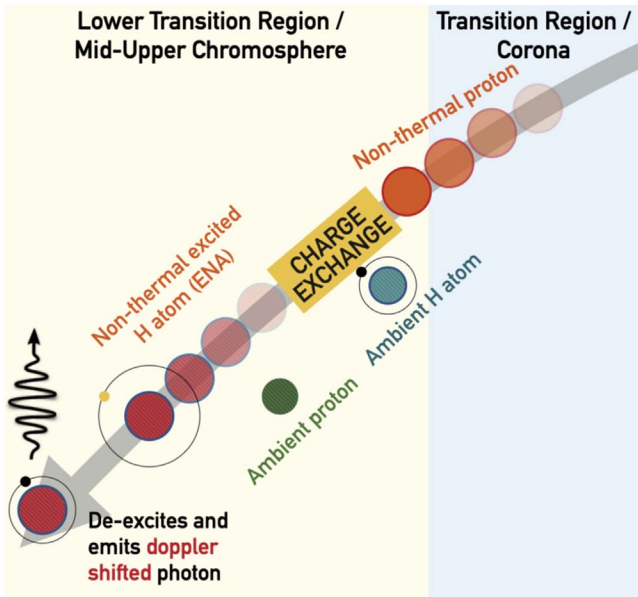


Figure 1. A schematic of the Orrall–Zirker (OZ) Effect, in which a distribution of nonthermal protons streams through the solar atmosphere until it reaches a population of neutral hydrogen atoms. Due to charge exchange with this ambient population, and subsequent collisions with ambient particles (hydrogen, protons, electrons) some fraction of the precipitating nonthermal protons is transformed into a nonthermal hydrogen atom (an energetic neutral atom, ENA) in an excited state. Spontaneous radiative de-excitation subsequently produces a highly Doppler-shifted photon (e.g., of Ly α , Ly β , H α).

While there is no track record of successful solar flare detections, it is important that we revisit the Orrall–Zirker effect now that we have more sophisticated numerical tools with which to attack the problem. Performing rigorous modern studies, of which this manuscript is the first, could lead to constraints on low-energy cutoffs of the nonthermal proton distribution, and thus total energetics and acceleration mechanisms. If our modern calculations confirm prior numerical studies that suggest that the OZ effect is, in principle, eminently observable then we must address the null detections. The solar physics community also has a renewed interest in solar Lyman line observations with several new and upcoming missions set to provide Ly α or Ly β data (including Solar Orbiter, Solar-C/EUVST, ASO-S, and the SNIFS sounding rocket). In this study we revisit the predictions of the strength of the nonthermal Lyman line emission over time in flare simulations, and importantly we compare this to the predicted ambient flare emission in the same wavelength range.

In Section 2 we describe our proton beam-driven flare models, and in Section 3 discuss in detail the processes relevant to the Orrall–Zirker effect. The results of applying our Orrall–Zirker numerical code are shown in Section 4, where we synthesize Ly α and Ly β emission. Our results confirm the earlier findings that nonthermal H α emission is far too weak to be detected (Canfield & Chang 1985; Brosius & Woodgate 1999), so we do not discuss that line in detail here. Finally, in Section 5 we degrade our predictions to the resolution of the SPICE instrument on board Solar Orbiter.

2. Numerical Flare Experiments

The field-aligned radiation hydrodynamics (RHD) code RADYN (Carlsson & Stein 1992, 1995, 1997; Abnett & Hawley 1999;

Allred et al. 2005, 2015) is a well-established resource that models the solar atmosphere’s response to energy injection. The coupled, nonlinear equations of hydrodynamics, charge conservation, radiation transfer, and nonequilibrium atomic level populations are solved on an adaptive grid (capable of resolving the steep gradients and shocks that form during flares) spanning the subphotosphere through the corona. In the preflare atmosphere the $z = 0$ point is defined as where the standard optical depth is unity, $\tau_{5000} = 1$. Crucially, an NLTE, nonequilibrium chromosphere is modeled, with feedback between radiative heating and cooling, the hydrodynamics, and on the energy transport mechanisms. RADYN has been used extensively to study many aspects of flares, from details of energy transport, driving of mass flows, and the evolution of the flaring plasma, to the formation of radiation and sources of energy losses (e.g., Kuridze et al. 2015; Kowalski et al. 2015; Kerr et al. 2016; Simoes et al. 2017; Kowalski et al. 2017; Polito et al. 2018; Brown et al. 2018; Kerr et al. 2019a, 2019b; Polito et al. 2019; Graham et al. 2020; Kerr et al. 2020, 2021; Kowalski et al. 2022; Allred et al. 2022). For a complete description of the base level of the code see Allred et al. (2015). RADYN has been modified somewhat since Allred et al. (2015), for example to include suppression of thermal conduction (Allred et al. 2022), more accurate Stark broadening of Balmer lines (Kowalski et al. 2022), and to model the transport of energetic particles via the FP (Fokker–Planck) code (Allred et al. 2020). Of those, we employ the latter in this work. For a comprehensive review of flare loop modeling, including RADYN, in the context of observations from the Interface Region Imaging Spectrograph (IRIS; De Pontieu et al. 2014) see Kerr (2022, 2023).

When simulating solar or stellar flares driven by nonthermal particles RADYN now uses FP to model the propagation, evolution, and thermalization of that particle population. FP is a standalone, open-source code that has been merged with RADYN, that while being similar to the approach used in Allred et al. (2015) offers an upgraded treatment of the solution of the Fokker–Planck kinetic theory, in particular more accurate warm-target physics, the inclusion of return currents, and the ability to model particles of arbitrary mass and charge. A nonthermal particle distribution is injected at the loop apex, which propagates through the loop, with energy lost via Coulomb collisions transferred to the plasma. This distribution is a power law defined by the total energy flux F [$\text{erg s}^{-1} \text{cm}^{-2}$] above some low-energy cutoff E_c [keV], with a spectral index δ . FP (and hence RADYN+FP) now has the ability to model a multispecies beam, but here we restrict ourselves to a proton-only distribution to simplify this initial study. Also, the only flare footpoints imaged at MeV γ -ray energies by RHESSI (from which we infer the presence of nonthermal protons) have an offset relative to the centroid positions of hard X-rays (i.e., the electron population). With the caveat that these are only a few events, such a separation may suggest that different species were accelerated along different flux tubes. In addition to energy losses, nonthermal particle distributions can cause collisional ionization or excitation of ambient species. For nonthermal collisional rates from the ground state of neutral hydrogen impacted by the proton beams we follow the treatment of Henoux et al. (1993).

RADYN+FP provides us with the time-dependent NLTE ionization stratification alongside the full spectral distribution of nonthermal protons, and self-consistently models the feedback between them. Armed with these we can make

predictions of the production of energetic neutral atoms in a more accurate manner than previous codes. In this paper we focus on studying how detectable the nonthermal Lyman line emission might be for different injected energy fluxes. As mentioned in the Introduction, we do not have a firm grasp of the total energy flux carried by nonthermal protons and so, for the time being, we elect to study the impact on the various physical processes of three regimes of flare heating where the energy flux varies over three orders of magnitude, $F = [10^9, 10^{10}, 10^{11}] \text{ erg s}^{-1} \text{ cm}^{-2}$. The smallest energy flux could represent the initial stages of energy injection into a flare footpoint before some ramp up to a higher energy flux, or it could represent the total flux injected in some small-scale heating event.

2.1. Proton Beam-driven Flares

Three proton beam-driven flares were simulated using RADYN+FP, varying the injected flux $F = [10^9, 10^{10}, 10^{11}] \text{ erg s}^{-1} \text{ cm}^{-2}$ with fixed $E_c = 150 \text{ keV}$ and $\delta = 5$. In this initial study we explore the potential of detection based on energy flux, and a fuller exploration of the parameters defining the proton distributions can be performed at a later date. Hereafter we refer to these experiments as 1F9, 1F10, 1F11, for the $F = [10^9, 10^{10}, 10^{11}] \text{ erg s}^{-1} \text{ cm}^{-2}$ models, respectively. Energy was injected at a constant rate for $t_{\text{inj}} = 20 \text{ s}$, with each loop allowed to cool over a further $t_{\text{decay}} = 80 \text{ s}$. Return-current effects were not included in these models. A return current would appear as a downward propagating electron beam in order to neutralize the current introduced by the nonthermal proton beam. As discussed in Allred et al. (2020), this could result in energy losses of the proton beam itself, or in Joule heating (operating primarily in the corona or upper transition region). Since return-current losses are typically on the order of 10 keV (Allred et al. 2020), which is a small fraction of the energy of the protons, it is safe to neglect return-current-induced energy losses of the nonthermal protons in our experiments. Any Joule heating of the corona or upper transition region would also likely be small and not impact our conclusions regarding the chromospheric footpoints.

The evolution of the stratification of temperature, electron density, and H ionization fraction is shown in Figure 2 for each experiment, where color represents time. It is clear that increasing the magnitude of the energy flux carried by the nonthermal protons has a more dramatic effect on the atmospheres.

In the 1F10 and 1F11 flares the chromosphere is strongly compressed, and the transition region is pushed deeper in altitude, with a steeper transition region in the 1F11 case. The electron density increases by orders of magnitude, with a narrow region exceeding $n_e = [2.5 \times 10^{13}, 2.5 \times 10^{14}] \text{ cm}^{-3}$ for F10 and 1F11, respectively, with the location of the peak being pushed ever deeper as the chromospheres compress. There is a more extended region of more moderately elevated electron density toward greater depths, of a few $\times 10^{12-13} \text{ cm}^{-3}$. Due to both the increased temperatures and the presence of nonthermal collisional ionizations the ionization fraction rapidly increases, particularly so for the 1F11 simulation. Even by $t = 1 \text{ s}$ there is a “wall” of ionization present in both of those simulations, where the ionization climbs from $\chi_H \sim 0.5$ to $\chi_H \sim 1$ in only a few meters, which occurs $z \sim 100 \text{ km}$ deeper for the 1F11 simulation compared to the 1F10. As time goes on the location of this ionization front pushes increasingly deeper, such that it is located at 0.71 and 0.96 Mm in the 1F11 and

1F10 simulations, respectively, by the end of the heating phase. The gradient from $\chi_H \sim 0.1$ to $\chi_H \sim 1$ is steeper for the 1F11 simulation, so that for $t \gtrsim 10 \text{ s}$ this change in ionization fraction occurs over a span of only $\Delta z \sim [10-30] \text{ km}$, compared to $\Delta z \sim [50-130] \text{ km}$ for the 1F10 simulation.

Compared to those simulations, the dynamics in the 1F9 simulation are much more modest. The flare transition region does not really change location substantially, eventually residing $z \sim 100 \text{ km}$ higher in altitude than it did before the flare. However, it is more extended, with $40 \text{ kK} < T < 1 \text{ MK}$ over a span of $\Delta z \sim 130 \text{ km}$. Given the modest temperature rise and a smaller rate of nonthermal collisions, electron densities were commensurately smaller than the other simulations, peaking at only $n_e \sim 4.5 \times 10^{12} \text{ cm}^{-3}$, and with a tail of only a few $\times 10^{11} \text{ cm}^{-3}$ through the lower chromosphere. Consequently, the ionization stratification is less sharp, with the “wall” feature being softened to occur over a span of $\Delta z \sim 10-20 \text{ km}$, with a similarly more gentle gradient down to $\chi_H \sim 0.1$.

Since we inject nonthermal protons with energies $\sim 150 \text{ keV}$, their speed is intermediate between the thermal electron and ion speeds for plasma with a temperature greater than $\sim 1 \text{ MK}$. This is the warm-target regime described in Tamres et al. (1986). Warm-target collisions are much less effective at slowing nonthermal particles than cold-target collisions (Tamres et al. 1986; Allred et al. 2020). Therefore, the protons are able to transport through the corona relatively unimpeded until they reach the transition region where they encounter cooler plasma and are stopped, thereby heating the transition region and upper chromosphere, evaporating more plasma into the corona and forcing the transition region increasingly deeper.

We briefly note that the momentum of a nonthermal proton is larger than that of a nonthermal electron, which should be accounted for. With FP we include the particle momentum self-consistently (Tamres et al. 1986; Allred et al. 2020). Comparing the acceleration induced to that of mass flows resulting from plasma heating indicates that plasma heating dominates, and that it is plasma heating that ultimately results in pushing the transition region to greater depths.

2.2. Ambient Lyman Emission in Flares

Since RADYN truncates the Lyman line profiles at 10 Doppler widths in order to mimic the effects of partial frequency redistribution (PRD) for the purposes of energetics (to avoid overestimating radiative losses in the wings), we use RH15D (Uitenbroek 2001; Pereira & Uitenbroek 2015) with our RADYN flare atmospheres as input to synthesize the Lyman lines and the region around them.

RH15D solves the radiation transport equation and atomic level populations for a desired species, given an atmosphere (the depth scale, temperature, electron density, gas velocity, and hydrogen populations). We have modified RH15D to keep the hydrogen and Ca II populations from RADYN fixed so that nonequilibrium effects and nonthermal collisions are accounted for, while still solving the statistical equilibrium atomic level populations of other species. Full NLTE radiation transport was solved for the following species (with PRD for certain transitions as appropriate, including Ly α and Ly β , using the hybrid scheme of Leenaarts et al. 2012): H I, C I+C II, O I, Si I +Si II, Ca II, He I+He II & Mg II. Other species were included as sources of background opacity. The opacity “fudge” factors of Bruls et al. (1992), which are typically included in RH and RH15D by default to mimic the UV line haze, were not used as

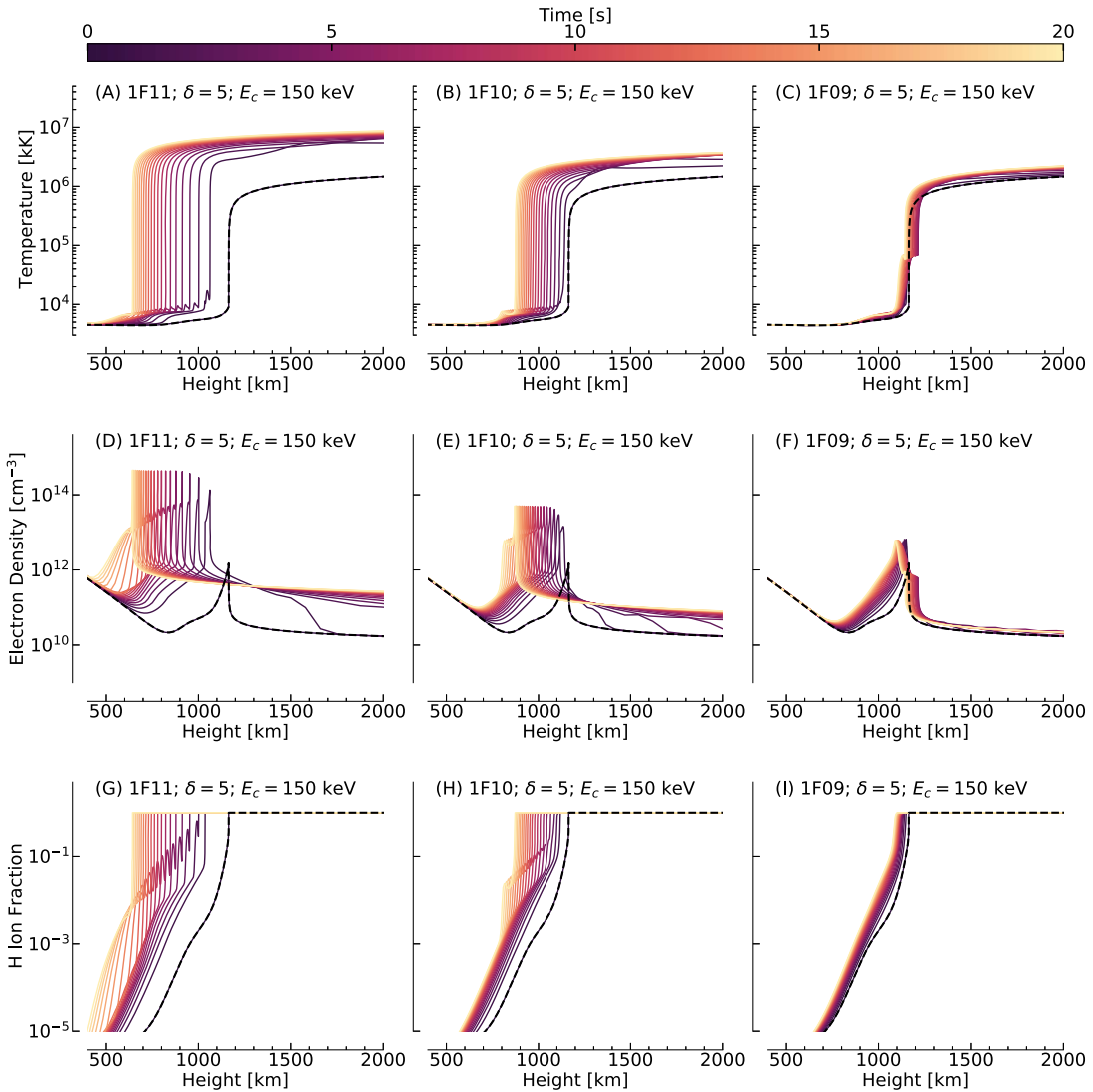


Figure 2. The atmospheric stratification for each simulation in our study. The top row (A)–(C) shows temperature, the middle row (D)–(F) shows electron density, and the bottom row (G)–(I) shows ionization fraction. Color represents time, where we show lines at $t = 1$ s cadence during the heating phase of the flares. The first column shows the $F = 1 \times 10^{11} \text{ erg s}^{-1} \text{ cm}^{-2}$ simulation, the middle column shows the $F = 1 \times 10^{10} \text{ erg s}^{-1} \text{ cm}^{-2}$ simulation, and the final column shows the $F = 1 \times 10^9 \text{ erg s}^{-1} \text{ cm}^{-2}$ simulation.

these are rather specific to a VAL-C type atmosphere (Vernazza et al. 1981) and likely not appropriate for more complex flare-like atmospheres. Instead we include opacity from the myriad of bound–bound lines not solved in detail via the Kurucz line lists,¹⁰ where hundreds of thousands of lines were included in the range $\lambda = [20\text{--}8000] \text{ \AA}$. The intensities of those lines surrounding the Ly α and Ly β lines were estimated in LTE (though with two-level scattering) with a wavelength sampling of at least 0.1 \AA (finer sampling was present where there was overlap with the model atoms included in the solution). Microturbulent broadening of $v_{\text{nthm}} = 8 \text{ km s}^{-1}$ was included, and lines were Doppler shifted as appropriate.

In addition to the RH15D solution we synthesize the OVI doublet at $\lambda = 1031.9 \text{ \AA}$ and $\lambda = 1037.6 \text{ \AA}$, which are two strong optically thin lines that appear in the vicinity of the peak of the nonthermal emission in the red wing of Ly β . The contribution functions, $G(n_e, T)$, which encapsulate various atomic processes that populate the relevant levels, were

computed for each line from the CHIANTI atomic database (V8.07; Dere et al. 1997; Del Zanna et al. 2015) for various values of temperature and electron density. These were tabulated with grid spacings of $\delta \log T [\text{K}] = 0.05$ and $\delta \log n_e [\text{cm}^{-3}] = 0.5$, and at each grid cell in the RADYN atmospheres we interpolate $G(n_e, T)$ to the local values of temperature and electron density. The intensity emitted within each cell, $I_{\lambda,z}$ is then:

$$I_{\lambda,z} = A_{\text{O}} G(n_e, T) n_{e(z)} n_{\text{H}}(z) \delta z, \quad (1)$$

where A_{O} is the elemental abundance relative to hydrogen,¹¹ δz is the length of the cell, and n_{H} is the hydrogen density. The

¹¹ We used the abundance value from Schmelz et al. (2012), $A_{\text{O},\log} = 8.61$, defined on the usual logarithmic scale, where $A_{\text{H},\log} = 12$ (the abundance is then $A_{\text{O}} = 10^{A_{\text{O},\log} - A_{\text{H},\log}} = 4.1 \times 10^{-4}$). There is much debate over whether to use the coronal or photospheric abundance values during flares, with the intensity varying by some factor based on the choice. Since we are synthesizing these lines for the purpose of determining if there is a region between them that permits the identification of nonthermal emission, it will not affect our conclusions if their intensity is over- or underestimated by some relatively small factor.

¹⁰ <http://kurucz.harvard.edu/linelists.html>

spectral lines were thermally broadened according to the local temperature, and Doppler shifted as appropriate. The total emergent intensity was then $I_\lambda = \sum_z I_{\lambda,z}$. Note that we only include these lines to help assess whether the nonthermal emission can be identified at wavelengths surrounding them, and that a more rigorous synthesis would include resonant scattering of chromospheric radiation.

From RADYN+FP+RH15D we obtained the ambient Lyman line plus surrounding emission as a function of time in our flares, on top of which the nonthermal OZ emission would appear.

3. Orrall–Zirker Effect

3.1. Overview of the Problem

As discussed in Section 1 and sketched in Figure 1, the nonthermal protons streaming through the solar atmosphere may undergo charge exchange with ambient hydrogen once they reach the chromosphere. These charge exchange interactions may occur both in the ground or excited states. Subsequent collisions can also excite a nonthermal hydrogen atom into an excited state, or cause transitions between bound excited states.

Following the discussions laid out in Orrall & Zirker (1976) and Canfield & Chang (1985), we assume that the nonthermal hydrogen atoms created through charge exchange share the same velocity (i.e., energy) as their parent nonthermal proton, and that the de-excitation of a nonthermal hydrogen atom occurs over a very short distance. This latter point means that the creation and destruction of nonthermal hydrogen is an inherently local process, such that advection terms can be ignored along with time-dependent terms (see the demonstration that de-excitation occurs much faster than deceleration in Orrall & Zirker 1976). These two assumptions allow us to write the population equation of nonthermal hydrogen levels in the form of a statistical equilibrium equation, with the total number of nonthermal particles (protons plus hydrogen atoms) within some energy range $[E, E + \delta E]$, located within some height range $[z, z + \delta z]$, conserved. This also means that we can assume any emitted photon is redshifted by an amount equal to the velocity of the original nonthermal proton that underwent charge exchange.

The number density of a nonthermal hydrogen atom, with energy E at height z , is $\eta_j(E, z)$ [cm^{-3}], where $j = (1, \dots, m)$ is the atomic level. Similarly, the nonthermal proton density is $\eta_p(E, z)$ [cm^{-3}]. Here we follow the convention in Canfield & Chang (1985) that η symbols refer to number densities in (E, z) space as opposed to the thermal population with number densities of particles defined solely on (z) space, for example the ambient electron density $n_e(z)$ or ambient hydrogen densities $n_j(z)$.

Assuming statistical equilibrium for a total creation term C (e.g., charge exchange, collisional excitation, etc.) and total destruction term D (e.g., spontaneous radiative de-excitation, collisional de-excitation, etc.):

$$\sum_{j \neq i}^m \eta_j C_{ji} - \eta_i D_i = 0. \quad (2)$$

Collating creation and destruction terms (detailed below) to a total rates matrix P , and including $j = (1, \dots, m)$ then:

$$P \cdot N = X, \quad (3)$$

$$\begin{pmatrix} P_{11} & P_{12} & P_{13} & P_{1p} \\ P_{21} & P_{22} & P_{23} & P_{2p} \\ P_{31} & P_{31} & P_{32} & P_{33} \\ P_{p1} & P_{p2} & P_{p3} & P_{pp} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_p \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

where we dropped the (E, z) dependence for clarity. To solve this system we replace one of the equations with the particle conservation equation so that the total number of nonthermal particles at (E, z) is conserved. We choose the final equation so that $P_{p1} = P_{p2} = P_{p3} = P_{pp} = 1$ and $X_p = \eta_{p,FP}$, for $\eta_{p,FP}$ the nonthermal proton distribution number density from RADYN+FP.

Once we know the nonthermal populations η , the photon emissivity, $\Phi(\Delta\lambda, z)$, of transition $j \rightarrow i$, at some $\Delta\lambda \text{ \AA}$ from line center, at height z in the atmosphere, in photons $\text{s}^{-1} \text{cm}^{-3} \text{sr}^{-1} \text{\AA}^{-1}$ is given by

$$\Phi(\Delta\lambda, z) = \frac{\sqrt{2m_p c^2}}{4\pi\lambda_{ji}} E^{1/2} \eta_j(E, z) A_{ji}, \quad (4)$$

where A_{ji} is the Einstein coefficient for spontaneous emission and m_p is the proton mass (or particle mass in the more general sense). In the above we have used the fact that $d(\Delta\lambda) = \lambda_{ji} dv/c = \lambda_{ji} dE/(c\sqrt{2m_p E})$. This becomes an emissivity in terms of energy, $\psi_{ji}(\Delta\lambda, z)$ $\text{erg s}^{-1} \text{cm}^{-3} \text{sr}^{-1} \text{\AA}^{-1}$, via the conversion,

$$\psi_{ji}(\Delta\lambda, z) = \frac{hc}{\lambda_{ji}} \Phi_{ji}(\Delta\lambda, z). \quad (5)$$

the emergent intensity is then simply,

$$I_{\text{nthm},ji}(\Delta\lambda) = \int_{z_1}^{z_2} \psi_{ji}(\Delta\lambda, z) dz, \quad (6)$$

where z_1 and z_2 are two preselected height ranges (e.g., $z_1 = 0$ and $z_2 = \infty$ if we sum over the full atmosphere). If we assume the chromosphere is optically thin with a few angstroms from the Lyman line cores, then this nonthermal intensity can be added as a straight sum to the emergent intensity from the ambient flaring emission. Note also the assumption here that the protons are streaming almost vertically near disk center. In the event that the flare is located with an increasing heliocentric angle toward the solar limb then the $\Delta\lambda$ would be somewhat different due to the dependence on the nonthermal particle velocity, v . Despite the ideal nature of the setup we are presently investigating, including angular effects would not be expected to markedly reduce the peak intensity of the emission.

3.2. Atomic Processes and Cross Sections

3.2.1. Populating Nonthermal Hydrogen Excited Levels

Here we lay out the relevant atomic processes that populate a three-level nonthermal hydrogen atom, allowing us to then synthesize Ly α and Ly β emission. These processes are: charge exchange (both to ground and excited levels); collisional excitation of bound levels by ambient particles; collisional ionization by ambient particles; collisional de-excitation by ambient particles; spontaneous radiative de-excitation of an excited level; stimulated radiative de-excitation of an excited level; radiative recombination of $\eta_p \rightarrow \eta_j$; three-body recombination of $\eta_p \rightarrow \eta_j$.

Canfield & Chang (1985) looked in detail at these processes, noting that charge exchange predominantly occurred at the ground levels, and that collisional excitation of $\eta_1 \rightarrow \eta_2, \eta_3$ was an important pathway to creating excited nonthermal hydrogen. Charge exchange was orders of magnitude more efficient than both three-body and spontaneous radiative recombination so we ignore those processes going forward. Spontaneous radiative de-excitation from the excited levels was far more rapid than any other destruction process so we additionally ignore collisional de-excitation. Radiative excitation of η_1 is also safely neglected. So, we are left with charge exchange, collisional excitation and ionization, and spontaneous radiative de-excitation as the important atomic processes to include in our model.

In the expressions that follow, terms such as C_{ij} refer to rates of collisional processes from level $i \rightarrow j$, in s^{-1} . For our three-level model hydrogen atom $j = (1, 2, 3)$, and p refers to a nonthermal proton. Charge exchange rates in s^{-1} are labeled CX_{pj} , where the electron is captured to the j -th level. Radiative terms are listed in terms of the Einstein coefficients for spontaneous emission rates A_{ji} in s^{-1} . Collisional rates are individually comprised of interactions between the beam particle, with velocity v in $cm\ s^{-1}$, and a target of ambient protons (density n_p), hydrogen atoms (density n_H), and electrons (density n_e). Atomic cross sections are defined as $Q_{ij}^{P,E,H,CX}$ cm^2 where the index refers to a collision with a proton, electron, or neutral hydrogen respectively, or that a charge exchange interaction takes place.

The ground state of nonthermal hydrogen η_1 is populated by charge exchange, and depopulated by collisional excitation or ionization;

$$\eta_1 \rightarrow \eta_p CX_{p1} - \eta_1(C_{12} + C_{13} + C_{1p}) = 0, \quad (7)$$

where in the above

$$\begin{aligned} CX_{p1} &= n_H Q_{p1}^{CX} v \\ C_{12} &= (n_p Q_{12}^P + n_H Q_{12}^H + n_e Q_{12}^E) v \\ C_{13} &= (n_p Q_{13}^P + n_H Q_{13}^H + n_e Q_{13}^E) v \\ C_{1p} &= (n_p Q_{1p}^P + n_H Q_{1p}^H + n_e Q_{1p}^E) v \end{aligned} \quad (8)$$

The first excited state of nonthermal hydrogen η_2 is populated by direct charge exchange and by collisional excitation from η_1 , and depopulated by collisional excitation or ionization plus spontaneous emission;

$$\eta_2 \rightarrow \eta_p CX_{p2} + \eta_1 C_{12} - \eta_2(C_{23} + C_{2p} + A_{21}) = 0, \quad (9)$$

where in the above

$$\begin{aligned} CX_{p2} &= n_H Q_{p2}^{CX} v \\ C_{23} &= (n_p Q_{23}^P + n_H Q_{23}^H + n_e Q_{23}^E) v \\ C_{2p} &= (n_p Q_{2p}^P + n_H Q_{2p}^H + n_e Q_{2p}^E) v \end{aligned} \quad (10)$$

The second excited state of nonthermal hydrogen η_3 is populated by direct charge exchange and by collisional excitation from η_1 and η_2 , and depopulated by collisional ionization plus spontaneous emission;

$$\eta_3 \rightarrow \eta_p CX_{p3} + \eta_1 C_{13} + \eta_2 C_{23} - \eta_3(C_{3p} + A_{31} + A_{32}) = 0, \quad (11)$$

where in the above

$$\begin{aligned} CX_{p3} &= n_H Q_{p3}^{CX} v \\ C_{3p} &= (n_p Q_{3p}^P + n_H Q_{3p}^H + n_e Q_{3p}^E) v \end{aligned} \quad (12)$$

3.2.2. Cross Sections Used in this Study

The cross sections are an important element in the prediction of the nonthermal hydrogen level populations, so we compiled more up-to-date values than used previously, where available, particularly for the charge exchange cross sections. As seen in Figures 3 and 4, both charge exchange and collisional excitation/ionization peak in the deka-keV to 100 keV range, so that the relevant wavelengths for the Lyman lines are roughly $\Delta\lambda \approx [5-50]$ Å. For certain cross sections we fit polynomial functions to data (sometimes from multiple sources), allowing us to subsequently evaluate the cross sections at any energy within the bounds of the data, for example at the nonthermal energies used by the RADYN+FP grid. Those polynomial fits were performed in $\log Q$ and $\log E$ space, for Q in 10^{-17} cm^2 and E in keV, and have the form:

$$\log Q = a_0 + \sum_i a_i (\log E)^i. \quad (13)$$

For other cross sections we used functions from the International Atomic Energy Agency's (IAEA) supplement series IAEA (1993), which provide functions fit to data that they compiled from multiple sources for each process (see their references). As discussed below, where relevant we extrapolated beyond the energy bounds quoted by various authors but fit the decay so that extrapolations are a straight line in \log - \log space, which seems justified by the $Q(E)$ values at high energies. We did not fit beyond the IAEA functions but did check that extrapolating beyond the energy bounds that they quote provided sensible results. This is only really required when considering collisions with ambient electrons, and for other processes we were able to locate $Q(E)$ values in the energy range appropriate for the problem at hand. We were unable to locate cross sections for Q_{2p}^H , Q_{23}^H , or Q_{3p}^H , at suitably high energies, and so omitted those in the rates matrix. This should not have a major impact as we do consider the same collisional processes where the targets are protons or electrons. Those missing processes were not included in previous models. Below we show the cross sections, and comment on their sources. For those that we fit polynomial functions to, their coefficients are listed in Table 1. These data, either fit by us or by the IAEA, come from both experiments and theoretical calculations.

Cross Sections for Charge Exchange: Charge exchange cross section data were taken from Winter (2009) for $E = [1-100]$ keV, and Belkić et al. (1992) for $E = [125-8000]$ keV. Those energy ranges were combined and an 8° polynomial fit from $E = [1-8000]$ keV was performed, the results of which are shown in Figure 3. Prior studies of the OZ effect used data from various sources including Bates & Dalgarno (1953), Massey et al. (1974), Stier & Barnett (1956), Cheshire et al. (1970), Shakeshaft (1978), and Ludde & Dreizler (1982). For comparison, on each panel of Figure 3 we also show the polynomial fits by Fang et al. (1995) and by Brosius & Woodgate (1999), the latter of which are comparable to those

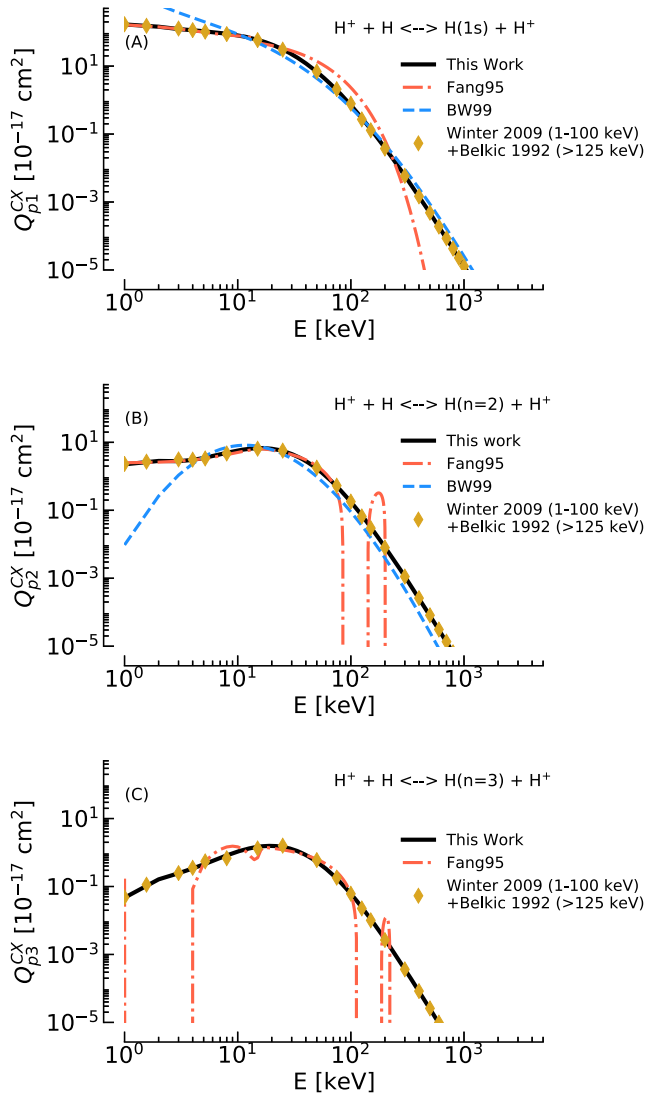


Figure 3. Charge exchange cross sections used in this study, where the black line shows the fit to data (diamonds). The functions used by Fang et al. (1995) are shown as dotted–dashed red lines, and by Brosius & Woodgate (1999) as blue dashed lines. Panel (A) is charge exchange to the ground state, panel (B) is to the first excited state ($n=2$), and panel (C) is to the second excited state ($n=3$).

used by Orrall & Zirker (1976) and Canfield & Chang (1985), to illustrate the differences with the more up-to-date cross sections (which can be large at times). Though the functions from Fang et al. (1995) are not ideal at certain energies, the most important energies were mostly adequately fit.

Cross Sections for Collisional Ionization of η_1 (Q_{1p}^I): Cross sections for collisional ionization of η_1 by ambient protons (Q_{1p}^I) were combined from Shah & Gilbody (1981); Shah et al. (1987, 1998), with $E = [38–1500]$ keV, $E = [9–75]$ keV, $E = [1.25–9]$ keV, respectively. An 8° polynomial fit from $E = [1–1500]$ keV was performed. Since in $\log Q - \log E$ space the range $E = [500–1500]$ keV is a straight line, a 1° polynomial was fit to that range, so that an extrapolation to $E > 1500$ keV can be performed if required (though in practice we really only care about $E < 1$ MeV). Q_{1p}^I is shown in Figure 4(a), alongside the underlying data.

Cross sections for collisional ionization of η_1 by ambient hydrogen (Q_{1p}^H) are the recommended values from Cariatore & Schultz (2021), in the energy range $E = [0.0362–10,000]$ keV.

An 8° polynomial fit was performed over that range, the result of which is shown in Figure 4(a), alongside the underlying data.

Cross sections for collisional ionization of η_1 by ambient electrons (Q_{1e}^E): The values from the IAEA. We used their functional form, which is shown in Figure 4(a), and can be found on page 28 of IAEA (1993).

Cross Sections for Collisional Excitation of $\eta_1 \rightarrow \eta_2$ (Q_{12}): Cross sections for collisional excitation of $\eta_1 \rightarrow \eta_2$ by ambient protons (Q_{12}^P) are the values from the IAEA. We used their functional form, which is shown in Figure 4(b), and can be found on page 46 of IAEA (1993).

Cross sections for collisional excitation of $\eta_1 \rightarrow \eta_2$ by ambient hydrogen atoms (Q_{12}^H) are taken from Hill et al. (1979) and McLaughlin & Bell (1983). For the $2s$ level the Hill et al. (1979) experimental results are used for $E = [2–25]$ keV with McLaughlin & Bell’s (1983) calculations used for $E = [36–100]$ keV. For the $2p$ level, McLaughlin & Bell (1983) used the full range 2–100 keV. The $2s$ and $2p$ cross sections are summed to obtain the total $n=2$ values. These data were fit with an 8° polynomial within the range $E = [2–100]$ keV. Since there is a linear decay (in log space) at higher energies, a 1° polynomial was fit to $E = [36–100]$ keV allowing extrapolation to $E > 100$ keV. The fit results are shown in Figure 4(b) alongside the underlying data.

Cross sections for collisional excitation of $\eta_1 \rightarrow \eta_2$ by ambient electrons (Q_{12}^E) are the values from the IAEA. We used their functional form, which is shown in Figure 4(b), and can be found on page 6 of IAEA (1993).

Cross Sections for Collisional Excitation of $\eta_1 \rightarrow \eta_3$ (Q_{13}): Cross sections for collisional excitation of $\eta_1 \rightarrow \eta_3$ by ambient protons (Q_{13}^P) are the values from the IAEA. We used their functional form, which is shown in Figure 4(c), and can be found on page 48 of IAEA (1993).

Cross sections for collisional excitation of $\eta_1 \rightarrow \eta_2$ by ambient hydrogen atoms (Q_{12}^H) are taken from McLaughlin & Bell (1987), in the energy range $E = [1–1024]$ keV. Data from the $3s$, $3p$, $3d$ levels were summed to obtain the cross sections for $n=3$. These data were fit with an 8° polynomial. Since there is a linear decay (in log space) at higher energies, a 1° polynomial was fit to $E = [324–1024]$ keV allowing extrapolation to $E > 1024$ keV. The fit results are shown in Figure 4(c) alongside the underlying data.

Cross sections for collisional excitation of $\eta_1 \rightarrow \eta_3$ by ambient electrons (Q_{13}^E) are the values from the IAEA. We used their functional form, which is shown in Figure 4(c), and can be found on page 14 of IAEA (1993).

Cross Sections for Collisional Ionization of η_2 (Q_{2p}): Cross sections for collisional ionization of η_2 by ambient protons (Q_{2p}^P) are the values from the IAEA. We used their functional form, which is shown in Figure 4(d), and can be found on page 70 of IAEA (1993).

Cross sections for collisional ionization of η_2 by ambient electrons (Q_{2e}^E) are the values from the IAEA. We used their functional form, which is shown in Figure 4(d), and can be found on page 34 of IAEA (1993).

Cross Sections for Collisional Excitation of $\eta_2 \rightarrow \eta_3$ (Q_{23}): Cross sections for collisional excitation of $\eta_2 \rightarrow \eta_3$ by ambient protons (Q_{23}^P) are the values from the IAEA. We used their functional form, which is shown in Figure 4(e), and can be found on page 56 of IAEA (1993).

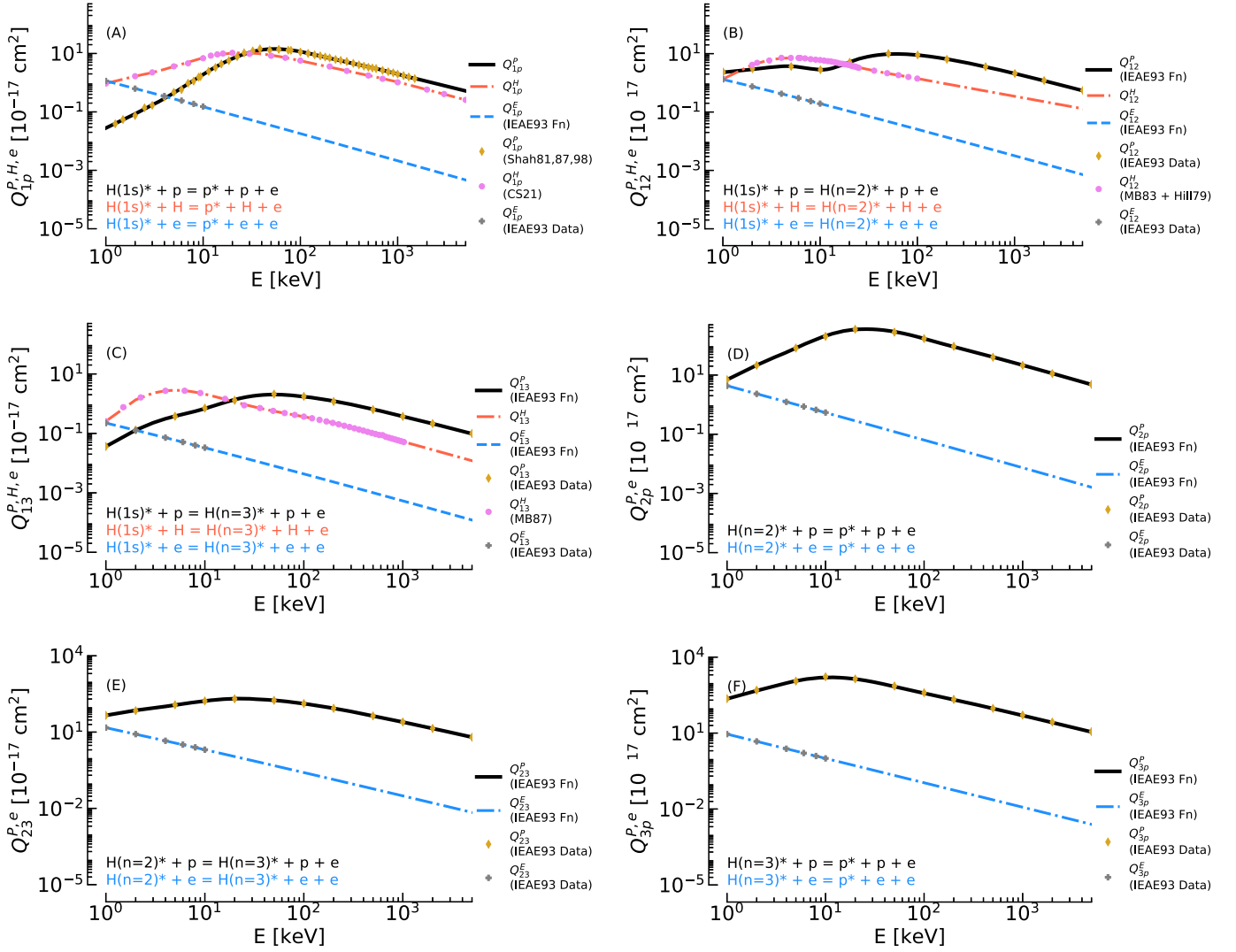


Figure 4. Cross sections for collisional excitation and ionization with ambient electrons, protons, and hydrogen (E, P, H superscripts, respectively) used in this study. Solid lines are our own polynomial fits, or are the functions from the IAEA (IAEA93 on each panel; IAEA 1993), as indicated. The underlying data are shown as symbols, and are as described in the text. Shah81,87,98 refers to Shah & Gilbody (1981); Shah et al. (1987, 1998). MB83 refers to McLaughlin & Bell (1983). Hill79 refers to Hill et al. (1979). MB87 refers to McLaughlin & Bell (1987). An asterisk in the transition descriptions indicates a nonthermal particle.

Cross sections for collisional excitation of $\eta_2 \rightarrow \eta_3$ by ambient electrons (Q_{23}^E) are the values from the IAEA. We used their functional form, which is shown in Figure 4(e), and can be found on page 22 of IAEA (1993).

Cross Sections for Collisional Ionization of η_3 (Q_{3p}): Cross sections for collisional ionization of η_3 by ambient protons (Q_{3p}^P) are the values from the IAEA. We used their functional form, which is shown in Figure 4(f), and can be found on page 72 of IAEA (1993).

Cross sections for collisional ionization of η_3 by ambient electrons (Q_{3p}^E) are the values from the IAEA. We used their functional form, which is shown in Figure 4(f), and can be found on page 36 of IAEA (1993).

3.3. Comments on OrrallZirkerPy

The model described above was implemented via a python package that we call `OrrallZirkerPy`, which can be used to postprocess flare simulations such as `RADYN+FP` to obtain the nonthermal HI emission resulting from the bombardment of the chromosphere by nonthermal protons. This code takes as

input a model atmosphere, from any source not just from `RADYN+FP`, that is the ambient hydrogen, electron, and proton number densities in cm^{-3} . Those can be defined either as a point, or on a depth scale. If input as a depth scale, the height of each grid cell must be provided in also (in cm). A single snapshot, or a series of snapshots can be input, where the time of each snapshot should be provided (in seconds). Additionally, the distribution of precipitating nonthermal particles must be input, (in units of particles $\text{cm}^{-3} \text{sr}^{-1} \text{keV}^{-1}$), along with the energy grid on which that distribution is defined (in keV), and the pitch angle grid if the particle distribution's pitch angle is resolved. The nonthermal particle distribution should be defined on the same height and time grids as the ambient atmosphere.

A default set of cross sections is included (those described in Section 3.2.2), defined for either a two- or three-level hydrogen atom alongside other necessary atomic information (particle mass, wavelengths, Einstein coefficients, etc.) Those defaults can be straightforwardly substituted to any cross sections contained within the `CrossSections` module, to which the

Table 1
Parameters of Polynomial Fits of Certain Cross Sections Used in This Study

	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	E (keV)
Q_{p1}^{CX}	2.22694	2.19952(−1)	−2.61594	4.88315	−4.06255	1.49092	−2.50756(−1)	1.43542(−1)	3.20664(−4)	1 – 8000
Q_{p2}^{CX}	3.52822(−1)	1.64356	−7.98958	1.55848(1)	−1.33785(1)	5.69352	−1.28934	1.48437(−1)	−6.77411(−3)	1 – 8000
Q_{p3}^{CX}	−1.33340	1.81849	−3.20438	8.73242	−8.72014	3.83963	−8.43943(−1)	8.84133(−2)	−3.29639(−3)	1 – 8000
Q_{1p}^p	−1.55347	1.36699	1.85672	−5.69725	9.64059	−8.12511	3.44612	−7.16542(−1)	5.84099(−2)	1 – 1500
	2.76008	−8.22412(−1)								>1500
Q_{1p}^{H}	−2.64545(−2)	1.52462	−4.62208	1.03086(1)	−1.04104(1)	5.36558	−1.49664	2.15950(−1)	−1.26636(−2)	1 – 10^4
Q_{12}^{H}	1.29154(−1)	1.59680	2.19750	−1.06375(1)	1.50168(1)	−1.13679(1)	4.78532	−1.01709	8.03237(−2)	2 – 100
	1.35631	−6.05696(−1)								>100
Q_{13}^{H}	−5.91530(−1)	3.09401	−2.40878	7.40410(−1)	−2.21064	2.89666	−1.53333	3.67348(−1)	−3.33341(−2)	1 – 10^{24}
	1.49527	−9.23204(−1)								> 10^{24}

Note. Numbers in parentheses are the indices of the powers of 10 to which the values are raised. For example $2.19952(−1) = 2.19952 \times 10^{-1}$. The valid energy ranges for each fit are indicated in the final column. Fits were performed in $\log Q$ & $\log E$ space, for Q in $\times 10^{-17} \text{ cm}^2$ and E in keV.

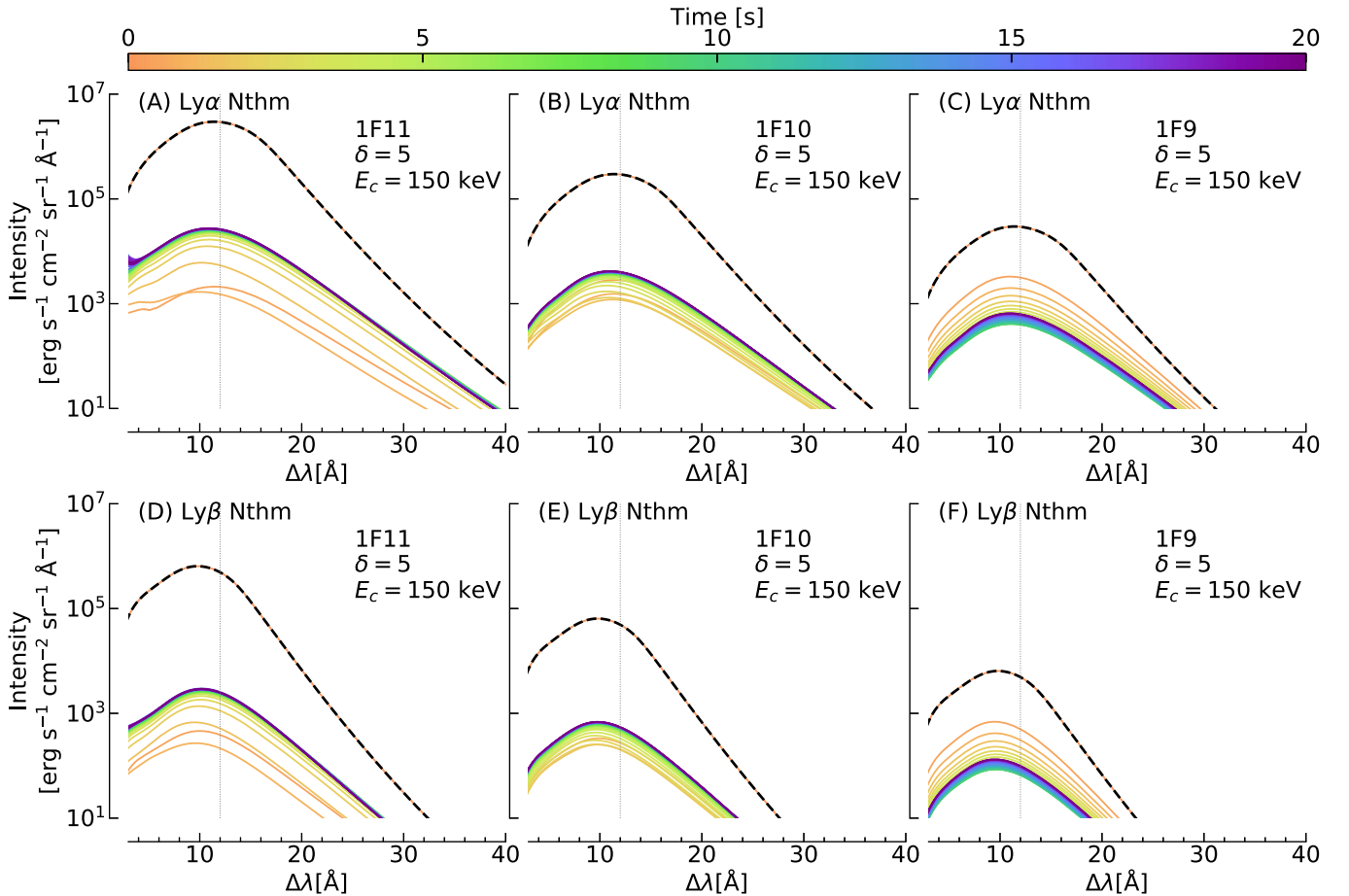


Figure 5. Nonthermal spectra of Ly α (top row, panels (A)-(C)) and Ly β (bottom, panels (D)-(F)) as a function of time where color represents the time at 0.5 s cadence. The black dashed line is the emission at $t = 0$ s which is the moment that the proton beam first impacts the chromosphere. Each column represents a different injected energy flux (i.e., flare strength). Panels (A), (D) are the 1F11 simulations, panels (B), (E) are 1F10, and panels (C), (F) are 1F9. The x-axis is the wavelength from the core of each line. The dotted vertical line on each panel shows $\Delta\lambda = 12$ Å, which is the wavelength used as a reference when showing various formation properties in other figures.

user can append any additional cross sections. That module also contains methods to fit the data from the various sources of cross sections.

From the input atmosphere and nonthermal particle distribution, and defined cross sections, `OrrallZirkerPy` solves the set of Equations (3) to obtain the nonthermal hydrogen atom level populations, where by default the final equation is replaced by the nonthermal particle conservation equation though this can be defined by the user. Using those populations the emissivity is calculated as a function of energy, and converted to a Doppler shift (in Å and km s^{-1}). Currently `OrrallZirkerPy` can model the emissivity of Ly α , Ly β , and H α . There are further methods to return either the intensity within each grid cell, or to integrate through the entire loop to obtain the total emergent intensity. Of course, the user can integrate the emissivity over any desired height range outside of `OrrallZirkerPy`.

There are plans to extend this to model nonthermal He II 304 Å emission, created in the same manner as nonthermal H I emission, but by a precipitating α -particle beam rather than protons. `OrrallZirkerPy` is available online¹².

¹² <https://github.com/grahamkerr/OrrallZirkerPy/releases/tag/v1.0.1>

4. Nonthermal Lyman Emission

Outputs from each of the `RADYN+FP` flare simulations were used as input to `OrrallZirkerPy`, producing the nonthermal Ly α & Ly β emission as functions of height and time, that is a nonthermal emissivity produced by the proton beam. The nonthermal intensity produced by the beam is then the integral of the emissivity through the full extent of the chromosphere. Here the general characteristics and formation properties of the nonthermal emission are presented, along with a comparison to the ambient thermal Lyman line emission.

4.1. Characteristics of Nonthermal Ly Emission

Nonthermal spectra are shown in Figure 5, as functions of $\Delta\lambda$ from the line center of Ly α (A-C; top row) and Ly β (D-F; bottom row). Color represents time. Though there is a slight drift over time, the wavelength of peak intensity is relatively stable. For Ly α the peak is initially $\Delta\lambda = 11.4$ Å. In the 1F11 simulation this initially rises to $\Delta\lambda = 11.6$ Å, then shortens to $\Delta\lambda = 9.6$ Å for the first few seconds, before settling to $\Delta\lambda \sim 11$ Å. The weaker simulations show less scatter, with the 1F10 simulation initially exhibiting a shortening to $\Delta\lambda = 10.6$ Å, before settling to $\Delta\lambda = 11$ Å by $t = 5.1$ s. The 1F9 shows even less scatter, though settles somewhat to a longer wavelength at $\Delta\lambda = 11.15$ Å.

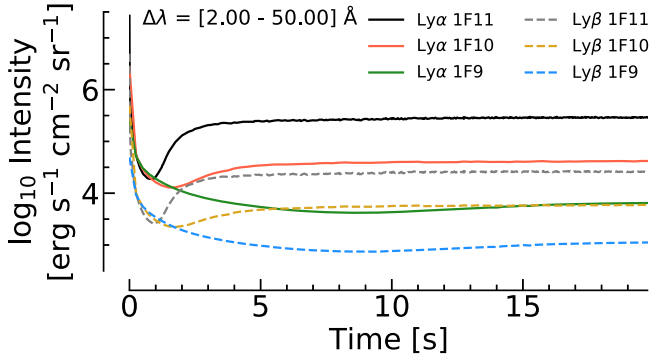


Figure 6. Lightcurves of nonthermal Ly α (solid lines) and Ly β (dashed lines) integrated over the range $\Delta\lambda = [2-50]$ Å. Black and gray are the 1F11 simulations, red and gold are the 1F10 simulations, and green and blue are the 1F9 simulations.

Since the line core of Ly β is around ~ 190 Å shorter than Ly α the nonthermal Ly β emission peaks with a somewhat shorter displacement, initially at $\Delta\lambda = 9.8$ Å. As before, in the 1F11 flare this initially lengthens (within the first $t = 1$ s) to $\Delta\lambda = 10$ Å, before shortening to a minimum value of $\Delta\lambda = 9.3$ Å, finally settling at around $\Delta\lambda = 10.25$ Å by $t = 5$ s. The 1F10 simulation initially has a decreasing peak to $\Delta\lambda = 9.5$ Å, but by $t = 5$ s has settled back to $\Delta\lambda = 9.8$. Finally, the 1F9 simulation settles quickly to $\Delta\lambda = 9.5$.

While the variations in the Doppler shift of the peak emission are small, there is a significant change in the intensity over time, with the strongest emission occurring immediately at flare onset. At that time the nonthermal protons have reached the chromosphere and some portion of the beam is neutralized before the ambient flaring Lyman emission has brightened and before the ionization fraction increases. Nonthermal Ly α emission peaks at $I_{\text{nthm,Ly}\alpha} = [3 \times 10^6, 3 \times 10^5, 3 \times 10^4]$ erg s $^{-1}$ cm $^{-2}$ sr $^{-1}$ Å $^{-1}$, for the 1F11, 1F10, 1F9 flares respectively. Just 0.25 s later this emission has dropped significantly so that each simulation actually has a comparable peak intensity of $I_{\text{nthm,Ly}\alpha} \approx 5 \times 10^3$ erg s $^{-1}$ cm $^{-2}$ sr $^{-1}$ Å $^{-1}$. Over the next few seconds the 1F10 and 1F11 flares continue to drop in intensity, but then both increase to peak values of $I_{\text{nthm,Ly}\alpha} = [4 \times 10^3, 2.6 \times 10^4]$ erg s $^{-1}$ cm $^{-2}$ sr $^{-1}$ Å $^{-1}$, respectively. The 1F9 simulation instead continues to decrease, such that by $t = 10$ s, $I_{\text{nthm,Ly}\alpha} = 4.25 \times 10^2$ erg s $^{-1}$ cm $^{-2}$ sr $^{-1}$ Å $^{-1}$. This temporal evolution can be more clearly seen in Figure 6, which shows lightcurves integrated over the range $\Delta\lambda = [2-50]$ Å.

Nonthermal Ly β emission behaves similarly, with the following peak intensities: at $t = 0$ s for 1F11, 1F10, and 1F9, $I_{\text{nthm,Ly}\beta} = [6.4 \times 10^5, 6.4 \times 10^4, 6.4 \times 10^3]$ erg s $^{-1}$ cm $^{-2}$ sr $^{-1}$ Å $^{-1}$; at $t = 0.25$ s all flares have a peak near $I_{\text{nthm,Ly}\beta} \approx 1.16 \times 10^3$ erg s $^{-1}$ cm $^{-2}$ sr $^{-1}$ Å $^{-1}$; at $t = 10$ s for 1F11, 1F10, and 1F9, the peak intensities are $I_{\text{nthm,Ly}\beta} = [2.9 \times 10^3, 6.5 \times 10^2, 8.8 \times 10^1]$ erg s $^{-1}$ cm $^{-2}$ sr $^{-1}$ Å $^{-1}$. Lightcurves of Ly β are shown as dashed lines in Figure 6, indicating that nonthermal Ly β emission follows the temporal pattern of Ly α , but is weaker.

4.2. Formation Properties of Nonthermal Ly Emission

To better understand the nonthermal emission characteristics we look at where in the atmosphere this emission forms, and what the plasma properties are at those times. Figures 7 and 8 show the emissivity of nonthermal Ly α and Ly β , respectively,

for each simulation (1F11, 1F10, 1F9 for each column left-to-right), as functions of height and wavelength, at $t = 0$ s (top rows), and $t = 8$ s (bottom rows). Integrating through height for each wavelength yields the total emergent intensity. Initially, there is a broad emitting region around the peak wavelengths, spanning a few tens of km (note the log scale on the color bar of these images). Very quickly the intensity drops, and the altitude of peak emissivity shifts deeper, more so for the stronger flare. By the latter stages of the simulation, emission from the stronger flares forms over a vanishingly narrow extent, and form progressively deeper, whereas the 1F9 simulation has a broader emitting region located at a higher altitude.

Selecting a reference wavelength of $\Delta\lambda = 12$ Å, Figure 9 shows a more detailed view of the stratification of emissivity. Here we see that the emissivity initially drops through the mid-upper chromosphere, but the lower atmosphere does not vary greatly. In the stronger simulations the peak emissivity rebounds, regaining some strength while the peak height occurs deeper and deeper. Once it has rebounded, though, the peak region is vanishingly narrow compared to the broader emitting region of the 1F9 simulation.

Following Kowalski et al. (2017) we construct the normalized cumulative distribution function of the emissivity through height $\text{NCDF}_j(z)$, and define the emitting region as the region bounded by $\text{NCDF}_j(z) = 0.05$ and $\text{NCDF}_j(z) = 0.95$; that is, the region where the bulk of the emergent intensity originates. Labeling these heights as z_{upp} and z_{low} , the mean formation height can be found by

$$\langle z \rangle = \frac{\int_{z_{\text{low}}}^{z_{\text{upp}}} j(z) z dz}{\int_{z_{\text{low}}}^{z_{\text{upp}}} j(z) dz}, \quad (14)$$

where we weight by the emissivity. This is shown for $\Delta\lambda = 12$ Å in Figure 10, where we see that nonthermal emission for the stronger flares forms significantly deeper in the atmosphere compared to the 1F9 simulation, which barely changes in altitude. Both Lyman lines form essentially at the same altitude. If we then take the width of the formation region to be $\Delta z = z_{\text{upp}} - z_{\text{low}}$, shown in panel (B), we see that for the two stronger flares when the nonthermal Lyman emission intensity drops in strength, the formation region widens up to 100–150 km. Over time, the widths of the formation region drop to only a few km, or down to meters, at which times the magnitude of the nonthermal intensity has risen somewhat but is not as strong as at the early phase of the flare. The weaker flare, however, exhibits a gradually increasing Δz such that the nonthermal Ly β emission originates from a somewhat broader region.

Performing a similar calculation to obtain the mean ionization fraction in the nonthermal emission forming region we can understand the patterns we see above. The mean ionization fraction is:

$$\langle \chi_{\text{H}} \rangle = \frac{\int_{z_{\text{low}}}^{z_{\text{upp}}} j(z) \chi_{\text{H}} dz}{\int_{z_{\text{low}}}^{z_{\text{upp}}} j(z) dz}. \quad (15)$$

These results are shown in Figure 11. For the 1F11 simulation, the mean ionization fraction initially drops before peaking close to unity. Recall, though, that this is the mean over the full

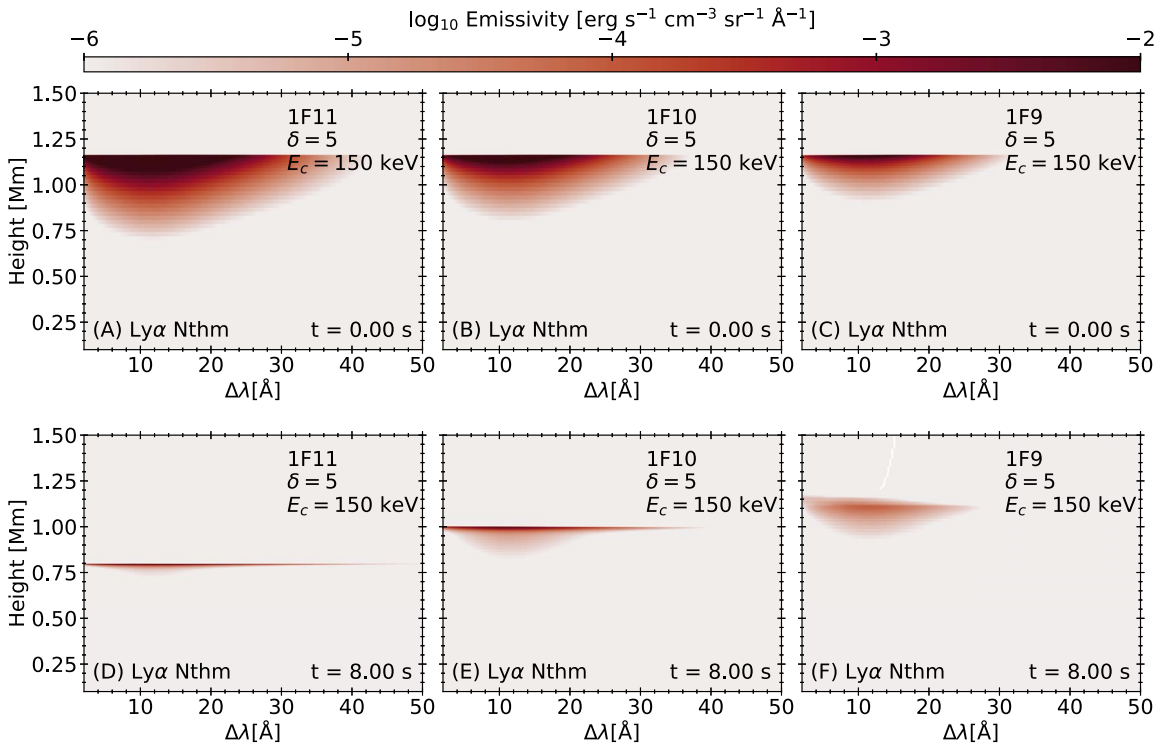


Figure 7. The emissivity on a \log_{10} scale (represented by color) of nonthermal Ly α emission at $t = 0$ s (top row, (A)-(C)) and $t = 8$ s (bottom row, (D)-(F)) as functions of both wavelength from line center and height in the atmosphere. Integrating through height yields the total emergent nonthermal Lyman line intensity shown in Figure 5. Each column is a different flare simulation: (A), (D) are the 1F11; (B), (E) are the 1F10; and (C), (F) are the 1F9 flares. The $t = 0$ s panels show the first instant of nonthermal emission caused by the proton beams. The sharp upper boundary represents the location of the flare transition region, above which there are insufficient neutrals to allow significant charge exchange interactions to take place (i.e., there is little to no nonthermal emission). The $t = 8$ s shows a time during the heating phase when atmospheric dynamics have greatly changed the interaction of the nonthermal distribution and ambient plasma.

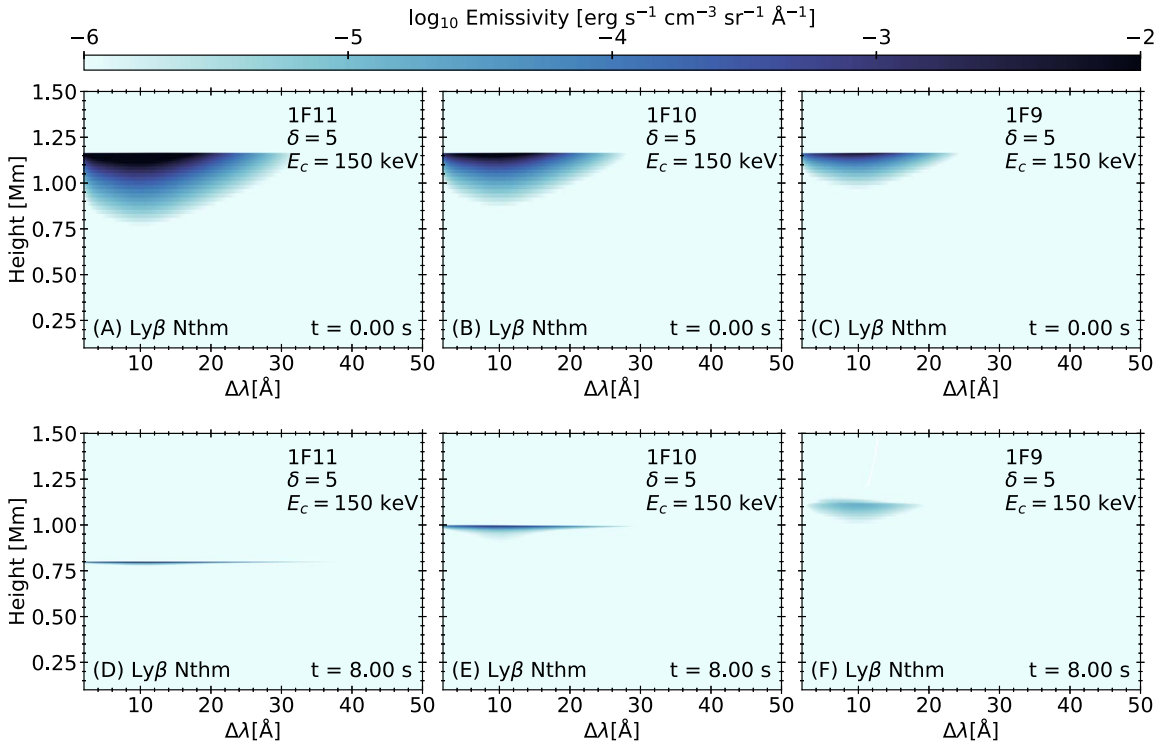


Figure 8. Same as Figure 7 but showing nonthermal Ly β emission.

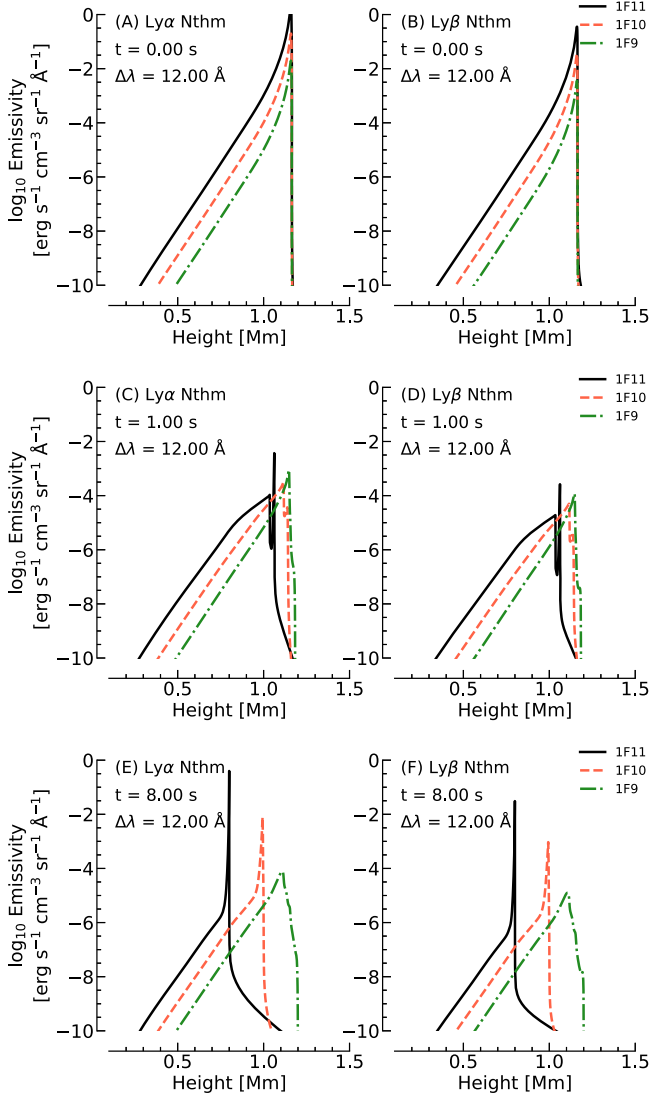


Figure 9. The emissivity as a function of height for a wavelength $\Delta\lambda = 12$ Å from line cores of Ly α (first column, panels: (A), (C), (E)) and Ly β (second column, panels: (B), (D), (F)) for three different times. Color represents the flare strength, with the black solid line being the 1F11 simulation, the red dashed line being the 1F10 simulation, and the green dotted–dashed line being the 1F9 simulation.

formation region, and that the formation region of the 1F11 nonthermal emission rapidly broadened, such that a larger fraction of the lower atmosphere influences the mean. The ionization fraction in the upper chromosphere has, however, increased. Consequently there are fewer charge exchange interactions, which require a supply of neutrals, and the nonthermal emissivity drops. As the temperature in the mid-upper chromosphere climbs, the ionization fraction climbs and emissivity continues to fall. However, as we noted before, there is a rebounding of nonthermal emission in the 1F11 simulation. This occurs due to the severe compression of the chromosphere. The proton beam encounters a fresh supply of neutrals once the transition region is pushed deeper, which picks up pace after a few seconds. At these times $\langle\chi_H\rangle$ is only somewhat larger than at the time of peak emission, but crucially, the formation region width Δz is significantly reduced (itself due to where protons are predominantly

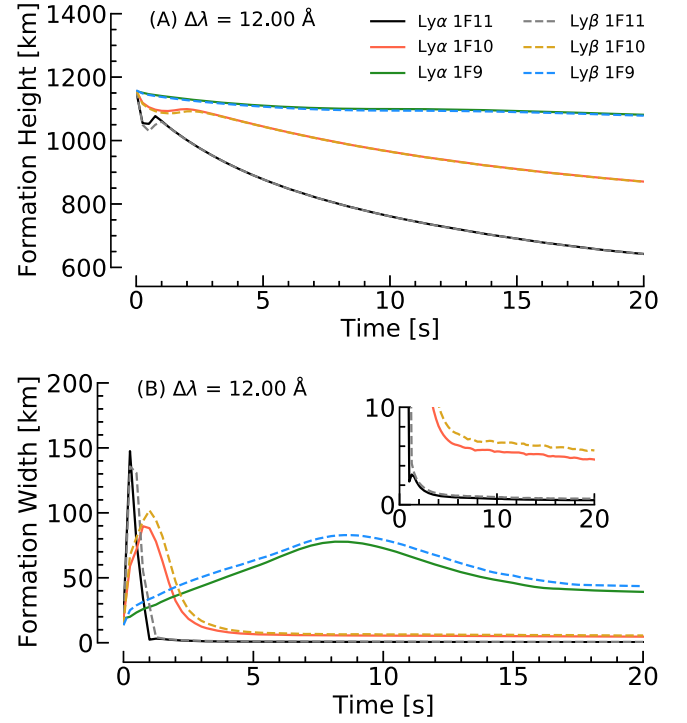


Figure 10. Panel (A) shows the mean formation height of $\Delta\lambda = 12$ Å from line cores of Ly α (solid lines) and Ly β (dashed lines) as a function of time. The black and gray lines show the 1F11 simulation, the red and gold lines show the 1F10 simulation, and the green and blue lines show the 1F9 simulation. Panel (B) shows the widths of the formation region, where colors and lines are the same as panel (A).

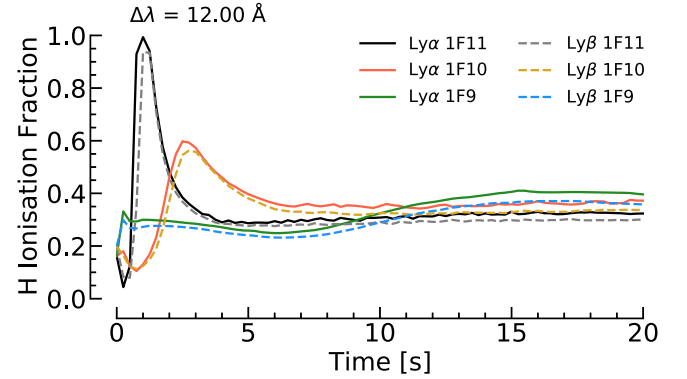


Figure 11. The mean hydrogen ionization fraction in the formation region of $\Delta\lambda = 12$ Å from line cores of Ly α (solid lines) and Ly β (dashed lines) as a function of time. The black and gray lines show the 1F11 simulation, the red and gold lines show the 1F10 simulation, and the green and blue lines show the 1F9 simulation.

thermalized by the increase in atmospheric density there). Consequently, integrating the emissivity through height produces only a modest total intensity. The 1F10 simulation behaves similarly and with the same end result, though in a less dramatic fashion, only achieving a maximum ionization fraction of $\langle\chi_H\rangle \approx 0.6$.

Since the atmospheric dynamics are less extreme in the 1F9 flare, the reduction in emissivity is smaller compared to the drop seen in the strong flares, owing to the smaller increase in ionization fraction. While the peak emissivity continues to drop due to the increasing ionization fraction, the formation region

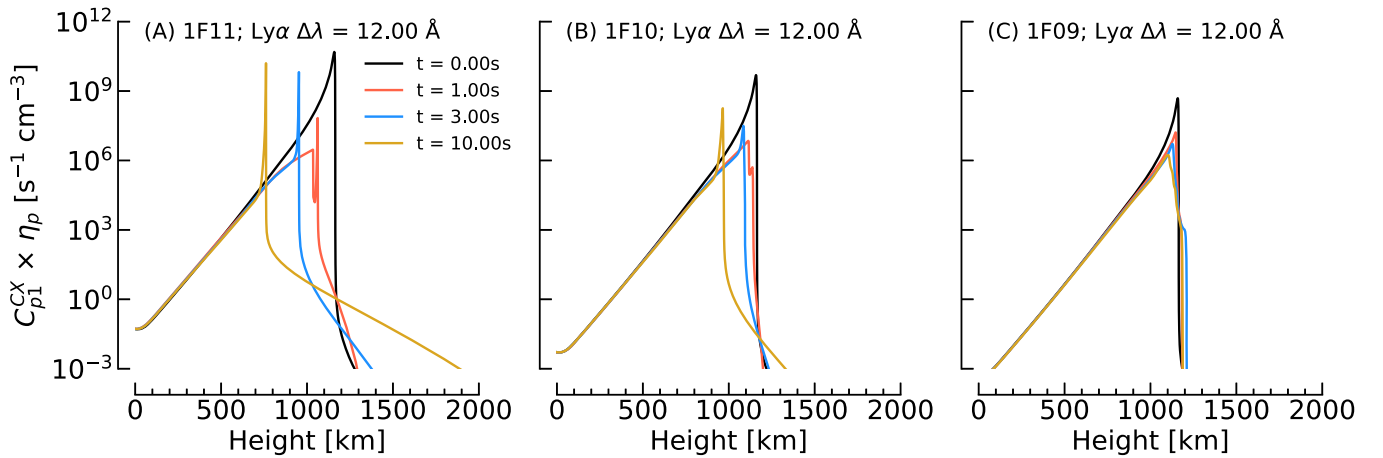


Figure 12. The rate of charge exchange interactions to the ground state, at the energy responsible for producing a Doppler shift of $\Delta\lambda = 12 \text{ \AA}$ from Ly α , as a function of height, for four times (black is $t = 0 \text{ s}$, red is $t = 1 \text{ s}$, blue is $t = 3 \text{ s}$, and gold is $t = 10 \text{ s}$) in each simulation. Panel (A) is the 1F11 simulation, (B) is 1F10, and (C) is 1F9.

widens due to a lack of compression in the chromosphere, which combats the decreasing peak. A gradually declining emergent intensity results.

Finally, to confirm that the rising ionization fraction is met by a reduction in the number of charge exchange interactions (and hence emissivity of nonthermal emission) we show, in Figure 12, the product of the charge exchange rate to the ground state and the nonthermal proton population at an energy equivalent to a Doppler shift of $\Delta\lambda = 12 \text{ \AA}$. Each panel represents a different simulation and the color of each line is a different snapshot within that simulation. As expected, we see the same patterns as found for the emissivities.

In summary, although the nonthermal emission is initially very strong, with intensity directly related to the strength of the flare (due to the number of nonthermal protons injected), the increase in ionization fraction rapidly reduces the intensity. In the case of the stronger flares, the compression of the atmosphere results in a slight rebounding of intensity as the beam encounters new neutrals, but since this happens to a very narrow extent, the emergent intensity never reaches that of the onset.

4.3. Comparison to Ambient Solar Flare Ly Emission

Now that we understand the temporal characteristics of the nonthermal Lyman line emission, the natural question is can we observe it over and above the ambient Lyman line emission? Summing thermal and nonthermal Lyman line emission, and comparing it to the thermal by itself, shows us that this is likely going to be a very difficult detection to make. Figure 13 shows those comparisons, where colored lines show the sum (gold is 1F9, red is 1F10, and blue is 1F11) and greyscale shows the ambient emission. The left column is Ly β , and the right is Ly α , with each row showing a different time. At flare onset the emission is orders of magnitude above the background, but this is only true for a very short time. By $t = 0.5 \text{ s}$ in the 1F11 simulations the combination of the strong enhancement of the background emission and the reduction in nonthermal emission makes it difficult to distinguish the nonthermal feature, which is now only a small factor larger than the thermal emission, not an order of magnitude. Nonthermal emission can still be distinguished in the 1F10 simulation and in the 1F9 simulation. It is only in the weakest flare simulation (1F9), however, that the nonthermal emission can be seen easily against the

background by $t = 3\text{--}5 \text{ s}$. It is easier to do so for the Ly β line owing to the smaller wings and less intense nearby emission. At later times the difference between thermal and nonthermal emission in the 1F9 simulation is also likely too small to easily detect.

Integrating over a region of the spectrum where nonthermal emission appears, which is free of strong lines, and comparing it to the same region in the simulations without nonthermal OZ emission illustrates these timescales. Figure 14 shows these lightcurves, with the same color scales as Figure 13, where panel (A) shows Ly β and panel (B) shows Ly α . Those figures compare the following pairs of lines: blue solid with black dashed (F11), red solid with dark gray dashed (1F10), and gold solid with light gray dashed (1F9). After a short time each solid line is only some small factor above its greyscale counterpart. In all cases the Ly β nonthermal emission shows a larger contrast to the thermal emission alone. Figure 14(c) shows the contrast introduced by the nonthermal emission. That is, $(I_{\text{total}} - I_{\text{ambient}})/I_{\text{ambient}}$. The scale on that figure has been chosen to illustrate the contrasts present at later times in the simulations ($>10 \text{ s}$), where the 1F11 is only around $\sim 2.5\%$, compared to $\sim 15\%$ for 1F10 and 30% for 1F9.

If the intensity in the region near the Lyman lines is only a small factor larger than it would otherwise be due to the presence of nonthermal emission then it is problematic to unambiguously observe, especially if the signal-to-noise of the photometry is small. Asymmetries present between the red and blue line wings, caused by the presence of nonthermal emission only redward of the line cores, may help us more clearly identify the nonthermal emission beyond the first few seconds of heating in each footprint. To quantify these asymmetries we define the asymmetry as

$$A = \frac{I_R - I_B}{I_R + I_B}, \quad (16)$$

where I_R and I_B are the intensities integrated over some $\Delta\lambda$ from line center in the red and blue wings, respectively. This is illustrated for Ly β in Figure 15, for two values of $\Delta\lambda$. Those two regions are mostly free of strong lines: (1) $\Delta\lambda \pm [12.5\text{--}13.5] \text{ \AA}$ and (2) $\Delta\lambda \pm [8.5\text{--}9.5] \text{ \AA}$.

In the $\Delta\lambda \pm [8.5\text{--}9.5] \text{ \AA}$ region, very quickly ($t > 0.5 \text{ s}$) the large asymmetry present in the strongest flare footprint (F11)

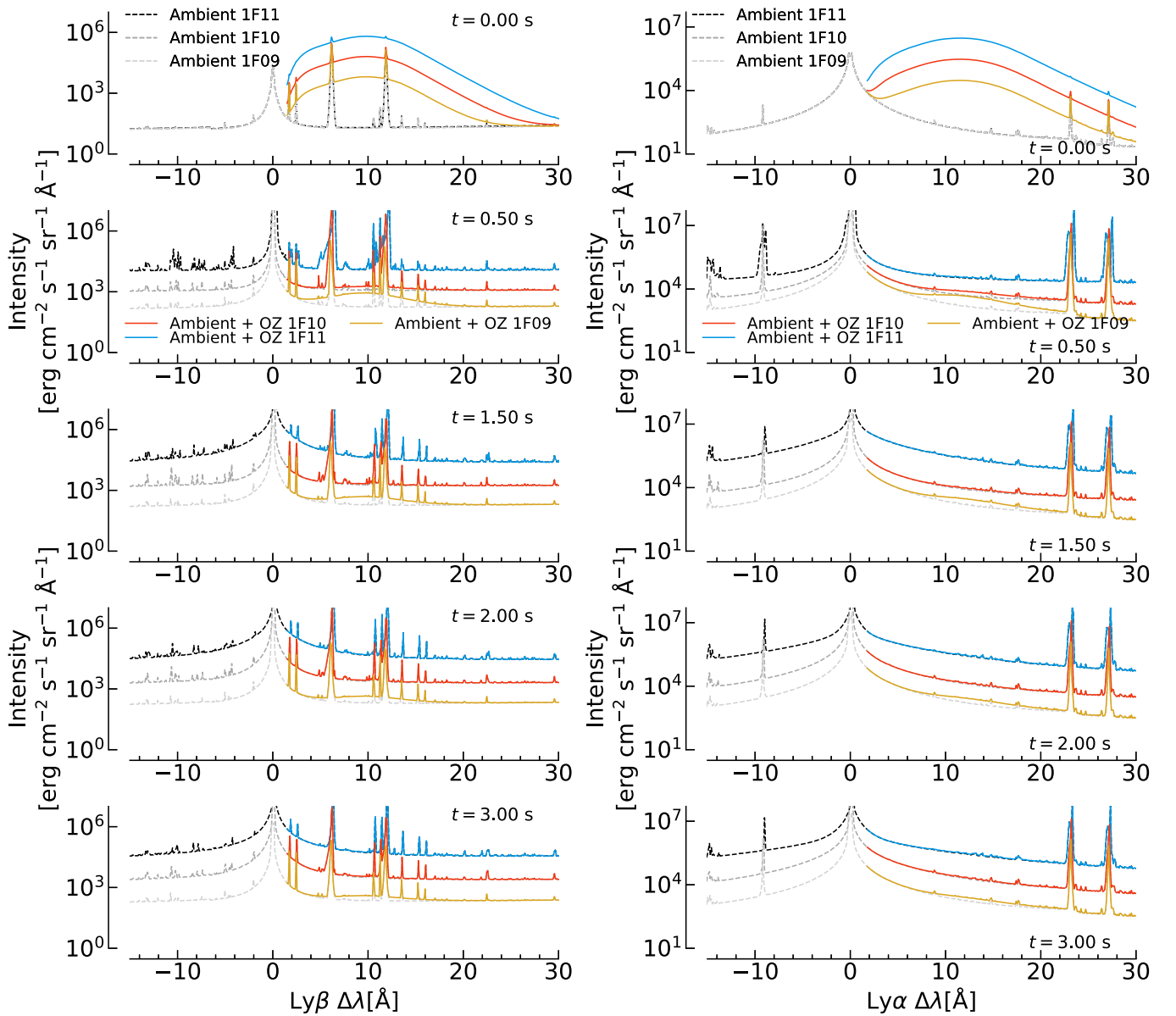


Figure 13. Ambient flare emission (greyscale; darkest is 1F11, lightest is 1F9) surrounding the Ly β (left column) and Ly α (right column) lines, compared to the total emission (ambient summed with the nonthermal Lyman lines from the OZ effect), shown in colored lines (blue is 1F11, red is 1F10, and gold is 1F9). Each row shows a snapshot during the flare.

disappears, and approaches the value to be expected from ambient Ly β emission. Similarly, the 1F10 asymmetry decreases sharply after $t \sim 1\text{--}2$ s, but does remain somewhat above the asymmetry present in the ambient emission. The asymmetry of the 1F9 flare shows a more gradual decrease out to $t \sim 5\text{--}6$ s, and does lie above the background throughout the heating phase. The maximum asymmetry in the ambient flare emission around $\Delta\lambda \pm [8.5\text{--}9.5]$ Å was $A \sim [0.02\text{--}0.04]$. After 10 s or so the asymmetries caused by the addition of nonthermal emission were on the order $A \sim [0.14, 0.08, 0.06]$ for 1F9, 1F10, and 1F11 simulations, respectively. For the two weaker flares these do seem comfortably above the typical ambient asymmetries. A careful study of the typical asymmetries in observations is needed to indicate if these small excess asymmetries caused by the nonthermal emission at later times are indeed detectable.

In the $\Delta\lambda \pm [12.5\text{--}13.5]$ Å region, the initial very large asymmetries are also present, but some small lines in the region make the asymmetry more influenced by ambient flare emission. A more rigorous study when using actual observations could remove these lines, and we show our simple asymmetry calculation just to indicate that the very large asymmetry at the onset of particle precipitation into each footpoint should appear in multiple locations along the spectrum, even with the influence of these other features.

5. Solar Orbiter / SPICE Predictions

5.1. The Spectral Imaging of the Coronal Environment (SPICE) Instrument

The SPICE instrument (SPICE Consortium et al. 2020) on board Solar Orbiter observes the Sun spectroscopically in the EUV in two passbands, the short-wavelength channel (SW;

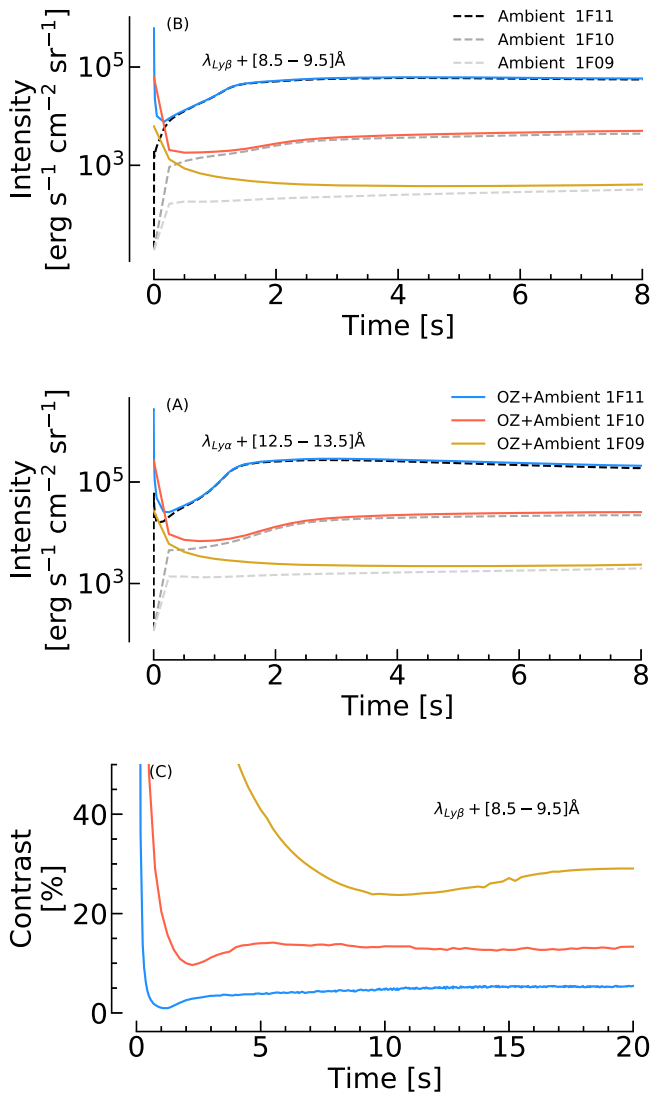


Figure 14. The temporal evolution of nonthermal OZ emission + thermal emission (colored lines; blue is 1F11, red is 1F10, and gold is 1F9), compared to the evolution of thermal emission alone (greyscale; darkest is 1F11, palest is 1F9). Clearly, strong OZ enhancements are short-lived, persisting only a fraction of a second in the strongest flare (1F11), and only a few seconds in the weakest simulation (1F9). Panel (A) shows Ly α integrated between [12.5 and 13.5] Å, and panel (B) shows Ly β integrated between [8.5 and 9.5] Å. Panel (C) shows the contrast between nonthermal and thermal emission for Ly β integrated between [8.5 and 9.5] Å, scaled to focus on the latter part of the simulations, illustrating that after the initial onset, the differences drop substantially.

$\lambda = [700\text{--}792]$ Å) and long-wavelength channel (LW; $\lambda = [970\text{--}1053]$ Å), the latter of which includes the Ly β and O VI doublet. Though flare ribbons have, at the time of writing, not yet been observed by SPICE, this instrument does offer the best hope of detecting the OZ effect. SPICE offers relatively high temporal resolution, with typical exposure times for dynamic studies of $\tau_{\text{exp}} = [1, 5]$ s, and can also provide spatial information within the flare ribbons, unlike the Sun-as-a-star Ly β observations from the EVE (Woods et al. 2012), on board SDO. Solar Orbiter has tight telemetry and observing constraints, owing to the uniqueness of its orbit, and most remote-sensing observations take place during 30 day windows during each perihelion. These constraints limit the time available for flare observations. However, Solar Orbiter’s observing program does include targeting active regions and

flares, making it worth determining if SPICE would be able to detect the transient OZ effect.

5.2. Degrading to SPICE Resolution

The ambient plus nonthermal emission in the region around the Ly β line in each simulation was degraded to SPICE level-2 count rates, assuming two exposure times of $\tau_{\text{exp}} = [1, 5]$ s, and that a sit-and-stare observing mode was used (i.e., continuous observation of a single source). Here, level-2 refers to the science-ready data provided to the community by the SPICE team, in which a common wavelength dispersion, spectral resolution, and other properties are applied. The data for the following steps were obtained from both SPICE Consortium et al. (2020) and from the information provided by the SPICE consortium following the second data release (Auchère 2022).

The preflare ambient Ly α emission was repeated for 10 s at a cadence of $\delta t = 0.25$ s, and added to the original simulation time series, so that the flare footpoint heating phase in our synthetic SPICE observations was between $t_{\text{spice}} = [10\text{--}30]$ s. This new time series was degraded to SPICE resolution as follows. For each 0.25 s snapshot, the wavelength scale was recast to the SPICE spectral plate scale (0.09623 Å pixel $^{-1}$), and the emission smeared by a spectral point-spread function (PSF) assumed to be a Gaussian with full width at half maximum (FWHM) of 9.4 SPICE wavelength pixels. This was converted from an energy flux to a photon flux (using hc/λ as the energy per photon for each SPICE wavelength bin). Our simulated intensities are per solid angle, which was removed by multiplying out the solid angle subtended by a pixel along the SPICE slit, where we assumed the 2'' wide slit was being used (SPICE also has 4'', 6'', and 30'' slits). The spatial plate scale along the slit is 1''/098. We did not apply any spatial PSF at this stage as this would require making a number of assumptions about the ribbon width and elongation. The emission was multiplied by the effective area of the LW channel, which can be seen in Figure 24 of SPICE Consortium et al. (2020). From this output in photons s $^{-1}$ pixel $^{-1}$ we then integrated the emission in time for each of our modeled exposure times (ignoring the readout time, which is up to 0.42 s if the full detector is read, but shorter if only a subset of lines and slit pixels are read). Poisson noise was added to the intensities in photon pixel $^{-1}$, which were then converted to DN pixel $^{-1}$ assuming 7.5 DN photon $^{-1}$.

5.3. Can SPICE Detect OZ Emission?

Performing the same exercise as we did earlier to compare the sum of thermal to nonthermal emission to just the ambient emission, but this time degrading each to SPICE resolution, indicates that SPICE does offer the potential for a detection of the OZ effect. Figure 16 shows our experiment assuming $\tau_{\text{exp}} = 1$ s. The top row is the 1F11 flare footpoint, the middle row is the 1F10 flare footpoint, and the bottom row is the 1F9 flare footpoint. Each column shows a different snapshot (recall that here the flare starts at $t = 10$ s). Since the nonthermal emission at flare onset is so much larger than the ambient emission, the redshifted feature is quite evident. It is only clearly present for one frame in the strongest flare, but persists for two frames in the 1F10 and 1F9 flares. Even if we use a longer exposure time of $\tau_{\text{exp}} = 5$ s (Figure 17), the emission is clearly visible in each flare for one frame. While

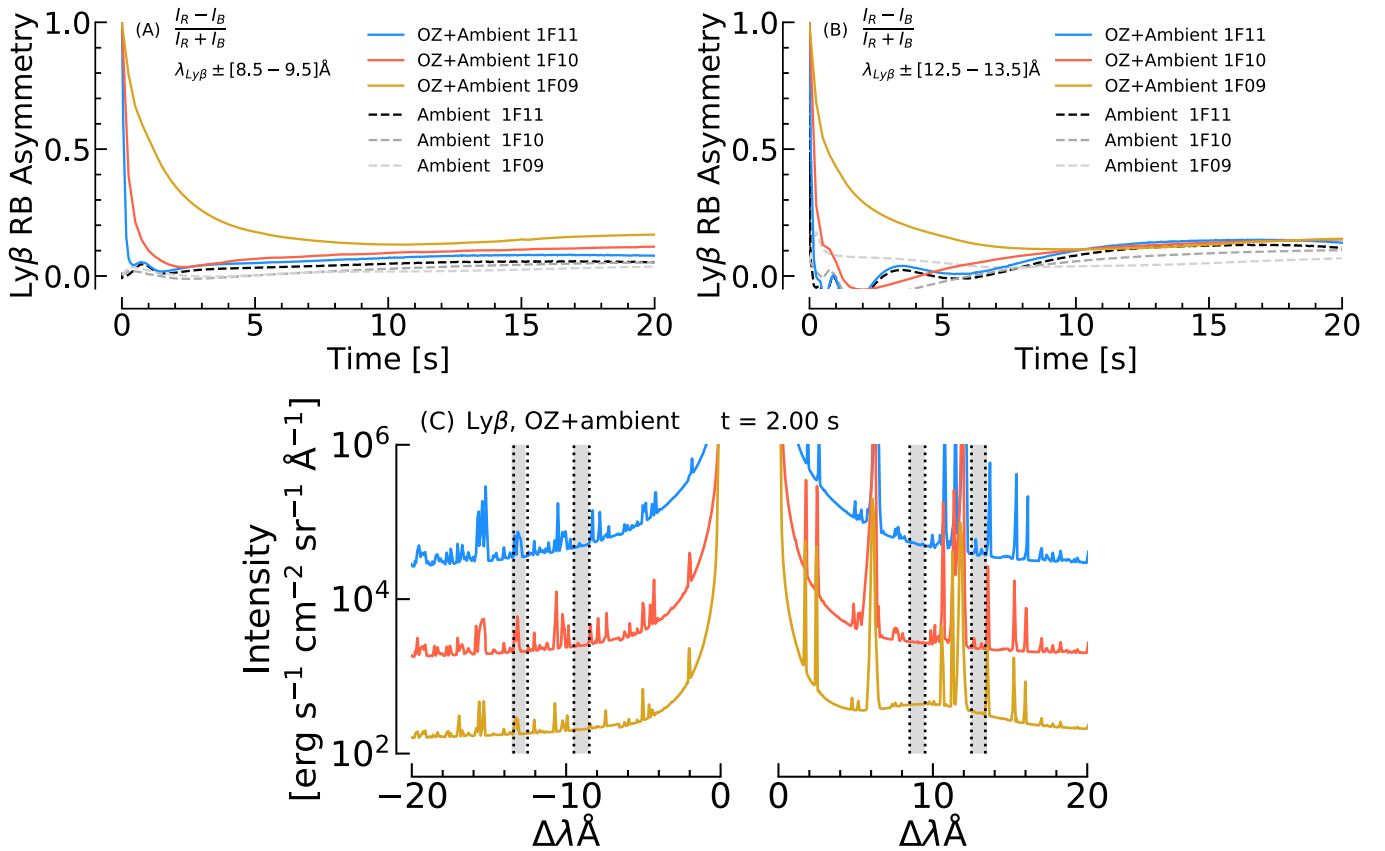


Figure 15. The red-blue (RB) wing asymmetries, A , of Ly β with (solid colored lines; gold is the 1F9 flare, red is the 1F10 flare, and blue is the 1F11 flare) and without (dashed greyscale; darkest is 1F11, palest is 1F9) nonthermal emission. Panel (A) shows A determined from $\Delta\lambda \pm [8.5-9.5] \text{ \AA}$, and panel (B) shows A from $\Delta\lambda \pm [12-13] \text{ \AA}$. Panel (C) shows the line emission at $t = 2 \text{ s}$, indicating the regions selected (shaded gray regions bounded by dotted lines) for the asymmetry calculation.

acknowledging the very transient nature of the OZ features, these results are encouraging. Depending on the orientation of the slit to the flare ribbon propagation direction we might expect to see a propagation of the redshifted feature in consecutive frames, in effect tracking the ribbon front (where energy is first deposited when new field lines reconnect).

For this experiment we varied the exposure time, but there are other considerations such as the slit width. Increasing the slit width will gather more photons, increasing the signal-to-noise, which might be necessary as the instrument response degrades over the course of the mission. The obvious downside is a reduction of spatial resolution but this might become a necessary tradeoff. Similarly, we did not bin in the spectral direction, but since we are looking for a broad feature we can likely safely increase count rates that way.

6. Summary and Conclusions

New and upcoming observatories are set to provide Lyman line observations in flares, for example from Solar Orbiter/EUI (Ly α ; though imaging without spectra), Solar Orbiter/SPICE (Ly β), Solar-C/EUVST (Ly α), ASO-S Ly α , and the SNIFS sounding rocket (Ly α). Further, Ly α profiles from line scans observed by SORCE/SOLTICE are now available, and SDO/EVE Sun-as-a-star observations cover the Ly β lines but have been relatively little studied.

Given the potential of new solar flare Lyman line observations, and the existence of untapped data sets, alongside

recent important improvements to flare radiation hydrodynamic modeling (namely proton beam-driven flares including warm-target effects), we have revisited the possibility of using the Ly α and Ly β lines to detect the presence of deka-keV nonthermal protons in the flaring chromosphere. To achieve this revisit of the Orrall-Zirker effect, we employ modern atomic cross sections, and state-of-the-art numerical flare models, initially focusing on the impact of varying the magnitude of energy deposited into each footpoint. Our results have demonstrated that while a very difficult detection, accelerated protons present in the flaring chromosphere could produce transient nonthermal Lyman line emission lying above the ambient flare emission, following neutralization of some fraction of the proton beam via charge exchange.

In our model, a precipitating population of nonthermal protons is injected at the apex of a flare loop, assumed to be vertical and near disk center. The transport and thermalization of these particles are modeled using the FP code, with energy lost through Coulomb collisions heating the plasma, the evolution of which was modeled using the radiation hydrodynamics code RADYN. Together RADYN+FP provided the nonequilibrium ionization stratification and nonthermal proton distribution function over time during our flare simulations, vital improvements over previous attempts to model the OZ effect. These were used as input to a new package developed by us, `OrrallZirkerPy`, that models the effects of the neutralization of some fraction of the nonthermal

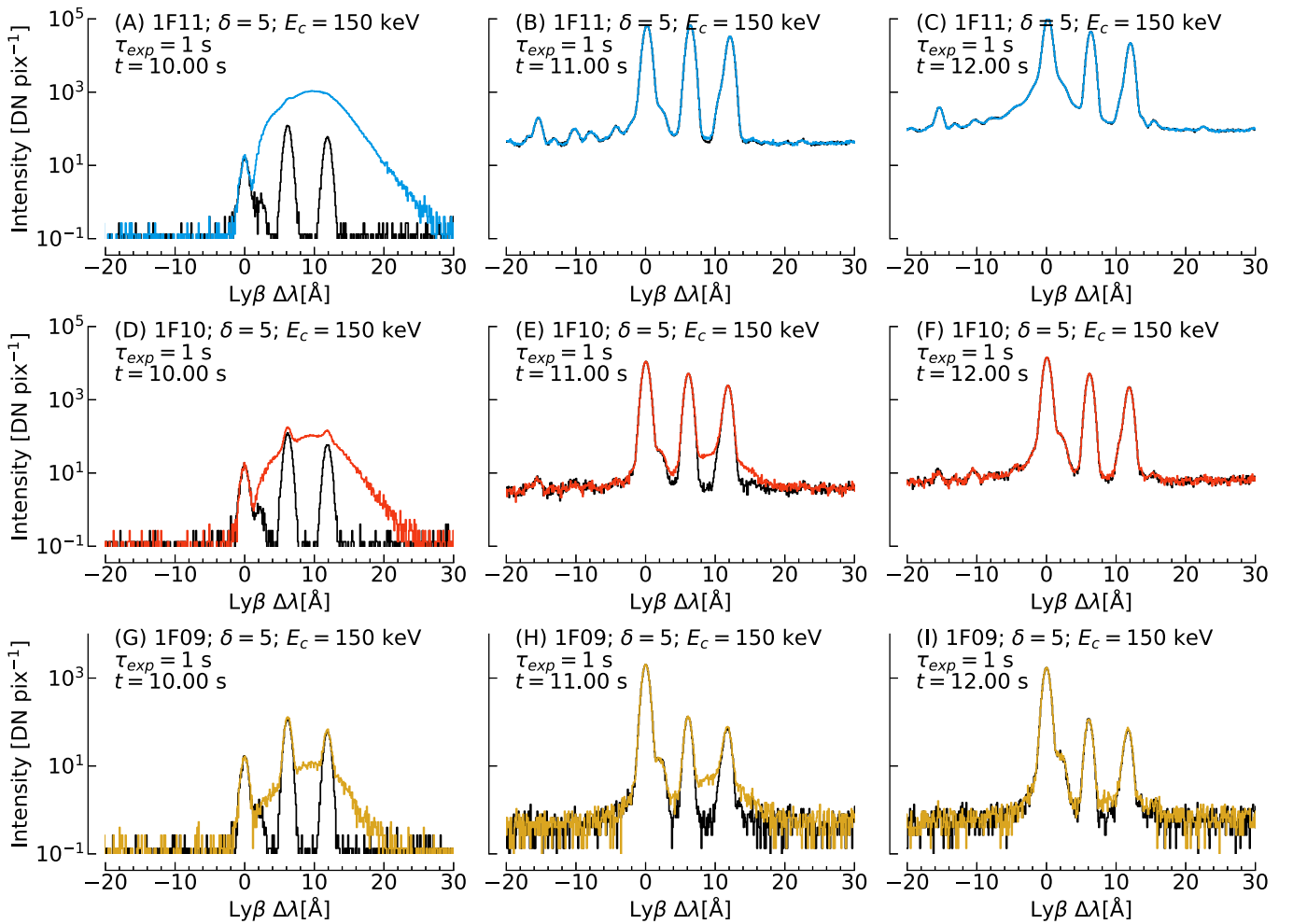


Figure 16. Comparing the level of nonthermal Ly β emission to ambient emission around the Ly β line for each of our three simulations, degraded to SPICE resolution assuming $\tau_{\text{exp}} = 1$ s, and a slit width of $2''$. In each panel black is the ambient emission (i.e., without OZ emission), and the colored lines are the sum of ambient plus OZ emission. The top row shows 1F11 (blue), the middle show shows 1F10 (red), and the bottom row shows 1F9 (gold). As discussed in the text, in these synthetic observations there are 10 s of preflare to facilitate averaging over exposure time, so that the flare starts at $t = 10$ s here. Three consecutive exposures, starting at flare onset, are shown demonstrating the very transient nature of nonthermal emission, which does outshine the ambient emission for a brief time around and between the O VI doublet at 1032 and 1038 Å.

proton distribution via charge exchange interactions between the beam and ambient plasma. This new population of energetic neutral atoms subsequently emits extremely redshifted photons of Ly α and Ly β . Our model is open source and freely available, and can be used to study OZ hydrogen emission given any ambient plasma stratification and nonthermal proton distribution. That is, it is flexible and can be used with inputs resulting from codes other than RADYN+FP, or from toy models, and could potentially be adapted to investigate similar processes on Martian aurorae as observed by MAVEN (e.g., Hughes et al. 2019).

It was found that, contrary to expectations based on the first calculations of Orrall & Zirker (1976) and Canfield & Chang (1985), this broad nonthermal “bump” in the red wings of Ly α and Ly β will be notable only for an extremely short time in strongly heated flare footpoints (subsecond), but notable emission will persist for somewhat longer ($t \sim 3\text{--}5$ s) in weaker flare footpoints. This is due to the rapid ionization of the atmosphere, quenching charge exchange interactions. Though in the strongly heated flare footpoints the precipitating protons push ever deeper (due to warm-target effects), thus encountering a fresh supply of ambient neutrals, the

compression of the atmosphere means that the emitting layer is narrow and the emergent intensity is weak compared to the initial burst of nonthermal emission. Coupled with the fact that emission from the thermal Ly α and Ly β wings and nearby continuum becomes enhanced, the nonthermal emission does not stand out from the background after a short time. That said, the asymmetries caused by nonthermal emission did persist, and for the 1F9 and 1F10 simulations they are $\gtrsim 2$ that of the largest ambient asymmetries. These could potentially be observable by high-cadence instruments with high signal-to-noise.

Degrading our synthetic emission to the resolution of the SPICE instrument on board Solar Orbiter indicates that for short exposure times ($\tau_{\text{exp}} \leq 5$ s, though ideally $\tau_{\text{exp}} \leq 1$ s) it should be possible to observe the presence of nonthermal Ly β emission, albeit for only one or two consecutive frames. Perhaps focusing on the leading edge of flare ribbons in small-to-moderate flares offers the best chance of detection. These are the locations of initial energy deposition, representing a “pristine,” not yet ionized target, and recent modeling results indicate that flare ribbon fronts likely undergo injection by a relatively weak energy flux (e.g., Polito et al. 2023). Since the

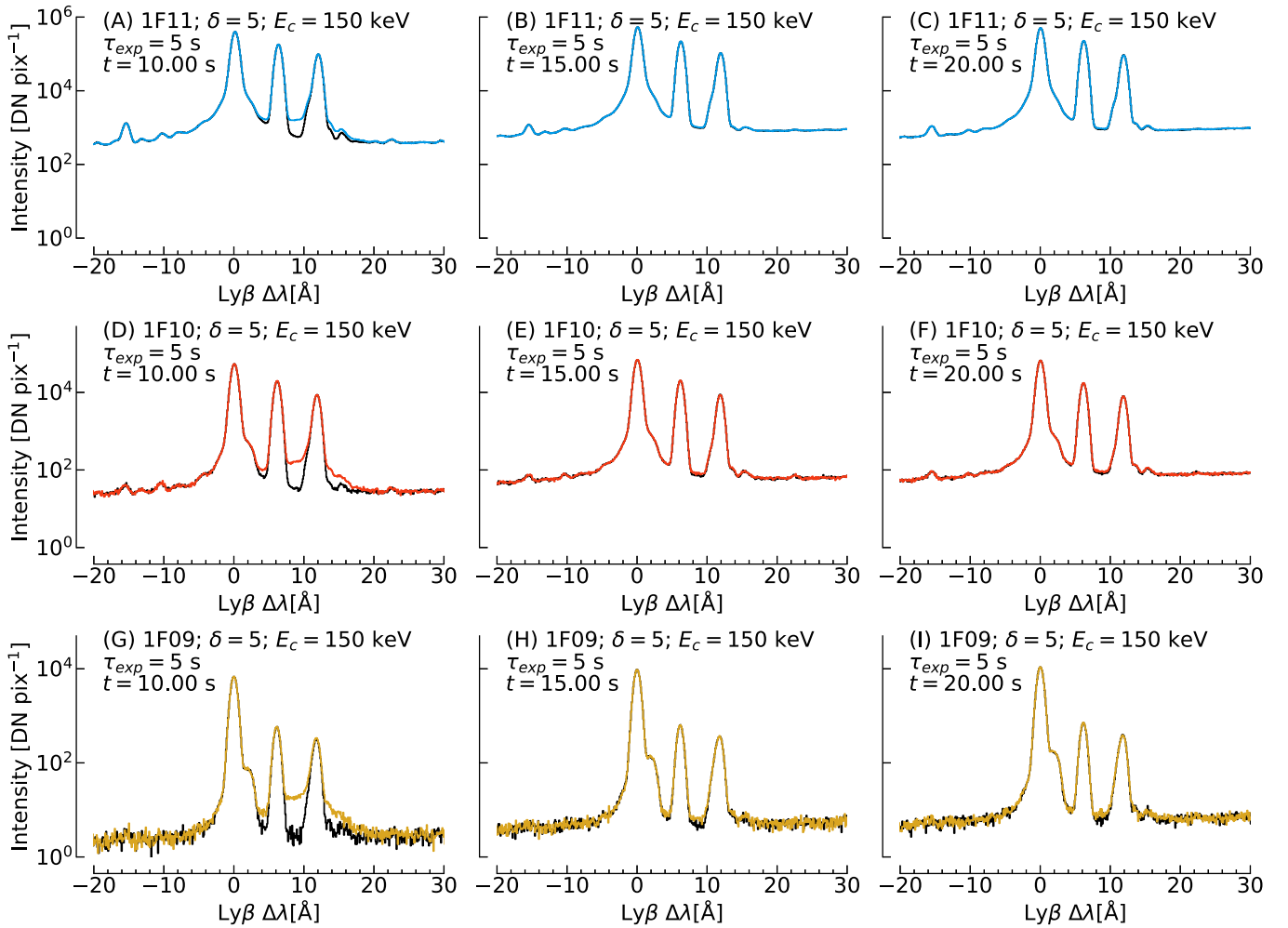


Figure 17. Same as Figure 16, but for SPICE exposures with $\tau_{\text{exp}} = 5$ s.

initial onset of nonthermal emission is so large compared to the preflare, even a stronger flare source should be detectable in one SPICE exposure at each ribbon location (though obviously this presents a difficult observation to obtain). Very high-cadence imaging, if a sufficiently broad passband is used, might also reveal this nonthermal emission as a transient flash, followed by a decay and further brightening. In general, to detect and fully exploit these signatures as diagnostics during flares we need an instrument capable of high signal-to-noise, exceptional cadence (0.1 s) and very high spatial resolution (certainly subarcsecond).

There are a number of fruitful avenues for investigation beyond this exploratory initial study. Here we have compared the magnitude of injected energy, but not the other parameters that define the nonthermal proton distribution. As indicated by prior investigations (Canfield & Chang 1985; Brosius & Woodgate 1999) varying E_c and δ should result in changes to the appearance of the broad nonthermal feature. Therefore using `OrrallZirkerPy` and a large number of `RADYN+FP` flare models, the influence of those parameters can be investigated. Extension to stellar flares, including modeling synthetic Hubble Space Telescope observables is also an exciting direction, especially given that the only confirmed detection of the OZ effect thus far has been on a dMe star (Woodgate et al. 1992) and that follow on searches produced null detections (e.g., Robinson et al. 1993; Feinstein et al.

2022). There are of course model limitations and idealizations that can be improved upon, such as modeling multispecies particle beams,¹³ and moving away from the assumptions of a near disk-center vertical beam.








Finally, we speculate that the ratio of the O VI doublet may vary in the presence of highly redshifted Ly β emission. These lines are formed by both collisional processes as well as resonant scattering of chromospheric radiation (e.g., see discussions in Strachan et al. 2000; Raymond & Ciaravella 2004). Pumping of the O VI 1032 Å line by Ly β during fast coronal mass ejections (CMEs), and the resulting change to the $I_{1032}:I_{1038}$ ratio has been exploited to estimate the electron density during a CME (Raymond & Ciaravella 2004). A similar effect may occur in our OZ scenario, with pumping of O VI 1032 Å by the proton beam-produced Ly β emission observable as a transient variation of the O VI doublet ratio. Characteristics of this change to the doublet ratio could also be diagnostically useful, since it presumably would vary depending on properties of the nonthermal component, which itself varies with properties of the injected nonthermal proton distribution. Further study of this would require more advanced modeling of the

¹³ If both energetic electrons and protons bombard a chromospheric footpoint then ionization may be more rapid. Multispecies beams are novel experiments that have yet to be performed, but are now possible with very recent updates to `RADYN+FP`.

O VI doublet than we have performed here, but it would be a very interesting endeavor.

We thank the anonymous referee for their insightful comments. G.S.K. acknowledges the financial support from a NASA Early Career Investigator Program award (Grant# 80NSSC21K0460). G.S.K. and R.O.M. acknowledge financial support from the Heliophysics Supporting Research program (Grant# 80NSSC21K0010). J.C.A., N.Z.P., T.A.K., and J.W.B. acknowledge NASA funding for the SPICE instrument team at GSFC. N.Z.P. and J.W.B. were funded through cooperative agreement 80NSSC21M0180. J.C.A. acknowledges funding through NASA's Heliophysics Supporting Research and Heliophysics Innovation Fund programs. HSH is grateful for the hospitality of the School of Physics and Astronomy at the University of Glasgow. This manuscript benefited from discussions held at a meeting of the International Space Science Institute team: "Interrogating Field-Aligned Solar Flare Models: Comparing, Contrasting and Improving," led by Dr. G. S. Kerr and Dr. V. Polito. The authors thank Dr. William Thompson for helpful discussions regarding SPICE. Solar Orbiter is a space mission of international collaboration between ESA and NASA, operated by ESA. The development of SPICE has been funded by ESA member states and ESA. It was built and is operated by a multinational consortium of research institutes supported by their respective funding agencies: STFC RAL (UKSA, hardware lead), IAS (CNES, operations lead), GSFC (NASA), MPS (DLR), PMOD/WRC (Swiss Space Office), SwRI (NASA), UiO (Norwegian Space Agency). Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center.

ORCID iDs

Graham S. Kerr  <https://orcid.org/0000-0001-5316-914X>
 Joel C. Allred  <https://orcid.org/0000-0003-4227-6809>
 Adam F. Kowalski  <https://orcid.org/0000-0001-7458-1176>
 Ryan O. Milligan  <https://orcid.org/0000-0001-5031-1892>
 Hugh S. Hudson  <https://orcid.org/0000-0001-5685-1283>
 Natalia Zambrana Prado  <https://orcid.org/0000-0001-6395-7115>
 Therese A. Kucera  <https://orcid.org/0000-0001-9632-447X>

References

- Abbett, W. P., & Hawley, S. L. 1999, *ApJ*, 521, 906
 Allred, J. C., Alaoui, M., Kowalski, A. F., & Kerr, G. S. 2020, *ApJ*, 902, 16
 Allred, J. C., Hawley, S. L., Abbett, W. P., & Carlsson, M. 2005, *ApJ*, 630, 573
 Allred, J. C., Kerr, G. S., & Gordon Emslie, A. 2022, *ApJ*, 931, 60
 Allred, J. C., Kowalski, A. F., & Carlsson, M. 2015, *ApJ*, 809, 104
 Auchère, F. 2022, SPICE Data Release v2.0, IDOC, doi:10.48326/idoc.medoc.spice.2.0
 Bates, D. R., & Dalgarno, A. 1953, *PPSA*, 66, 972
 Belkić, D., Gayet, R., & Salin, A. 1992, *ADNDT*, 51, 59
 Brosius, J. W. 2001, *ApJ*, 555, 435
 Brosius, J. W., & Woodgate, B. E. 1999, *ApJ*, 514, 430
 Brown, J. C. 1971, *SoPh*, 18, 489
 Brown, S. A., Fletcher, L., Kerr, G. S., et al. 2018, *ApJ*, 862, 59
 Bruls, J. H. M. J., Rutten, R. J., & Shchukina, N. G. 1992, *A&A*, 265, 237
 Canfield, R. C., & Chang, C.-R. 1985, *ApJ*, 295, 275
 Cariatore, N. D., & Schultz, D. R. 2021, *ApJS*, 252, 7
 Carlsson, M., & Stein, R. F. 1992, *ApJL*, 397, L59
 Carlsson, M., & Stein, R. F. 1995, *ApJL*, 440, L29
 Carlsson, M., & Stein, R. F. 1997, *ApJL*, 481, 500
 Cheshire, I. M., Gallaher, D. F., & Taylor, A. J. 1970, *JPhB*, 3, 813
 Chupp, E. L., Forrest, D. J., Ryan, J. M., et al. 1981, *ApJL*, 244, L171
 De Pontieu, B., Title, A. M., Lemen, J. R., et al. 2014, *SoPh*, 289, 2733
 Del Zanna, G., Dere, K. P., Young, P. R., Landi, E., & Mason, H. E. 2015, *A&A*, 582, A56
 Dere, K. P., Landi, E., Mason, H. E., Monsignori Fossi, B. C., & Young, P. R. 1997, *A&AS*, 125, 149125
 Emslie, A. G. 1978, *ApJ*, 224, 241
 Emslie, A. G., Dennis, B. R., Shih, A. Y., et al. 2012, *ApJ*, 759, 71
 Fang, C., Feautrier, N., & Henoux, J. C. 1995, *A&A*, 297, 854
 Feinstein, A. D., France, K., Youngblood, A., et al. 2022, *AJ*, 164, 110
 Fletcher, L., Dennis, B. R., Hudson, H. S., et al. 2011, *SSRv*, 159, 19
 Graham, D. R., Cauzzi, G., Zangrilli, L., et al. 2020, *ApJ*, 895, 6
 Henoux, J. C., Fang, C., & Gan, W. Q. 1993, *A&A*, 274, 923
 Hill, J., Geddes, J., & Gilbody, H. B. 1979, *JPhB*, 12, 2875
 Holman, G. D., Aschwanden, M. J., Aurass, H., et al. 2011, *SSRv*, 159, 107
 Hudson, H. S. 1972, *SoPh*, 24, 414
 Hudson, H. S., Fletcher, L., MacKinnon, A. L., & Woods, T. N. 2012, *ApJ*, 752, 84
 Hughes, A., Chaffin, M., Mierkiewicz, E., et al. 2019, *JGRA*, 124, 10,533
 Hurford, G. J., Krucker, S., Lin, R. P., et al. 2006, *ApJL*, 644, L93
 Hurford, G. J., Schwartz, R. A., Krucker, S., et al. 2003, *ApJL*, 595, L77
 IAEA 1993, in Atomic and Plasma-Material Interaction Data for Fusion, ed. R. K. Janev (Vienna: International Atomic Energy Agency) No. 4
 Kane, S. R., & Anderson, K. A. 1970, *ApJ*, 162, 1003
 Kerr, G. S. 2022, *FrASS*, 9, 1060856
 Kerr, G. S. 2023, *FrASS*, 9, 1060862
 Kerr, G. S., Allred, J. C., & Carlsson, M. 2019a, *ApJ*, 883, 57
 Kerr, G. S., Carlsson, M., & Allred, J. C. 2019b, *ApJ*, 885, 119
 Kerr, G. S., Allred, J. C., & Polito, V. 2020, *ApJ*, 900, 18
 Kerr, G. S., Fletcher, L., Russell, A. J. B., & Allred, J. C. 2016, *ApJ*, 827, 101
 Kerr, G. S., Xu, Y., Allred, J. C., et al. 2021, *ApJ*, 912, 153
 Kontar, E. P., Brown, J. C., Emslie, A. G., et al. 2011, *SSRv*, 159, 301
 Kowalski, A. F., Allred, J. C., Carlsson, M., et al. 2022, *ApJ*, 928, 190
 Kowalski, A. F., Allred, J. C., Daw, A., Cauzzi, G., & Carlsson, M. 2017, *ApJ*, 836, 12
 Kowalski, A. F., Hawley, S. L., Carlsson, M., et al. 2015, *SoPh*, 290, 3487
 Kuridze, D., Mathioudakis, M., Simoes, P. J. A., et al. 2015, *ApJ*, 813, 125
 Leenaarts, J., Pereira, T., & Uitenbroek, H. 2012, *A&A*, 543, A109
 Lin, R. P. 2011, *SSRv*, 159, 421
 Lin, R. P., Dennis, B. R., Hurford, G. J., et al. 2002, *SoPh*, 210, 3
 Ludde, H. J., & Dreizler, R. M. 1982, *JPhB*, 15, 2703
 Massey, H. S. W., Burhop, E. H. S., & Gilbody, H. B. 1974, *Electronic and Ionic Impact Phenomena* (Oxford: Clarendon Press)
 McLaughlin, B. M., & Bell, K. L. 1983, *JPhB*, 16, 3797
 McLaughlin, B. M., & Bell, K. L. 1987, *JPhB*, 20, L217
 Najita, K., & Orrall, F. Q. 1970, *SoPh*, 15, 176
 Orrall, F. Q., & Zirker, J. B. 1976, *ApJ*, 208, 618
 Pereira, T. M. D., & Uitenbroek, H. 2015, *A&A*, 574, A3
 Peter, T., Ragozin, E. N., Urnov, A. M., Uskov, D. B., & Rust, D. M. 1990, *ApJ*, 351, 317
 Polito, V., Kerr, G. S., Xu, Y., Sadykov, V. M., & Lorincik, J. 2023, *ApJ*, 944, 104
 Polito, V., Testa, P., Allred, J., et al. 2018, *ApJ*, 856, 178
 Polito, V., Testa, P., & De Pontieu, B. 2019, *ApJL*, 879, L17
 Ramaty, R., Kozlovsky, B., & Lingenfelter, R. E. 1975, *SSRv*, 18, 341
 Ramaty, R., Mandzhavidze, N., Kozlovsky, B., & Murphy, R. J. 1995, *ApJL*, 455, L193
 Raymond, J. C., & Ciaravella, A. 2004, *ApJL*, 606, L159
 Robinson, R. D., Carpenter, K. G., Woodgate, B. E., & Maran, S. P. 1993, *ApJ*, 414, 872
 Robinson, R. D., Linsky, J. L., Woodgate, B. E., & Timothy, J. G. 2001, *ApJ*, 554, 368
 Schmelz, J. T., Reames, D. V., von Steiger, R., & Basu, S. 2012, *ApJ*, 755, 33
 Shah, M. B., Elliott, D. S., & Gilbody, H. B. 1987, *JPhB*, 20, 2481
 Shah, M. B., Geddes, J., McLaughlin, B. M., & Gilbody, H. B. 1998, *JPhB*, 31, L757
 Shah, M. B., & Gilbody, H. B. 1981, *JPhB*, 14, 2361
 Shakeshaft, R. 1978, *PhRvA*, 18, 1930
 Shih, A. Y., Lin, R. P., & Smith, D. M. 2009, *ApJL*, 698, L152
 Simnett, G. M. 1995, *SSRv*, 73, 387
 Simoes, P. J. A., Kerr, G. S., Fletcher, L., et al. 2017, *A&A*, 605, A125
 SPICE Consortium, Anderson, M., Appourchaux, T., et al. 2020, *A&A*, 642, A14
 Stier, P. M., & Barnett, C. F. 1956, *PhRv*, 103, 896
 Strachan, L., Panasyuk, A. V., Dobrzycka, D., et al. 2000, *JGR*, 105, 2345
 Švestka, Z. 1970, *SoPh*, 13, 471

Tamres, D. H., Canfield, R. C., & McClymont, A. N. 1986, [ApJ](#), 309, 409
Uitenbroek, H. 2001, [ApJ](#), 557, 389
Vernazza, J. E., Avrett, E. H., & Loeser, R. 1981, [ApJS](#), 45, 635
Vilmer, N. 2012, [RSPTA](#), 370, 3241
Winter, T. G. 2009, [PhRvA](#), 80, 032701

Wood, B. E., Redfield, S., Linsky, J. L., Müller, H.-R., & Zank, G. P. 2005, [ApJS](#), 159, 118
Woodgate, B. E., Robinson, R. D., Carpenter, K. G., Maran, S. P., & Shore, S. N. 1992, [ApJL](#), 397, L95
Woods, T. N., Eparvier, F. G., Hock, R., et al. 2012, [SoPh](#), 275, 115