

Can ion-neutral damping help to form spicules?

S. P. James¹, R. Erdélyi¹, and B. De Pontieu²

¹ Space and Atmosphere Research Center, Department of Applied Mathematics, University of Sheffield, The Hicks Building, Hounsfield Road, Sheffield S3 7RH, UK

² Lockheed Martin Solar & Astrophysics Lab, 3251 Hanover St., O/L9-41, Bldg. 252, Palo Alto, CA94304, USA
e-mail: s.p.james@shef.ac.uk; robertus@shef.ac.uk; bdp@lmsal.com

Received 19 March 2003 / Accepted 7 May 2003

Abstract. The possible mechanism of generation of spicules by Alfvénic waves is studied in dissipative MHD where dissipation is mainly caused by ion-neutral collision damping, as suggested by Haerendel (1992).

Ion-neutral damping becomes non-negligible at the high cyclic frequencies involved, typically greater than 0.1 Hz, and the potential role played by this effect in both forming and supporting solar spicules is investigated.

The propagation of high frequency Alfvén waves on vertically open solar magnetic flux tubes is considered. The flux tubes are taken to be axisymmetric and initially untwisted with the field strength declining from 1600 G in the photosphere to 10–40 G in the corona. Their propagation is investigated by numerically solving a set of fully nonlinear, dissipative 1.5D MHD equations with the waves being generated by a continuous sinusoidal driver introduced into the equation of angular momentum in the low atmosphere of the Sun.

Spicule-like structures with heights of up to 7000 km were formed. The formation was found to be caused by the impact of a series of slow shocks generated by the continuous interaction between the upward propagating driven wave train and the downward propagating train of waves created by reflection off the transition region and aided by the increased thermal pressure gradient caused by Joule heating due to ion-neutral collisions. The adiabatic results suggest that *ion-neutral damping may not support spicules* as described by Haerendel (1992). However, the effect is highly sensitive to the level of ionisation and therefore to the energy balance. Including the effects of thermal conduction and radiation may well lead to different results and thus it would be premature to dismiss the mechanism completely at this point. In addition, the relatively high chromospheric temperatures obtained, even at frequencies for which ion-neutral damping and heating might be expected to be unimportant, suggest intriguing possibilities for combining the mechanism with others that are better able to recreate spicule dynamics but suffer from unrealistically low temperatures.

Key words. magnetohydrodynamics (MHD) – waves – Sun: transition region – Sun: faculae, plages – Sun: atmosphere – Sun: chromosphere

1. Introduction

Solar spicules are long, thin cylindrical jet-like structures seen above the solar limb and best observed in strong chromospheric emission lines. The majority of observations have been obtained in H α and it is these spicules that the following refers to, unless otherwise specified. They appear to be guided along the intense magnetic flux tubes at supergranule boundaries. As the spatial resolution of observing instruments has improved, estimates of their width have reduced. Lynch et al. (1973) obtained a mean width of 950 km with over 80% having widths in the range 500–1250 km. Later, Nishikawa (1988) found an average of 615 km with the majority below 900 km. It seems likely that a sizable number of “wide” spicules are actually unresolved bundles of thinner spicules. Whatever their true

dimensions, it is certainly true to say that spicule widths are of a similar order to the spatial resolution of observational instruments and this is probably a major factor in the disagreement in observed spicule properties. Lynch et al. (1973) reports a decrease in spicule width with height which is also reported by Nishikawa (1988) but in both cases the decrease is so slight that the spicule is to all intents and purposes constant in diameter. Whilst magnetic flux tubes must widen considerably in going from photospheric field strengths in excess of 1000 G to coronal values of a few tens of Gauss, most of this widening has already occurred by the transition region heights at which spicules can first be separated from the surrounding continuum. Therefore a constant diameter spicule is not in contradiction with the assumption that spicules are guided along magnetic flux tubes.

Dynamically, spicules are short-lived events comprising an apparent upward motion of chromospheric material at speeds

Send offprint requests to: R. Erdélyi,
e-mail: robertus@shef.ac.uk

of around 25 km s^{-1} to heights of 6500–9500 km after which most are seen to descend with similar velocities but some fade from view. Typical lifetimes (for which the spicule is observable above the continuum) are around 10 mins. That some spicules fade from view rather than descend is noteworthy. One possible explanation is that the plasma is heated and “evaporated” into the transition region and corona, thus vanishing from $H\alpha$ lines. Since the globally averaged upward mass flux of spicules is some 100 times the coronal mass-loss rate (Beckers 1972), it is likely spicules play an important role in the solar atmospheric mass-balance. The possibility of spicular evaporation has been investigated by Budnik et al. (1998) who conclude that at least some spicular material is heated to transition region temperatures and evaporates into the corona.

The time evolution of the spicule velocity is unclear. The tendency in the literature is to assume the rise is non-ballistic in nature. Nishikawa (1988) shows that observational evidence is actually inconclusive on this point because the regions of high velocity lie below the observable level at the solar limb. However, one strong argument against a ballistic model is that the required initial velocities of 80 km s^{-1} have not been observed on the disk. Both Hasan & Keil (1984) and Pasachoff et al. (1968) report seeing variations in the velocity profiles propagating at greater than 300 km s^{-1} between about 2000 km and 4000 km in height. This is much greater than the chromospheric sound speed but similar to the Alfvén speed, suggesting acceleration of spicular material is driven by a magnetic mechanism.

Many spicule emission spectra are inclined to the spectrograph dispersion (Beckers 1972) which can be attributed to rotation of the spicule. Pishkalo (1994) found that about half of all spicules showed spectral tilts at some point in their lifetimes and that rotation must be in the range of $20\text{--}35 \text{ km s}^{-1}$ to fit the observed spectral profiles.

The radiation field in spicules is photospheric whereas the temperature can be above 10 000 K and electron density of order 10^{17} m^{-3} . Under these conditions, deviations from LTE are significant. This is another source of disagreement between observers as involved non-LTE calculations are required to interpret spicule line information. Beckers (1972) found temperatures steadily increased from 9000 K at 2000 km to 16 500 K at the top of the spicule. Matsuno and Hirayama (1988) find much lower temperatures of 5000–8000 K and also find a decrease in temperature from 2200–3000 km in height before the temperature starts to increase. It is probably reasonable to say that spicule temperatures are of typical chromospheric values and that the temperature rises along most of the spicule length. There is less disagreement about spicule densities. Beckers (1972) finds electron densities of $3\text{--}15 \times 10^{16} \text{ m}^{-3}$ with the value decreasing over the whole spicule length. Mass densities are about $3 \times 10^{-10} \text{ kg m}^{-3}$, which is typical of upper chromospheric densities. Both the temperature and the density profiles are very flat; the density scale height is several thousand kilometres compared with a few hundred for the upper chromosphere when in hydrostatic equilibrium.

Spicule velocities are comparable with the local sound speed so that any model must necessarily take account of non-linear effects. As a consequence, the focus has shifted toward

numerical modeling as computers have become sufficiently advanced to handle the non-linear MHD equations. Sterling (2000) gives an excellent review of the current state of progress of numerical models. Almost exclusively, these models involve a deposition of energy and/or some artificial driving force in the lower atmosphere of a rigid flux tube. Suematsu et al. (1982) investigated the effect of a sudden pressure enhancement in the low chromosphere in a flux tube of constant cross-section. The initial pressure pulse nonlinearly steepens into a gasdynamic shock which impacts on the transition region, pushing it upward. This model is ballistic in nature and predicts initial velocities of around 60 km s^{-1} . As mentioned above, such large velocities have not been observed on the disk. The model recreates other spicule properties reasonably well although the temperatures are very much at the lower end of observations. The model is adiabatic and it seems likely that inclusion of radiation losses would reduce temperatures below observational constraints.

Hollweg (1982) investigated a weak velocity pulse in a flux tube with expanding cross section. The velocity perturbation results in an acoustic gravity wave train consisting of a wave front followed by an oscillating wake as described in the linear regime. Both the initial wave front and the oscillating wake steepen into a series of slow shocks, referred to as *rebound shocks* since they form when falling plasma uplifted by previous shocks rebounds back upward. The initial and rebound shocks continually impact the transition region resulting in intermittent rises. Since the shocks are weaker than in the pressure pulse model of Suematsu et al. (1982), the high velocities are not seen. It is argued that the intermittent velocity would fit the non-ballistic approximately constant velocity profile observed under low temporal resolution. A number of follow up investigations of this rebound shock model have been conducted by various authors; Sterling & Hollweg (1988) extended the original calculations in time and parameter space, Sterling & Mariska (1990) added radiation and heat conduction and Cheng added the effects of ionisation and variable molecular weight in a series of papers (Cheng 1992a–c). Incorporation of radiative losses tends to reduce the height of the structures formed. As yet, observations are unable to either confirm or refute the testable prediction of these models, namely the intermittent velocities of the spicule tops. Future higher spatial and temporal resolution observations will hopefully clarify this issue.

The first MHD spicule simulations were carried out by Hollweg et al. (1982) who studied the evolution of a low-frequency Alfvénic pulse launched from the photosphere. This pulse evolved into an MHD fast-mode shock which moved the transition region upward. The structure formed was too short and cool to be identified with a spicule before the inclusion of radiative losses (which have been shown for other formation mechanisms to impede spicule formation). Kudoh & Shibata (1999) studied the effect of a series of randomly generated Alfvénic pulses. They found that spicular structures with heights of 5000 km could be generated but again their model did not incorporate radiation losses which can be expected to reduce heights.

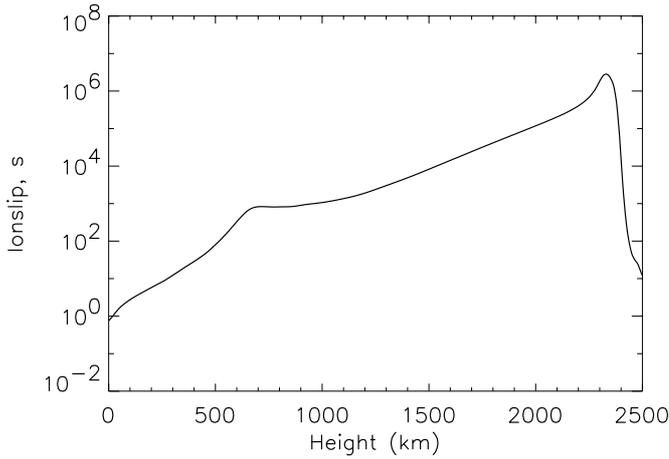


Fig. 1. Ion slip profile for the case of 20 G coronal field.

2. The possible role of ion-neutral damping in spicule formation

The presence of a magnetic field forces electrons and ions to gyrate about that field. As a result of this reduction in mobility, the conductivity perpendicular to the magnetic field is reduced. Following Cowling (1976), a more generalised form of Ohm's law can be derived:

$$\mathbf{E} = -\mathbf{v} \times \mathbf{B} + \frac{1}{\sigma_0} (\mathbf{J} + \beta_e \mathbf{J} \times \mathbf{e}_B + s \mathbf{e}_B \times (\mathbf{J} \times \mathbf{e}_B)), \quad (1)$$

where σ_0 is the electrical conductivity. $\beta_e = \omega_e / \nu_{eH}$ being the electron gyrofrequency divided by the collision frequency between electrons and heavy particles, is a measure of the typical number of times an electron can gyrate around the magnetic field between collisions. $s = (\rho / \rho_n)^2 \beta_e \beta_i$ is the ion slip and is purely due to the interaction between the ion and neutral fluids. Changes in the magnetic and electric fields impact on the ions directly but are only felt by the neutral fluid due to collisional coupling. The neutral and ion fluids thus “slip” over each other, potentially causing dissipation and heating. This mechanism becomes important if the time scale for changes in the electric and magnetic fields is comparable with the ion-neutral collision time, τ_{in} .

Figure 1 shows a typical ion slip profile as calculated by our code (details below). Just below the transition region, ion-neutral damping results in a resistivity enhancement of more than 1 million times. It is in the upper chromosphere that the effect first becomes important. One interesting feature is the small region of enhanced ion slip just below the transition region. At this point, the temperature is high enough for the hydrogen to be mostly, but not completely, ionised whilst the helium is predominantly neutral, the most efficient conditions in the solar atmosphere for the mechanism. The ion slip drops off rapidly once the temperature becomes high enough to completely ionise hydrogen in the transition region.

Work by Haerendel (1992) and later by De Pontieu & Haerendel (1998) showed, using the WKB approximation applied to plane parallel linear Alfvén waves, that ion-neutral

damping results in a wave period averaged force in the direction of propagation of

$$\overline{F_z} = \rho_n \overline{v_{1y}^2} \frac{\omega^2}{2\nu_{ni}v_A}, \quad (2)$$

with an associated wave period averaged energy dissipation of

$$\overline{\epsilon} = \rho_n \overline{v_{1y}^2} \frac{\omega^2}{\nu_{ni}} = 2v_A \overline{F_z}, \quad (3)$$

where ω is the angular wave frequency, ν_{ni} the neutral-ion collision frequency, v_{1y} is the transverse wave velocity, v_A is the Alfvén speed and ρ_n is the neutral density. They speculated that Alfvén waves with frequencies around 0.5 Hz could provide enough upward momentum by this mechanism to support structures with typical spicular parameters against gravity.

De Pontieu (1999) carried out numerical simulations to test this hypothesis. He used a set of essentially 1D *hydrodynamic* wave-period averaged equations which incorporated the vertical magnetic field implicitly by following the evolution along a flux tube of height-dependent cross-section. Transport and damping of Alfvén waves was included without having to resort to 1.5D MHD equations by the addition of an equation of wave action density. The resulting structure and its dynamics showed good agreement with observations. However, in the transition region the density drops and the temperature increases by roughly two orders of magnitude in just 2–300 km whilst the Alfvén speed is around 100 km s⁻¹. The WKB approximation is clearly violated, even at the high frequencies involved. In particular, the steep temperature and density gradients will lead to strong reflection of the Alfvén waves at the transition region and these reflected waves will interact with the upcoming wave train. No account of these reflections was taken by De Pontieu's model. The natural progression of the model, which is followed in this paper, is to move to full 1.5D MHD equations to see how the results are modified by the relaxation of the WKB assumption. Preliminary results using this model have already been presented in James & Erdélyi (2002). These suggested that ion-neutral damping could provide significant heating of spicule-like structures but there was little evidence of the momentum transfer predicted by the WKB analysis. This paper expands the narrow parameter range which was considered in the previous work.

3. Governing equations

Following Hollweg (1982), we consider a rigid, vertical, axisymmetric and initially untwisted flux tube. We define a local orthogonal curvilinear coordinate system by z , the distance along a field line, θ , the azimuthal angle about the axis of symmetry and ξ , a coordinate measured in the $\hat{z} \times \hat{\theta}$ direction. We assume axial symmetry,

$$\frac{\partial}{\partial \theta} \equiv 0. \quad (4)$$

Additionally, all motions and gradients perpendicular to the flux tube are suppressed;

$$\frac{\partial}{\partial \xi} \equiv 0, \quad v_\xi = B_\xi = 0. \quad (5)$$

This latter assumption, equivalent to assuming horizontal homogeneity since the flux tube we consider is approximately vertical, requires that we study the evolution along field lines close to the tube axis. In this first paper, we consider the system to be adiabatic with the exception of a heating term resulting from ion-neutral damping. The basic MHD equations reduce to the following system:

$$\frac{\partial}{\partial t} \left(\frac{\rho}{B_z} \right) + \frac{\partial}{\partial z} \left(\frac{\rho v_z}{B_z} \right) = 0, \quad (6)$$

$$\frac{\partial}{\partial t} \left(\frac{\rho r v_\theta}{B_z} \right) + \frac{\partial}{\partial z} \left(\frac{\rho r v_\theta v_z}{B_z} \right) = \frac{1}{\mu_0} \frac{\partial}{\partial z} (r B_\theta), \quad (7)$$

$$\begin{aligned} \frac{\partial}{\partial t} \left(\frac{\rho v_z}{B_z} \right) + \frac{\partial}{\partial z} \left(\frac{\rho v_z^2}{B_z} \right) = \\ - \frac{1}{B_z} \left[\frac{\partial p}{\partial z} + \rho g - \frac{v_\theta^2}{r} \frac{\partial r}{\partial z} + \frac{B_\theta}{\mu_0 r} \frac{\partial}{\partial z} (r B_\theta) \right], \end{aligned} \quad (8)$$

$$\begin{aligned} \frac{\partial}{\partial t} \left(\frac{B_\theta}{r B_z} \right) + \frac{\partial}{\partial z} \left(\frac{B_\theta v_z}{r B_z} \right) = \\ \frac{\partial}{\partial z} \left(\frac{v_\theta}{r} \right) + \frac{1}{\mu_0} \frac{\partial}{\partial z} \left(\frac{s}{B_z r^2 \sigma_0} \frac{\partial}{\partial z} (r B_\theta) \right), \end{aligned} \quad (9)$$

$$\begin{aligned} \frac{\partial}{\partial t} \left(\frac{e}{B_z} \right) + \frac{\partial}{\partial z} \left(\frac{e v_z}{B_z} \right) = \\ - p \frac{\partial}{\partial z} \left(\frac{v_z}{B_z} \right) + \frac{s}{B_z \mu_0^2 \sigma_0 r^2} \left[\frac{\partial}{\partial z} (r B_\theta) \right]^2, \end{aligned} \quad (10)$$

$$e = \frac{p}{\gamma - 1} + n_e \chi_H, \quad (11)$$

where e is the internal energy, the sum of the thermal and ionisation energies. The ionisation energy is only an approximation as it assumes that all electrons result from ionised hydrogen. However, this is a very good approximation as hydrogen is by far the most abundant element in the solar atmosphere and the ionisation energies of the other elements are of the same order as that of hydrogen. It is important to include the ionisation energy as it dominates the thermal energy in the upper chromosphere and in conditions typical of spicules.

4. Details of the numerical simulation

4.1. Equilibrium conditions

The atmosphere is assumed to be initially in hydrostatic equilibrium:

$$\frac{\partial p}{\partial z} = -\rho g. \quad (12)$$

We take the perfect gas law in the form

$$p = \alpha_1 \alpha_2 n_H k T, \quad (13)$$

where n_H is the total hydrogen number density, both ions and neutrals, $\alpha_2 n_H$ is the number density of all heavy particles

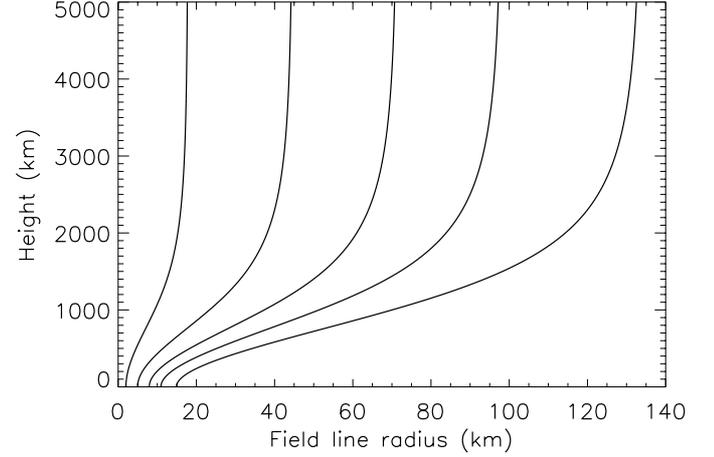


Fig. 2. Field lines of various initial radius for a model magnetic field of 20 G coronal field strength.

$\alpha_1 = 1 + n_e/\alpha_2 n_H$. For the solar atmosphere, $\alpha_2 = 1.0852$. The density can then be expressed by

$$\rho = \frac{\mu m_p p}{\alpha_1 k T}, \quad (14)$$

where $\mu = 1.34$ is the mean molecular mass for the solar atmosphere and $m_p = 1.673 \times 10^{-27}$ kg is the mass of a proton. Equations (13) and (14) are then combined to give

$$\frac{\partial p}{\partial z} = - \frac{\mu m_p p g}{\alpha_1 k T}. \quad (15)$$

We introduce regions of suppressed gravity in both the upper and lower regions of the computational domain in order to make Eq. (15) compatible with the flow-through boundary condition, $\frac{\partial}{\partial z} \equiv 0$. This is achieved by fitting the gravitational acceleration profile using cubic functions to ensure continuity up to and including the second order differential. Both of these regions are placed well outside the domain of interest. To obtain the initial atmosphere, Eq. (15) is integrated by a Runge-Kutta method with the temperature and α_2 profiles being taken from the VAL IIIc quiet Sun reference atmosphere of Vernazza et al. (1981).

For the vertical magnetic field, we take the potential field model for a flux tube at the edge of several supergranular cells from De Pontieu (1996):

$$B = B_{\text{corona}} + \frac{B_{\text{photosphere}}}{(1 + (z/R_s)^2)^{3/2}}, \quad (16)$$

where B_{corona} and $B_{\text{photosphere}}$ are the coronal and photospheric magnetic fields respectively. R_s is a parameter depending on the area of the hexagonal supergranular cells. Figure 2 shows the profile of a number of field lines with $B_{\text{photosphere}} = 1600$ G and B_{corona} with initial radius (radius at $z = 0$) varying from 2–15 km. Finally, we run our code on the model atmosphere generated by the process above until the velocities become negligible. This final state is then used as our equilibrium.

The grid spacing is determined in such a way that there are initially at least 50 gridpoints per wavelength. Because we

expect chromospheric material to be ejected upwards, we over-resolve above the transition region to ensure sufficient resolution at all subsequent times in the simulation. We also under-resolve in the low photosphere since we are unconcerned with wave evolution in this region. As well as significantly decreasing the number of gridpoints required and thus the computational time, the resultant numerical damping helps to reinforce the flow-through boundary condition and ensures negligible reflection at the lower boundary. Runs on finer grids were carried out for many parameter choices to check convergence. Whilst the grid spacing changes significantly in absolute terms, the variation between adjacent grid points is always less than 2% so that the grid is close to uniform.

4.2. Ion slip and the ionisation state

The ion slip term is given fully (e.g. Mitchner 1973) by:

$$v_{\text{in}} = \sum_{\text{ions}, j} \sum_{\text{neutrals}, k} \frac{\rho_{ij} \rho_{nk}}{\rho_i \rho_n} v_{ij, nk}, \quad (17)$$

where ρ_{ij} is the density of ions of species j , ρ_{nk} is the density of neutrals of species k , ρ_i is the total ion density, ρ_n is the total neutral density and $v_{ij, nk}$ is the momentum transfer collision frequency between ions of species j and neutrals of species k . This latter is defined, assuming the species have equal temperature, by:

$$v_{12} = n_2 \left(\frac{8kTm_2}{\pi m_1 (m_1 + m_2)} \right)^{1/2} \sigma_{12}, \quad (18)$$

where σ_{12} is the collisional cross-section for collisions between the two species. At most temperatures, hydrogen is by far the most abundant ion and also the most abundant neutral. Since hydrogen ionises at lower temperatures than helium, at sufficiently high temperatures helium becomes important in the balance of neutrals. Only the hydrogen and helium neutrals need be considered in calculating the ion slip. Since metals in the Sun ionise at lower temperatures than hydrogen, at sufficiently low temperatures hydrogen is no longer the dominant ion. However, at these low temperatures the ion slip is negligible so only hydrogen ions are included in calculating the ion slip. Equation (17) reduces to:

$$v_{\text{in}} = \left[\frac{1}{\sqrt{2}} n_{\text{H}^{\text{I}}} \sigma_{\text{H}^{\text{I}}\text{H}^{\text{I}}} + \sqrt{\frac{4}{5}} n_{\text{He}^{\text{I}}} \sigma_{\text{H}^{\text{I}}\text{He}^{\text{I}}} \right] \left(\frac{8kT}{\pi m_{\text{H}}} \right)^{1/2}. \quad (19)$$

We have taken the values of $\sigma_{\text{H}^{\text{I}}\text{H}^{\text{I}}} = 5 \times 10^{-19} \text{ m}^2$ and $\sigma_{\text{H}^{\text{I}}\text{He}^{\text{I}}} = 8 \times 10^{-20} \text{ m}^2$ from Osterbrock (1961).

In calculating the ionisation state we closely follow the technique described by De Pontieu (1999). Since $s/\sigma_0 \propto 1/n_e$, we include some of the heavy elements in order to provide electrons and prevent the damping and heating becoming unrealistically large at low temperatures. Our model atmosphere consists of hydrogen, helium, magnesium, silicon and iron. With the exception of hydrogen, we assume LTE and use the Saha equation;

$$\frac{n_{j+1}}{n_j} n_e = \frac{u_{j+1}}{u_j} 2 \frac{(2\pi m_e k_B T)^{3/2}}{h_3} e^{-\chi_j/k_B T}, \quad (20)$$

where n_j is the number density of ions in the j th ionisation state, m_e is the electron mass and χ_j is the ionisation energy needed to ionise the j th state. The partition functions are equated with the lowest ground term,

$$u_j = g_{j,0}. \quad (21)$$

For singly ionised elements, Eq. (20) yields for each element, k :

$$n_{i_k} = \frac{n_k}{1 + (n_e/a_k)}, \quad (22)$$

where n_k is the total number density of element k , n_{i_k} is the number density of singly ionised element k , n_e is the total electron density and a_k is the right hand side of Eq. (20).

Both singly and doubly ionised helium are included and some simple algebra yields:

$$n_{\text{He}^{\text{II}}} = \frac{n_{\text{He}}}{(n_e/a_{\text{He}^{\text{II}}}) + 1 + (a_{\text{He}^{\text{III}}}/n_e)} \quad (23)$$

and

$$n_{\text{He}^{\text{III}}} = \frac{a_{\text{He}^{\text{III}}}}{n_e} n_{\text{He}^{\text{II}}}. \quad (24)$$

A non-LTE modification is applied for hydrogen following Brown (1973):

$$\frac{n_{j+1}}{n_j} n_e = \frac{a}{b_1}, \quad (25)$$

where

$$b_1 = \frac{T}{WT_R} e^{\left[\frac{\chi_{\text{H}^{\text{I}}}}{4kT} (T/T_R - 1) \right]}. \quad (26)$$

Here $T_R = 5700 \text{ K}$ is the temperature of the photospheric radiation field and $W = 0.5$ is the dilution factor.

Since the plasma is electrically neutral, we obtain an implicit equation for the total electron density which is a function of temperature and density (equivalently total hydrogen number density since we assume no variation with height of relative elemental abundances):

$$n_e = \sum_k n_{i_k} q_{i_k}, \quad (27)$$

where $q_{i_k} = 2$ for doubly ionised helium and 1 for all other species. Equations (11), (27) and the perfect gas law form a closed system of equations in the unknowns p , n_e and T given the known quantities e and n_{H} . This system is solved iteratively at the beginning of each time step using the quantities in Table 1 for the various atomic parameters.

4.3. Alfvén wave driver

Alfvénic disturbances are launched by the addition of a localised artificial driver to the right hand side of Eq. (7), taking the following form:

$$F(z) = A \sin(\omega t) e^{-(z-z_0)/L}. \quad (28)$$

There are four parameters associated with this driver; A governs the initial amplitude of the Alfvén waves, z_0 is the input height of the driver in the atmosphere, ω is the angular

Table 1. Atomic parameters of elements in our model from Allen (1963).

Element name	Atomic weight (m_H)	\log_{10} (abund)	χ (eV)	g_0	g_{i0}
H	1	12	13.598	2	1
He	3.97	10.93	24.587	1	2
He ⁺	3.97		54.416	2	1
Mg	24.11	7.6	7.646	1	2
Si	27.86	7.6	8.151	9	6
Fe	55.41	7.6	7.87	25	30

frequency of the waves and L controls the spatial extent over which the driver acts. L is held fixed at 10% of the initial Alfvén wavelength at the input height since tests showed the variation of this parameter to have a negligible effect. z_0 has a more pronounced but still small impact. Thus the emphasis is on the effect of varying ω and A .

4.4. Numerical technique

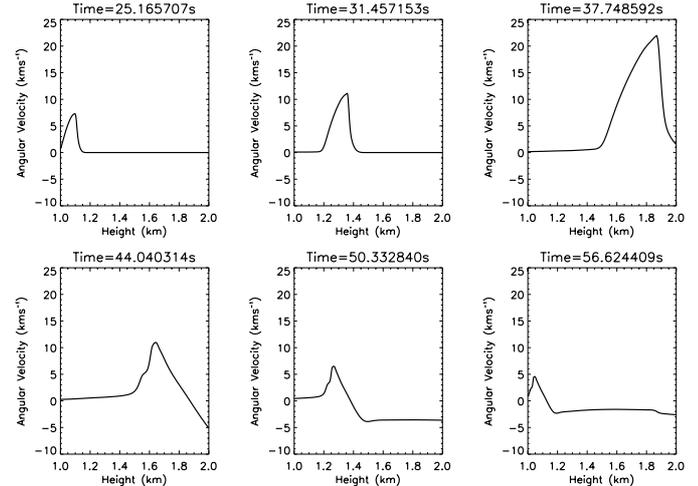
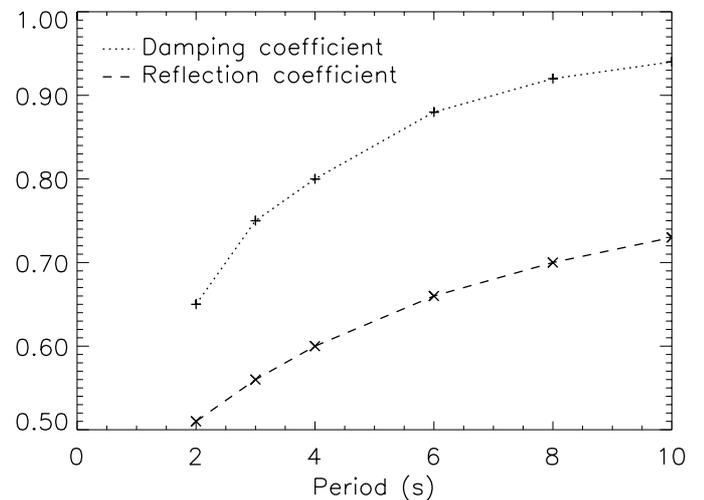
Equations (6)–(10) are solved using the Versatile Advection Code (VAC) developed by Tóth (1996). Of the numerous numerical schemes implemented by this code, we have chosen to use the two flux corrected transport (FCT) schemes. In essence, FCT schemes proceed by adding a weighted combination of a low- and high-order numerical solution at each time step. The weighting attempts to retain as much of the accuracy of the high order scheme as possible whilst avoiding the trailing ripples such schemes tend to produce behind steep gradients. They are very effective in coping with sharp gradients and are thus excellent for non-linear modeling in general and shock capturing in particular. The interested reader is directed to Book (1981) for further details.

5. Results

5.1. Transition region reflections

Before presenting the main results, it is useful to present a preliminary study of the reflection of Alfvén waves at the transition region. In order to test this, we switch off the driver described by Eq. (28) after half a wave period. This results in the generation of a single Alfvénic pulse. Figure 3 shows an example with 8 s wave period and an initial amplitude 20% of the background Alfvén speed at the launch height of 600 km. As the pulse propagates upward, it grows in amplitude as the background Alfvén speed increases and the waveform steepens non-linearly. On hitting the transition region, located slightly above the range displayed on the plots, much of the incident wave is reflected. The bottom row shows the downward propagation of this reflected wave.

In order to quantify the extent of this reflection, we calculate a reflection coefficient, R , defined as the ratio of the amplitude of the reflected wave to the initial upward propagating wave. This is calculated in the case where damping is switched off since we wish to isolate the reflection effect. The calculation is carried out at the same (arbitrary) point in the atmosphere

**Fig. 3.** Evolution of an Alfvénic pulse and reflection off the transition region.**Fig. 4.** Variation of reflection and damping coefficients with wave period for 20% amplitude and 20 G coronal field.

for each case to ensure consistency. Values of R are shown by the dashed line in Fig. 4. One can see that the reflections are significant. At all periods more than half the incident Alfvén wave energy is reflected with the strength of reflection increasing monotonically with wave period. This trend is exactly what we would expect – it is well known that high frequency (short period) waves are better able to penetrate into the solar corona.

5.2. Damping

Next, it is useful to determine the variation of the damping with waveperiod. According to the linear WKB analysis of De Pontieu & Haerendel (1996), the wave period averaged force is proportional to the square of the angular frequency. To examine this, we ran a series of simulations either including or neglecting the ionslip damping and heating terms with otherwise identical parameters. Figure 5 shows the Alfvén wave train shortly before the initial impact with the transition region for 3 different wave periods (2, 8 and 20 s). The undamped results are shown by the dotted lines and the damped results

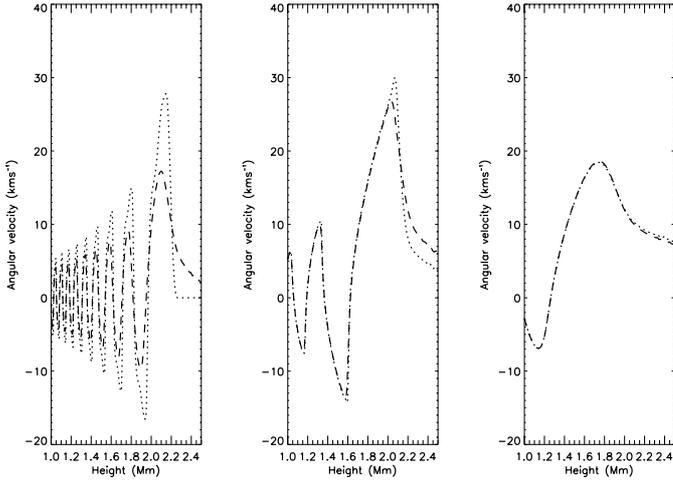


Fig. 5. Effect of wave period on damping. Dotted lines show undamped waves, dashed lines show damped waves. From left to right, plots show waveperiods of 2 s, 8 s and 20 s respectively.

by the dashed lines. One can clearly see the progression from quite strong damping at 2 s period to negligible damping at 20 s period. The dotted line in Fig. 4 shows the trend more quantitatively. The damping coefficient is calculated from the same single pulse runs from which the reflection coefficients were obtained and is defined as the ratio of the amplitude of the damped to the undamped pulse. Again, it is calculated at a constant and arbitrary point in the atmosphere, in this case just below the transition region at 1800 km.

5.3. Shock formation by reflections

The continuous wave driver results in continuous interaction between the driven upwardly propagating wave train and the downward propagating reflected wave train. Once the first reflected wave reaches the wave source, a partially standing wave pattern will be formed. This will not be a true standing wave since the amplitude of the reflected waves are reduced compared to the driven waves and also the point of reflection, the transition region, is not stationary. To study the implications of this, we consider an extremely simplified analysis; a superposition of two waves propagating in opposite directions in an homogeneous atmosphere:

$$B_{\theta} = B_1 \cos(\omega t - kx) + B_2 \cos(\omega t + kx). \quad (29)$$

It is trivial to show that in this case the Lorentz force is given by

$$\frac{\partial B_x}{\partial \theta} = kB_1^2 \sin(2(\omega t - kx)) - kB_2^2 \sin(2(\omega t + kx)) + 2kB_1B_2 \sin(2kx). \quad (30)$$

The first two terms produce no net force over one wave period. The last term produces a periodic force with spatial wavelength of half the Alfvén wavelength. Since the plasma beta is small, this force would be expected to drive plasma toward the nodes and away from the troughs of the standing wave pattern until it is countered by the plasma pressure gradient. Whilst this analysis is extremely simple, it serves to explain what we observe. Figure 6 shows the process in action. At the same time, the

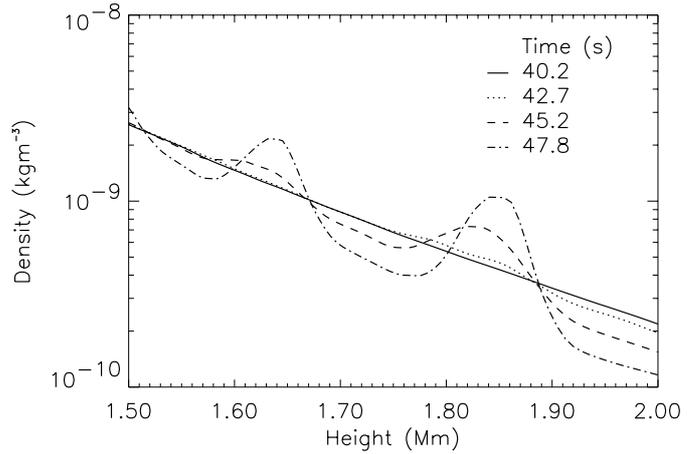


Fig. 6. Development of density profile caused by interaction of upward propagating driven waves with downward propagating waves reflected at the transition region.

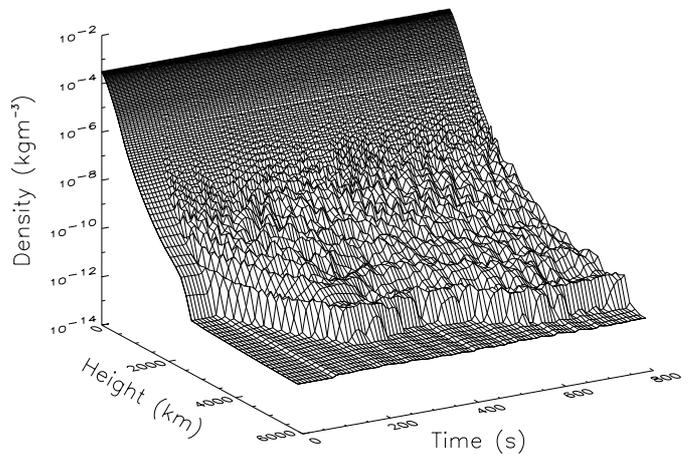


Fig. 7. 3D surface showing the evolution of the density profile with time and height.

plasma expelled from the troughs of the standing wave pattern develops into a series of slow shocks.

5.4. Time and spatial evolution of a typical case

A continuous train of waves is generated that propagate upward at the Alfvén speed in the initially undisturbed atmosphere. As the waves propagate they increase in amplitude due to the increasing Alfvén speed and the waveform steepens due to the finite amplitude. When the leading edge of this wavetrain strikes the transition region it imparts upward momentum and the transition region starts to rise, typically at speeds of 20–30 km s⁻¹. A significant proportion of the incident wave is reflected and the atmosphere becomes stratified as previously described as the reflected waves propagate downwards. We shall refer to this process as *chromospheric deformation* with the *strength* of the deformation referring to the relative change in density from peak to trough. During the initial rise, the transition region moves more quickly than the slow shocks, which are typically formed within a few wave periods of the initial impact, and their passage has no impact on the dynamics during this

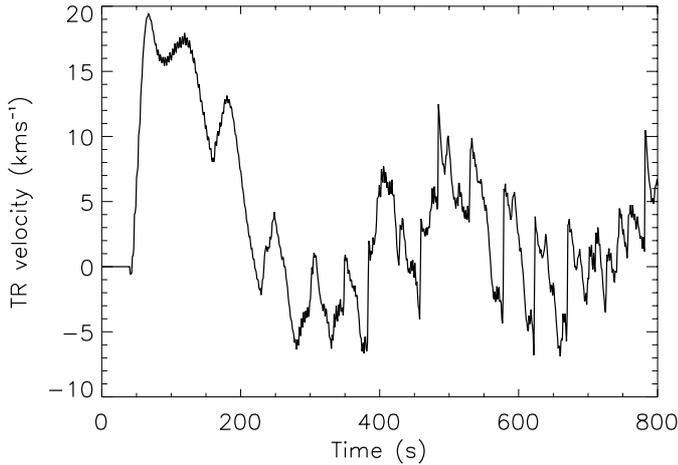


Fig. 8. Transition region velocity against time for 6 s period, 20 G magnetic field and 20% amplitude.

stage. The initial rise is eventually halted by gravity and the up-risen matter starts to fall. The slow shocks are now able to impact with the transition region causing approximately periodic increases in transition region velocity, as illustrated in Fig. 8. Whilst this process is occurring, there is continual energy input into the upper chromosphere from Joule heating. This results in an area immediately below the transition region where temperatures are enhanced to a little over 10 000 K. This temperature is a particular consequence of the mechanism – wherever hydrogen is significantly but not completely ionised, heating is significant. This ionises more hydrogen which increases the damping and heating still further. This process of positive feedback continues until the hydrogen is completely ionised at which point the ionslip drops dramatically. This region of enhanced temperature slowly broadens. It helps to support the spicular structure against gravity by means of an increased thermal pressure gradient. The enhanced temperature region also enables significant shock heating – simulations neglecting ion-neutral damping and heating do show increased temperatures with the passage of the slow shocks but some of the energy is absorbed as ionisation energy. This energy buffer is much reduced following the ionisation of hydrogen resulting in temperatures as high as 20 000–30 000 K around the shock fronts.

On switching the driver off, there is nothing to maintain the standing wave pattern and it slowly ceases to exist. The Lorentz forces maintaining the chromospheric deformation also vanish and the atmosphere relaxes to a more standard profile. This also brings an end to the generation of slow shocks. Once the last of the slow shocks has passed out of the system, the atmosphere oscillates with decreasing amplitude about a new equilibrium height determined by the level of Joule and shock heating that took place since the energy from both of these remains in the system due to lack of radiation treatment.

5.5. The contribution of ion-neutral damping

The evolution of the spicule is strongly influenced by the generation of slow shocks, a mechanism that occurs with or without ion-neutral damping. A natural question is what effect

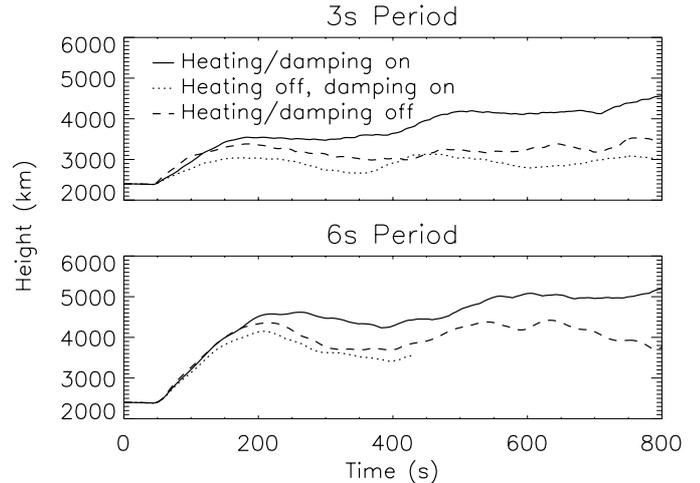


Fig. 9. Transition region height against time – impact of ion-neutral damping and subsequent heating. For the case 20% amplitude, coronal magnetic field of 20 G and launch height of 600 km. Top plot shows 3 s period, bottom plot shows 6 s period.

ion-neutral damping has on this process. In order to examine this, it would be ideal to run simulations neglecting the effect of heating (removing the last term in Eq. (10)) and ion-neutral damping (the last term in Eq. (9)). The inclusion of magnetic diffusivity results in a further constraint on the time step in addition to the Courant condition. When heating is included, regions of locally high ion slip cannot persist for long as the plasma is quickly heated, leading to ionisation which reduces the ion slip and increases the time step. When heating is not included, these regions of high ion slip persist and the time step often becomes prohibitively small. Unfortunately, computing constraints mean we are unable to obtain full runs with damping but no heating in the majority of cases. This makes it difficult to assess the contribution purely due to the momentum transfer predicted by Haerendel (1992). Figure 9 shows the position of the transition region throughout the simulation for two of the more complete cases. We consider waveperiods of 3 s, being a relatively high frequency case with strong damping and heating, and 6 s, being an intermediate case for which we can still obtain a reasonably long run with damping on and heating off. Comparing the evolution when no heating is included, the initial rise is reduced when damping is on as a result of the reduced amplitude of the initial impacting wavefront. For the 3 s case, after the initial rise, the difference between the two cases remains fairly constant (as an average over long times). Damping continues after the chromospheric deformation, as evidenced by a continuing temperature rise in the chromosphere in the case where heating is included, and the slow shocks formed, which are dependent on wave amplitudes, are smaller with damping switched on. This being the case, we might expect the difference between the spicule height in the damping on and off cases to increase steadily with time. That it does not suggest there may be a small contribution from the predicted momentum transfer. Unfortunately, the 6 s case is incomplete but seems to be following a similar pattern. The difference between damping on and off is smaller in this case, exactly as one would expect as damping strength decreases

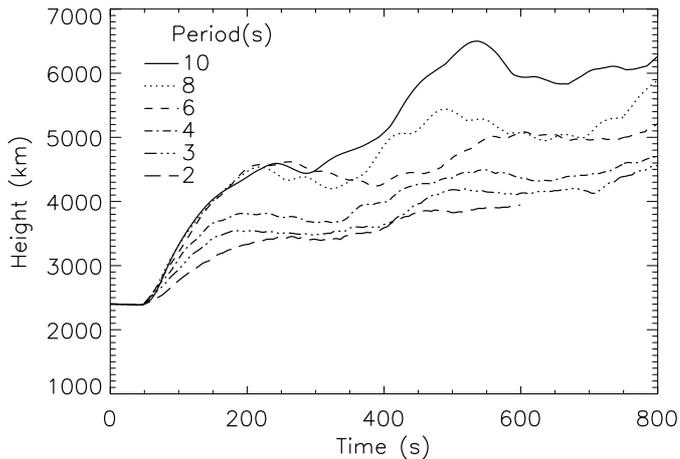


Fig. 10. Transition region height against time – variation by wave period. All simulations have amplitude of 20% and coronal magnetic field of 20 G.

with decreasing wave frequency in the linear regime. There is a marked difference when heating is switched on. The increased thermal pressure helps to support the structure against gravity and the maximum heights attained (measured relative to the initial position) are substantially higher.

In conclusion, ion-neutral damping does seem to aid spicule evolution, but it does so by heating rather than the momentum transfer mechanism proposed by Haerendel (1992).

5.6. Parameter dependence

Figure 10 shows the time evolution of the top of the transition region (hereafter described as spicule height) for various wave periods. The spicule rises to heights of between 4000 km and 6300 km. As has already been discussed, the spicule evolution is dominated by the slow shocks formed following the chromospheric deformation. The strength of shocks formed depends on the strength of the chromospheric deformation. Partly, the increased reflection at longer periods accounts for increasing deformation strength since the force described by Eq. (30) depends on the product of the amplitudes of the driven and reflected waves. Probably the major factor is related to the initial density profile being monotonically decreasing. The spatial scale of chromospheric deformation is one half the local Alfvén wavelength. In our initial atmosphere, density decreases by a factor of just 1.5 over one half Alfvén wavelength at 3 s period whilst for 10 s wave period the decrease is more than a factor of 10. Thus the ratio between the density at (what becomes) the peaks and troughs initially increases with wave period and this difference is retained when the chromospheric deformation process is complete. In turn this results in stronger shocks and higher spicules.

Table 2 shows the dependence of spicule height on initial amplitude. At 10 s period the trend is to increasing height with increasing amplitude, which is unsurprising. At 4 s period, the trend is also increasing but in this case the maximum height is capped at 4600 km. However, the time at which this maximum height is achieved decreases with increasing amplitude. It seems that 4600 km represents the maximum height which can

Table 2. Showing dependence on the amplitude parameter. The table shows maximum spicule heights in Mm for a variety of initial amplitudes and wave periods with 20 G coronal field and launch height of 600 km.

	10%	15%	20%	25%	30%
3 s	3.4	4.2	4.5	4.6	4.6
10 s	3.5	4.7	5.4	6.7	

Table 3. Showing dependence on the coronal magnetic field. The table shows maximum spicule heights in Mm for various coronal magnetic field and wave periods with 20% initial amplitude and launch height of 600 km.

	10 G	20 G	30 G	40 G
3 s	3.1	4.6	4.7	4.9
10 s	3.6	6.2	6.2	7.1

be supported against gravity by the increased thermal pressure gradient. Since the chromospheric deformation is very weak at short wave periods, the shocks are also not very strong and shock heating does not have a major effect on temperatures. The enhanced temperature region, and therefore the supporting pressure gradient, is generated mostly by Joule heating. This has as an upper limit the temperature at which hydrogen is completely ionised. This “maximum” temperature state is reached more quickly the greater the amplitude but is reached by all amplitudes (with the exception of 10% amplitude) by the end of the simulation, hence the uniformity of results.

Table 3 shows the dependence of spicule height on the coronal magnetic field. Heights increase with increasing magnetic field strength. Once again, the results for 3 s period show some degree of uniformity, for much the same reason as for the amplitude dependence.

6. Conclusion

Many proposals for the generation of solar spicules have been put forward. Haerendel (1992), suggested that already formed spicules can be supported against gravity by momentum transfer of Alfvén waves into the vertical direction via ion-neutral damping. Further theoretical calculations under the WKB approximation on plane polarised linear waves (De Pontieu & Haerendel 1998) strengthened this as a possibility. De Pontieu (1999) subsequently tested this hypothesis numerically, again under the WKB assumption, and found that spicule-like structures could not only be supported by this mechanism as theorised, but also generated. The obvious extension is to remove the WKB assumption, which is invalid in the transition region at the frequencies concerned. Advances in computing power have now made such simulations practical and have been carried out in this paper.

On relaxing the WKB approximation, the Alfvén waves can be reflected by the transition region. This turns out to be crucial in the subsequent evolution. Interaction between the driven upward propagating wave train and downward propagating wave train reflected by the transition region results in

stratification of the atmosphere into high and low density regions. Slow shocks are generated which subsequently impact on the transition region, periodically pushing it upward. Under these conditions, which are substantially different from those under which the theoretical model is derived, there is little evidence of the proposed momentum transfer mechanism in the spicule dynamics. However, dissipation due to ion-neutral damping does contribute significantly to the energetics of the system, heating the plasma to the point at which hydrogen ionises (approx. 10 000 K). Interestingly, significant heating occurs even at relatively low frequencies which suggests more careful consideration may be required before deciding to neglect the effects of ion-neutral damping in models involving the upper chromosphere. It should be noted that the momentum transfer mechanism is strong only in a narrow band of temperatures where hydrogen is almost, but not quite, completely ionised. Before dismissing the mechanism, it is essential to carry out simulations with more realistic energetics, particularly the effects of radiation.

Spicular densities were, on average, reasonably well created although a little on the low side, as were electron densities. The stratification probably occurs on lengthscales too short to be observable. This stratification is primarily a result of the rigid sinusoidal form of the wave driver. A more randomised source of energy input should help to reduce this aspect and reduce the dominance of slow shocks in the subsequent evolution. Temperatures fall within the upper range of observational constraints and heights can be recreated with a reasonably broad set of parameters, particularly favouring higher coronal magnetic field strength. Whilst the velocity of the initial rises matches observations for a number of parameter sets, the typical velocity over the spicule lifetime does not. This is the biggest problem with the model as the velocity is probably one of the stronger observational constraints. On balance, the current model is promising in some aspects but ultimately fails to convincingly recreate spicule conditions. One positive element is the temperatures obtained. Many previous spicule models have struggled with low temperatures, even before including radiation losses. If anything, temperatures obtained by this model are too high – momentum transfer due to ion-neutral damping is most efficient at temperatures at which positive feedback takes place and consequently the timescales on which the mechanism acts in the current simulation are short, probably the reason for its negligible impact. With radiation losses included, it may be that the optimum conditions can be obtained for longer periods and that this additional momentum will increase velocities to better match observations. This will be the target of future work. Even if this is not the case,

the significant heating raises possibilities for incorporating ion-neutral damping and an Alfvénic component into other models which more successfully recreate the dynamics. Another advantage of this idea could be to help explain the rapid rotations observed in many spicules.

Acknowledgements. This work was supported by Particle Physics and Astronomy Research Council (PPARC) grant PPa/S/S/2000/02983. RE acknowledges M. Kéray for patient encouragement and the financial support obtained from the NSF, Hungary (OTKA, Ref. No. TO43741) and The Nuffield Foundation (NAL/99-00).

References

- Allen, C. W. 1963, *Astrophysical Quantities* (Athlone Press)
- Beckers, J. M. 1972, *ARA&A*, 10, 73
- Book, D. L. 1981, *Finite-Difference Techniques for Vectorized Fluid Dynamics Calculations* (Springer-Verlag), 29
- Brown, J. C. 1973, *Sol. Phys.*, 29, 421
- Budnik, F. 1998, *A&A*, 334, L77
- Cheng, Q.-Q. 1992a, *A&A*, 266, 537
- Cheng, Q.-Q. 1992b, *A&A*, 266, 549
- Cheng, Q.-Q. 1992c, *A&A*, 262, 581
- Cowling, T. G. 1976, *Magnetohydrodynamics* (Hilger)
- De Pontieu, B. 1996, Ph.D. Thesis, University of Ghent, Belgium
- De Pontieu, B. 1999, *A&A*, 347, 696
- De Pontieu, B., & Haerendel, G. 1998, *A&A*, 338, 729
- Haerendel, G. 1992, *Nature*, 360, 241
- Hasan, S. S., & Keil, S. L. 1984, *ApJ*, 283, 75
- Hollweg, J. 1982, *ApJ*, 257, 345
- Hollweg, J., Jackson, S., & Galloway, D. 1982, *Sol. Phys.*, 75, 35
- Kudoh, T., & Shibata, K. 1999, *ApJ*, 514, 493
- James, S. P., & Erdélyi, R. 2002, *A&A*, 393, L11
- Lynch, D. K., Beckers, J. M., & Dunn, R. B. 1973, *Sol. Phys.*, 30, 63
- Matsuno, K., & Hirayama, T. 1988, *Sol. Phys.*, 117, 21
- Mitchner, M., & Kruger Jr, H. 1973, *Partially Ionised Gases* (Wiley-Interscience)
- Nishikawa, T. 1988, *Publ. Astron. Soc. Japan*, 40, 613
- Osterbrock, D. E. 1961, *ApJ*, 134, 3470
- Pasachoff, J. J., Noyes, R. W., & Beckers, J. M. 1968, *Sol. Phys.*, 5, 131
- Pishkalo, M. I. 1994, *Astron. Nachr.*, 315, 391
- Ruždjak, V. 1977, *Bull. Astron. Inst. Czech.*, 28, 198
- Sterling, A. C. 2000, *Sol. Phys.*, 196, 79
- Sterling, A. C., & Hollweg, J. V. 1988, *ApJ*, 327, 950
- Sterling, A. C., & Mariska, J. T. 1990, *ApJ*, 349, 647
- Suematsu, Y., Shibata, K., Nishikawa, T., & Kitai, R. 1982, *Sol. Phys.*, 75, 99
- Tóth, G. 1996, *Astrophysical Lett. Comm.*, 34, 245
- Vernazza, J. E., Avrett, E. H., & Loeser, R. 1981, *ApJ*, 45, 635