Astron. Astrophys. 355, 751–758 (2000)

ASTRONOMY AND ASTROPHYSICS

Electron cyclotron maser emission from double footpoints in solar flares

A.J. Conway[1,2] **and A.J. Willes**[3]

[1] Department of Physics & Astronomy, University of Glasgow, Glasgow, UK
[2] Now at Department of Physics & Astronomy, The Open University, Milton Keynes, MK7 6AA, UK
[3] Department of Theoretical Physics and Research Center for Theoretical Astrophysics, School of Physics, University of Sydney, NSW 2006, Australia

Received 4 June 1999 / Accepted 17 January 2000

Abstract. It is now known from Yohkoh Hard X-ray Telescope observations that double (or even multiple) hard X-ray sources in flares are a common occurrence. These sources, which are positioned at the feet of coronal soft X-ray loops, are synchronised to within 0.1s and have similar spectra, strongly suggesting that they are produced by a single population of electrons accelerated/injected at some point in the loop. As this electron population is reflected from the converging footpoint magnetic fields, it develops a loss cone and an electron-cyclotron maser instability may ensue. The frequency and intensity of such emission depends on the relative strengths and orientations of the footpoint magnetic fields. In this paper, we investigate the case of an almost symmetric loop to assess whether observable maser emission from *both* footpoints can result. In particular, we relate this theory to existing observations of solar microwave spike bursts which have two distinct frequency bands that are of non-integer ratio and comparable intensities. We conclude that differing footpoint magnetic field inclinations cannot explain the observations (specifically the comparable intensities), but that it is possible for slightly differing footpoint magnetic field strengths to explain the observations. The pros and cons of this 'geometric' model are then compared with a previous model of these events, which explained them in terms of the growth and then coalescence of Bernstein waves. We conclude that both interpretations seem plausible given current observations, but present a list of observable features that might be used discriminate between them in flare observations of the next solar maximum.

Key words: masers – Sun: flares – Sun: radio radiation – Sun: corona

1. Introduction

During the impulsive phase of solar flares, when hard X-rays from non-thermal electrons are observed, spikes are sometimes detected at microwave frequencies. These spikes generally have a high brightness temperature ($> 10^{15}$K), and have very narrow profiles in both time (~ 10 ms) and frequency (1 MHz). The frequencies of these spikes (0.3-8 GHz) suggest a magnetic field of a few hundred Gauss. All these facts are consistent with the observed emission being from an electron cyclotron maser (Melrose & Dulk 1982) operating in a coronal loop associated with a solar flare. The relation between such radio emission and the impulsive stage as observed in hard X-rays (the latter appear to occur 2-5s before the former) is discussed by Aschwanden & Güdel (1992). It is desirable to improve understanding of the narrowband microwave spike burst emission, as it is a complementary diagnostic to hard X-rays in probing the nature of energy release in some flares.

The electron cyclotron maser occurs when the resonance between electrons spiraling around a magnetic field and circularly polarised waves leads to growth of the waves. The condition for such growth requires the distribution to have a certain form of anisotropy. In solar coronal loops the lack of small pitch angle electrons, because of magnetic mirroring at the loop's footpoints where the magnetic field converges, provides the necessary conditions for growth to occur. In this paper, we are primarily concerned with the electromagnetic x- and o-modes of these waves, as they can, if unabsorbed, propagate away from the source region and then out to the Earth.

Some narrowband microwave spike burst events are observed to have two distinct frequencies bands of emission (Krucker & Benz 1994). The bands are generally very highly correlated in time (to within 0.1s) and are comparable in intensity. The most notable feature is that the frequencies cannot in general be accounted for by a ratio of integers. At non-relativistic electron energies, relevant to solar flares, the integer harmonic spectrum of conventional x- and o-mode maser emission cannot account for such a non-integer frequency ratio. Willes & Robinson (1996) proposed a model for these events that successfully accounts for them by a two stage process. Assuming that w_p/Ω_0 (the plasma frequency over the electron gyro-frequency) is greater than unity, Bernstein waves can grow by the electron cyclotron maser mechanism. By their nature Bernstein waves have frequencies between the harmonics of the gyrofrequency: it is this that leads to the non-integer frequency ratios. However,

Send offprint requests to: A.J. Conway (a.j.conway@open.ac.uk)

these waves cannot propagate away from the source region, so a second step is needed: the coalescence of two Bernstein waves to yield the x- and o-mode waves above the plasma frequency, which may then propagate away from the source region to be observed.

Another suggested model (Fleishman & Yastrebov 1994) is based on the assumption of a two-sided loss cone, which produces two peaks on either side of the gyrofrequency or its harmonic, with emission angle $\cos\theta_1 < 0$ and $\cos\theta_2 > 0$ for the two peaks, corresponding to emission on the surface of two hollow cones with axes parallel and antiparallel to $\mathbf{B}$. Although this model can produce noninteger harmonics, we believe that assuming a two sided loss cone is difficult to reconcile with commonly observed flare loop geometries. In particular, since the loss cone formed at a footpoint will be quickly filled as the maser rapidly grows and saturates, there is no time for the loss cone population of electrons to propagate through the loop to meet with electrons from the other footpoint to form a two-sided loss cone. In addition, for a distant observer to detect both peaks requires that both emission cones have opening angles almost perpendicular to $\mathbf{B}$. However, the required separation in frequency between the two peaks requires that the two emission angles are also well separated, on either side of $\theta = 90°$. It is thus unlikely that a distant observer could simultaneously detect both peaks.

One further possibility which has not yet been explored for this particular application is the upper-hybrid resonance, where the growth rate is strongly enhanced at the upper-hybrid frequency, where the upper-hybrid frequency is equal to an integer multiple of the electron cyclotron frequency (e.g. Kuijpers 1975). It is possible that the plasma density and magnetic field strength vary along the loop in the source region such that several harmonic resonances are excited. As for the Bernstein wave model, a nonlinear coalescence or decay process is then required to produce observable radiation.

We begin by stating the relevant maser theory, and introducing some new useful mathematical expressions for loss cone distributions in the full relativistic case in Sect. 2. In Sect. 3 we describe our model and investigate the implications of the observational constraints. In Sect. 4 we weigh the pros and cons of our geometrical description with respect to the Willes & Robinson (1996) model and discuss potential future observations that may discriminate between the two models.

2. Maser theory

In this section we will state the relevant results describing the loss cone driven electron cyclotron maser. We also show how fully relativistic expressions for maximum growth can be derived. More complete explanations of the theory are given by Conway & MacKinnon (1998) and Melrose & Dulk (1982) and references therein.

In Conway & MacKinnon (1998) expressions were derived to relate the frequency ω, expressed in ratio to the gyrofrequency as $Y = \Omega/\omega$, to the angle of maser emission to the magnetic field θ and the electron speed v corresponding to electron-cyclotron resonance. All of these expressions were obtained in the weak relativistic limit where $\gamma^{-1} = (1-v^2/c^2)^{1/2} \simeq 1 - v^2/2c^2$. In this regime the resonance condition can be expressed as a circle in $(v_\parallel, v_\perp)$-space, centred on a point on the $v_\parallel$-axis. The location of the centre, and the radius of the circle both depend on Y and θ. The condition for maximum growth requires that the (assumed sharp) edge of the loss cone, that is the line $v_\perp = v_\parallel \tan\alpha$, where α is the loss cone angle, just touches the resonance circle. This means that for a given Y, there is a unique value of θ and a unique value of v for maximum growth. Conceptually, as far as such resonance conditions are concerned, the only change in going to the full relativistic case is that the resonance circle becomes an ellipse. We now show how the full relativistic equations yield expressions relating Y, θ and v that are actually *simpler* in form than the weak relativistic expressions given by Conway & MacKinnon (1998).

The resonance condition (for the fundamental frequency) can be expressed as

$$\omega - \frac{\Omega}{\gamma} - k_\parallel v_\parallel = 0 \tag{1}$$

Using the variables introduced above, this can be re-expressed, without approximation, as

$$Y^2\left(1 - \frac{v^2}{c^2}\right) = \left(1 - N\frac{v_\parallel}{c}\cos\theta\right)^2 \tag{2}$$

where $N = kc/\omega$ is the refractive index, and will in general be a function of both Y and θ.

This can then be expanded to yield the equation for the resonance ellipse:

$$(v_\parallel - v_c)^2 + \eta v_\perp^2 - v_R^2 = 0 \tag{3}$$

where

$$\begin{aligned} v_c &= \frac{N\cos\theta}{Y^2 + N^2\cos^2\theta}c \qquad \eta = \frac{Y^2}{Y^2 + N^2\cos^2\theta} \\ v_R &= \frac{Y\sqrt{Y^2 + N^2\cos^2\theta - 1}}{Y^2 + N^2\cos^2\theta}c \end{aligned} \tag{4}$$

The condition for maximum growth is that the loss cone edge just touches the ellipse. This means that $dv_\perp/dv_\parallel$, as derived from the ellipse Eq. (3), must be equated with $\tan\alpha$, the gradient of the loss cone edge. Doing this, using the fact that $v_\parallel = v\cos\alpha$ and $v_\perp = v\sin\alpha$, substituting for v_c and v_R from (4) yields

$$\frac{v}{c} = \frac{N\cos\theta\cos\alpha}{Y^2 + N^2\cos^2\theta\cos^2\alpha} \tag{5}$$

Replacing v/c using this expression in (2) yields the result that

$$Y^2 = 1 - N^2\cos^2\theta\cos^2\alpha \tag{6}$$

and replacing $N\cos\theta\cos\alpha$ in (5) gives

$$Y = \sqrt{1 - \frac{v^2}{c^2}} = \gamma^{-1} \tag{7}$$

We can see from this equation that maser emission arising from resonance with electrons at typical hard X-ray energies cannot

give a Y that is much different from unity. Also, because the number of such electrons falls off as a power-law in energy, the reduction in the number of electrons from $Y = 0.95$ to $Y = 0.8$ (corresponding to a range of electron kinetic energy from 27keV to 128keV) is typically well in excess of two orders of magnitude. So, we can conclude that the intensity of emission will be greatly reduced for Y values that are far from unity for the typical hard X-ray distributions we will consider in this paper.

We will assume the distribution function f is of the form (per unit velocity volume, per unit spatial volume):

$$\begin{aligned} f(v,\alpha') &= f_0 E^{-a} & \alpha' > \alpha \\ &= f_0 E^{-a}\left(1+\tfrac{\alpha'-\alpha}{\Delta\alpha}\right) & \alpha > \alpha' > \alpha - \Delta\alpha \\ &= 0 & \alpha' < \alpha - \Delta\alpha,\ E < E_c \end{aligned}$$

This is a power law distribution, with a cutoff at E_c, which has developed a loss cone of angle $\sin\alpha = \sqrt{B/B_1}$ where B and B_1 are the magnetic field strengths in the corona and at the footpoint respectively. Since the distribution here is "per unit velocity volume", the relationship between a and the more conventional "per unit energy" number flux distribution index, often denoted by δ, is simply $\delta = a - 1$. As typical values are $E_c = 20$keV, and $a = 5$ ($\delta = 4$), it is clear that the majority of electrons are well below the relativistic energies, so in what follows we work in the weak or non-relativistic regime.

Before proceeding we will discuss issues surrounding the above choice of the initial distribution function. The functional form of the distribution chosen to represent the loss cone has important consequences whether, for instance, emission occurs above or below the cyclotron frequency (Kuijpers 1980). However, to date, a satisfactory self-consistent model for how loss-cone distributions form and evolve in a coronal loop has not been developed. So to some extent the form we have assumed is arbitrary. Nevertheless, the main conclusions of this paper should not be too strongly dependent on the choice of electron distribution, because we are only interested in the relative properties between of two footpoints, and not absolute values of frequency or intensity.

We now wish to calculate the energy density of radiation directed towards the observer (as discussed above the direction specifies a frequency, and a particle velocity through the resonance conditions), as this will be proportional to the observed intensity of emission. Adapting the arguments given in Conway & MacKinnon (1998) for the (minimum) energy density to the case of a power law distribution of electrons with a loss cone (see Appendix), yields an energy density:

$$W = nE_c \frac{a(2a-3)}{4(a-1)(a-2)} \left(\frac{E}{E_c}\right)^{\frac{5}{2}-a} g(a,\beta)\beta^{-\frac{1}{2}}(\Delta\alpha)_i \quad (8)$$

where E is the energy of the the resonant particles, n is the density of all the fast particles (i.e. integrated over all energies) and $(\Delta\alpha)_i$ is the width of the edge of the loss cone in the initial distribution. $\beta = B/B_1 = \sin^2\alpha$ is the ratio of coronal to footpoint magnetic field strengths, and is always less than 1. The exact form for the function g is given in the Appendix. For the purposes of this paper, however, we will only require the following approximation to g

$$g(a, b = 1-\beta) = \frac{a}{50}\left[1 - \frac{1}{2}b(1+b^4)\right] \quad (9)$$

which remains accurate to within 10% for $a = 5 \rightarrow 7$ for $0 \le b \le 0.99$.

The above expression for W differs from that given in Conway & MacKinnon (1998) who calculated the total amount of energy emitted by integrating over all possible $v_\parallel$. Here we are interested in just the energy emitted across the small range of $v_\parallel$ corresponding to the finite width of the loss cone edge at the resonance velocity $(v\cos\alpha, v\sin\alpha)$ for the given viewing angle θ. So instead of performing an integral in $v_\parallel$ we simply multiply by the factor $\Delta v_\parallel = v\Delta\alpha/\sin\alpha$. In calculating W we have neglected terms involving $(\Delta\alpha)_i$. As long as $(\Delta\alpha)_i$ is much smaller than α (see below), then the above expression is valid and not inconsistent with having to assume a finite $(\Delta\alpha)_i$ elsewhere, e.g. in $\Delta v_\parallel$.

The appropriate value of $(\Delta\alpha)_i$ is difficult to identify, but it seems likely that it will be set by the effect of Coulomb collisions on the electrons in the short time between the appearance of the loss cone and the growth of the maser instability to the level where it dominates Coulomb pitch angle scattering. The diffusive change in the pitch angle due to collisions whilst electrons are at the mirror point (where the pitch angle is $\pi/2$) can be estimated without recourse to numerical simulations using the theory in Conway et al. (1998). However, since an initial distribution of pitch angles will lead to electrons mirroring at a range of heights, where the density can be a rapidly changing function of height, such treatments are sensitive to assumptions about the assumed loop footpoint structure. For our purpose we need not be too concerned with these issues, because although the value of $(\Delta\alpha)_i$ is important when considering the absolute intensity, we are concerned with electrons that originate from one population, incident on two footpoints which have similar structures. For this reason, we will assume that $(\Delta\alpha)_i$ is the same for each footpoint, and that it cancels out in taking the ratio of intensities. Even if this is not true in reality, the effect of variations of $(\Delta\alpha)_i$ between footpoints can be thought of as being included in the distributions of other quantities introduced into the model later on, e.g. in the distributions of angle or the magnetic field strength.

3. The geometric model

The model we propose here is simple: the two different frequencies arise from maser instabilities occurring at the two footpoints of a coronal loop, where fast electrons are undergoing reflection in a magnetic mirror. The frequency of the radiation received from a particular footpoint depends on the strength of the magnetic field at the footpoint and the angle between the observer's line of sight and the direction of the footpoint magnetic field. It is easy to envisage that the intensity and timing of individual spikes from the two footpoints can be closely related if it is supposed that the fast electrons at both footpoints have a common

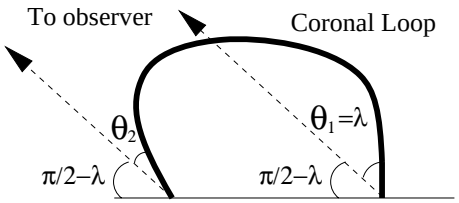

Fig. 1. A coronal loop seen by an observer with viewing angle $\pi/2-\lambda$ to the local solar surface, where λ can be thought of as the observed angular distance of the loop from the centre of the solar disk. θ_i is the angle, at footpoint i, between the magnetic field and the maser emission directed along the observer's line of sight. Although the magnetic field at footpoint 1 is shown as being normal to the solar surface, the model remains general if λ is re-interpreted as an angle combining both the tilt of the loop and its location the solar disk.

origin. This is discussed further in Sect. 4. So, in our model, we proceed by considering the electron distributions entering both footpoint regions to be the same. Fig. 1 illustrates the scenario we wish to consider, and defines the various parameters referred to below.

Before proceeding further, it is important to note that modern observations make it clear that it is highly unlikely that we are dealing with a simple "tube" of magnetic field lines that descends at some constant angle into the photosphere. In reality, one would expect a flare footpoint to be composed of many bundles of field lines, that enter the photosphere with curved, twisted and interwoven paths. So when we talk of angles and strengths of the magnetic fields, we are really dealing with average values at a given footpoint. However, we can be confident that the spread about these values is not great, because if it were, the electron cyclotron maser emission would not be confined to such a narrowband of frequencies as is observed. This also means that our distribution of values amongst loops, can, and should, also be interpreted as a distribution of values in a single footpoint. Admitting this kind of complexity does not necessarily lessen the viability of the geometrical model, in fact, it may strengthen it. The reason being that if there is a distribution of field lines at each footpoint, even if the "average" values are somewhat different, it is still possible that some subset of the field lines at the two footpoints are similar enough to produce emission at the slightly different frequencies. Indeed if this were the case, then two bands could conceivably arise from the same footpoint. Adaption of the theory of this paper to the case where the two sources were in the same footpoint would require no change in mathematical description, but would require some careful re-analysis of the assumptions behind the model. For example, why would having only two sources be so common? In the two footpoints model this can be easily explained by supposing the electrons are accelerated in the corona and then propagate down the two legs of the loop. For clarity in what follows, and because we have insufficient knowledge of loop magnetic field complexities at present, we will discuss the two sources in terms of being from two simple footpoints, with the understanding that such alternative interpretations do exist.

Another related point is that recent observations from TRACE (Transition Region And Coronal Explorer) show that flare loops can be significantly tilted from the vertical, with field convergence mainly confined to lower corona, or upper chromosphere. As depicted in Fig. 1, the variable λ represents the angle between one loop leg and the line of sight to the observer, where the loop legs are assumed to be nearly vertical. In this case λ also represents the angle between the solar disk's centre and the observed location of the flare. If the loop is tilted from the vertical, as the observations of TRACE suggest, then the following theory is still applicable, but λ has to be more strictly defined as the angle between the loop leg labeled 1 and the line of sight, having no reference to the line of sight. Although this does not affect the model described here, it does have implications for the observed distribution of such events across the solar disk.

We will use the subscripts 1 and 2 to refer to the two footpoints. The ratio of emitted frequencies r is therefore

$$r = \frac{\omega_1}{\omega_2} = \frac{B_1}{B_2}\frac{Y_2}{Y_1} = \frac{\beta_2}{\beta_1}\frac{Y_2}{Y_1} \tag{10}$$

Let q be the ratio of intensities of the two observed bands. Since the intensity is proportional to the saturated energy density, from (8) and using (7) and (9), q can be expressed in the following two forms:

$$\begin{aligned} q &= \frac{W_1}{W_2} = \left(\frac{E_1}{E_2}\right)^{\frac{5}{2}-a}\left(\frac{\beta_1}{\beta_2}\right)^{-\frac{1}{2}}\frac{1-\frac{1}{2}b_1(1+b_1^4)}{1-\frac{1}{2}b_2(1+b_2^4)} \\ &= \left(\frac{(1-Y_1)Y_2}{(1-Y_2)Y_1}\right)^{\frac{5}{2}-a}\left(\frac{\beta_1}{\beta_2}\right)^{-\frac{1}{2}}\frac{1-\frac{1}{2}b_1(1+b_1^4)}{1-\frac{1}{2}b_2(1+b_2^4)} \end{aligned} \tag{11}$$

where $b = 1-\beta$. Here Y can be regarded as a function of θ and β, as specified by (6). If r is explained by the Y ratio in (10), then it is clear that q becomes very sensitive to the ratios Y_2/Y_1 and $(1-Y_1)/(1-Y_2)$. As was discussed above, Y cannot deviate far from unity at the hard X-ray electron energies of interest here. So although q will not be sensitive to the Y ratio, it will be sensitive to the $1-Y$ ratio. We can therefore immediately state that any observation of two frequencies in ratio r, which are of comparable intensity, can only be accounted for if $r \simeq B_1/B_2$. In other words, only differences in the magnetic field strength, and not its direction, can account for non-integer microwave bursts of comparable intensity in the geometric model.

To proceed further we will assume that $N \simeq 1$, which is the case if Ω is sufficiently larger than the plasma frequency. We will also imagine the coronal loop is at an angular distance λ from the solar disk centre, with the field at footpoint 1 being perpendicular to the local solar surface. The field at footpoint 2 is at an angle to the perpendicular, and gives rise to a difference $\Delta\theta = \theta_2 - \theta_1$. In what follows, we will assume that $\Delta\theta$ has some random distribution amongst solar flare loops. Strictly, the correct method would be to assume such a distribution for the difference in magnetic field directions and then convolve this distribution with the distribution in the orientations of coronal loops. However, given that we are only interested in small $\Delta\theta$

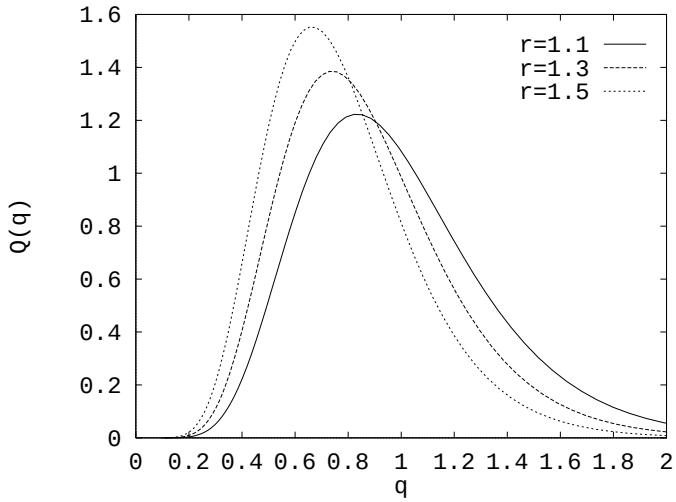

Fig. 2. This plot shows the distribution function $Q(q)$ for the ratio of intensities q from the two footpoints. It is calculated for parameters $\beta_1 = 0.1$, $a = 5$, $\lambda = 60°$ and $\sigma = 2°$.

(otherwise q cannot be near 1), and that we do not have sufficiently detailed information on the required distributions, we will simply assume $\Delta\theta$ to be Gaussian. Note that the description of the loop geometry solely in terms of λ and $\Delta\theta$ is quite general, because λ can be equivalently interpreted as the tilt angle of the field at footpoint 1.

With the distribution of $\Delta\theta$ given by

$$P(\Delta\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(\Delta\theta/2\sigma)^2}$$

σ^2 being the variance of the distribution, the observed distribution of q values will be $Q(q) = P(\Delta\theta)/|dq/d\Delta\theta|$. Fig. 2 plots $Q(q)$ for $a = 5$, $\lambda = 60°$, $\beta = 0.1$ and $\sigma = 2°$, with a curve for each of $r \simeq B_1/B_2 = 1.1, 1.3, 1.5$. In all cases events of $q \sim 1$ are not only possible, but favoured. The important features of each curve in Fig. 2 can be summarised as: the mean value of q, q_m, the standard deviation, q_s, and the probabilities of q lying within a factor of 2 and 10 of unity, q_2 and q_{10} respectively. q_2 is a measure of how likely it is to see the two frequency bands being of comparable intensity, and q_{10} is measure of how likely it is that one band is overlooked in the observations, because it is more than an order of magnitude less intense. Fig. 3 plots q_m, q_s q_2 and q_{10} as functions of r for $a = 5$, $\lambda = 30°, 80°$, $b = 0.1, 0.3, 0.5$ and $\sigma = 2°$.

In the parameter ranges of interest, $r\beta_1 < 1$ and $a = 5 \to 7$, the following expression can be obtained from (11) in the regime $(2a-5)\tan\lambda\Delta\,\theta \ll 1$:

$$q \simeq (1 - (2a-5)\tan\lambda\,\Delta\theta)\left(\frac{1-\beta_1 r}{1-\beta_1}\right)^{a-5/2} r^{-1/2} \qquad (12)$$

In fact, numerical evaluation of (11) shows that this expression is good to within 10% for most parameter ranges of interest here. The exception to this is as λ nears $90°$, but even there (12) still exhibits the correct trends, even if the absolute value is incorrect.

A particularly useful summary of the results displayed in Fig. 3 can be obtained by taking the expectations of q and q^2 using (12), given that the expectations of $\Delta\theta$ and $(\Delta\theta)^2$ are 0 and σ respectively. Doing so yields the following expression for the mean and standard deviation of q

$$q_m \simeq \left(\frac{1-\beta_1 r}{1-\beta_1}\right)^{a-\frac{5}{2}} r^{-1/2}\,,\; q_s \simeq (2a-5)\tan\lambda\,\sigma q_m \qquad (13)$$

Note that q_m is independent of λ and σ and will remain meaningful if $r\beta_1 < 1$. The expression for q_s is valid for $(2a-5)\tan\lambda\sigma \ll 1$. From these expressions we can see that we expect comparable intensities only if r is not too far from 1, β_1 is not too close to 1, and $\sigma < 57°/(2a-5)\tan\lambda$ (σ is expressed in degrees here). For example, with $a = 5$, $\beta_1 = 0.1$, and $\lambda = 60°$, $\sigma < 6°$ will ensure $q \sim 1$ for all $r < 2$.

Having established that bursts from the two footpoints can be of comparable intensity, we now consider whether the number distribution of bursts in r can be explained by the geometric model. In Krucker & Benz (1994), this is shown for 22 frequency pairs in the range $r = 1.0 \to 1.6$.

The $\lambda = 30°$ plots in Fig. 3 show that q_2 and q_{10} can rapidly drop to zero at particular values of r. This 'cutoff' effect can explain the lack of events with $r > 1.5$ in the observed distribution. However, there is no reason in the geometrical model for why this cut-off should be around $r = 1.5$. Also, it does not predict any peak in the distribution, because in all cases as $r \to 1$, $q \to 1$. So, as it stands, the geometrical model does not naturally explain the observed distribution in r.

There are numerous plausible *ad hoc* complications to the geometric model that could result in a peak appearing in the predicted r distribution. For example, it could simply be that the r distribution reflects the distribution in B_1/B_2 amongst solar flare loops. Alternatively, there could be a joint distribution between B and the angle of the field to the solar surface, having the effect of making σ a function of r, so that similar magnetic field strengths are more likely to be parallel. Another possibility is that events with r close to one have been mistaken for single band events. Given the spread in observed frequencies (e.g. Fig. 1a in Krucker & Benz 1994), this could well be happening for events of $r < 1.1$, but it would be difficult to use this to explain a peak at $r = 1.4$. Without more detailed observational knowledge of related distributions, crucially for B_1/B_2, these possibilities are merely speculative and do not warrant more detailed consideration.

Finally, we wish to comment on the small number statistics of the observed distribution. There were actually only 17 individual events. The 22 pairs presented by Krucker & Benz (1994) are from 13 individual double band events, 3 triple band events and 1 quadruple band event. This means that there are really only 17 independent values on the plot. The Kolmogorov-Smirnov test, using only the 17 lowest frequency ratios, suggests that there is a 10% probability of obtaining the observed data by chance if the real distribution is flat. This being the case, it would seem that some caution is required in drawing conclusions from the observed distribution in r; in particular that there

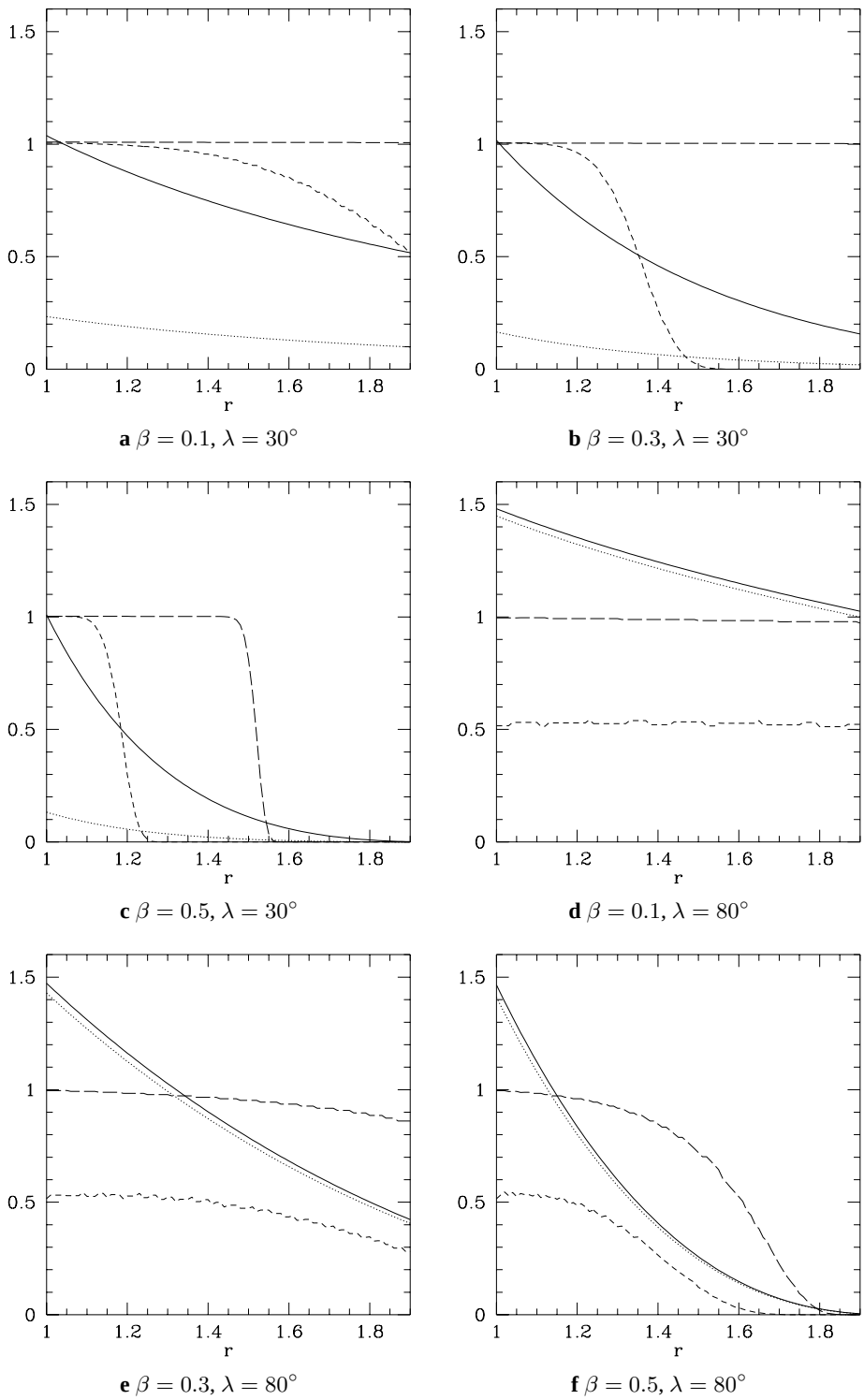

Fig. 3. Each plot displays four properties of the $Q(q)$ intensity distribution function: the mean q_m (solid), the standard deviation q_s (dotted), and the probabilities of q lying within a factor 2 and 10 of unity, q_2 (short dashed) and q_{10} (long dashed) respectively. It is calculated for parameters $a = 5$ and $\sigma = 2°$.

definitely is a peak at 1.39 ± 0.01 (quoted from Krucker & Benz 1994).

4. Conclusions and comparisons

As stated in Krucker & Benz (1994), any theory explaining the observations must account for four key points. 1) The observed non-integer frequency ratios (1.2 to 1.5, 1.06 was observed in one event, but it was also atypical because the frequencies drifted); 2) both bands should have comparable intensity; 3) the microwave emission must be able to propagate out to be observed; 4) the simultaneity of the two bands.

The simultaneity (to within 0.1s) of hard X-ray footpoints observed in flares (Sakao 1994; Sakao et al. 1996) mean that

the fast electron populations producing the hard X-rays must be synchronised somehow. This could be, for example, because fast electrons are injected or accelerated near the top of the loop in a symmetrical fashion. Whatever the reason, for our purposes, requirement 4) is satisfied because these hard X-ray observations suggest that the electron populations at the two footpoints responsible for the radio emission are produced simultaneously. However, we note that the time delay of several seconds between the hard X-ray emission and the maser emission (Aschwanden & Güdel 1992) remains unexplained.

The model discussed in this paper accounts for requirements 1) and 2) if the difference in magnetic fields at the two footpoints of a loop explain the frequency difference, i.e. $\omega_1/\omega_2 \sim B_1/B_2$. This also requires that the angles of the magnetic field at the footpoints are the same (to within a few degrees, depending on the loop's location on the solar disk and strength of magnetic field convergence). It is also possible to explain 1) in terms of different angles of the magnetic field to the solar surface at the two footpoints, but it is then not possible to account for 2), except with very contrived scenarios.

Meeting requirement 3) invokes a long standing problem for all solar electron cyclotron maser models: the escape of maser emission past the second harmonic layer, where, according to existing theory, it should be strongly absorbed. However, since we do often see microwave radiation that must be of electron cyclotron maser origin (high brightness temperature and narrowband) it is clear that the emission does somehow escape. A likely possibility (Kuncic & Robinson 1993) is that o-mode radiation escapes through the absorption layer, because the "windows" through which escape of radiation is permitted are wider for the o-mode than for the x-mode.

The Willes & Robinson (1996) model also satisfies all four requirements: 1) The non-integer frequency ratios are a natural consequence of the frequency of peak Bernstein wave growth being shifted away from each cyclotron harmonic. The peak in the frequency ratio just below 1.4 corresponds to a preponderance of excited $s = 2$ and $s = 3$ Bernstein modes. 2) Comparable intensities between simultaneous bands requires comparable maser growth rates in adjacent Bernstein modes, and similar emission rates for the nonlinear coalescence processes for each excited Bernstein mode. These conditions are often satisfied for typical coronal parameters. 3) Product transverse radiation from the Bernstein wave coalescence process is (for harmonic number $s \geq 2$) emitted above the second and third harmonic absorption layers and hence avoids significant damping. 4) The simultaneity of the multiple bands is a direct consequence of the simultaneous excitation of multiple Bernstein modes.

In the events described in Krucker & Benz (1994) there are three triple-band events and one quadruple-band event. The present model can be extended to explain such events if the flare region is more complicated than we have supposed. For example, if several interacting loops were involved in a flare, with fast electrons able to move from one loop to another, then there could be several footpoints with incident fast electrons. An alternative explanation, in terms of just a single loop, is in terms of solutions to the resonance condition for more than one frequency. In this paper, we have just assumed that the frequency ω is sufficiently far from the cut-off so that $N \simeq 1$. However, if this were not true, and in particular, if the functional dependence of N on Y was important, then (6) might well permit other physically realistic resonance frequencies for a given viewing angle θ and loss cone angle α. At present the observations do not warrant further detailed analysis of these possibilities.

Finally, we discuss ways in which the present geometric model and the model of Willes & Robinson (1996) may be distinguished observationally. The most obvious test of the present model would be to identify the two sources in spatially resolved observations. The first VLA observations of narrowband spike events (Krucker et al. 1995) achieve angular resolutions of about 50" to 100". To see distinct footpoints in flare loops requires an angular resolution of about 10". Although such resolutions are possible in principle (e.g. with the VLBI, see Tapping et al. 1983), as yet no observations of non-integer type events have been made.

The geometric model implies that there should be a distribution in the relative intensities of the bands due to the distribution in the magnetic field angles at the footpoints of loops. As future observations are obtained it should be possible to construct a distribution of the relative intensities and compare it with theoretical distributions (referred to as $Q(q)$ earlier), like those in Fig. 2. Also future observations of the footpoint magnetic field strengths in flare loops could be used to construct a distribution of expected r (frequency ratio) values. From this distribution it would be possible to estimate how many events should have r in a given range, say 1.0 to 1.6, in a given period of time and so check if this is consistent with the observed number of such non-integer events.

Another way to distinguish between these two models involves polarisation measurements. The geometric model is likely to produce high degrees of polarisation in favour of the o-mode, because the o-mode can more readily escape from the source region, whereas the Willes & Robinson (1996) model favours low to moderate degrees of polarisation in favour of the x-mode (Willes 1999). However, at this stage, insufficient definitive polarisation observations have been published to distinguish between these models (see discussion by Willes 1999). Also, even when such observations become available, the issue will remain complicated because the polarisation may be altered during a wave's passage through the solar atmosphere outside the source region.

Appendix: derivation of the energy density

As argued in Conway & MacKinnon (1998), the (minimum) energy released by a maser during the time from initial growth to saturation can be calculated by subtracting the energy of a final distribution which has maximum energy, but $df/dv_\perp \leq 0$ everywhere, from the initial distribution. For the purposes of this paper, we are interested only in the energy released in a narrow range of frequencies around the resonance frequency for a given emission angle (corresponding to the direction of the observer). The range of frequencies corresponds to the range of resonant

velocities for which there is significant growth, which is set by the finite width of the loss cone edge $(\Delta\alpha)_i$ (discussed further in Sect. 2). As we are working in the non-relativistic regime, the release of energy is dominantly accounted for by loss of perpendicular energy in the electron distribution. So, we can write that the emitted energy density W is given by

$$W = \Delta v_{\parallel} \quad 2\pi \int_0^{\infty} (f_i - f_f) E v_{\perp} dv_{\perp}$$

where, according to the discussion above, the small range of parallel velocities corresponding to the observed frequency width is $\Delta v_{\parallel} = v(\Delta\alpha)_i / \sin\alpha$. Using f_i as given in Sect. 2 of this paper, and using Eq. (11) from Conway & MacKinnon (1998) for f_f then yields the final form given for W in (8). The exact expression for the function g is given by

$$g(a, \beta) = 1 + x^{a-1}(a-2) - x^{a-2}(a-1)$$

where x is the solution of

$$\frac{x^{1-a} - a}{a-1} + x(1-\beta) = 0$$

This expression relating x to a and β comes from the condition that the number density of particles in the distribution does not change, and is equivalent to the equation after (11) in Conway & MacKinnon (1998) which was for a Maxwellian distribution with a loss cone (note: there is an error in that equation, there should be a $+1$ on the righthand side).

We now compare our expression, which estimates the minimum energy density of radiation, with the expression derived in Melrose & Dulk (1982) which derived the saturated energy density, obtained by assuming that the loss cone has been filled by the enchanced pitch angle diffusion associated with the maser. In our notation, Eq. (14) in Melrose & Dulk (1982) becomes:

$$W_M = nE_c(2a-3)\pi(\Delta\alpha)_i \beta(1-\beta) \left(\frac{E}{E_c}\right)^{5/2-a}$$

The ratio of this to W as given in (8) is

$$\frac{W_M}{W} = 4\pi \frac{\beta^{3/2}(1-\beta)}{g(a,\beta)} \frac{(a-1)(a-2)}{a}$$

In the case where β is not close to unity, which is of interest to us here (and in any case a condition for either treatment to remain meaningful), then, using the approximation to g (9) we find

$$\frac{W_M}{W} \simeq 200\beta^{1/2} \frac{(a-1)(a-2)}{a^2}$$

In this paper, we are only interested in a ratio of intensities, and so the 200 factor and the factor involving a would not change our results (we assume the electrons at both footpoints to originate from the same population). Although in principle the extra factor of $\beta^{1/2}$ can change our results, we have found it only effects details of our results, and does not change our overall conclusions in any way.

To improve upon either Melrose & Dulk (1982)'s method or our method would require not only a more detailed time-dependent modelling of the maser, such as that of Aschwanden (1995), but a more detailed consideration of how the maser began to grow during the time in which the loss cone formed. Such a treatment is beyond the scope of this paper.

Acknowledgements. AJC would like to thank PPARC for funding, and provision of the Starlink computer facilities, and thanks Alec MacKinnon for many useful discussions. AJW acknowledges support from a University of Sydney U2000 postdoctoral fellowship. AJC also thanks the Research Center for Theoretical Astrophysics, University of Sydney for funding his visit there. We also thank the referee, Jan Kuijpers, for his constructive criticism of the paper.

References

Aschwanden M.J., Güdel M. 1992, ApJ 401, 736
Aschwanden M.J. 1995, A&AS 85, 1141-1117
Conway A.J., MacKinnon A.L. 1998, A&A 339, 298
Conway A.J., MacKinnon A.L., Brown J.C., MacArthur G.K. 1998, A&A 331, 1103
Fleishman G.D., Yastrebov S.G. 1994, Sol. Phys. 154, 361-369
Kuijpers J. 1975, A&A 40, 405
Kuijpers J. 1980, IAU Symp., Radio Physics of the Sun, pp. 341-361
Kuncic Z., Robinson P.A. 1993 Sol. Phys., 145, 317
Krucker S., Benz A.O. 1994, A&A 285, 1038-1046
Krucker S., Aschwanden M.J., Bastian T.S., et al. 1995, A&A 302, 551-563
Melrose D.B., Dulk G. 1982, ApJ 259, 844
Sakao T. 1994, Ph.D. thesis, University of Tokyo
Sakao T., Kosgui T., Masuda S., Yaji K., Inda-Koide M., Makishima K. 1996, Adv. Space. Res. 17, 67-70
Tapping K.F., Kuijpers J., Kaastra J.S., et al. 1983, A&A 122, 177-180
Willes A.J., Robinson P.A. 1996, ApJ 467, 465-472
Willes A.J. 1999, Sol. Phys. 186, 319