EXTREMELY RAPID RADIO SPIKES IN FLARES  
(Review)  

A. O. Benz  
Institute of Astronomy, ETH  
Zurich, Switzerland  

ABSTRACT  

Radio spikes of a few to tens of milliseconds of the solar radio emission have recently seen a surge of interest of theoreticians who are fascinated by their high brightness temperature of up to $10^{15}$ K, their association with hard X-ray bursts, and a possibly very intimate relation to electron acceleration. Their bandwidth and global distribution in frequency have quantitatively been measured only recently. This review is intended to emphasize the considerable extend of old and new observational knowledge which is hardly touched upon by theory. The wide range of spike observations is summarized and brought into the perspective of recent models. It is concluded that spikes yield a considerable potential for the diagnostics of energetic particles, their origin, and history in astrophysical plasmas.  

1. Introduction  

Millisecond radio spikes are a rapidly growing field of solar radio astronomy. Although their role and diagnostic capabilities for flare theory, nor even their emission mechanism are clear, considerable progress in our understanding has been achieved over the last few years. Spikes today are generally agreed to be a non-thermal, coherent emission closely connected with particle acceleration and energy release in flares.  

Radio bursts with durations of less than 100 ms have first been noted by Dröge and Riemann (1961) and Elgarøy (1961). They have been studied by de Groot (1962), Elgarøy (1962), and later by Eckhoff (1966) and de Groot (1966). The first major articles on the subject did not appear until Dröge (1967) and Malville, Allen and Jansen (1967) summarized their observations. This work was extended by de Groot (1970) and Tarnstrom and Philip (1972a,b). The pioneers used various names for the new phenomenon: knots, pips, rain, flash bursts, etc. We will use the word "spike" introduced by de Groot, which today is well established in the community, and restrict it to narrowband peaks of less than 100 ms total duration. Note that in hard X-ray and high frequency microwave data the same expression sometimes is used for entirely different phenomena such as peaks with ten times or longer durations or tiny fluctuations of a background emission.  

In the first decade of spike observations the observing frequencies were in the range from 200
to 350 MHz. This was extended by more than an order of magnitude in both directions in the following years. Barrow and Saunders (1972) have found spikes at 18 - 26 MHz associated with type III radio bursts. Their observation, however, has never been confirmed by spectral measurements. At microwaves, spikes were observed up to 1420 MHz by Dröge (1967 and 1977), at 2650 and 2840 MHz by Slottje (1978) and Zhao and Yin (1982). They have recently been discovered up to 5200 MHz by Stähli and Magun (1986). These authors did not find spikes (as defined above) at higher frequencies. It is clear today that spikes are most abundant in the decimetric range, i.e. from 300 to about 3000 MHz.

How many types of spikes are there? The report of Slottje (1978) of fully polarized spikes in a microwave event at the time was considered evidence for a species of spikes entirely different from the intermediately polarized kind at lower frequency. However, subsequent observations by Slottje (1980) and Stähli and Magun (1986) of a larger set of spikes at microwaves showed that the event was exceptional and the general polarization behaviour is similar to the one deduced by Benz, Zlobec, and Jaeggi (1982) at 300 MHz. Secondly, spike emissions may have different origins if they occur in different contexts as manifested by other radiations. Spikes have been found to be associated with metric type I storms (Elgarsøy, 1962; Eckhoff, 1966), type III bursts (Elgarsøy and Rødberg, 1963; Tarnstrom and Philip, 1972b), and type IV events (Dröge, 1961; Elgarsøy, 1961; de Groot, 1962, and later authors). Malville et al. (1967) measuring only total flux at two frequencies could not find any difference between spikes and type I bursts except in duration. Elgarsøy and Eckhoff (1966) noted a smooth transition from type I bursts to spikes and back during a noise storm. However, spectrographic observations by de Groot (1970) demonstrated that spikes associated with type III and type IV bursts preferentially occurred at higher frequencies than type I bursts. Finally, Benz et al. (1982) found significant differences to type I bursts in polarization, bandwidth, and spatial distribution on the solar disk. In conclusion, it seems that presently only two species of spikes can safely be distinguished: spikes in noise storms at metric frequencies, which seem to be identical to type I bursts except for their shorter duration, and "real" spikes, which extend to much higher frequencies and are associated with flares. This review concentrates on the second kind. Whether it is a homogeneous set of phenomena or needs to be divided along frequencies (such as decimetric vs. microwaves) or associated metric activity (type III vs. type IV) needs to be investigated. Nevertheless, this review follows the rule that phenomena have to be considered as manifestations of one type until shown to be different.

This article is the first summary of 25 years of spike research. The main emphasis is on a complete discussion of observations and their theoretical implications. The theory of spike emission is also briefly reviewed. The goal is to draw the attention of the larger community interested in solar flares to the rapidly growing set of spike observations, and most of all to bridge the gap which sometimes seems to separate theoreticians and observers.
2. The Spike Phenomenon

2.1. Time Profile

The duration of spikes, orders of magnitude shorter than any other type of radio emission, led to their discovery as soon as appropriate instruments were in operation. Early observations are shown in Figure 1. Many authors have reported contradictory values for the duration of single spikes. Limited instrumental resolution may explain some of the discrepancies. Other authors emphasized extremely short values which were, however, exceptional cases. In addition there seems to be a trend to shorter duration at higher frequencies noted already by Dröge (1967) and Tarnstrom and Philip (1972b).

![Graph showing spike phenomenon](image)

Fig. 1: Full Disk radio observations (total flux vs. time) of a solar flare at different frequencies. Top: Spikes superimposed on major type IV event in microwaves. Bottom: Spikes at 460 MHz associated with metric type III bursts at 240 MHz (from Dröge, 1967).

Considering only measurements with sufficient resolution, typical durations of single spikes around 250 MHz are 50 - 100 ms (Dröge, 1967, Benz et al., 1982). Barrow et al. (1984) measuring with 0.3 ms resolution noted structure down to 5 ms. The typical duration decreases to 10 - 50 ms at 460 MHz and to 3 - 7 ms at 1420 MHz (Dröge, 1967). It seems to be below 10 ms around 3000 MHz (Zhao and Yin, 1982; Stähli and Magun, 1986).

Tarnstrom and Philip noted that the duration of spikes is comparable to the electron-ion collision time interactions,

\[
r = \frac{0.18 \, T^{3/2}}{n_i \, \ln \Lambda},
\]

(Zheleznyakov, 1970) assuming equal electron and ion temperature \( T \). With \( \ln \Lambda \approx 11.2 \) and for fundamental plasma emission
\[ r \approx 4.0 \left( \frac{1}{\nu_{GHs}} \right)^2 \left( \frac{T}{2 \cdot 10^6} \right)^{3/2} \text{ms} \]

where \( \nu_{GHs} \) is the observing frequency in GHz. Several authors have derived upper limits on the source size, requiring it to be smaller than the duration divided by the speed of light. In the light of the above correlation of duration and collision frequency, it does not seem plausible that the source size decreases with frequency. More likely is the duration determined by some collision time, and the upper limit of the size derived from duration at high frequency is closest to the actual dimension.

2.2. Spectrum

A better estimate of the source size can be derived from the bandwidth of single spikes. Early spectras (e.g. de Groot, 1970) have already revealed that spikes are very narrow-banded. Reported observations of the bandwidth vary between 0.5 and 15 MHz. They have been measured with various methods and need to be considered with caution. Film recordings yield total bandwidth above threshold. The measured values thus depend on peak flux minus threshold. The firstquantitative spectra of spikes have been published only very recently (Benz, 1985). The half-power width at practically instantaneous time is typically 10 MHz, or 1.5 %, at a center frequency of 600 MHz. Figure 2 is an instructive comparison between spike and type III bursts in frequency and time. This extremely narrow width is a powerful restriction on possible emission processes.

![Fig. 2: Three-dimensional representation of spikes (front) and type III bursts (back): time increases to the right (total of 4 seconds is shown). Frequency decreases with depth (370-250 MHz), and flux is shown logarithmically in vertical direction. The data was recorded by the digital spectrometer (IKARUS) in Zurich on 1980, September 24, 0731 UT.](image)
Spikes have escaped detection by routine film-recording spectrographs for a long time. For this reason the total bandwidth of spike activity and the total number per event remained unclear. Single frequency observations by Dröge (1967) suggested total bandwidths of a few hundred MHz for spike activity in typical events. This has been confirmed with the digital spectrometer in Zurich. Using this instrument Benz (1985) has estimated the total number of spikes per event between 8200 and more than 13200 in 4 rich spike events. The multitude of spikes is evident in Figure 2 showing the contrast between type III and spike bursts. Spikes associated with type IV bursts may after some time of random occurrence arrange themselves to patterns in the frequency-time plane, which may resemble broadband pulsations or parallel drifting bands (Kuijpers et al., 1981). A global shift of spike activity from 3 GHz to $\leq 1$ GHz has been noted by Fu et al. (1983) in the 1981, May 16 event. This may reflect a general shift of the spike sources to lower density and possibly higher altitude.

2.3. Polarization

De Groot (1962), Chernov (1976, 1978) and Slottje (1978) reported "strong" circular polarization of spikes. More recent measurements (Slottje, 1980; Benz et al., 1982; Stähli and Magun, 1986; Nonino et al., 1986) agree that the polarization is generally higher than e.g. for type III bursts, but it can vary from 0 to 100 %. It is interesting to note that these observers measured

![Calibrated spectrogram of impulsive phase of a flare on 1980 September 24 observed by the digital spectrometer in Bleien (Zurich). Top: total flux. Bottom: polarization spectrum of the same time interval showing separation of type III and spike bursts. Left circular polarization is represented bright, right circular polarization dark, and zero polarization is gray (from Benz and Kane, 1986).](image-url)
different frequencies (from 0.2 to 3.2 GHz) and associated with different metric activity (type III and type IV). An example of a polarization measurement is shown in Fig. 3. For this rare case the sense of polarization of the type III is opposite to the spikes. The polarization averaged over many spike events is between 25 and 30 %. Surprisingly, the value does not vary between 0.238 GHz (Zurich and Trieste observations) and 3.2, resp. 5.2 GHz (unpublished Bern observations).

2.4. Position

The center-to-limb variation of the rate of occurrence of spikes has been investigated by statistics on associated Hα flare positions. No longitudinal effect has been noted by Benz et al. (1982) at 0.3 GHz and Stähli and Magun (1986) at 3.2 GHz. It may thus be concluded that propagation effects do not play a major role in the spike process.

Only one direct measurement of the position of a spike event has been reported (Heyvaerts et al., 1978). The sources were found separated from the associated type III bursts by about 1 arcmin. Therefore it seems that not only the emission mechanisms of the two radiations are different, but also the source environment.

3. Phenomena Associated with Spikes

The timing of spike emission in relation of the flare process is an important indicator for the interpretation of spikes.

![Graph showing correlation of frequency-averaged spike flux in the frequency band 580-640 MHz (middle) with type III emission in the 250-310 MHz band (top) and HXR (bottom).](image)

Fig. 4: Correlation of frequency-averaged spike flux in the frequency band 580-640 MHz (middle) with type III emission in the 250-310 MHz band (top) and HXR (bottom). The radio data has been recorded with the Zurich digital spectrometer (IKARUS), the HXR observations were made by HXRBS/SMM (from Benz, 1985).

Provided by the NASA Astrophysics Data System
3.1. Other Radio Emissions

Spikes most frequently appear at times of type III bursts, the radio signature of electron beams in the corona. Even with a modern film recording spectrograph spikes are observed near the starting frequency of type III bursts in 10% of all cases (Benz et al., 1982). They are generally at higher frequency (and thus higher source density) than the associated type III bursts. Examples of type III-spike associations are given in Figs. 2 - 4. Figure 4 shows a relatively close correlation of the time variations. Details of type III and averaged spike emission sometimes, but not always, correlate. Some examples of detailed correlation of single type III bursts with clusters of spikes have been given by Benz et al. (1982). It seems very likely that spikes are caused by energetic electrons or their acceleration process.

Karlicky (1984) has analyzed spikes in big outbursts (usually type II and IV). He finds the spikes not always related to type III bursts. Some appeared shortly before the start of a type II or another manifestation of mass ejection. Spikes generally occurred before pulsations, which have been proposed to be caused by energetic particles trapped in magnetic loops. These observations suggest that spike emission neither requires streaming, nor trapped particles.

Stähli and Magun (1986) find from single frequency measurements at 3.2 GHz that 10% of all events show temporal fine structure possibly caused by spikes. In agreement with Slottje (1978), Zhao and Yin (1982) and others, they find the spike activity to generally occur in the rise and maximum phase of the impulsive microwave (synchrotron) emission. An example of the phasing of spikes in relation to the impulsive microwave emission is given in Fig. 5.

![Graph](image)

Fig. 5: Observation of spikes at 5.2 GHz by the Institute of Applied Physics in Bern on 1982, February 10. The spikes are superposed on the smoother, impulsive synchrotron emission (courtesy of M. Stähli).

3.2. X-Ray Emission

Hard X-ray (HXR) emission originates from bremsstrahlung and provides reliable information on the energy of fast electrons. Benz and Kane (1986) find enhanced X-ray emission above 26 keV in 71% of well developed type III/spike events. All major spike events are accompanied by enhanced HXR. Benz and Kane (1986) have noted that HXR emission associated with spikes
tends to be more impulsive and shorter in duration than the average HXR burst. The correlation of HXR and spikes can be very close as e.g. in Fig. 4. However, it is generally not as good and reliable as between HXR and microwave emissions. The occurrence of spikes seems to require additional conditions on the source or on the exciter. The occasionally close association of spikes with impulsive HXR emission suggests that spikes are intimately related to the energization of fast electrons. This is supported by the observed concentration of spikes in the rise and maximum phase of HXR bursts (Benz and Kane, 1986), which is contrary to the timing of all other decimetric emissions except type III. The understanding of spikes may thus yield information on the primary energy release in flares.

Soft X-ray observations of the 1980 August 31 flares by Strong et al. (1984) yield preflare densities with plasma frequencies in the range of the frequencies of the spikes observed later during the flare (Benz, 1985). It is generally believed that spike emission occurs at a frequency which is within a factor of two of the local plasma frequency. The observations then indicate that spikes occur near the flare site before the density increase by evaporation of chromospheric material. If shown to be generally true, the range of spikes in frequency would limit the density in the primary energy release region to about $10^9 - 10^{11} \text{cm}^{-3}$.

If the spike sources were located in an isothermal atmosphere, the derived density range would correspond to an extent of 5.5 scale heights or 550000 km at $2 \times 10^6$ K. The close correlation with hard X-rays (known to originate mostly from $\lesssim 2500$ km, Kane, 1981) and the absence of drift in spike clusters clearly exclude such a possibility. The spread of spike activity in the spectrum thus seems to be caused by a density (or magnetic field) inhomogeneity other than gravitational. Compared with the much smaller inhomogeneity observed in type III or U bursts, the inhomogeneity of spike sources seems to be due to a gradient in perpendicular direction to the magnetic field which, in addition, is steeper at higher frequency (lower altitude). Benz and Kane (1986) have concluded that spikes emission (and thus acceleration) take place in a highly inhomogeneous region with density variations of about one order of magnitude.

4. Theory

Recent observations at decimeter and microwave frequencies have shown that millisecond spikes are a phenomenon associated with the impulsive phase of primary energy release in flares. They often correlate with HXR and type III radio emission, both manifestations of 10 - 100 keV electrons. Occasional absence of correlation has been interpreted in terms of unfavorable source conditions (Benz and Kane, 1986): Type III emission requires electrons streaming on quasi-open field lines, HXR have a high threshold for detection, and the conditions for spikes are unknown. It is generally agreed today that spikes are signatures of energetic electrons.

4.1. Source Size and Brightness Temperature

Estimates of the source size of spikes yield small values and thus lead to enormous brightness temperatures of spike radiation. They are a challenge to theory and have in the past attracted
the attention of theoreticians. Upper limits based on duration may not be very meaningful, since
the duration seems to depend on frequency and decreases approximately with the mean collision
time (§ 2.1.). Estimates using the bandwidth seem to be more reliable. Let us assume that the
emission frequency depends on a characteristic frequency (such as the local plasma frequency or
gyrofrequency). The source dimension $l$ of a spike is then determined by the scale length $\lambda$ of the
characteristic frequency and the bandwidth $\Delta \omega$ of the spike:

$$l = \lambda \frac{\Delta \omega}{\omega}$$  \hspace{1cm} (3)

If the natural width of the emission frequency cannot be neglected, equ. (3) only gives an
upper limit on the size. Quantitative measurements of all variables in equ. (3) yield $l \lesssim 200$ km
(Benz, 1985). This is an order of magnitude smaller than the "speed of light dimension" derived
from the duration of spikes at 600 MHz, but comparable to that upper limit at 3 GHz. It is
interesting to note that the source size derived from equ. (3) agrees with the possible observation
of a spike by VLBI technique yielding a diameter of approximately 50 km (Tapping et al., 1983).

With a diameter of 200 km and for a circular source the brightness temperature of spikes is
up to $10^{15} K$. Only coherent emission processes can reach such an intensity.

4.2. Emission Process

Early ideas on the emission mechanism included plasma emission and electron cyclotron emis-
sion and were based on analogies to other impulsive radio emissions (Malville et al., 1967; Tarn-
strom and Philip, 1972b). A plasma wave model was first presented by Zheleznyakov and Zaitsev
(1975). They proposed emission at the harmonic of Langmuir waves generated by unstabilized
electron beams. As soon as the "gentle-beam" instability stabilizes, the beam emits ordinary type
III radiation. Chernov (1978) developed the model further and realized that such beams would
have to be small in size (500 km) and nearly monoenergetic. Although plasma emission is still used
today in modelling spike emission (Kuijpers et al., 1981, Karlicky, 1984) its predicted similarity
to type III emission contradicts the observations. Spikes have a much smaller intrinsic bandwidth,
higher polarization and, most of all, a 4 orders of magnitude higher brightness temperature. Spikes
probably have a different emission mechanism.

Langmuir waves may still be the cause of spikes. Their transformation into radio emission,
however, would have to be an extraordinary process (as examples we mention strong turbulence
or direct conversion on density gradients). A further possibility has been studied by Vlahos et al.
(1983) who considered the coherent wave-wave coupling of two antiparallel upper-hybrid waves.
The random initial phase and finite coherence length produce a spiky radio emission. This emission
process however still needs to be shown to agree with the wealth of observations summarized in
sections 2 and 3.

Cyclotron emission is today's most favored process for spike radiation. Cyclotron waves grow
exponentially in loss-cone velocity distributions of electrons. Such a distribution may be the result
of trapping (or just one reflection) of energetic particles in magnetic mirrors. Of particular interest
is the cyclotron maser instability, in which electrons with velocity $v$ are in resonance with transverse
electromagnetic waves \((\omega, k)\) if

\[
\omega - s \Omega_e - k v_{\parallel} = 0
\]  

(4)

where the index \(\parallel\) is the component parallel to the magnetic field, \(s\) the harmonic number of the wave, and \(\Omega_e\) the relativistic electron gyrofrequency. Equation (4) describes the equality of wave and particle gyrofrequency in the Doppler-shifted frame of the electron. The term "maser" was given to this instability, since it generally occurs for electron distributions depleted of particles with low perpendicular velocity constituting a reversed population. The instability directly converts particle energy into radiation and is able to produce very high brightness temperatures. For this reason it was proposed as the emission process of spikes by Holman et al. (1980).

Melrose and Dulk (1982a) have worked out the details of the growth and energetics of the maser emission. The effects of the ambient plasma have been included by Sharma et al. (1982). Growth and escape of the various modes and harmonics have recently been discussed for coronal conditions by Sharma and Vlahos (1984) and for auroral kilometric radiation by Melrose et al. (1984). It seems that the maser mechanism operates only in strong magnetic fields \((\omega_p/\Omega_e < 0.9)\) and mainly emits on the fundamental \((s = 1)\). Then it may not only be a strong radio source but can even considerably heat the ambient medium and thus may redistribute the flare energy (Melrose and Dulk, 1984) or accelerate particles (Sprangle and Vlahos, 1983).

While a considerable effort has been made to theoretically understand the maser instability and interpret the high brightness temperature, very little has been done to explain other features of spikes. In a model discussed by Vlahos and Sharma (1984) the bandwidth of spike emission is given by the inhomogeneity of the magnetic field (equ. 3). The short duration of spikes has been interpreted by Li (1986) in terms of injection of small beams at skew angles and fast relaxation of the anisotropic electron distribution.

It is concluded that the emission mechanism is still unclear. Although cyclotron masering looks attractive, other possibilities are still open and should be investigated.

4.3. Spikes and the Flare Process

Since spikes appear during the primary energy release in flares, it is most interesting to view spikes in the general context of flares. The close agreement of the source density of spikes with the preflare density of flare loops as derived from soft X-rays suggests that spikes are emitted from a source close to the primary acceleration region. Unless some novel coherent radiation mechanism is at work the exciter of spikes must be fragmented into 10000 or more single elements. This is usually assumed for maser models. Then, the simplest assumption is that the flare energy, i.e. at least its part taken up by fast electrons, is released in ten thousands of elements (microflares). This scenario has been considered by Benz (1985). These flare elements may be the result of a global MHD instability of the flare region.

Each flare element should not be much larger than the spike source (less than 200 km in diameter). The acceleration process in such a region may be caused by a constant electric field or double layer (run-away, considered by Kuijpers et al., 1981), stochastic acceleration (Benz, 1984),
or small shocks (Vlahos and Sprangle, 1985). In any case, the extent of spike emission over more than an octave in frequency and the compactness of flare kernels suggests that the large range of different magnetic fields or densities is due to strong inhomogeneities.

The suggested fragmentation of flare energy release needs confirmation. It may be difficult to observe in hard X-rays, since most of the electrons may loose their energy far from the acceleration site at higher density. Also type III bursts may be the combined result of many elementary accelerations. They are well-known to often consist of superposed fine structures.

5. Conclusions

Spikes are an intriguing emission in the impulsive phase of flares. They are the most fragmented flare radiation consisting of ten thousands of individual elements. The most important question is whether this fragmentation is original or the result of a secondary process. If original, it will have a major impact on flare theory reducing flare time scales by several orders of magnitude.

The fact that spikes are not seen in every flare should not be overemphasized. Observations have mainly been done on frequencies below 1 GHz and only on a few single frequencies above. Many spike events have remained unobserved. Complete coverage from 0.3 to 5 GHz is urgently needed. In addition, propagation conditions in the source region may often prohibit the escape of spike radiation. This is particularly critical for cyclotron maser emission being strongly absorbed at skew angles to the magnetic field.

Future spike observations should be compared with other flare radiations. Spatial resolution and location in relation to the HXR flare are of great importance. Theoretical studies should proceed from merely considering the emission process to construction of models of spikes in the frame of current flare theory.

Acknowledgements

This work was partially supported by a grant from the Swiss National Science Foundation (No. 2.460-0.82).

References