CYCLOTRON MASER INSTABILITY AND ITS APPLICATIONS

C. S. Wu

Institute for Physical Science and Technology,
University of Maryland,
College Park, MD 20742, USA

Abstract

In view of the numerous possible applications of the cyclotron maser theory to a variety of radio sources, a discussion of the basic principle of the maser instability in simple terms is desirable. The present review is written for those who are interested in the topic but are not concerned with sophisticated theoretical details. The theory of the auroral kilometric radiation (AKR) proposed by Wu and Lee (1979) and subsequent developments are briefly summarized. Finally other possible applications of the maser instability are outlined.

1. Introduction

Solar system radio astronomical observations have made impressive progress and achievements during the last several decades. First the Sun was found as an intense radio emitter in the 1940s. Various types of sporadic radio bursts have been recorded and classified. (See review articles by Kundu, 1965, 1982; and Zheleznyakov, 1970). Subsequently, Jupiter's intense decametric radiation was discovered in the 1950s (Burke and Franklin, 1955a; Carr and Gulkis, 1969). Later the fascinating phenomenon of Io modulation was realized by Bigg (1964) which has stimulated a considerable amount of research effort. Ten years later, the Earth was identified as another strong emitter of the so–called auroral kilometric radiation (AKR) (Gurnett, 1974; Kaiser and Stone, 1975). More recently, observations of radio emissions from the planets Saturn and Uranus have also been reported (Kaiser et al., 1984; Warwick et al., 1981, 1986). The impressive recent achievements are attributed to high resolution and sophisticated measurements aboard several spacecraft such as the Radio Astronomy Explorer, Pioneer 10 and 11, Voyager, etc. .

However, despite of the rapid advances made on the observational side, the developments on the theoretical front are far slower. For instance, the mysteries of most of the well–known phenomena such as the metric solar radio bursts and Jovian decametric radiation are still not completely understood. Nevertheless, the gist of the current opinion is that spontaneous emission processes definitely cannot explain the intense nonthermal radiation and the phenomena must be attributed to induced emissions, a relatively new concept in radiation theory (Bekefi, 1966; Melrose, 1980).

Inspired by the auroral kilometric radiation (AKR) in the terrestrial magnetosphere which is generated in the high latitude auroral regions and correlated with inverted–V electron precipitation (Gurnett, 1978; Kaiser and Stone, 1975; Kurth et al., 1975; Benson and Calvert, 1979; Green et al., 1979), Wu and Lee (1979) proposed a theory which offers a novel and attractive explanation of the generation mechanism of AKR. The Wu–Lee model differs from the previous theories suggested by other authors in two important respects. First, the theory stresses the importance of mirror–reflected auroral electrons; whereas, other authors attribute the emission to the precipitating electrons. Second, the theory demonstrates the existence of an electromagnetic instability which can lead to the direct amplification of the extraordinary and ordinary mode radiation; whereas, the previous theories relied on multi–step processes which are generally inefficient, ineffective, and unconvincing. The Wu–Lee theory has been very well received in the scientific community and it is frequently called the cyclotron maser theory (Wu, 1985) in the literature.

In view of the growing interest and many other possible applications of the maser theory, this discussion intends to present a brief review of the subject, particularly for those readers who are not familiar with the theory but are interested in the essence of the maser mechanism. In the following presentation, emphasis will be placed on the basic physical principle expressed in simple terms rather than sophisticated theoretical details.

The organization of the paper is as follows: In Section 2, the basic principle of the cyclotron maser instability is presented. Then, the model proposed by Wu and Lee (1979) concerning AKR and subsequent developments are described in Section 3. Finally, a brief summary of the applications of the maser theory to cases other than AKR is given in Section 4.

2. Principle of cyclotron maser instability

In the article by Wu and Lee (1979), it is demonstrated that a kinetic instability exists when a population of suprathermal electrons possess a loss–cone distribution function. The instability theory has been subsequently studied extensively by a number of authors (Lee and Wu, 1980; Wu et al., 1981b; Dusenbery and Lyons, 1982; Hewitt et al., 1982, Wong et al., 1982; Winglee, 1983; Wu and Qiu, 1983; Le Quéau et al., 1984a). Although the sophisticated versions of the theory are complicated, the essence of the instability can be described in simple terms. The important point is that the inclusion of the relativistic effect on the usual cyclotron resonance condition is necessary because it can lead to a highly dramatic influence on the wave–particle interaction process. A succinct explanation is presented below.

For simplicity let us consider the case in which the energetic suprathermal electrons have a density lower than that of the background cold plasma and the radiation frequency is close to the electron cyclotron frequency Ω_e where $\Omega_e = |eB_0/mc|$. Radiation with this frequency range is very significant in a plasma in which the electron plasma frequency ω_{pe} is much lower than the cyclotron frequency Ω_e. If we write the radiation frequency $\omega = \omega_r + i\omega_i$ and assume that the dispersion relation $\omega_r = \omega_r(\tilde{k})$ is dictated by the background plasma, it can be shown from the linearized Vlasov theory that, when all perturbation quantities are proportional to $\exp[i\tilde{k} \cdot \tilde{r} - i\omega t]$, the quantity ω_i associated with the RX mode depends upon the integral (Wu and Lee, 1979)

$$\omega_i \sim \int d^3\tilde{u} \;\; u_\perp \frac{\partial F_e}{\partial u_\perp} \delta(\gamma\omega_r - \Omega_e - k_\parallel u_\parallel) \tag{1}$$

where $\tilde{u}$ denotes momentum per unit mass; F_e is the unperturbed distribution function of the suprathermal electrons; γ denotes the usual Lorentz factor, i.e. $\gamma = (1 + u^2/c^2)^{1/2}$; ω_r is the real part of the wave frequency; $u_\perp$ and $u_\parallel$ denote the component of $\tilde{u}$ perpendicular and parallel to the ambient magnetic field $\tilde{B}_0$; and $k_\parallel$ is the component of $\tilde{k}$ parallel to the ambient field. When the integral appearing in (1) is positive, waves are amplified. In that case ω_i is called a growth rate.

Physically, the delta function in the integrand signifies a resonant condition in the momentum space. It is instructive to point out that the resonance condition

$$\left(1 + \frac{u_2}{c^2}\right)^{1/2} \omega_r - \Omega_e - N \cos\theta \, \omega_r \frac{u_\parallel}{c} = 0 \tag{2}$$

is an ellipse if $N \cos\theta < 1$ where $N = kc/\omega_r$ is the usual refractive index and $\cos\theta = k_\parallel/k$. In $(u_\parallel, u_\perp)$ space this ellipse may be expressed as

$$\frac{u_\perp^2}{A^2} + \frac{(u_\parallel - u_0)^2}{B^2} = 1 \tag{3}$$

where

$$\begin{aligned}
\frac{u_0}{c} &= \frac{\Omega_e}{\omega_r} \frac{N\cos\theta}{(1 - N^2\cos^2\theta)} \\
\frac{A^2}{c^2} &= \frac{\Omega_e^2}{\omega_r^2} \frac{1}{(1 - N^2\cos^2\theta)} - 1 \\
\frac{B^2}{c^2} &= \frac{A^2}{c^2} \frac{1}{(1 - N^2\cos^2\theta)}
\end{aligned}$$

One important remark should be made at this point. Returning to (1) we see that, if the relativistic effect is neglected by setting $\gamma = 1$, the resonant condition reduces to

$$u_\parallel = \frac{\omega_r - \Omega_e}{k_\parallel}$$

which represents a straight line in the momentum space. In that case, regardless of the functional form of F_e, the integral is always negative and thus no instability occurs. However, the relativistic effect can qualitatively change the nature of the resonance condition. For instance, if the resonance ellipse is situated in the momentum space where $\partial F_e/\partial u_\perp > 0$, instability can occur. (This is a sufficient condition for instability.)

The preceding discussions explain why in the AKR theory by Wu and Lee (1979), even for electrons with energies of several keV, the relativistic effect in the stability analysis is vital.

Another relevant and very important point is that in the maser theory the ratio of ω_{pe}/Ω_e often appears as a critical parameter which may control the emission of the RX mode. This point can also be addressed, based on the resonance condition. In the following, a discussion is in order.

It is useful to introduce a polar coordinate system in the momentum space such that in a gyrotropic case we consider (u, ϕ) to replace $u_\parallel, u_\perp$. Here $u = (u_\parallel^2 + u_\perp^2)^{1/2}$ and $\phi = \tan^{-1} u_\perp/u_\parallel$.

In this representation the resonance velocity u which satisfies the equation

$$\left(1 + \frac{u^2}{c^2}\right)^{1/2} \omega_r - \Omega_e - N\cos\theta\, \omega_r \cos\phi \frac{u}{c} = 0$$

can be written as

$$\frac{u_{1,2}}{c} = \frac{N\cos\theta\cos\phi\,\frac{\Omega_e}{\omega_r} \pm \left(N^2\cos^2\theta\cos^2\phi + \frac{\Omega_e^2}{\omega_r^2} - 1\right)^{1/2}}{1 - N^2\cos^2\theta\cos^2\phi} \tag{4}$$

Thus, for a given ϕ there are two values of u when $\omega_r > \Omega_e$ and $N\cos\theta < 1$. However, when $\omega_r < \Omega_e$ and $N\cos\theta < 1$, u has only one meaningful root. (Remember, $u > 0$ is required.) The primary reason is that in this case, the resonance ellipse includes the origin.

Now let us return to the case in which $\omega_r > \Omega_e$. It is clear that for a given ω_r, N, and θ there exists a pitch angle ϕ such that

$$\cos^2\phi = \frac{1}{N^2\cos^2\theta}\left(1 - \frac{\Omega_e^2}{\omega_r^2}\right) \tag{5}$$

and u_1 and u_2 coincide into one value, i.e., a tangent is defined. Now we define $u_1 = u_2 = u_m$. From (4) and (5) we obtain

$$\begin{aligned} \frac{u_m}{c} &= \frac{N\cos\theta\cos\phi}{1 - N^2\cos^2\theta\cos^2\phi}\left(\frac{\Omega_e}{\omega_r}\right) \\ &= N\cos\theta\cos\phi\left(\frac{\omega_r}{\Omega_e}\right) \\ &= \left(\frac{\omega_r^2}{\Omega_e^2} - 1\right)^{1/2} \end{aligned} \tag{6}$$

Let us pay special attention to the RX mode which has a cutoff frequency $\omega_x \simeq (1 + \omega_{pe}^2/\Omega_e^2)\Omega_e$ when $\omega_{pe}^2 \ll \Omega_e^2$. Thus for $\omega_r > \omega_x$, the resonant wave frequency ω_r must satisfy the following condition

$$\frac{\omega_r - \Omega_e}{\Omega_e} > \frac{\omega_{pe}^2}{\Omega_e^2} \tag{7}$$

From (6), we can see that the momentum u_m at the point of tangency is

$$\frac{u_m}{c} = \frac{(\omega_r^2 - \Omega_e^2)^{1/2}}{\Omega_e} \gtrsim \sqrt{2}\frac{\omega_{pe}}{\Omega_e} \tag{8}$$

For purpose of illustration, let us consider the case in which the electron distribution function F_e has the form

$$F_e = A \exp\left(-\frac{u^2}{\alpha^2}\right) \qquad \textit{for} \;\; \phi \geq \phi_c \tag{9}$$

$$F_e = 0 \qquad \textit{for} \;\; \phi < \phi_c$$

where A is the normalization coefficient and ϕ_c denotes a loss–cone angle. In this case, we can readily see from (8) and (9) that if

$$\frac{u_m^2}{c^2} = 2\frac{\omega_{pe}^2}{\Omega_e^2} > \frac{\alpha^2}{c^2} \tag{10}$$

the population of resonant electrons is small and thus we anticipate that the growth rate may be insignificantly small. Hence, significant amplification can occur only if

$$\frac{\alpha^2}{c^2} \gtrsim 2\frac{\omega_{pe}^2}{\Omega_e^2} \;\; , \tag{11}$$

and (11) may be considered as a qualitative criterion for the amplification of the RX mode radiation. For example, if $m_e\alpha^2/2 \sim 10$ keV, we find $\omega_{pe}/\Omega_e \lesssim 0.14$ is required.

Of course, the preceding discussion has assumed that the background electrons dictate the dispersion relation of the RX mode waves. If the energetic electrons prevail, then the cutoff frequency ω_x of the RX mode can be lower than the gyrofrequency Ω_e (Wu et al., 1981b, 1982; Wong et al., 1982; Pritchett, 1984) and can lead to enhanced growth.

3. The theory of AKR

In this section we review the theoretical model of AKR *á la* Wu and Lee (1979). There are two important assumptions which form the foundation of the theory. These assumptions were made *a priori* at the time the model was proposed but were later justified by available observational results. In the following they are described separately.

Loss–cone distribution function of the reflected electrons

It is postulated by Wu and Lee (1979) that only a small fraction of the plasma–sheet electrons can precipitate while they are injected into the high–latitude polar region. The reason is that all electrons with pitch angles greater than the atmospheric loss–cone angle are expected to be reflected by the mirror effect associated with the convergent magnetic field lines. As a result, the upgoing auroral electrons are anticipated to possess a loss–cone distribution function which serves as a "pump" for the maser mechanism in our theory. This hypothesis was later confirmed by the S3–3 observations (Fennel et al., 1981). There are several characteristic features shown by the S3–3 measurements. First, overall the distribution is fairly isotropic outside the loss–cone region. Second, the loss–cone for the upcoming electrons is partially filled. Third, the loss–cone distribution appears to be modified by a parallel electric field beneath and above the point of observation. The loss–cone distribution function can give rise to a cyclotron maser instability and plays a primary role in the theory of AKR, as we have already discussed earlier.

Here we remark that, as pointed out by Dusenbery and Lyons (1982), Omidi et al. (1984), and Le Quéau et al. (1984a), the measured distribution functions often show a "hole" region for the downgoing electrons and a "bump" region associated with a population of trapped electrons. The "hole" region may be unstable for AKR but does not seem to be important, whereas the "bump" region may play a crucial role for the generation of the Z mode radiation, as shall be commented on later.

Depletion of low–energy electrons by the parallel electric field

Another hypothesis made by Wu and Lee (1979) is that the presence of a parallel electric field in the source region of AKR can significantly reduce the electron density, because most of the low–energy electrons (including the secondary electrons) are expected to be removed due to the potential drop. This postulation was later attested by the ISIS–1 observations. Benson and Calvert (1979), Benson et al. (1980), and Calvert (1981b) have shown that the predicted plasma cavity indeed occurs in the AKR source region.

The "density depletion" hypothesis was actually motivated by the fact that AKR correlates with the inverted–V events (Green et al., 1979). In the theory concerning the cyclotron maser instability, the ratio of the electron plasma frequency to the electron cyclotron frequency represents a crucial parameter. It turns out that only when this ratio is smaller than 0.3 the emission of AKR with frequencies close to the electron gyrofrequency is possible (Lee et al., 1980; Melrose et al., 1982; Hewitt et al., 1982; and Wu and Qiu, 1983). In short, the Wu–Lee model implies that the parallel electric field which is responsible for the inverted–V event can trigger the maser instability via the "density depletion" process.

The theory of AKR based on the maser mechanism may be summarized in three aspects.

A. Fundamental AKR

Since the article by Wu and Lee (1979) was published, a number of subsequent discussions have emerged (Lee et al., 1980; Dusenbery and Lyons, 1982; Melrose et al., 1982; Omidi and Gurnett, 1982; Wu et al., 1982; Le Quéau et al., 1984a,b; Omidi et al., 1984; Winglee, 1985). These discussions have further elaborated and improved on the cyclotron maser theory of AKR.

There are two important points which deserve special attention. First, Omidi and Gurnett (1982) and Omidi et al. (1984) have numerically calculated the growth rate of the maser instability suggested by Wu and Lee (1979) by utilizing a distribution function measured by the S3–3 satellite rather than an idealized theoretical model. The results show that, indeed, the distribution function is unstable and can give rise to the emission of AKR in the RX and LO modes. Second, it has been pointed out by Dusenbery and Lyons (1982), Omidi et al. (1984), and Le Quéau et al. (1984a) that the actually measured distribution functions indicate other sources of free energy which may also contribute to the emission of AKR, particularly the O mode and Z mode.

B. Second harmonic AKR

The topic of second harmonic AKR has stimulated much discussion in recent years. Benson (1982, 1984, 1985) has cited a number of cases observed by the ISIS–1 spacecraft. These cases indicate convincingly that the harmonic AKR are very likely of natural origin.

Theoretically, the second harmonic AKR has been investigated by several authors (Lee and Wu, 1980; Lee et al. 1980; Melrose et al., 1982; Hewitt et al., 1982; and Wu and Qiu, 1983). These authors have calculated the growth rate for the second harmonic AKR within the context of the cyclotron maser theory. The consensus is that, although the emission can indeed occur, the growth rate is approximately one order of magnitude smaller than that of the fundamental AKR. "Why do the observed second harmonic AKR emissions appear very effective although weaker than the fundamental AKR?" appears to be an interesting question for theorists.

One possible answer to this question is that spontaneous emission of the R–X mode at the second harmonic of the cyclotron frequency (Freund et al., 1984) is comparatively much higher than that at the fundamental when the wave normal angle θ is sufficiently large, say greater than 70°. This result implies that the induced emission of the second harmonic AKR can be significant although the growth rate is small.

C. Generation of Z mode radiation

A full discussion of the DE–1 observations of the Z mode radiation is given by Gurnett et al. (1983a). Omidi et al. (1984) and Omidi and Wu (1985) have demonstrated that Z mode radiation with perpendicular propagation (i.e., $\theta = 90°$) can have a sizable growth rate which is attributed to several sources of free energy; the loss–cone, the "hole", and the "bump". Moreover, the unstable waves have frequencies below the gyrofrequency Ω_e in the source region. Both the feature $\theta - 90°$ and the feature $\omega_r < \Omega_e$ are highly consistent with the DE–1 observations reported by Gurnett et al. (1983a). Thus the cyclotron maser mechanism appears rather plausible for the generation of the Z mode radiation observed along the auroral field lines. More recently Lin et al. (1986) have discussed that "hole" distribution may also give rise to the excitation of the Z mode.

To summarize the current understanding of the generation of various modes of radiation and their principal sources of free energy, we present the following table.

Free – Energy Sources of AKR and Z mode Radiation

Radiation Mode	Propagation Direction	Principal Free–Energy Sources
Fundamental AKR		
RX mode	Upward	Loss–cone and "bump"
LO mode	$\theta - 90°$	"Bump", loss–cone, and "hole"
Second Harmonic AKR		
RX mode	$\theta - 90°$	"Bump", loss–cone, and "hole"
Z mode	$\theta - 90°$	"Bump", loss–cone, and "hole"
	Downward	"Hole"

4. Other applications

The cyclotron maser theory of AKR is attractive and appealing in several respects. First, it provides a direct and efficient amplification process. The required condition $\partial F_e / \partial u_\perp > 0$ in certain domains in the momentum space may be satisfied in many regions within or beyond the solar system. Second, the basic theory is simple and unsophisticated. Third, the instability can be attested to by measured distribution functions (Omidi and Gurnett, 1982; Omidi et al., 1984).

The successful discussion of AKR has stimulated considerable interest. Numerous possible applications of the cyclotron maser mechanism to other radio emission processes have been suggested or speculated. It has been proposed that the microwave spike bursts often observed during solar flares may be attributable to the cyclotron maser instability (Holman et al., 1980; Sharma et al., 1982; Willson, 1985). The cyclotron maser mechanism has also been suggested as being responsible for Jupiter's decametric radiation (Hewitt et al., 1981, 1982) and the broad band kilometric radiation (Leblanc and Daigne, 1988). Recently the kilometric radio emission from Uranus has attracted considerable theoretical interest. Models concerning Uranus kilometric radiation in the context of the cyclotron maser theory have been presented by Curtis (1985), Curtis et al. (1987), and Desch and Kaiser (1987).

In astrophysics literature the cyclotron maser instability has been thought to be responsible for the radiation bursts from flare stars and close binaries (Dulk et al., 1983; Gary et al., 1982; Lang and Willson, 1986; Bastian and Bookbinder, 1987). These aforementioned discussions represent a growing literature on the subject and indicate clearly that the maser mechanism can be pervasive.

Up to this date, the cyclotron maser mechanism is the only induced emission process which has been definitely identified in a natural plasma. Of course, this does not mean that it is the only induced emission process operative. We believe that there is still much room for further theoretical development in the subject area of radio astronomy.

Acknowledgements. This work was supported by grants from the National Aeronautics and Space Administration and the Office of Naval Research.