Mon. Not. R. astr. Soc. (1978) **183**, 491–500

Time variability of astrophysical masers

M. Salem and M. S. Middleton *Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE*

Received 1977 October 21; in original form 1977 February 11

Summary. The time-dependent equations of radiative transfer are solved for a maser source much longer than it is wide. Transitions are assumed to occur between two energy levels only (this has previously been shown to be a justifiable approximation for the four-level hydroxyl (OH) maser, and is discussed briefly). It is found that the intensity may vary rapidly although the physical conditions in the masing region, including the rates of pumping and relaxation, are constant. Under certain circumstances the observed intensity of a saturated maser may vary quasi-periodically with large amplitude and a period (in years) equal to the maser length (in light years); the maser may seem to appear and disappear. All calculations are carried out for two of the four levels of the hydroxyl ground rotational-state maser, but the results are qualitatively valid for any system which is approximately equivalent to a two-level maser.

Introduction

The intensities of OH and H_2O masers have been observed to vary on a timescale ranging from days to months. The variations could be due to changes in the masing region or to the process of radiative transfer. In the former case, we might expect to find correlated variations in other radiation from the same region as was, indeed, suggested and found for infrared, OH and H_2O variability in several sources (Schwartz, Harvey & Barrett 1974). The variability in the maser can then be attributed to changes in the pumping mechanism. In several other cases no correlation has been found between variations of the maser radiation and of other observables (Coles, Rumsey & Welch 1968; Buhl *et al.* 1969; Rydbeck, Kollberg & Ellder 1970; Zuckerman *et al.* 1972; Lo & Bechis 1974; Rickard, Zuckerman & Palmer 1975; Sullivan & Kerstholt 1976; Gruber & de Jager 1976).

The observed variations may be due to changes in the number density of the hydroxyl (OH) molecules involved in the maser, in the pumping mechanism, or in the direction in which the maser is pointing. Another possibility discussed in detail below, is that a single rapid change in the physical conditions can cause quasi-periodic variations of the observed brightness temperature of a saturated maser on a timescale much longer than the original perturbation. Conversely, a maser which can be shown to vary due to the mechanism described in this paper must be saturated.

The calculations

In the present work, numerical calculations have been carried out only for transitions between sublevels of the ground rotational state of hydroxyl, although the theory is valid for any system equivalent to a two-level maser. The rotational states of OH are split by Λ-doubling, and again by hyperfine splitting, giving four sublevels in all. Four transitions are allowed between sublevels of this quartet (discussed in detail by ter Haar & Pelling 1974). Only one of the four possible transitions is observed from any given position on a source (Harvey *et al.* 1974). Cook (1975b, private communication) showed that if only one transition is masing, then the four-level maser equations may be equivalent to those describing a two-level maser, with parameters which are complicated functions of the original four-level parameters. Accordingly, we solve the equations of radiative transfer for a two-level maser. As there is no reliable information on physical conditions in OH masers, we have carried out calculations for several sets of parameters.

We assume that the maser is much longer than it is wide; that the phsyical conditions depend only upon the distance along the maser and that stimulated emission occurs into a constant, small solid angle. We shall speak of a constant-angle maser, which differs from the tubular maser of Goldreich & Keeley (1972) in that the maser is not restricted to the interior of a cylinder. While we consider the constant solid-angle model to be the most appopriate, we do not expect the general results of our work to depend critically upon the model chosen.

The time-dependent equations of radiative transfer and the energy-level population equations for a constant-angle maser as formulated by Cook (1969) can be written in the form:

$$\frac{\partial I^+}{\partial x}+\frac{1}{c}\frac{\partial I^+}{\partial t}=(b_{12}+b_{21})\,\delta n I^+ + a_{21}\delta n + a_{21}n_2^0, \tag{1}$$

$$-\frac{\partial I^-}{\partial x}+\frac{1}{c}\frac{\partial I^-}{\partial t}=(b_{12}+b_{21})\,\delta n I^- + a_{21}\delta n + a_{21}n_2^0, \tag{2}$$

$$\begin{aligned}\frac{d\delta n(t;x)}{dt} &= -\left[A_{21}+p_{12}/\alpha+(\beta_{21}+\beta_{12})\left(I^+ + I^- + \frac{4\pi}{\delta\omega}I^{\mathrm{bck}}\right)\right]\delta n\\ &\quad - A_{21}n_2 + p_{12}n/[1+(g_2/g_1)],\end{aligned} \tag{3}$$

$$I^+(0,t)=I_0^+=\text{constant},$$

$$I^-(l,t)=I_l^-=\text{constant},$$

$$\delta n(0;x)=\text{given function of } x.$$

In the above equations, t is time, x is the distance coordinate along the maser, measured from the point furthest from the observer, l is the length of the maser, subscripts 1 and 2 refer to the lower and upper levels, $I^+(x,t)$ is the intensity per unit bandwidth per unit solid angle (specific intensity) propagating in the direction of increasing x (Joules $\mathrm{s^{-1}\,m^{-2}\,ster^{-1}\,Hz^{-1}}$), $I^-(x,t)$ is the specific intensity propagating in the opposite direction. I^{bck} is the background specific intensity (assumed isotropic), g_1, g_2 are the statistical weights of levels 1 and

2, n is the total number density of molecules involved in the maser, n_2 is the population of level 2, in molecule/m^3,

$$n_2^0 = n(g_2/g_1)/[1 + (g_2/g_1)], \tag{4}$$

$$\delta n(t; x) = n_2 - n_2^0, \tag{5}$$

c is the velocity of light, A_{21} is the Einstein coefficient for spontaneous emission, $\delta\omega = \pi(\text{radius})^2/(\text{length})^2$ is the (constant) solid angle into which stimulated emission occurs,

$$a_{21} = h\nu A_{21}/4\pi\Delta\nu, \tag{6}$$

$$b_{21} = c^2 A_{21}/\nu^2 8\pi\Delta\nu = b_{12}g_2/g_1, \tag{7}$$

$$\beta_{21} = \delta\omega c^2 A_{21}/8\pi h\nu^3 = \beta_{12}g_2/g_1. \tag{8}$$

$p_{12} = C_{12} - C_{21}g_2/g_1$ is the net pump rate, i.e. the net rate coefficient for non-radiative transitions from level 1 to level 2, C_{ij} is the rate coefficient for transitions from level i to level j for all processes except direct radiative transitions (intermediate states will in general be involved); $\alpha = p_{12}/(C_{12} + C_{21})$ is the ratio of the net pump rate to the total non-radiative transition rate, and takes values from almost zero (downward relaxation, but no pumping) to 1 (the only downward path is via the masing transition), ν is the frequency of the transition, $\Delta\nu$ is the width of the line and h is Planck's constant. We emphasize that C_{ij} is more general than the rate coefficient for collisional excitations or de-excitations from level i to level j.

It is sometimes convenient to express specific intensities in terms of the brightness temperature T_b, defined at radio frequencies by $I = 2kT_\mathrm{b}/\lambda^2$. We may speak of $T_\mathrm{b}^+(x, t)$, $T_\mathrm{b}^-(x, t)$, T_0^+, T_l^- and T^bck.

The intensities and populations are functions of position and time, but we assume that the level populations at a given position do not depend explicitly upon the populations at other positions; in other words we assume that no mixing occurs along the maser. We also assume that the sum of the population densities of the two levels is constant and equal to the total hydroxyl population at all points along the maser. These assumptions and that of constant solid angle are not essential to the theory and should not affect the qualitative form of the results.

The principal difficulty in solving equations (1)–(3) lies in the very great increase in intensity along the length of the maser and in time – typically ten orders of magnitude. Additionally, the major part of the increase may occur over a very small interval in space or time, a property that makes an approach based on straightforward integration over a uniform grid practically impossible, as a vast grid would be required. The difficulty can be overcome by integrating along the characteristic lines of the transfer equations, $x \pm ct = 0$, and taking into account the exponential behaviour of the intensities. Details of the methods used will be published later.

A computer program was written which solved the system of equations (1)–(3) numerically. The following checks of the numerical results were made for a case with extremely rapid changes of intensity:

(a) Calculations were carried out using 129 and 65 points. All intensities agree to within 1 per cent.

(b) Calculations were carried out using IBM single (32 bit) and double (64 bit) precision, using 129 points. All intensities agreed to within 1 per cent.

(c) The program was allowed to run until the intensities were constant in time along most of the length of the maser. These intensities were then compared with steady-state intensities calculated by numerical solution of the time-independent equations (to be published). These entirely different approaches agreed to within 1 per cent.

(d) Calculations were carried out for a limited number of cases using a conventional integration method. Agreement to within 1 per cent was obtained, but at great expense in computer time.

Results

For any two-directional maser at any time, there will be a point at which the forward- and backward-going intensities will be equal. In the vicinity of this point neither of the beams will be saturated. Nearer the ends of the maser the intensity of one of the beams will dominate, typically by many orders of magnitude. The dominant beam at each end may become saturated. Fig. 1 shows the intensities along a saturated steady-state maser as functions of position. This result is in agreement with the results of Goldreich & Keeley (1972) for tubular and spherical masers. In the case of a saturated maser we speak of an unsaturated core. This core is always present, although it may occupy only a small fraction of the maser length.

In the unsaturated core, the intensity increases exponentially with position; the increase is linear where the maser is saturated. The mechanism for time variability discussed below occurs only in saturated masers; unsaturated masers will accordingly not be discussed further. Due to the very large uncertainties in the physical parameters, we feel that quantitative fitting to observations is premature, but we do note that brightness temperatures similar

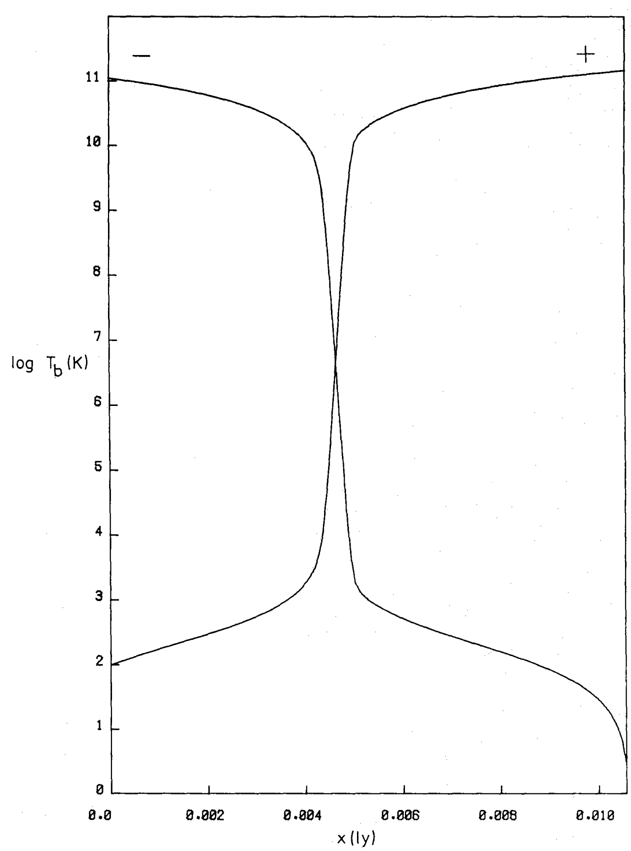

Figure 1. $T_b^+(x, t_0)$ and $T_b^-(x, t_0)$ for the 1.665-GHz OH maser of Fig. 2, at a time t_0 shortly after the output intensity has stabilized. Compare with figure 1 of Litvak (1971).

to those observed are obtained with what we consider to be not unreasonable values. The qualitative nature of the solutions is, however, interesting. We shall give a few examples with arbitrary values of the parameters; qualitatively similar results are obtained for a very wide range of values. In the following cases the initial value of δn is zero everywhere. All the cases studied below become saturated along at least part of their length; the output intensities are consequently only very weakly dependent upon the input intensities. In the entirely unsaturated case, the intensities increase exponentially with position; the forward- and backward-going beams are not coupled to each other or to variations in the energy-level populations.

The following types of behaviour were found:

(a) The most striking phenomenon is a quasi-periodical behaviour for sufficiently large values of the length, pumping rate and density. The period, in years, is equal to the length of the maser in light years. The output intensity is plotted as a function of time for a typical case (Fig. 2). Note that the amplitude of the oscillation is initially several orders of magnitude – the source would appear to switch on and off and then to vary with decreasing amplitude, finally reaching a steady state. Oscillations of smaller amplitude can, of course, occur. The maximum intensity decreases slightly from one cycle to the next. The cause of the oscillatory behaviour is clear – the intensity I^+ increases so far as to reduce the inversion along at least part of the maser very considerably; the intensity will then drop until the inversion is restored by the pump. This type of behaviour is characteristic of a saturated maser: any sharp change which tends to increase the intensity (e.g. switching on the pump)

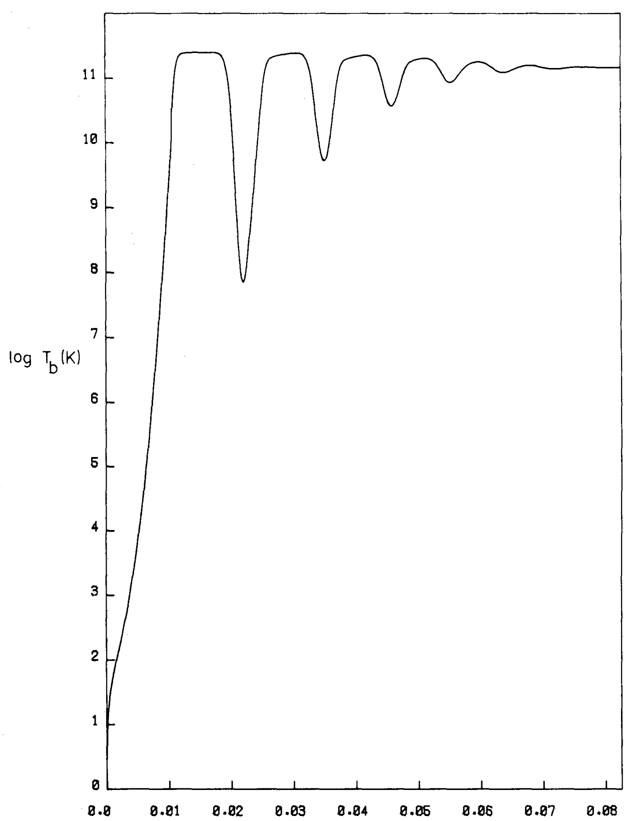

Figure 2. Output brightness temperature as a function of time in light years $T_b^+(l, t)$, for model of the 1.665-GHz OH maser. Physical parameters: $n = 5 \times 10^6\,\mathrm{m}^{-3}$; $l = 10^{14}\,\mathrm{m}$; $\delta\omega = 10^{-6}\,\pi$ ster (equivalent to observed width of $10^{11}\,\mathrm{m}$); $p_{12} = 10^{-8}/\mathrm{s}$; $\alpha = 1$ (no relaxation); $T_0^+ = 100\,\mathrm{K}$; $T_l^- = 3\,\mathrm{K}$; $T^{\mathrm{bck}} = 3\,\mathrm{K}$; $\Delta\nu = 4\,\mathrm{kHz}$. 129 integration points.

may cause oscillations of the output intensity. Fig. 1 shows the intensities of the beams as functions of position after the output intensity has stabilized; subsequently the boundaries between the unsaturated core and the saturated regions continue to become sharper.

(b) The previous result corresponds to $\alpha = 1$, i.e. molecules can leave the upper level only by spontaneous or stimulated transitions to the lower level. As α decreases (downward non-radiative relaxation increases), the length of the unsaturated core increases. The behaviour of the output intensity is changed only slightly so long as there is a saturated portion near the end of the maser. In particular, the oscillations do not die out more rapidly, although the first minima become shallower. Eventually a critical value of α is reached, for which the whole maser becomes unsaturated; further decrease causes the output intensity to decrease. The output intensity of the now unsaturated maser does not, of course, oscillate. The critical value of α depends upon the extent of the unsaturated core for $\alpha = 1$; the smaller the core in the absence of relaxation, the greater the change in α required to unsaturate the maser along its entire length. In the case plotted in Fig. 3, the behaviour of the output intensity is little changed by a reduction of α from 1 to 5×10^{-3} (an increase of C_{21} from 0 to $100 p_{12}$).

(c) If the length, pumping rate and density are too low to allow the intensity to build up to the saturation point, the intensity increases steadily and monotonically to the equilibrium value. The maser is always unsaturated along its entire length.

(d) If the pumping rate varies, say sinusoidally, the output intensity will be modulated similarly. No resonant phenomena have been found for pumping-rate frequencies harmonically related to the frequency of oscillation of the maser output (Fig. 4). The amplitude of the oscillations of the output intensity increases with the period of the modulating function.

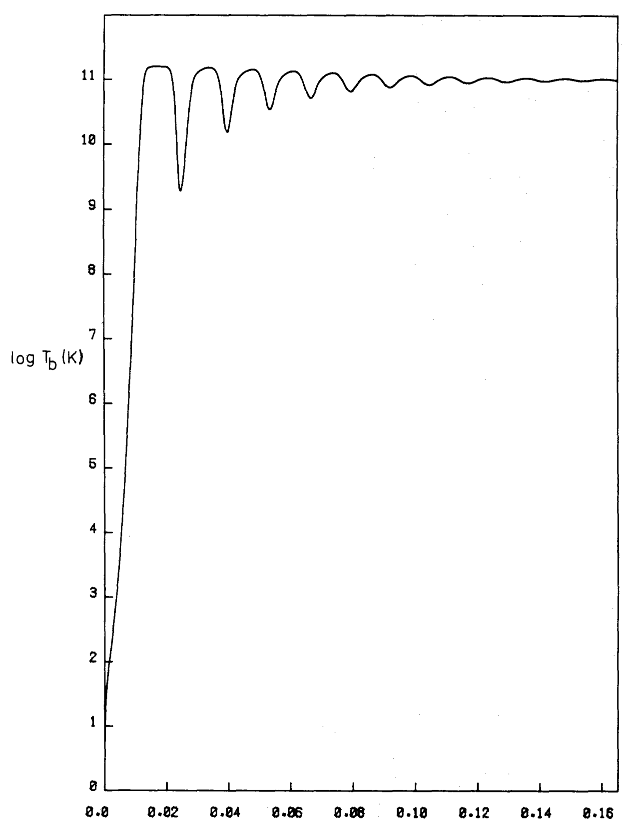

Figure 3. As Fig. 2, but $\alpha = 5 \times 10^{-3}$ ($C_{21} = 100 p_{12}$). Note different scale.

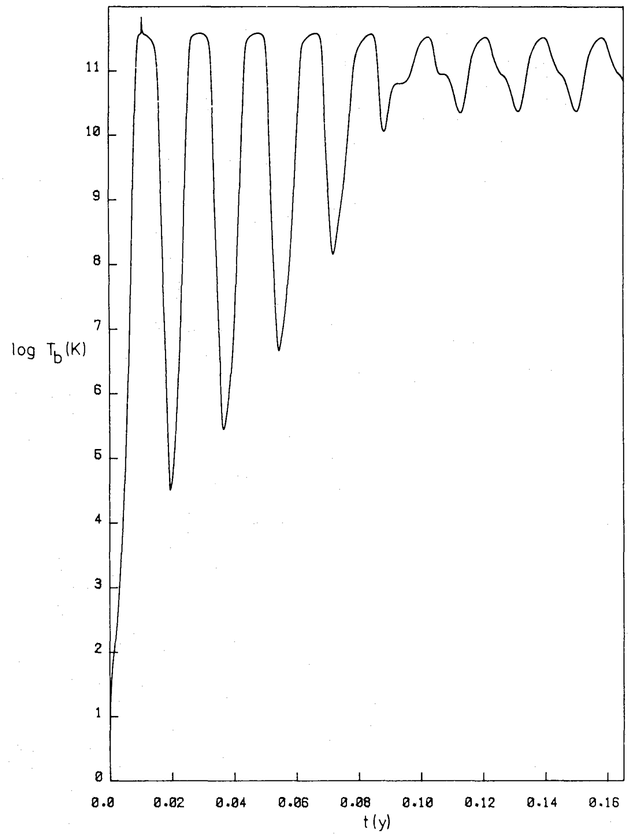

Figure 4. As Fig. 2, but p_{12} varies sinusoidally with amplitude 10^{-8}/s around the mean value 10^{-8}/s, with wavelength = 1.77 maser length. Pip on curve corresponds to maser length (light years). Scale as Fig. 3.

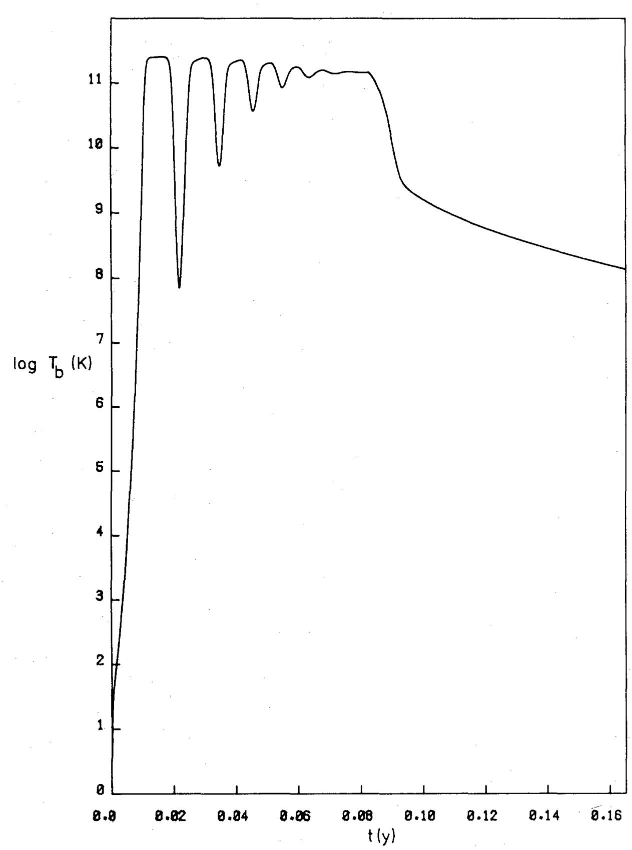

Figure 5. As Fig. 2, but pump is switched off ($p_{12} = 0$) at $t = 0.083$ yr. Scale as Fig. 3.

(e) If there is a significant degree of population inversion before the conditions for maser amplification are satisfied (a 'prepumped' maser), the intensity rises sharply to a value which may be very much larger than the final steady-state value and then drops rapidly. Subsequent behaviour follows the patterns described above.

Discussion

The physical problem of the OH maser must, of necessity, be simplified for mathematical treatment to be possible. Several questions can be raised.

How does a maser start? Do molecules with a significantly inverted population exist beforehand and then become aligned in velocity (a 'prepumped' maser), or does the velocity alignment occur before significant inversion has taken place? One would expect an intermediate situation to be the best approximation to reality. Do inhomogeneities and fluctuations in physical quantities play a significant part?

We have simplified the OH problem by assuming that only one transition dominates, excluding the other three. All maser sources which have been interferometrically observed in more than one of the OH lines do indeed radiate in one line only. Are the other three lines eliminated by some physical filtering mechanism, or has one of the lines simply become dominant during the evolution of the source? Cook (1975a) has shown that if the velocity of OH in a region is correlated with the magnetic field, one Zeeman component of one transition could be selected and amplification of lines from all other transitions would be excluded. The lines would be polarized, as is observed. If such a filtering effect exists, then the two-level approximation should be sufficient to explain observed behaviour. If, on the other hand, two or more transitions can be amplified significantly during the evolution of the source (with one eventually becoming dominant), it is not correct to use a two-level model with the parameters of the transition which is observed. It was shown by Cook (1975b) that a four-level maser in which one transition dominates is formally equivalent to a two-level maser, but the coefficients that appear are complicated functions of the four-level ones; in fact the equations become integro-differential. Numerical solution of the four-level equations for a few cases has shown no significant departure from the types of behaviour described above.

Whatever the answers to the questions formulated in this section, the results obtained are qualitatively meaningful. Although the calculations were carried out for the ground rotational-state quartet of OH, similar conclusions hold for other systems equivalent to two-level masers, in particular for transitions between the sublevels of rotationally excited levels of OH, and for the water and silicon monoxide masers.

Two extreme types of variability can be distinguished. If the physical conditions in a masing region change slowly, the time variability will reflect the change in physical conditions (quasi-static case). If the conditions change rapidly, the behaviour of the maser will initially be determined by the process of radiative transfer (transient case) and will slowly become dominated by the new physical conditions. By 'slowly (rapidly)' we mean 'with a timescale in years much longer (shorter) than the maser length in light years'.

The possible types of transient behaviour are: monotonic increase or decrease to a steady state, or oscillatory increase to a steady state. It may be difficult to distinguish the monotonic transient modes from quasi-static or intermediate behaviour. The oscillatory increase is distinctive and is specific to saturated masers.

It is clear that transients can only occur if the physical conditions in a large part of the maser change rapidly. Let us consider the most favourable case of some disturbance (a shock wave say) which propagates in a direction perpendicular to the maser axis. A transient will occur if the disturbance propagates across the width of the maser in a time not much longer than that required for light to travel the length of the maser. In other words, the disturbance must travel at a velocity $v \gtrsim c\rho/l$ perpendicular to the maser axis, where ρ is the maser diameter. The relative width ρ/l is unlikely to be less than 10^{-3}, requiring speeds of not less than 300 km/s. Such speeds appear unrealistically high for matter in a molecular cloud; only a radiation field can propagate fast enough.

Where transient behaviour is unequivocally observed, the physical conditions in the region must have been changed by an abrupt variation in a radiation field. We stress that a single sharp variation can cause many oscillations – it is not necessary to postulate a continually variable cause. When slower (quasi-static) changes are observed, the timescale may give information on the velocity of propagation of a shock front or other disturbance moving at right angles to the axis of the maser, if an estimate of the diameter of the maser is available.

Very rapid variations have been observed. Zuckerman *et al.* (1972) observed a group of features of the $^2\Pi_{3/2}$, $J = 5/2$, $F = 3 \rightarrow 3$ maser in NGC 6334N which was not detectable on 1970 June 2 ($T_A < 0.3$ K), was strong on June 4 ($T_A = 4$ K), and was no longer detectable on June 5. Rickard *et al.* (1975) looked for, but did not detect this group of rapidly varying features in 1972 March. They did find a feature of the same transition in Sgr B2 which varied in intensity by an order of magnitude in 24 hr.

The state of saturation of the maser strongly influences the sensitivity to changes in physical conditions. Broadly speaking, the output of a maser which is saturated along a portion of its length will change in proportion to any change in parameters along the saturated portion, being insensitive to changes in the unsaturated part. A maser which is unsaturated along its entire length will respond much more strongly; if the unperturbed gain is e^G, a change in conditions along the entire length of the maser will produce a proportional change in G, and a very much larger relative change in the gain and hence in the intensity. If G is say 25, a 3 per cent change to 25.7 will double the output intensity. The transient oscillations described above are positive proof of saturation. The question of saturation of OH masers cannot be considered to be solved, although there is some evidence that observed masers may be saturated (Davies, Booth & Perbet 1977).

Movement of the maser (rotation, etc) is another possible cause of rapid changes of intensity at Earth, which we shall not consider further here.

Acknowledgments

We thank Professor A. H. Cook for suggesting this work and for many helpful discussions. It is also a pleasure to thank Professor R. D. Davies for a useful discussion. This work was supported by a grant from the SRC (MS) and by a Clerk Maxwell Scholarship from the University of Cambridge (MSM).

References

Buhl, D., Snyder, L. E., Schwartz, P. R. & Barrett, A. H., 1969. *Astrophys. J.*, **158**, L97.
Coles, W. A., Rumsey, V. H. & Welch, W. J., 1968. *Astrophys. J.*, **154**, L61.
Cook, A. H., 1969. *Physica*, **41**, 1.
Cook, A. H., 1975a. *Mon. Not. R. astr. Soc.*, **171**, 605.
Davies, R. D., Booth, R. S. & Perbet, J.-N., 1977. *Mon. Not. R. astr. Soc.*, **181**, 83.
Goldreich, P. & Keeley, D. A., 1972. *Astrophys. J.*, **174**, 517.
Gruber, G. M. & de Jager, G., 1976. *Astr. Astrophys.*, **50**, 313.

Harvey, P. J., Booth, R. S., Davies, R. D., Whittet, D. C. B. & McLoughlin, W., 1974. *Mon. Not. R. astr. Soc.*, **169**, 545.
Litvak, M. M., 1971. *Astrophys. J.*, **170**, 71.
Lo, K. Y. & Bechis, K. P., 1974. *Astrophys. J.*, **190**, L125.
Rickard, L. J., Zuckerman, B. & Palmer, P., 1975. *Astrophys. J.*, **200**, 6.
Rydbeck, D. E. H., Kollberg, E. & Ellder, J., 1970. *Astrophys. J.*, **161**, L25.
Schwartz, P. R., Harvey, P. M. & Barrett, A. H., 1974. *Astrophys. J.*, **187**, 491.
Sullivan, W. T. & Kerstholt, J. H., 1976. *Astr. Astrophys.*, **51**, 427.
ter Haar, D. & Pelling, M. A., 1974. *Rep. Prog. Phys.*, **37**, 481.
Zuckerman, B., Yen, J. L., Gottlieb, C. A. & Palmer, P., 1972. *Astrophys. J.*, **177**, 59.

Note added in proof

The examples discussed in detail in this paper all involved drastic changes in a region containing OH. The mechanism causing the changes is equally applicable to other masers (including specifically water and silicon monoxide) and to less drastic changes – any rapid increase in the rate of population inversion may cause oscillations of smaller amplitude than the cases considered above.