Physical characteristics of astronomical masers

Moshe Elitzur

Department of Physics and Astronomy, University of Kentucky, Lexington, Kentucky 40506

The radio line emission of interstellar molecules routinely shows deviations from thermal equilibrium which culminate with strong maser radiation in some sources. Like its laboratory counterpart, the maser radiation is amplified through the effect of induced processes in a region where a population inversion exists and displays many of the basic features of laboratory lasers. This review is intended to explain the basic theory of astronomical masers and to survey the theoretical models which were developed for specific sources.

CONTENTS

I. INTRODUCTION

A. Overview

The hydroxyl radical (OH) was the first interstellar molecule detected at radio frequencies (Weinreb *et al.*, 1963). Each of its rotation states is split into four levels by Λ doubling and hyperfine interactions (Fig. 1). The resulting allowed transitions within the ground state lead to four radio lines at wavelengths of approximately 18 cm. The simultaneous detection of all four lines provides an idea of the population distribution among the four ground-state levels. It soon became evident that the populations were almost never in thermal equilibrium, leading to anomalous emission patterns. Soon thereafter it was found that some OH sources were so intense that their output could not be explained by noncoherent spontaneous emission and an explanation based on maser amplification through induced processes had to be invoked.

The first substantial wave of discoveries of radio lines of interstellar molecules occurred in the years 1968–69. One of the newly detected molecules was water vapor (H_2O), discovered by Cheung *et al.* (1969). It was detected in a radio transition between two excited rotation states which are accidentally close in energy (Fig. 2). This H_2O radiation always involves maser action and provides the most spectacular example of the effect. One of the water maser sources (W49) emits almost a full solar luminosity in a single line with a width of only 50 kHz.

The next strong maser effect was discovered in emission from the SiO molecule. This is a unique maser, occurring in rotational transitions of excited vibration states. The first line was discovered in the $J=2\rightarrow1$ tran-

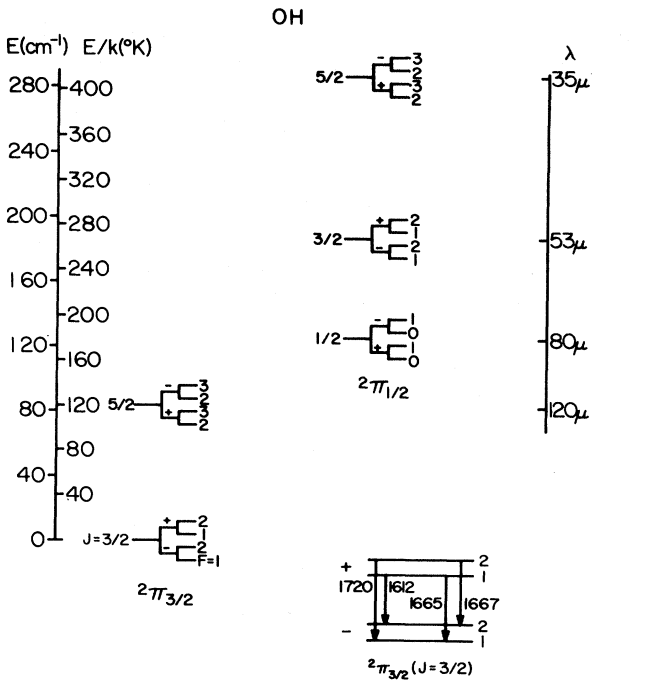

FIG. 1. The rotation levels of OH which couple radiatively to the ground state $^2\Pi_{3/2}(J=3/2)$. F is the total angular momentum, including nuclear spin. The splitting of each rotation state is not to scale. The ground state is plotted on an expanded scale at the right-hand corner with frequencies (in MHz) marked on the arrows. The F-conserving transitons (at 1665 and 1667 MHz) are called "main" lines, and the F-changing ones (at 1612 and 1720 MHz) are "satellite" lines. The wavelengths of the ground-state lines are approximately 18 cm. The scale on the right corresponds to wavelengths for transitions to the ground state.

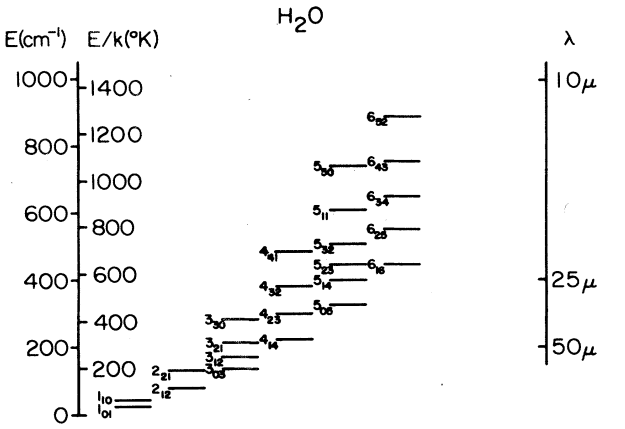

FIG. 2. Excited rotation levels of ortho H_2O with energies lower than 1000 cm^{-1} and $J \leq 6$. The notation is $J_{K_+K_-}$, where J is the total angular momentum and K_+ and K_- are its projections on two molecular axes. The maser transition is $6_{16} \rightarrow 5_{23}$. It has a frequency of about 22 GHz (wavelength of approximately 1.35 cm).

sition of the $v = 1$ vibration state (Snyder and Buhl, 1974), followed by discoveries of a large number of other lines (Fig. 3).

Although the OH, H_2O, and SiO are the only strong masers discovered to date, a large number of other molecules display weak maser effects in some of their lines. It is therefore evident that maser action occurs rather easily in interstellar space. The recognition of this fact led to a general disappointment with the maser phenomenon. Reading through the early literature on the subject, one senses a prevalent feeling that the maser radiation is such a spectacular effect that it alone should teach us a great deal about the emitting sources. The backlash of this unfulfilled anticipation, when it was realized that population

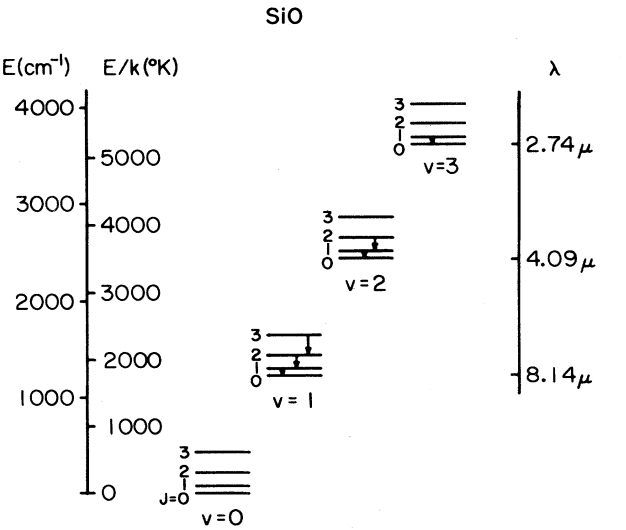

FIG. 3. The vibration levels of SiO with $v \leq 3$. For each level the rotation states with angular momentum $J \leq 3$ are plotted. The rotation energy separations were increased by a factor of 50. Transitions in which maser emission has been detected are marked with arrows.

inversion could be easily produced, led to the opposite extreme view that nothing can actually be learned from the masers.

A balanced approach has begun to emerge only recently. The mere fact that a particular source emits maser radiation cannot in itself provide too much useful information about it. However, in conjunction with other observations, maser radiation becomes a rather powerful tool for studying the emitting source. Some maser models are already detailed enough to make quantitative predictions which have been successfully verified by observations. The maser radiation probes regions in these sources and provides useful information about them unavailable through any other type of observation. By a lucky coincidence, maser action is associated with star formation and with late stages in the life of a star. The early and late stages of stellar evolution are currently considered the most interesting topics of research in this field, and both are probed by maser radiation.

Astronomical masers operate, of course, on the same principles as laboratory lasers, but are much simpler because neither a resonant cavity nor reflecting mirrors are involved. Using laboratory jargon they can be described as single-pass lossless gaseous lasers without feedback. It is evident that in order to make a useful tool out of them we must first realize what can and, sometimes even more important, what cannot be learned from maser radiation. The first part of this review, Secs. I through III, is therefore devoted to a detailed discussion of maser theory with this objective in mind. The presentation in these sections is rather detailed, with all assumptions spelled out clearly and all steps worked out. This approach was motivated partly by the surprisingly large number of mistakes permeating the literature, including very recent papers, and partly by the fact that a comprehensive summary of the basic theory is simply not available.

The second part of the review discusses the details of source models for the three strong maser molecules. Weak masers by and large are omitted, although some particular examples are occasionally discussed. This choice was made because it is much more difficult to pin down with reasonable certainty the conditions in these sources; there is usually a large degree of freedom in producing weak effects. The presentation in this part is much less extensive than in the first one and details are left to cited references. Experimental work is referenced only to the extent that is needed to explain the theoretical ideas. A rather extensive survey of the observations is available in the very recent review of Reid and Moran (1981).

B. Basic concepts

As mentioned above, the operation of a maser effect was recognized when the radio line radiation from certain sources turned out to be too strong for incoherent sponaneous emission. To find out what sets the scale for radiation intensity it will prove useful to introduce the concept of the radiation "brightness temperature."

The radiation intensity (more commonly referred to as "specific intensity" or "brightness" in radio astronomy) is the energy flux carried by the radiation per unit frequency bandwidth and solid angle. As a flux per solid angle it does not suffer any spatial dispersion in free space, and its only possible attenuation results from interaction with matter. For the radiation field of a blackbody at temperature T, the brightness at frequency ν is given by the familiar Planck distribution

$$B_\nu(T)=\frac{2h\nu^3}{c^2}\cdot\frac{1}{\exp(h\nu/kT)-1} . \tag{1.1}$$

In addition to a frequency distribution, the Planck function provides also an absolute scale for the brightness and as such can be used to measure the intensity of the radiation field. For a given intensity I_ν, the brightness temperature T_b is defined as the temperature of a blackbody that will emit the same amount of radiation at the given frequency ν, namely,

$$I_\nu=B_\nu(T_b) . \tag{1.2}$$

In the long-wavelength region (where $h\nu\ll kT_b$; the value of $h\nu/k$ is only 0.08, 1.06, and 2.1 K for the OH, H_2O, and $J=1\rightarrow0$ SiO maser lines, respectively), the Planck function is approximated by the Rayleigh-Jeans formula, and the following explicit expression for T_b is obtained:

$$kT_b=\frac{c^2}{2\nu^2}I_\nu . \tag{1.3}$$

The brightness temperature is obviously a formal quantity which in general will vary with wavelength for a given source. It is wavelength independent only for a source that emits according to the blackbody law.

To appreciate the significance of the brightness temperature, recall the equation of radiative transfer

$$\frac{dI_\nu}{dl}=-\kappa_\nu I_\nu+\varepsilon_\nu , \tag{1.4}$$

where l measures distances along the ray, κ_ν is the absorption coefficient, and ε_ν is the volume emission rate. The equation is usually cast in the form

$$\frac{dI_\nu}{d\tau_\nu}=-I_\nu+S_\nu , \tag{1.5}$$

where we have introduced the source function

$$S_\nu=\varepsilon_\nu/\kappa_\nu \tag{1.6}$$

and the optical depth element

$$d\tau_\nu=\kappa_\nu dl . \tag{1.7}$$

Consider now radiative transfer in a given line. The volume emission coefficient is

$$\varepsilon_\nu=N_2A_{21}\frac{h\nu_0}{4\pi}\phi(\nu) , \tag{1.8}$$

where N_2 is the volume density population of the upper level, A_{21} is the Einstein coefficient for spontaneous emission, $h\nu_0$ is the energy separation of the levels, and $\phi(\nu)$ is a normalized profile centered on ν_0, which is frequently approximated by $(\Delta\nu)^{-1}$, the inverse linewidth.

The line absorption coefficient is given by

$$\kappa_\nu=(B_{12}N_1-B_{21}N_2)\frac{h\nu_0}{4\pi}\phi(\nu) , \tag{1.9}$$

where B_{ij} is the appropriate coefficient for stimulated transitions. The line-source function is therefore given by

$$S_\nu=\frac{A_{21}N_2}{B_{12}N_1-B_{21}N_2} . \tag{1.10}$$

Introducing the populations per magnetic sublevel

$$n_i=N_i/g_i,\quad i=1,2 , \tag{1.11}$$

where g_i are the appropriate statistical weights, and utilizing Einstein's relations

$$g_1B_{12}=g_2B_{21};\quad A_{21}=B_{21}\frac{2h\nu^3}{c^2} \tag{1.12}$$

leads to

$$S_\nu=\frac{2h\nu^3}{c^2}\cdot\frac{1}{\frac{n_1}{n_2}-1} . \tag{1.13}$$

Introduce now the excitation temperature T_x of the two levels via

$$\frac{n_2}{n_1}=\exp(-h\nu/kT_x) . \tag{1.14}$$

The line-source function finally becomes

$$S_\nu=B_\nu(T_x) , \tag{1.15}$$

namely, it is the Planck function at the line excitation temperature.

With the expression for the source function just derived and the definition of the brightness temperature [Eq. (1.2)], the equation of radiative transfer [Eq. (1.5)] becomes

$$\frac{dI_\nu}{d\tau_\nu}=B_\nu(T_x)-B_\nu(T_b) . \tag{1.16}$$

The Planck function is a monotonically increasing function of temperature for any given frequency. Equation (1.16) therefore implies that for the radiation which is produced within the source, the brightness temperature cannot exceed the line excitation temperature, namely,

$$T_b\le T_x , \tag{1.17}$$

since otherwise the medium would absorb, rather than emit, its own radiation. The measured brightness temperature therefore provides a lower limit estimate for the line excitation temperature at the source.

Equation (1.16) is often written as an equation involving the appropriate temperatures, assuming they all obey the Rayleigh-Jeans limit,

$$\frac{dT_b}{d\tau}=-T_b+T_x . \tag{1.18}$$

Assuming a constant excitation temperature which is independent of the radiation field, the solution of this equation is

$$T_b = T_x(1-e^{-\tau}) + T_c e^{-\tau} , \tag{1.19}$$

where T_c is the brightness temperature of the radiation of a background continuum source which may be impinging on the medium, and the explicit expression for the optical depth τ is

$$\tau = \int \kappa \, dl = \frac{h\nu}{4\pi\Delta\nu} g_2 B_{21}(n_1 - n_2) l . \tag{1.20}$$

Equation (1.19) makes it obvious again that in the absence of a background source, the brightness temperature will be smaller than the excitation temperature and will approach it only for optically thick transitions ($\tau_\nu \gg 1$).

The significance of the brightness temperature, then, is that it provides an estimate of the possible excitation temperature in the emitting medium. Although in general excitation temperatures need not coincide with the kinetic temperature in the source, they usually have the same order of magnitude and in most cases are actually smaller. The discovery of OH brightness temperatures as high as 10^{12} K in some astronomical sources was therefore quite a spectacular event. The maser radiation which was later detected in a water-vapor transition brought brightness temperatures in excess of 10^{15} K. We might note in passing that these enormous brightness temperatures were inferred from the observations only after the dimensions of the relevant sources were resolved with interferometric techniques and sizes as small as a few milliarcseconds were deduced (Moran *et al.*, 1968; Burke *et al.*, 1970). The measured intensity is spectacular even though the radiation flux density at the radio telescope need not always be unusual, since the angular sizes of the maser spots are so small.

These spectacular brightness temperatures obviously cannot bear any resemblance to the kinetic temperature at the source, since molecules dissociate at temperatures of only a few thousand degrees. In addition, the linewidths of the strong OH lines are always narrow enough that, if interpreted as thermal widths, temperatures of less than about 100 K were obtained, in agreement with expected values for interstellar regions. It is therefore evident that the reasoning which led from Eq. (1.16) to Eq. (1.17) does not always apply.

Indeed, a hidden assumption in the previous discussion is that all the temperatures involved are positive, as is usually the case. Suppose, however, that, for one reason or another, the popultion of the two levels is inverted so that $n_2 > n_1$ and T_x becomes negative [Eq. (1.14)]. The Planck function of a negative temperature is negative, and since in this case the absorption coefficient κ_ν (and hence also τ_ν) is also negative, the equation of radiative transfer will no longer restrict the brightness temperature. This is best illustrated by the version which involves temperatures [Eq. (1.18), assuming the Rayleigh-Jeans limit for all temperatures], which now becomes

$$\frac{dT_b}{d|\tau|} = T_b + |T_x| . \tag{1.21}$$

Obviously, T_b need not now be smaller than $|T_x|$. In fact, if negative excitation temperature and optical depth are inserted in the solution for the brightness temperature [Eq. (1.19)], the absorption factor $\exp(-\tau)$ becomes an amplification factor, and the absolute value of the optical depth is therefore referred to as a "gain." A gain of more than 20 leads to amplification in excess of 10^8 and could explain the observed brightness temperatures.

One may wonder whether an experimental determination of T_x is possible at all, providing a direct proof for population inversion. Unfortunately, Eq. (1.19) includes two unknown quantities—the excitation temperature T_x and the optical depth τ, and both, obviously, cannot be determined simultaneously from a single equation without some further assumptions. Suppose, however, that an interstellar molecular cloud happens to lie in front of a radio continuum point source. If the cloud is extended enough, one may perform a radio measurement in a direction which includes the point source and a second measurement in a direction which excludes it, thus obtaining two brightness temperatures, with and without the term involving T_c, respectively [Eq. (1.19)]. Since the value of T_c can be deduced from measurements at frequencies outside the relevant line, one obtains in this way two independent equations which can be solved for the two unknowns T_x and τ.

This clever technique was utilized by Rieu *et al.* (1976) to study an interstellar cloud in front of the extragalactic radio source 3C123. Their results for three of the OH ground-state lines are displayed in Fig. 4. In the "off-source" measurements, those in which 3C123 is outside the telescope beam, all three lines appear in emission. In the "on-source" measurements, in the direction of 3C123, the two main lines appear in absorption, indicating that their excitation temperatures are lower than the brightness temperature of 3C123. The line at 1720 MHz, on the other hand, appears now in stronger emission, which means that the cloud is actually amplifying the background radiation in this line. This is one of the most spectacular maser effects observed in radio astronomy, even though it involves a very weak maser with a gain of only ~ 0.1, since it shows how a cloud can act as an amplifier for a background radiation. The inferred excitation temperature is about -10 K.

II. MASER THEORY

A. Generalities

It is instructive to ask first why the interstellar medium can produce maser radiation with such ease, whereas special efforts are required in the laboratory to achieve the same end. The answer lies in the great differences in densities and geometrical dimensions between the two cases.

Consider a two-level system and assume that the only excitation mechanism is collisions. With the neglect of

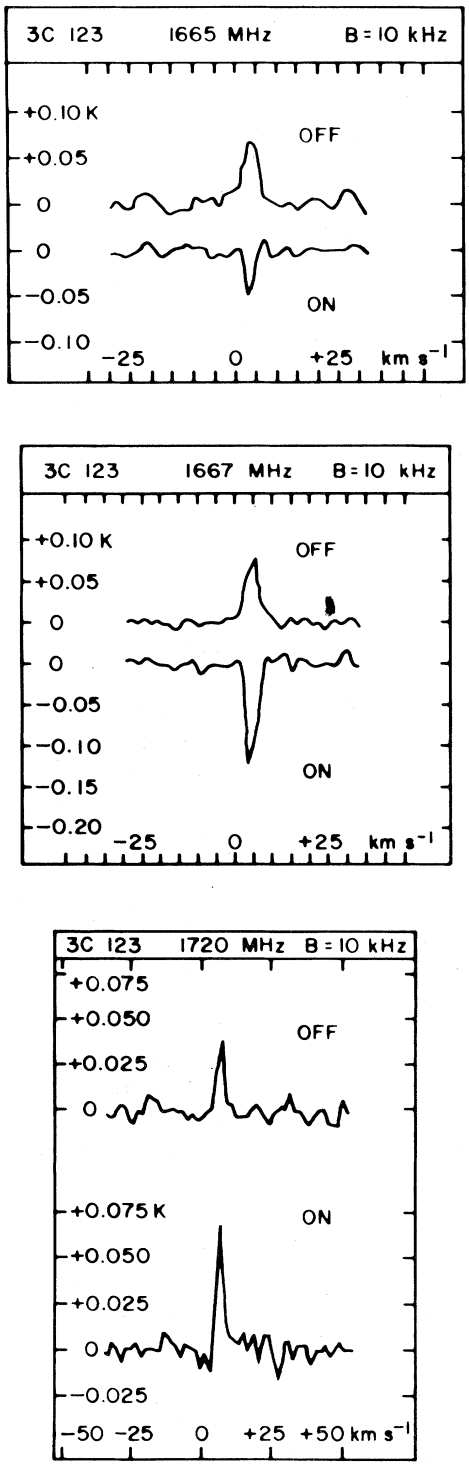

FIG. 4. OH spectrum toward a cloud in front of the extragalactic radio source 3C123. The "on source" measurements include the source in the radio-telescope beam, whereas the "off" measurements exclude it. The frequency scale was converted to equivalent Doppler shifts from the rest frequency at the local standard of rest (LSR)—a frame at rest with respect to galactic rotation at the position of the cloud (from Rieu *et al.*, 1976).

optical depth effects, which are discussed below, the steady-state level populations obey

$$N_1 C_{12} = N_2 (C_{21} + A_{21}) , \tag{2.1}$$

where C_{ij} are the appropriate collision rates obeying the detailed balance relation

$$g_1 C_{12} = g_2 C_{21} \exp(-h\nu/kT) , \tag{2.2}$$

and T is the kinetic temperature, namely, the temperature which characterizes the Maxwell-Boltzmann velocity distribution of the colliding particles. We are therefore assuming here that, in addition to the molecules that populate the two discussed levels, which are the potential maser, there is another species of colliding particles whose kinetic motions are in thermal equilibrium. This is usually the case in the interstellar medium, where the major constituent is hydrogen, in either molecular or atomic form. Possible violations of Eq. (2.2) when the motions of the colliding particles are not thermalized are discussed in Sec. IV.C, below.

The populations of the two-level system therefore obey

$$\frac{n_2}{n_1} = \exp(-h\nu/kT)/(1 + A_{21}/C_{21}) , \tag{2.3}$$

which coincides with the Boltzman distribution in the limit that $C_{21} \gg A_{21}$. The levels will therefore thermalize (i.e., $T_x = T$) when collisions dominate all the downward transitions, namely, when the densities are sufficiently high. To get an estimate for the required densities, consider a "typical" rotation transition of a diatomic molecule which (for $J = 1 \rightarrow 0$) has $A \simeq 10^{-6}$ sec^{-1}. Taking as representative values a geometric cross section of $\sim 10^{-15}$ cm^2 and a thermal velocity of $\sim 10^5$ cm/sec, the collision rate coefficient is about 10^{-10} cm^3/sec. The decays will therefore be collision dominated for densities in excess of 10^4 cm^{-3}. Although this would be considered an extremely low density in the laboratory, achieved only by the most advanced vacuum techniques (pressures of $\sim 10^{-12} - 10^{-13}$ Torr), this is quite a high density by interstellar standards. Line thermalization is therefore the rule in terrestrial circumstances but the exception in interstellar space due to the huge difference in relevant densities.

The above estimates were for "typical" parameters, but variations are quite large, espeically for the A coefficient. Note first that the parameter which measures the intrinsic line strength is the B coefficient, which is related to the transition dipole moment μ_{21} via

$$B_{21} = \frac{8\pi^2}{3\hbar^2 c} |\mu_{21}|^2 . \tag{2.4}$$

The A coefficient measures the transition probability and as such depends also on a phase-space factor [$\sim \nu^3$; see Eq. (1.12)] in accordance with Fermi's golden rule. Its range of variation is therefore quite large, reflecting transition frequency as well as line strength variations.

The CO moelcule, for instance, has an anomalously small dipole moment, and its A coefficients for rotational transitions are thus much smaller than typical ones (only

$6\cdot10^{-8}$ sec^{-1} for $J=1\rightarrow0$). The CO molecule therefore thermalizes quite easily and is considered a good thermometer for interstellar clouds. On the other hand, hydrogen-bearing molecules (including OH and H_2O) have large rotation energies resulting in high rotation frequencies and large values for A ($\sim10^{-1}-10^{-3}$ sec^{-1}) and can therefore maintain nonthermal populations at higher densities. Molecular vibration and electronic transitions, with typical A's of ~1 sec^{-1} and $\sim10^{7}$ sec^{-1}, respectively, require even higher densities to thermalize.

The above discussion follows a solution of a simple two-level model which has always $n_2<n_1$, as is evident from Eq. (2.3), and will therefore not invert. The discussion illustrates, however, the significance of the deexcitation transitions in establishing thermal equilibrium for the level populations. When we consider a larger energy-level diagram, of which the two relevant levels are a subsystem, and when all downward transitions are dominated by spontaneous decays, the populations deviate from thermal equilibrium, and any distribution is possible in principle, including inversion.

An appreciable maser effect requires large gains, which in turn mean large column densities [$\int n\,dl$; see Eq. (1.20)]. This is in conflict with the demand for small volume densities, which are required for the deviations from thermal population distribution. The laboratory solution to this problem is to bounce the laser light between mirrors, effectively increasing the linear dimensions of the system by accumulating gain through many passes of the beam in the resonance cavity. This introduces a high degree of mode selectivity, since phase coherence, which leads to constructive interference, is possible only inside a limited bandwidth such that $\Delta\nu/\nu$ is equal to the inverse of the Q value of the resonance cavity (with values easily in excess of 10^7 for optical resonators). In contrast, typical lengths of astronomical masers are at least $10^{12}-10^{13}$ cm (as much as the sun-earth distance), and the required gains are obtained for plain photon propagation in the source, as in a single-pass laser. There is therefore no selection of any field mode, unlike laboratory lasers, and astronomical masers are broadband with line profiles which have at least a full thermal width, except for possible line-narrowing effects which are mentioned below (Sec. II.D).

The small densities and large dimensions thus make interstellar space an ideal medium for maser emission. Astronomical masers also provide what may be described as the "purest" form of maser action, since strong amplification through stimulated emission occurs naturally as the radiation propagates in the inverted medium without the need for external aids such as good focusing or mode locking. As a result, the radiation does not maintain phase coherence across its front. Since the question of coherence has occasionally caused some confusion, it would be instructive to discuss it.

From the quantum-mechanical point of view, the phase Φ of any given state of the radiation field with n photons is subject to the uncertainty

$$\Delta\Phi\Delta n\geq\tfrac{1}{2}\ . \tag{2.5}$$

For an ideal maser amplifier, with all the population initially in the upper level, it is possible to show that the equality sign holds and that $\Delta n\simeq\sqrt{n}$ for large n (Serber and Townes, 1960). As the amplification increases the number of photons in the state, its phase becomes better defined, and an ideal maser can lead to the buildup of ideal coherent states. This conclusion applies, however, only to a given mode of the field, characterized by a given wave vector $\mathbf{k}$. Consider now radiation propagating at the same frequency along two close directions with an angle θ between them. The phase difference accumulated along a distance l will be

$$\frac{2\pi l}{\lambda}(1-\cos\theta)\simeq\pi\frac{l}{\lambda}\theta^2\ . \tag{2.6}$$

A good laboratory system will make this phase difference very small indeed, but with the dimensions and wavelengths appropriate to astronomical masers, the phase difference will be large unless $\theta\lesssim10^{-6}$. This is much smaller than any reasonable estimate for the emission angles of astronomical masers, which are usually in the range $\theta_m\simeq10^{-2}-10^{-1}$ (see below). Thus, even though it was suggested that coherent pulses might form in astronomical masers (Rosen, 1974), it is evident that this is impossible due to the large dimensions involved, and the maser radiation is thus phase incoherent across the wave front and should obey random statistics. This was correctly recognized in the early theoretical studies of Litvak (1970, 1973) and of Goldreich, Keeley, and Kwan (1973a, 1973b) and is fully corroborated by observations (Evans *et al.*, 1972; J. M. Moran, 1981).

Whereas wave-front phase coherence at the source is impossible for the radiation, velocity coherence for the emitting molecules is a must. The amplification is achieved by induced emission, and a maser photon therefore must be able to encounter a molecule with the right transition frequency that has not been Doppler shifted outside the linewidth. This imposes a restriction on maser action, since sizeable motions are common inside interstellar clouds, as is evident from molecular linewidths. The fullwidth at half maximum of a thermal Doppler profile for molecular mass m expressed in velocity units is

$$\Delta v_D=2\left[2\ln2\frac{kT}{m}\right]^{1/2}=0.2\left[\frac{T}{m/m_H}\right]^{1/2} \tag{2.7}$$

in km/sec. For the CO molecule, for instance, the linewidth at about 30 K (which is a reasonable estimate for the kinetic temperature inside a cloud) should therefore be about 0.2 km/sec. The observed CO lines, however, are always much broader, with typical widths of a few km/sec at least. It is quite clear that line broadening is caused by intercloud motions, although the origin of these motions is a matter of debate. The strong maser lines, on the other hand, are much narrower than either ordinary thermal lines or weak maser lines that do not have a large gain. This narrowness arises because the maser photons are seeking a path that maintains good coherence in the line-of-sight velocity to obtain a large amplification. The

masers therefore have the shape of elongated tubes, or cylinders, which are of course not well-defined physical entities, but rather directions which developed by chance the required velocity coherence. Maser radiation is beamed in the direction of the cylinder axis into an angle given by the ratio of its radius to its length. Estimates for this ratio lead to the above-mentioned values of emission angles. This picture for the structure of the sources is corroborated by interferometric observations, which show that maser sources are comprised of a number of spots, each having its own well-defined velocity Doppler shift.

B. Basic equations

In order to discuss the physical properties of astronomical masers we need to set up the equations that govern the level populations and the transfer of the maser radiation. The situation, fortunately, is simpler than that for laboratory laser radiation in that there is no interaction with a resonant cavity and the radiation is broadband and obeys random statistics, as mentioned above. The resulting equations are therefore quite simple and are basically the same as for nonmaser radiation. The phenomenological theory, which we now describe, was first worked out by Litvak (1970) and later extended by Goldreich, Keeley, and Kwan (1973a) to study polarization effects. It basically follows the formalism developed by Lamb (1964) for laser radiation.

The Poynting theorem follows directly from Maxwell's equations and it reads, in Gaussian units,

$$c\nabla\cdot(\mathbf{E}\times\mathbf{H})+\mathbf{E}\cdot\frac{\partial\mathbf{D}}{\partial t}+\mathbf{H}\cdot\frac{\partial\mathbf{B}}{\partial t}=-4\pi\mathbf{J}\cdot\mathbf{E}\ . \tag{2.8}$$

With

$$\mathbf{D}=\mathbf{E}+4\pi\mathbf{P}\ ,$$

and assuming $\mathbf{B}=\mathbf{H}$, the equation becomes, for a plane wave propagating in the direction $\mathbf{l}$ (so that $\mathbf{B}=\mathbf{l}\times\mathbf{E}$),

$$\left[\mathbf{l}\cdot\nabla+\frac{1}{c}\frac{\partial}{\partial t}\right]\frac{\mathbf{E}^2}{4\pi}=-\frac{1}{c}\mathbf{E}\cdot\left[\mathbf{J}+\frac{\partial\mathbf{P}}{\partial t}\right]\ , \tag{2.9}$$

which is basically the equation of radiative transfer. For a quasimonochromatic wave component, namely, with small variations in amplitude and phase over a wavelength scale, this leads to equations for the wave's amplitude and phase so that polarization effects can be studied.

The effect of free charges on the propagation of the radiation is described by the current $\mathbf{J}$ and that of the maser levels by the polarization vector $\mathbf{P}$. Let us now study the much simplified two-level model, neglecting the vector nature of the electric field, and considering only transfer of radiation centered on the transition frequency. This still preserves the essential features of the more detailed calculations.

The eigenfunctions of the levels are $\psi_i(i=1,2)$, and the system wave function is therefore

$$\psi(t)=\sum_{i=1}^{2}a_i(t)\psi_i\ . \tag{2.10}$$

From the Schrödinger equation we can derive the equations which govern the time behavior of the density matrix elements $\rho_{ij}(=a_i^*a_j)$. These are then supplemented by the usual phenomenological terms that describe pumping into the levels by production rates P_i and losses via a damping factor Γ, which for simplicity we take as being equal for both levels.

The equations for the diagonal elements of the density matrix then become

$$\frac{d}{dt}\rho_1=P_1-\Gamma\rho_1+\frac{i}{\hbar}(\rho_{12}V_{21}-\rho_{21}V_{12})\ , \tag{2.11}$$

where $\rho_i\equiv\rho_{ii}$ and the levels interact through their dipole moment with the electric field of the radiation $E(t)$ so that

$$V_{ij}=-\mu_{ij}E(t) \tag{2.12}$$

and μ_{ij} is the transition dipole matrix element. Since the tensorial structure of the electric field is neglected, the phases can be chosen such that $\mu_{ij}(=\mu_{ji})$ is real so that $V_{12}=V_{21}\equiv V$.

The equation for ρ_2 is similar to Eq. (2.11). These are basically the statistical rate equations which describe the level populations. For the interaction with the field we need the equation for the off-diagonal element $\rho_{12}(=\rho_{21}^*)$, which is

$$\frac{d}{dt}\rho_{12}=(i\omega_0-\Gamma)\rho_{12}-\frac{i}{\hbar}V(\rho_2-\rho_1)\ , \tag{2.13}$$

where $\hbar\omega_0$ is the levels' energy separation. The formal solution of this equation is

$$\rho_{12}(t)=-\frac{i}{\hbar}\int_{-\infty}^{t}[\rho_2(t')-\rho_1(t')]V(t')\times\exp[(i\omega_0-\Gamma)(t-t')]dt'\ . \tag{2.14}$$

Goldreich, Keely, and Kwan (1973a) argue that the diagonal elements (namely, ρ_1 and ρ_2) are approximately time independent. This approximation is discussed below. Granting that, for the moment, we get

$$\rho_{12}(t)=-\frac{i}{\hbar}\Delta\rho\int_0^\infty V(t-\tau)\exp[(i\omega_0-\Gamma)\tau]d\tau\ , \tag{2.15}$$

where $\Delta\rho\equiv\rho_2-\rho_1$.

The level population equations then become

$$\begin{aligned}0&=P_1-\Gamma\rho_1+U\Delta\rho\ ,\\ 0&=P_2-\Gamma\rho_2-U\Delta\rho\ ,\end{aligned} \tag{2.16}$$

where

$$U(t)=\frac{2\mu_{12}^2}{\hbar^2}\int_0^\infty E(t)E(t-\tau)\exp(-\Gamma\tau)\cos\omega_0\tau\,d\tau\ . \tag{2.17}$$

The meaningful quantity is the slowly varying part of $U(t)$, so this equation has to be averaged over a time which is long enough to include many field oscillations

but short enough that the amplitude does not vary appreciably. This simply leads to the autocorrelation function of the field E, which is just the Fourier transform of its power spectrum. Assuming random statistics, which is appropriate as discussed above, we take the power spectrum as a Gaussian centered on ω_0 with width $\Delta\omega$. The transform then leads to the standard result for a Gaussian wave packet,

$$\overline{E(t)E(t-\tau)}=\overline{E^2(t)}\cos\omega_0\tau\exp[-1/2(\tau\Delta\omega)^2]\ , \tag{2.18}$$

where the bar denotes average values. The expression for $\overline{U(t)}$ can now be evaluated explicitly, and with the assumption $\Delta\omega\gg\Gamma$ we finally obtain

$$\bar{U}=\left[\frac{\pi}{2}\right]^{1/2}\left[\frac{\mu\bar{E}}{\hbar}\right]^2 1/\Delta\omega\ , \tag{2.19}$$

where $\bar{E}\equiv(\overline{E^2})^{1/2}$. Since the radiation intensity is given by

$$I_\nu=\frac{c}{4\pi\Delta\nu}\overline{E^2}\ , \tag{2.20}$$

we see that, up to a factor of order unity, the interaction rate is the same as the usual stimulated rate

$$\bar{U}\simeq BI\ , \tag{2.21}$$

where B is the appropriate Einstein coefficient [Eq. (24)].

Having derived the equations for the level populations, we turn now to the equation of radiative transfer [Eq. (2.9)]. The dipole moment of the system is

$$P=\mu(\rho_{12}+\rho_{21})\ , \tag{2.22}$$

and to get the macroscopic polarization vector we have to integrate this expression with the number density distribution function. Here we neglect this step, assuming simply that all molecules have the same transition frequency ω_0. From Eq. (2.14) we then obtain

$$\frac{\partial P}{\partial t}=\frac{2\mu}{\hbar}\omega_0\Delta N\int_0^\infty V(T-\tau)\exp(-\Gamma\tau)\cos\omega_0\tau\,d\tau\ , \tag{2.23}$$

where ΔN is the difference in level population. The equation of radiative transfer then becomes

$$\frac{d\overline{E^2}}{dl}=-\frac{4\pi}{c}\hbar\omega_0\Delta N\bar{U}=-\kappa\overline{E^2}\ , \tag{2.24}$$

where the absorption coefficient κ so derived is again identical with that of Eq. (1.9), within a factor of order unity, and the differentiation is performed here as in a convective derivative.

The present formalism does not describe spontaneous emission because the field is treated classically. A similar calculation with field quantization included would produce the spontaneous emission term but otherwise yield identical results (e.g., Jaynes and Cummings, 1963), at the expense of greater complexity.

A major assumption made in the above derivation, following Goldreich *et al.* (1973a), is that the diagonal elements of the density matrix are time independent. Maser emission shows good time stability. Some sources have hardly changed since their discovery more than 15 years ago. Except for an H_2O flare-up in Orion, the fastest variations observed so far have time scales of about a few weeks — much larger than the inverse values of typical pump or decay rates in the pertinent sources. The assumption of steady state for astronomical masers is thus certainly justified for the averaged quantities. There remains the question of the fluctuations around the mean values. These, of course, cannot be evaluated in general without first solving the whole problem. However, an analysis which treats the interaction with the field as perturbation can be performed. Carrying such a perturbation calculation to third order, Goldreich *et al.* (1973a) report that they have obtained results which were independent of the assumption of constancy for the diagonal elements.

Another key point in the analysis is the assumption of a broad bandwidth for the radiation. It is important to ask what sets the scale for the bandwidth, since $\Delta\omega/\omega$ is very small for astronomical masers (only about 10^{-5}). The bandwidth is large, though, if

$$\bar{\tau}\Delta\omega\gg 1 \tag{2.25}$$

for each relevant time scale $\bar{\tau}$ in the rate equations. This condition is indeed obeyed when $\bar{\tau}$ is taken as either the mean time between collisions or the approriate radiative lifetime, and in this respect astronomical masers are always broadband. When the maser radiation becomes sufficiently strong, however, the shortest time scale in the problem is the interval between successive interactions with the radiation field, which in order of magnitude is given by

$$\bar{\tau}\simeq\left[\frac{\mu\bar{E}}{\hbar}\right]^{-1}\ . \tag{2.26}$$

From Eq. (2.19) it is evident that the condition for a large bandwidth Eq. (2.25) becomes equivalent in this case to

$$\xi=\frac{\Delta\omega}{\bar{U}}\gg 1\ , \tag{2.27}$$

and the question of the linewidth is linked to that of the strength of the radiation field. A maser source may be so intense that it will become effectively quasimonochromatic.

Before estimating the value of ξ for various maser sources, we have to be a bit more careful with regard to the question of the angular distribution of the radiation. The radiation can propagate in the source in all directions, and the interaction term $\bar{U}$ should therefore be integrated over all solid angles, so that instead of BI_ν, the stimulated transition term in the rate equations is actually BJ_ν, where

$$J_\nu=\int I_\nu\frac{d\Omega}{4\pi}\ . \tag{2.28}$$

When the radiation is isotropic, the two expressions are of course the same, but when the radiation is beamed into a

small solid angle $\Delta\Omega_m$, we have

$$J_\nu \simeq I_\nu \frac{\Delta\Omega_m}{4\pi} . \tag{2.29}$$

Notice that this expression covers also the isotropic case (where $\Delta\Omega_m = 4\pi$). The directional case is the approriate one for astronomical masers, since the radiation is beamed because of the requirement for velocity coherence, as discussed above. The expression for ξ is then

$$\xi = \frac{\Delta\omega}{BI_\nu} \cdot \frac{4\pi}{\Delta\Omega_m} . \tag{2.30}$$

Using Eqs. (1.3) and (1.12) we have

$$BI = A\frac{kT_b}{h\nu} , \tag{2.31}$$

and finally

$$\xi = \frac{T_\xi}{T_b} \cdot \frac{4\pi}{\Delta\Omega_m} , \tag{2.32}$$

where

$$T_\xi = \frac{2\pi h\nu^2 \Delta v}{kcA} ,$$

and Δv is the linewidth in velocity units.

The value of T_ξ is $6\cdot10^{11}$, $4\cdot10^{13}$, and $3\cdot10^{14}$ K for the SiO, OH, and H_2O masers, respectively, for a linewidth of 1 km/sec. Hence, $T_\xi > T_b$ for SiO and OH but not for H_2O, where T_b in excess of 10^{14} K was observed. The maser beaming factor, however, is

$$\frac{\Delta\Omega_m}{4\pi} = \frac{1}{2}(1-\cos\theta_m) \simeq \left[\frac{\theta_m}{2}\right]^2 , \tag{2.33}$$

and as mentioned above, θ_m is probably $\simeq 10^{-1}-10^{-2}$. Hence $\xi \gg 1$ for all astronomical masers.

The fact that $\xi > 1$ is of utmost importance for astronomical maser theory. As is evident from Eqs. (2.18) and (2.25), this ensures that field fluctuations will decay between successive interactions with the molecules. Because of this, various effects observed in the laboratory which require coherence over the radiation bandwidth do not apply to astronomical masers. In addition, the derivation of the rate equations outlined above omits the step which involves integration over the molecular frequency distribution. It can be shown that with $\xi > 1$ this additional step does not modify the final results, but when $\xi < 1$ the expression for the maser gain is different and involves nonlinearities in the radiation intensity (Litvak, 1970). It is then a fortunate coincidence that the masers are broadband and not strong enough so that $\xi > 1$ always applies.

The final form of the equation of radiative transfer is then identical to the standard one [Eq. (1.4)]. This can now be integrated over the line profile, and with the average intensity I defined as

$$I = \int I_\nu d\nu / \Delta\nu , \tag{2.34}$$

the equation is

$$\frac{dI}{dl} = -\kappa I + \varepsilon , \tag{2.35}$$

with κ and ε given by the same expressions as in Eqs. (1.8) and (1.9) with $\Delta\nu^{-1}$ replacing the profile function $\phi(\nu)$.

C. Phenomenological maser model

The equations of level populations will now be written in greater detail, including the various modes by which the maser levels can exchange population among themselves and with other levels. The simplest model involves the two maser levels and some structureless reservoir, approximating all other levels, which can exchange population with them. It would prove useful to visualize such a model in the context of the SiO maser. In this case, most of the molecules are in the ground vibration state which constitutes the basic pool (the "reservoir") out of which molecules are pumped into the maser levels. The maser levels would then be, for example, the $J=0$ and 1 states of $v=1$, with the reservoir being the entire $v=0$ state.

The pump rate per unit volume from the reservoir into each maser level is $\tilde{P}_i (i=1,2)$, and the loss rate for the reverse direction is Γ_i. In addition, the molecules exchange population among themselves via collisional and radiative (spontaneous and induced) transitions. The level populations therefore obey the equations

$$\frac{dN_2}{dt} = \tilde{P}_2 - \Gamma_2 N_2 - A_{21}N_2 - J(N_2B_{21} - N_1B_{12}) - (N_2C_{21} - N_1C_{12}) , \tag{2.36}$$

with

$$J = \int J_\nu \phi(\nu) d\nu$$

and a similar equation for N_1. Using the detailed balance relations which the B and C coefficients obey, the equation can be rewritten for the population per sublevel n_2, which is the relevant quantity for the gain, as

$$\frac{dn_2}{dt} = P_2 - \Gamma_2 n_2 - An_2 - BJ(n_2 - n_1) - C[n_2 - n_1 \exp(-h\nu/kT)] , \tag{2.37}$$

where

$$P_i = \tilde{P}_i / g_i$$

is the pump rate per sublevel. The indices of the exchange rate coefficients can be omitted without confusion. With the exception of a sign reversal for the exchange terms, the equation for n_1 is essentially the same but includes the ratio g_1/g_2, so to simplify matters further we set this equal to unity.

The last three terms in the rate equation involve exchange between the maser levels; and as long as the detailed balance relations are obeyed, they will never produce inversion. An inversion can then be achieved only when either the pump or loss rates for the two levels are different. I argue below that for the SiO masers the pump rates are in fact equal and the losses different. The algebra for this case, however, is slightly more complex

than for the opposite one of equal loss and different pump rates so I use the latter for the illustration here. In steady state, Eq. (2.37) and its counterpart for n_1 then yield for the overall population in the maser system

$$n_{12} \equiv n_1 + n_2 = (P_1 + P_2)/\Gamma = 2P/\Gamma \ , \tag{2.38}$$

where P is the average pump rate into the maser levels.

The A coefficients are $3 \cdot 10^{-6}$ sec^{-1} for the $J=1 \rightarrow 0$ transition of SiO, $2 \cdot 10^{-9}$ sec^{-1} for the H_2O maser, and about 10^{-11} sec^{-1} for the OH ground-state transitions. These rates are much smaller than all the other ones, and the spontaneous transitions can safely be neglected in the rate equations. To simplify the discussion further, let us neglect for the moment the effect of collisions which will be reintroduced later. With all of these simplifications the expression for the population difference $\Delta n \equiv n_2 - n_1$ can be written in either of the following forms (with $\Delta P \equiv P_2 - P_1$):

$$\Delta n = \frac{\Delta P}{\Gamma + 2BJ} = \frac{\Delta n_0}{1 + J/J_s} = \frac{\Delta n_0}{1 + I/I_s} \ , \tag{2.39}$$

where

$$J_s = \Gamma/2B; \quad I_s = J_s 4\pi/\Omega_m; \quad \Delta n_0 = \Delta P/\Gamma \ . \tag{2.40}$$

The expression for the absorption coefficient derived from Eqs. (1.9) and (2.39) is likewise[1]

$$\kappa = \frac{\kappa_0}{1 + J/J_s} = \frac{\kappa_0}{1 + I/I_s} \ . \tag{2.41}$$

The parameter J_s determines the radiation intensity at which the induced transitions begin to affect the level populations. For $J \ll J_s$ we have $\Delta n = \Delta n_0$, and the level populations are essentially independent of the radiation field. The equation of radiative transfer can be solved as before, and the solution in the linear case is, as in Eq. (1.19)

$$T_b = (T_x + T_c)\exp(\kappa_0 l) - T_x \tag{2.42}$$

assuming that all the temperatures are in the Rayleigh-Jeans regime. The range of intensities $J \ll J_s$ corresponds to an "unsaturated" maser, in contrast with the "saturated" one which is described shortly. The level populations are determined in this region by parameters which are inherent to the pumping scheme and are unaffected by the propagating radiation. The inverted population provides an amplifying medium, since stimulated emissions outnumber absorptions. The exponential growth of the flux density is due to the shower of induced emission photons which can be generated by one seed photon. The power-series expansion of the exponential function describes the successive generations of induced photons.

There are two potential sources of seed photons or "input radiation": an internal one, due to spontaneous transitions from the upper level; and an external one, the radiation of a background source which may impinge on the maser. The relative importance of these two sources and the question of what is the input radiation that the maser amplifies are determined by the relative magnitudes of T_c and T_x, as is evident from Eq. (2.42).

The exponential growth of the radiation cannot proceed indefinitely, because eventually the induced rates will become comparable to the intrinsic rates which determined the magnitude of the population inversion in the first place. The radiation will then begin to affect the level populations; and because of the excess of induced emission over absorption, the inversion will decrease, and the maser will begin to "saturate." This happens when J becomes comparable to J_s, as is evident from Eq. (2.39). This equation shows that although the population inversion is decreasing for $J > J_s$, the product $J\Delta n$ is approaching the constant value $J_s \Delta n_0$. We can therefore immediately determine the volume production rate of maser photons, Φ_m, from this equation as

$$\Phi_m = BJ\Delta n = \Delta P/2 = \eta P \ , \tag{2.43}$$

where

$$\eta = \Delta P/2P \ .$$

This is the central result of the discussion. It shows that *each pumping event leads to a maser photon, with an efficiency η which is determined by the details of the pumping scheme.* The flux density I can also be determined immediately without the need to solve the equation of radiative transfer, since from the definitions it is evident that,

$$I = \int \Phi_m \frac{h\nu}{\Delta\Omega_m \Delta\nu} dl \ , \tag{2.44}$$

and since, using Eq. (2.43), (2.40), (2.29), and (1.9) we find, for the linear case,

$$I = I_s \kappa_0 l \ . \tag{2.45}$$

The same result can be obtained from the equation of radiative transfer, which is therefore not an independent equation for a saturated maser. The linear growth of the flux density is not as spectacular as the exponential one, but note that now $I \gg I_s$, whereas the unsaturated amplification is restricted to $I \ll I_s$. The saturated regime thus provides the most efficient mode of maser operation with each pumping event leading to the production of $\eta (< 1)$ maser photons.

It is evident that in the saturated maser there is production, rather than amplification, of radiation, and that the linear growth is simply the result of the contribution of additional emitting volume, with constant photon production rate, in the line of sight. The expression for the production rate, Eq. (2.43), is quite easy to understand. A strong saturation (i.e., $J \gg J_s$) is set when the induced rates are much larger than the loss rate Γ [Eq. (2.39)]. A molecule pumped into one of the maser levels will therefore interact with the maser radiation field before decaying back to the reservoir, so that each pumping event

[1]From this point on, the notation for absolute value will be omitted from the expressions for κ, τ, and T_x of the maser levels. This should not cause any confusion, as it will always be clear which are the intrinsically negative quantities.

leads to an induced transition. Pumping into the lower level leads to an absorption and the removal of a maser photon, whereas pumping into the upper level leads to a stimulated emission and the production of a new maser photon. Production therefore exceeds absorption, per unit volume and time, by the amount $\Delta P=P_2-P_1$. The additional factor of $\frac{1}{2}$ in Eq. (2.43) reflects the fact that for a strongly saturated maser, the populations of the two levels are almost equalized by the radiation,[2] so a molecule will spend on the average only half of its time in the level into which it was initially pumped.

It is evident that maser operation in the saturated mode is entirely different from that in the unsaturated domain. A saturated maser is actually not an amplifier at all. Rather, it is a linear converter, converting the pumping events into maser photons with a specific efficiency η. The concept of an input source for amplification is therefore meaningless for a saturated maser. For some reason, this obvious point has caused considerable confusion in the literature.

To see how the dependence on the input source disappears for a saturated maser, let us consider in greater detail a partially saturated one-dimensional model with external radiation illuminating the unsaturated end. To simplify the discussion it will be assumed that all the temperatures are in the Rayleigh-Jeans domain and that the point $l=l_s$ is a sharp boundary at which the maser switches from unsaturated operation to complete saturation. If we denote by subscript 0 all the quantities in the unsaturated region ($l<l_s$), then the expression for the radiation brightness temperature is again

$$T_b=(T_{x,0}+T_c)\exp(\tau) , \tag{2.46}$$

with

$$\tau=\kappa_0 l$$

and assuming $\exp(\tau)\gg 1$. This expression holds as long as $I<I_s$, or, equivalently, $T_b<T_s$, where, using Eqs. (2.40), (1.12), and (1.3),

$$kT_s=\frac{1}{2}h\nu\frac{\Gamma}{A}\frac{4\pi}{\Delta\Omega_m} . \tag{2.47}$$

The location of the transition to saturation is therefore determined from

$$(T_{x,0}+T_c)\exp(\tau_s)=T_s \tag{2.48a}$$

or

$$\tau_s=\kappa_0 l_s=\ln\left[\frac{T_s}{T_{x,0}+T_c}\right] . \tag{2.48b}$$

In the saturated region, $l>l_s$, the equation of radiative transfer is

$$\frac{dI}{dl}=\kappa I=\frac{\kappa_0 I}{1+I/I_s}\simeq\kappa_0 I_s , \tag{2.49}$$

and its solution is

$$I=I_s[1+\kappa_0(l-l_s)] \tag{2.50}$$

or equivalently

$$T_b=T_s[1+\kappa_0(l-l_s)] .$$

The dependence on the temperatures of the input radiations is contained only in l_s [see Eq. (2.48b)]. However, even this weak logarithmic dependence disappears for a strong enough maser which operates at $l\gg l_s$, where $T_b=T_s\kappa_0 l$. The input radiation therefore: (1) is amplified in the unsaturated region, (2) determines the location of the transition to saturated operation, and (3) essentially disappears from the problem in the saturated region, which is just converting the pump operation to maser radiation.

The derivation here is for a linear geometry, but essentially the same results are obtained for the more realistic cylindrical maser, where the angle $\Delta\Omega_m$ introduces an additional l dependence. The relevant expressions for this case are available in the literature (e.g., Goldreich and Keeley, 1972; Reid *et al.*, 1980) and will not be reproduced here.

Some additional insight into the operation of a saturated maser can be obtained from the expressions derived above. Its optical depth τ is given by

$$\tau=\kappa_0 l_s+\frac{\kappa_0(l-l_s)}{1+I/I_s}=\tau_s+\frac{I-I_s}{I+I_s}\simeq\tau_s+1 , \tag{2.51}$$

and the contribution from the saturated region is therefore negligible (since $\tau_s\gg 1$, obviously). The saturated region is therefore effectively transparent, which explains why the equation of radiative transfer is redundant there, and the optical depth τ saturates at the value τ_s.

The excitation temperature T_x for the saturated maser can be obtained from Eq. (2.39) for the population difference Δn. Neglecting the variation of n_1 during saturation, the expression for T_x is

$$T_x=T_{x,0}(1+I/I_s)=T_{x,0}(1+T_b/T_s) ; \tag{2.52}$$

and the excitation temperature T_x increases linearly as the inversion decreases by the saturation effect (note that a weak inversion corresponds to a large negative temperature). This equation can be combined with Eq. (2.48), and for a strong maser ($T_b\gg T_s$) with no external radiation source the following result for T_b is obtained:

$$T_b=T_x\exp(\tau_s) , \tag{2.53a}$$

or, since $\tau=\tau_s$,

$$T_b=T_x\exp(\tau) . \tag{2.53b}$$

This expression is identical with the one for an unsaturated maser. Its meaning in the saturated region is entirely different, of course, since τ is now fixed at its saturation value, and since T_b increases only through the linear variation of T_x. The formal expression (2.53) is

[2]The population of each level is given, in the saturated case, by ($i=1,2$)

$$n_i=n_{1,2}\left[1+\frac{P_i}{2P}\frac{J_s}{J}\right]\Big/2 .$$

useful, however, for discussion of beaming effects, for instance, since it displays the variation of intensity with direction through its dependence on τ.

The expression given by Eq. (1.18) [or (2.42)], namely,

$$T_b=(T_x+T_c)\exp\left[\int \kappa\, dl\right]-T_x ,$$

is actually just the formal solution of the equation of radiative transfer under all circumstances. It demonstrates explicitly how the effects of a background source disappear when $T_x \gg T_c$. However, a background point source will still appear as a bright spot in the output of a saturated maser as long as T_c is not negligible in comparison with T_x. Nonetheless, this is not the result of background amplification. It simply reflects the fact that the direction which crosses the background source will saturate earlier and the saturated volume will therefore be larger. The contrast at the bright spot will diminish like T_c/T_x as the maser becomes stronger.

Finally, the effect of collisions has to be reintroduced. This can be easily done, with the result for the population difference Δn then becoming

$$\Delta n=\frac{\Delta P-n_{1,2}C[1-\exp(-h\nu/kT)]}{\Gamma+C[1+\exp(-h\nu/kT)]+2BJ} =\frac{\Delta n_0}{1+J/J_s} . \tag{2.54}$$

This is the same result as the one obtained previously [Eq. (2.39)] if the expressions for Δn_0 and J_s are redefined as

$$\Delta n_0=\frac{\Delta P\left[1-\frac{C}{\Gamma}\cdot h\nu/kT\right]}{\Gamma+2C};\quad J_s=(\Gamma+2C)/2B , \tag{2.55}$$

where we use the fact that $h\nu \ll kT$ for astronomical masers.

The previous discussion of the behavior of the maser radiation thus remains unchanged by the inclusion of collisions, although the magnitudes of Δn_0 and J_s are altered. The collisions do bring a new element into the discussion, though, namely, level thermalization. As is evident from the expression for Δn_0 [Eq. (2.55)], the inversion can disappear altogether if the collision rates become strong enough to dominate over the pump loss rate. This immediately provides a constraint on the flux density from a maser source where the pumping is based on collisions, since the pump rate cannot be increased indefinitely without running eventually into the collision thermalization effect (see also Sec. III.C below). Note, however, that the efficiency of this thermalization effect is reduced by the factor $h\nu/kT$, which is smaller than about 10^{-3} for most sources of interest.

D. Saturation; linewidth

It is evident that a key question in the analysis of any maser source is its degree of saturation, since it determines which properties of the source can be inferred from observations. Unfortunately, this is also a very difficult question, since the maser photons are not tagged by the saturation effects. The question is thus hard to tackle in a precise manner outside the framework of a theoretical model which identifies the underlying pump cycle in the source. With this caveat in mind, we now turn to arguments which suggest that the strong maser sources are probably all saturated.

The first suggestion that the strong masers are saturated comes from their time stability. An unsaturated maser responds exponentially to any fluctuation in the pump rate and would presumably display erratic time behavior. This is of course only a plausibility argument in general, but it can be refined if what is believed to be the underlying pump mechanism is subject to a smooth time variation and if the maser radiation is following this variation in a linear fashion. This is indeed the case for the OH radiation from late-type stars, discussed below in Sec. IV, where time variability (with a time scale of ~ 1 yr) shows that the maser responds linearly to the pump, as expected from a saturated maser. In addition to the linear time response, the maser radiation and the pump rate also obey a linear relation in their source-to-source variations. These sources therefore provide the best clear-cut example of saturated operation.

Another argument for saturation is provided by a number of sources which show no amplification of background radiation in spite of conditions which are favorable for it. As discussed in the previous section (II.C), this is expected in a saturated maser, in contrast with an unsaturated one, which would be amplifying the radiation of a background source. An example of such amplification in the case of a very weak, and certainly unsaturated, maser is provided by the observations of Rieu *et al.* (1976) (see Sec. I.B).

The theoretical analysis of strong maser sources leads invariably to the conclusion that they are saturated. The radiation is in fact so intense, in many cases, that it strains the potential pumps to their limits. Since saturated operation provides the highest efficiency in converting pumping events to maser photons, it would be impossible to understand these sources unless they were saturated.

It appears that maser sources divide into two broad classes: (1) large and relatively low-density ($\lesssim 10^4$ cm^{-3}) clouds, which lead to unsaturated and very weak maser action (gains much smaller than unity), and (2) compact and dense regions, with densities usually in excess of at least 10^5 cm^{-3}, and with very strong and saturated masers. It seems that not too many unsaturated strong masers, namely, with $\tau > 1$, have been detected so far. This rough dichotomy can perhaps be explained with a plausibility argument. Suppose the gain varies linearly with some underlying property of the source, such as the density. The maser output will then vary linearly at small values of this physical parameter, where the unsaturated exponential amplification behaves linearly for small gains, and also at large ones, where the maser saturates. The transition from one type of action to the oth-

er will occur over a short interval of the parameter due to the fast exponential variation at the strong ($\tau > 1$) but unsaturated region. The apparent absence of such sources is then simply the result of statistics, since there are many fewer of them.

A somewhat related question is the one of line narrowing. The analysis in the previous section was performed for the integrated intensity of the maser line, but it is a trivial matter to repeat it with the full frequency dependence retained. The unsaturated absorption coefficient $\kappa_0(\nu)$ is peaked at the line-center frequency, and the unsaturated exponential amplification $\exp[\kappa_0(\nu)l]$ will therefore enhance the center of the line much more than its wings. The amplified line will therefore be narrower than the input line. For a Gaussian frequency distribution it is easy to show that the linewidth is reduced by the amount $(1+\tau_0)^{1/2}$, where τ_0 is the gain at line center. This effect was first mentioned with relation to astronomical masers by Litvak *et al.* (1966). For a saturated maser, on the other hand, the emitted flux is simply proportional to the unsaturated absorption coefficient [Eq. (2.45)], and the linewidth is the same as that of the molecular velocity distribution. The line therefore broadens back to its original width during the saturation process, an effect which can be demonstrated in detail with exact numerical solutions of the line-formation problem (Litvak, 1970).

In practice it does not seem that many clear-cut cases of line narrowing have been established from observations. This is of course a somewhat tricky problem, since we know with certainty only the width of the observed line, not that of the original frequency distribution. As mentioned in Sec. I.B, the requirement of velocity coherence for the masering molecules leads to maser lines which are narrower than ordinary nonmaser lines in the first place. To establish with certainty the operation of the unsaturated amplification line-narrowing effect, one therefore needs to observe a linewidth which is definitely in conflict with thermal broadening. However, the temperatures which can be inferred from the maser lines are usually at least a few tens of degrees and can be reasonably interpreted as the kinetic temperatures in the sources.

A notable exception is a narrow feature in the spectrum of the strong OH source W3(OH) with a linewidth of only 0.1 km/sec or an equivalent temperature of 5 K (Barrett and Rogers, 1966).[3] A possible explanation for this narrow line is the suggestion by Goldreich and Kwan (1974a) that in some cases the maser lines need not rebroaden during saturation. This may happen under certain circumstances when the interaction with other levels maintains a thermal velocity distribution for the maser levels even in the presence of saturation. In the particular model of Goldreich and Kwan (1974a), the maser line remains narrow so long as its stimulated emission rate does not exceed the rate for cycling the molecules through other levels via optically thick transitions.

A problem with this model is the need for fast cycling times, which may require high temperatures. A specific example worked out by Goldreich and Kwan involves a temperature of 600 K. After narrowing by a factor of 5 or so, the linewidth would still correspond to a temperature of about 100 K. It is not clear whether a line as narrow as 5 K could be explained with variations on the standard line-narrowing theory.

The scarcity of clear cases of line narrowing may be related to the above-mentioned division of maser sources into two classes. Narrowing is not observed, because the strong masers are saturated, whereas the unsaturated ones are weak, such that the narrowing effect in them is too small to be detected.

E. Polarization

As in the case for laboratory lasers, maser radiation from astronomical sources is polarized in many, though not all, cases. The polarization sometimes approaches 100%. The only general method advocated so far for producing maser polarization, irrespective of the details of the pumping mechanism, is through the effects of a magnetic field which is assumed to permeate the maser source. The calculation is then performed along lines similar to those described in Sec. II.B, but account must now be taken of the full structure of the maser levels, of the propagation of different polarization modes, and of the effects of the magnetic field, which adds another time scale to the problem (the inverse gyrofrequency of the maser levels) as well as another characteristic direction. The complexity of the problem is evident from its description alone.

Goldreich, Keeley, and Kwan (1973a) undertook the heroic task of tackling this problem for a maser which operates between $J=1$ and $J=0$ levels. Although many specific numerical coefficients which they derived undoubtedly depend on this choice of angular momenta, the general characteristics of their results should probably be independent of the detailed level structure.

The equations are solved using different approximations according to all possible relations among the various characteristic time scales. The results are obviously quite involved and will not be reproduced here. The most general, and perhaps obvious, result is that when the Zeeman splitting of the levels is larger than the linewidth, the maser by and large amplifies the Zeeman pattern. This can be applicable to OH for magnetic fields in excess of about a milligauss, which are likely to exist in high-density regions, but not to SiO and H_2O, which are nonparamagnetic, with g factors of about 10^{-3}. The results of Goldreich *et al.* show, however, that even in this case it is possible to obtain linear polarization, provided the gyrorate exceeds BJ_m, the rate for induced transitions. Indeed, linear, but not circular, polarization has been detected in the maser radiation of

[3]The other features in the profile are broader with equivalent temperatures up to ~ 150 K.

these molecules, and an explanation with a magnetic field is a possibility.

In a subsequent paper, Goldreich, Keely, and Kwan (1973b) incorporated in their calculations the effects of interaction of the maser levels with other molecular levels. The photons which are trapped in such transitions, are cycling the population among the various magnetic substates of the maser levels, changing the polarization properties of the radiation. In particular, if maser radiation is beamed in a given direction which is chosen as the quantization axis, then only $\Delta m = \pm 1$ transitions can couple to the radiation. The interaction with the trapped photons [next section (III)] changes this and allows coupling between levels corresponding to $\Delta m = 0$, because the trapped radiation is not beamed.

A homogeneous medium with a magnetic field, as in the calculations of Goldreich *et al.*, cannot distinguish between left and right, leading to equal strength for the two components of circular polarization in the sources in which they are present. However, the OH masers which are associated with HII regions (see Sec. IV.C below) are comprised of components which are usually very strongly (close to 100%) circularly polarized in one sense only. Maser spots with a different sense of polarization are sometimes located close enough that they can be considered a Zeeman pair of circularly polarized radiation from the same region, but there are many more cases of maser spots which are clearly spatially removed from any possible Zeeman companion.

The only plausible explanation for this phenomenon, so far, is an ingenious filter mechanism which was proposed independently by Cook (1966b) and by Shklovskii (1969). The idea is as follows. The energy of the mth magnetic sublevel is shifted in a magnetic field by the amount $mg\hbar\Omega$, where g is the Landé factor and Ω is the gyrofrequency. Suppose the radiation is propagating in a medium where the magnetic field, and hence Ω, is increasing along the ray path. The separation between the Zeeman components will then increase continuously. Suppose also that the source is subject to a large-scale velocity field. If the velocity is also varying along the ray path, the Doppler and Zeeman shifts can be matched for only one of the polarization modes, if at all. It is possible, then, to construct a filter that will amplify only one polarization component.

The idea that the gradients of the velocity and magnetic fields are correlated may not be as *ad hoc* as it first sounds. The degree of ionization in interstellar regions is sufficient to ensure magnetic flux freezing which leads to a coupling of the motions and the magnetic field.

The mere fact that linear polarization is observed in a given source implies that there is some preferred direction in it, perpendicular to the line of sight. A magnetic axis is one obvious way to define such a direction, but not the only one. Suppose the maser molecules are inverted through interaction with a stream of either photons or other particles. The streaming direction can provide an axis of alignment for the magnetic sublevels, leading to polarized radiation. Indeed, models have been proposed to provide alignment and polarization through the interaction with a directional radiative flux (Burdyuzha and Varshalovich, 1972; Varshalovich and Burdyuzha, 1975) or an electron stream (Johnston, 1967). In each case, the polarization produced is linked to the inversion process itself.

III. PUMPING CONSIDERATIONS

A. Radiative transfer—the escape probability

The conditions in most strong astronomical masers are such that many transitions which connect the maser levels with other states are optically thick. These include, many times, the pump or loss transitions of the pumping cycle. The analysis of potential pumping schemes therefore often involves the problem of the transfer of radiation in optically thick lines. This has a significant effect on the basic pumping considerations, since the trapping of radiation in these lines slows down their effective transition rates. The significance of radiation trapping for pumping considerations of astronomical masers was first recognized by Litvak (1969b).

The difficulties introduced by the effects of radiative transfer stem from the coupling between the distribution of population between any given pair of levels and the intensity of the radiation in the line which connects them. Because of this coupling, which occurs through the effects of absorption and stimulated emission, a complete treatment of the problem requires a simultaneous solution of the equations of populations for all the relevant levels and the transfer of radiation in all the lines which couple them. This is practically impossible for any model other than the most simple ones. A practical treatment of the problem therefore requires some approximation method to decouple the equations of level populations and radiative transfer. The most popular such method is the formalism of escape probability, first introduced by Sobolev (1958), which we shall now examine.

Consider two levels denoted by i and j with $E_i > E_j$ and $g_i = g_j$. Consider, also, that part of the ith rate equation corresponding to population exchange between them due to radiative processes generated internally, namely,

$$\frac{dn_i}{dt} = -A_{ij}n_i - B_{ij}J_{ij}(n_i - n_j) , \tag{3.1}$$

where J_{ij} is the angle-integrated intensity produced within the source in the i-j line. Since this is a nonmasering line, the radiation is usually isotropic and $J_{ij} = I_{ij}$. The essence of the approximation is in replacing the radiation I_{ij}, which is of course determined by an integral over the entire source, with the locally produced radiation which is given by the source function S_{ij} [Eq. (1.13)], with the local values for n_i and n_j. The assumption is therefore that the strongest effect on the populations is from radiation produced in the immediate vicinity. But obviously only that fraction of the radiation

which is absorbed by the molecules is affecting their population, so that if β_{ij} denotes the fraction of the produced radiation which escapes the source, the final form of the approximation is

$$I_{ij} \simeq (1-\beta_{ij}) S_{ij} \ . \tag{3.2}$$

This leads to the following final result for Eq. (3.1):

$$\frac{dn_i}{dt} = -\beta_{ij} A_{ij} n_i \ , \tag{3.3}$$

where we use the expression for the source function [Eq. (1.13)] and the Einstein relation [Eq. (1.12)].

The result derived has the desired effect of decoupling the radiation from the rate equations, since it does not involve the intensity of the radiation, but only the probability that it escapes from the source. The intuitive meaning of the equation is rather simple: a spontaneous decay of level i does not necessarily lead to a decrease in its population, because the emitted photon may be absorbed elsewhere in the source and excite a molecule back to the same level. Only that fraction of the photons which escape the source leads to a change in the population. The rate of change is then the rate of decay ($A_{ij} n_i$) times the escape probability (β_{ij})—hence Eq. (3.3). Another way of looking at the problem is that the trapping of photons due to absorption and reemission in the line is effectively slowing down its transition rate by the number of such "scattering" events that a photon undergoes on the average, which is simply the inverse of the probability for a photon escape.

The effects of radiative transfer have essentially been lumped into the escape probability β, which has to be calculated once and for all in any given model. A much simplified calculation can again be performed in the one-dimensional model. For a radiation intensity I entering a slab with an optical depth τ, the intensity of radiation at exit is $I \cdot e^{-\tau}$ and the probability for radiation to escape from optical depth τ is therefore $e^{-\tau}$. For radiation from inside the source we do not know the precise point of production and would need to perform an average over optical depths. The mean escape probability is then

$$\beta = \langle e^{-\tau} \rangle = \frac{\int_0^\tau e^{-\tau'} d\tau'}{\int_0^\tau d\tau'} = (1-e^{-\tau})/\tau \ . \tag{3.4}$$

More formal derivations usually lead to a similar expression for the escape probability with possible differences arising from different averaging methods as prescribed by the specific model used.

A frequently used model which is especially suitable for the escape-probability method is that of a radial flow with a velocity gradient. A photon produced anywhere in the source will escape unless abosrbed locally, since the transition frequency elsewhere in the source is Doppler shifted because of either the velocity gradient for molecules on the same ray, or the different direction for molecules on other rays. This model was studied in an especially elegant manner by Castor (1970), and his result for the escape probability β is identical to the one derived above [Eq. (3.4)] with the addition of an averaging over directions.

The behavior of the escape probability in the limits of large and small optical depths is given by

$$\beta = \begin{cases} 1, & \tau \ll 1 \\ 1/\tau, & \tau \gg 1 \ . \end{cases} \tag{3.5}$$

These limits are easy to understand. In the optically thin case ($\tau \ll 1$) the medium is essentially transparent, and each photon escapes the source without interaction, so that $\beta = 1$. In the optically thick case the source can be thought of as τ slices with an optical depth of one. A photon will escape, on the average, if and only if it is produced in the outermost slice, and the probability for that is $1/\tau$.

The transition rate for an optically thick line is therefore effectively reduced to A/τ, which is independent of the line strength, since both A and τ are proportional to it [Eqs. (1.12) and (1.20)]. This happens because a transition with a larger dipole moment is more effective in both absorbing and reproducing photons and because the line-strength dependence of these two effects is the same.

The decrease in transition rates for optically thick lines has an important effect on considerations of line thermalization. A two-level system thermalizes when the downward transitions are collision dominated (Sec. II.A). This can thus happen at lower densities for optically thick lines where the effective radiative decay rate is reduced by $1/\tau$.

The escape probability can also be used to describe the interaction with external radiation. The probability of a photon emitted by an external source to reach a particular point inside the cloud is obviously equal to the probability of a photon emitted at that point to escape the cloud. The interaction with the external radiation, with intensity J_r, can therefore be described by adding the following term to the rate equations:

$$\beta_{ij} B_{ij} J_r(\nu_{ij}) (n_i - n_j) \ . \tag{3.6}$$

If the solid angle subtended at the source of interest by the external radiation source is $\Delta\Omega_r$, then

$$J_r = I_r \frac{\Delta\Omega_r}{4\pi} \ , \tag{3.7}$$

as in Eq. (2.29). The intensity I_r is usually described with the Planck function at the relevant temperature T_r, and the geometrical factor in Eq. (3.7) is usually referred to as a "dilution factor" W given by

$$W = \frac{\Delta\Omega_r}{4\pi} = \frac{1}{4} \left[\frac{R}{r} \right]^2 \ , \tag{3.8}$$

where R and r are the radius and distance, respectively, of the external source. The final form of the interaction with the external radiation is then

$$\frac{W \beta_{ij} A_{ij} (n_i - n_j)}{\exp(h\nu_{ij}/kT_r) - 1} \ . \tag{3.9}$$

The complete form of the ith rate equation, with col-

lisions and external radiation included, is then

$$\frac{dn_i}{dt}=-\sum_{i>j}\{A_{ij}\beta_{ij}[n_i+W\tilde{I}(\nu_{ij})(n_i-n_j)] +C_{ij}[n_i-n_j\exp(-h\nu_{ij}/kT)]\} +\sum_{j>i}\frac{g_j}{g_i}\{A_{ji}\beta_{ji}[n_j+W\tilde{I}(\nu_{ji})(n_j-n_i)] +C_{ji}[n_j-n_i\exp(-h\nu_{ji}/kT)]\} , \tag{3.10}$$

where

$$\tilde{I}(\nu_{ij})\equiv I_r(\nu_{ij})\frac{c^2}{2h\nu_{ij}^3}=[\exp(h\nu_{ij}/kT_r)-1]^{-1} . \tag{3.11}$$

For a given model with L levels, only $L-1$ of the equations (3.10) are independent. They are supplemented by

$$\sum_{i=1}^{L} g_i n_i = N_m , \tag{3.12}$$

where N_m is the total number density of the maser molecule. The steady-state level populations can thus be obtained from the solution of a set of algebraic equations without the need to solve the radiative transfer problem. Because of the dependence of β on the level populations, the equations are nonlinear and can be solved only numerically, in general. Once the population distribution is obtained, the intensity of emission in each line can easily be calculated, if desired.

The escape probability method applies also to the maser levels themselves, which are among the L levels whose populations are described by Eqs. (3.10) and (3.12). For an inverted transition, $\beta>1$, since in this case a single photon can lead to the production of many more.

B. Radiative pumps

For a pump dominated by radiative processes, the collision terms in Eq. (3.10) can be neglected. Suppose, first, that each of the maser levels exchanges populations with only one reservoir level. The rate equation then contains just one term and can be solved analytically, irrespective of the optical depths. If the population per sublevel of the appropriate reservoir level is denoted by n_0, the steady-state solution of Eq. (3.10) is

$$\frac{n_i}{n_0}=\frac{W\tilde{I}(\nu_{i,0})}{1+W\tilde{I}(\nu_{i,0})}=\frac{W}{\exp(h\nu_{i,0}/kT_r)-(1-W)} . \tag{3.13}$$

It is quite easy to show that if the population distribution among the reservoir levels is also determined by the radiation $I(\nu)$, the maser levels cannot invert so long as the radiation deviates from that of a blackbody by only a constant dilution factor. To produce inversion, in this scheme, the external radiation field must deviate from the frequency distribution of a blackbody in some specific manner.

It is evident that, in general, a pumping cycle which involves only radiative processes must involve different levels for the gain and loss of the maser levels, because when each maser level is losing molecules to the same level from which it was pumped, an inversion is impossible. The significance of this fundamental point is not well recognized, and a large number of radiative pumping schemes proposed for OH have involved population exchange with a single rotation state.

Another general result that follows from Eq. (3.13) is that for $W=1$, the level populations obey the Boltzmann distribution with temperature T_r. In fact, it is easy to see that with collisions neglected and $W=1$, the Boltzman distribution with temperature T_r is a solution of Eqs. (3.10) for an arbitrary number of levels and not for just a two-level model. This result is quite obvious on general grounds—if the molecules' only interaction is with a blackbody radiation field, the level populations obviously have to equilibrate with the temperature T_r. To produce an inversion with radiative pumping in a model which involves more than one reservoir level, it is still essential that the radiation field deviate from that of a blackbody. The deviation in this case, however, may involve only a constant dilution factor.

The simplest modification of the blackbody law is of course $W<1$, which is the result of geometrical dilution when the radiative pump source is external. Another common situation is when the pump radiation is due to internal emission by dust particles which are mixed with the source molecules. For a dust temperature T_d, the intensity of the dust radiation is

$$I_d(\nu)=B_\nu(T_d)\{1-\exp[-\tau_d(\nu)]\} , \tag{3.14}$$

where $\tau_d(\nu)$ is the dust's optical depth at the frequency ν. For optically thick dust, the intensity then follows the blackbody law with temperature T_d, but for optically thin dust [$\tau_d(\nu)<1$], the intensity is

$$I_d(\nu)=\tau_d(\nu)B_\nu(T_d) . \tag{3.15}$$

This corresponds to a blackbody which is not only diluted [since $\tau_d(\nu)<1$], but also distorted because of the frequency dependence of the dust absorption coefficient. The dust emission can be described by the diluted blackbody law used above if we agree to broaden the definition of W so that it can correspond to the dust's optical depth when necessary.

We have just derived two general conditions which must be obeyed by any radiative maser pump: (1) the radiation field cannot be a pure undiluted blackbody for any pumping scheme and (2) for a pumping cycle involving just a single level for the gain and loss, the deviation from blackbody of the external radiation field cannot amount to just a simple frequency-independent dilution factor but must involve some specific modification of the

frequency distribution. These two requirements are rather general and apply irrespective of the optical thickness of the various transitions in the pumping scheme.

We proceed now to estimate the number of photons that can be emitted from a radiatively pumped maser source. For a maser intensity I_m the quantity of interest is F_m, the net maser flux out of the source, given by

$$F_m = \int_{(2\pi)} I_m \cos\theta \, d\Omega \ , \tag{3.16}$$

where the integral is performed only over the outward direction. For maser radiation which is beamed into $\Delta\Omega_m$ the expression for F_m is

$$F_m = I_m \Delta\Omega_m = 4\pi J_m \ , \tag{3.17}$$

whereas for isotropic radiation

$$F = \pi I = \pi J \ . \tag{3.18}$$

The number of photons, N_m, emitted by a source with surface area ΔS and length Δl is

$$N_m = \Delta S \frac{F_m}{h\nu_m} \Delta\nu_m = \Delta S 4\pi J_m \frac{\Delta\nu_m}{h\nu_m} \ . \tag{3.19}$$

But J_m is related to the maser volume production rate Φ_m via [see also Eq. (2.44)]

$$J_m = \frac{\Phi_m}{4\pi} \Delta l \frac{h\nu_m}{\Delta\nu_m} \ , \tag{3.20}$$

so that

$$N_m = \Phi_m \Delta S \, \Delta l = \Phi_m \Delta V \ , \tag{3.21}$$

where ΔV is the volume of the maser source. If the maser were emitting isotropically, the expression for N_m would be reduced by a factor of 4.

For a saturated maser, Φ_m is simply proportional to the volume pump rate P [Eq. (2.43)], so that finally

$$N_m = \eta P \Delta V = \eta P_t \ , \tag{3.22}$$

where P_t is the total pump rate in the source. This result is a general one—the nature of the pump has not yet been specified.

Consider now a maser which is pumped by a radiation source at a distance r. Denote the population per sublevel at the maser reservoir by n_0. Using the results of the previous section, the pump rate P is given by

$$P = W_p \beta_p B_p I_p n_0 \ , \tag{3.23}$$

where the subscript p specifies quantities corresponding to the pump transition.

Consider first the case where the maser source is optically thick at the pump transition, namely, $\tau_p > 1$. The escape probability β_p is then $1/\tau_p$. If we insert the explicit expressions for τ_p [Eq. (1.20)] and for the geometric dilution factor W_p [Eq. (3.8)] and neglect the population of the maser levels in comparison with that of the reservoir, the expression for P becomes

$$P = \frac{\pi I_p}{h\nu_p} \Delta\nu_p \left[\frac{r_p}{r}\right]^2 1/\Delta l \ , \tag{3.24}$$

where r_p is the radius of the pump radiation source. If the solid angle subtended by the maser at the pump source is denoted by $\Delta\Omega (= \Delta S/r^2)$, the total pump rate can be written as

$$P_t = P r^2 \Delta\Omega \, \Delta l = N_{s,p} \frac{\Delta\Omega}{4\pi} \ , \tag{3.25}$$

where

$$N_{s,p} = 4\pi r_p^2 \frac{\pi I_p}{h\nu_p} \Delta\nu_p \ . \tag{3.26}$$

Since the pump source is emitting isotropically, $\pi I_p = F_p$ and $N_{s,p}$ is just the total number of pump photons emitted by the pump source per unit time. Obviously, then, P_t is simply the number of pump photons emitted toward the maser source per unit time. Since the maser is optically thick in the pump transition, it absorbs each pump photon impinging upon it, and the result just obtained can be written as

$$P_t = N_p \ , \tag{3.27}$$

i.e., the total pump rate is equal to the number of pump photons absorbed by the maser per unit time, N_p.

For a maser source which is optically thin at the pump transition, namely, $\tau_p < 1$, the escape probability β_p is 1. We can still insert $\beta_p = (1/\tau_p)\tau_p$ and proceed as before. The final result for the total pump rate P_t is then identical to Eq. (3.25), only multiplied by τ_p. Since τ_p is now smaller than 1, it is also equal to the fraction of impinging pump photons that the maser actually absorbs. To see that, notice that the equation of radiative transfer can also be written as

$$d\tau = -\frac{dI}{I} \ , \tag{3.28}$$

so that τ is simply the absorption probability for small optical depths. Equation (3.27) is therefore applicable for $\tau_p < 1$, as well, and we have just derived the following general result, which holds irrespective of the values of optical depths: the pump rate at a radiatively pumped maser is equal to the number of pump photons absorbed by it per unit time. When this is combined with Eq. (3.22), we obtain

$$N_m = \eta N_p \ , \tag{3.29}$$

and the number of maser photons emitted cannot exceed the number of pump photons absorbed.

It is clear that the most efficient maser operation is provided by an optically thick spherical shell around the pump source, since it absorbs all the pump photons emitted. If in addition the maser is beamed outward and is saturated, it will produce η photons for each pump photon emitted by the pump source, which is therefore its highest possible output.

This provides a very powerful constraint on any potential radiative pump model. If a certain pump is advocated for a given maser source, the number of photons emitted by the pump source in the proposed pump lines must exceed the number of maser photons, which is

directly determined from observations. An even stronger constraint can be obtained if we have reason to believe that any of the above efficiency requirements are not met, namely, that the maser is optically thin, does not fully surround the pump source, emits isotropically, or is unsaturated. Although the constraint on the pump model obtained in such a case would be tighter, the confidence in it would have to be smaller because of the extra assumptions made.

The results derived above can be used to calculate the brightness temperature of the maser radiation. To do that, Eq. (3.21) can be written as

$$N_m = B_{\nu_m}(T_m)\Delta\Omega_m \Delta S \frac{\Delta\nu_m}{h\nu_m} . \tag{3.30}$$

For an optically thick maser, Eqs. (3.25), (3.26), and (3.27) lead to

$$N_p = B_{\nu_p}(T_p) 4\pi W_p \Delta S \frac{\Delta\nu_p}{h\nu_p} ; \tag{3.31}$$

and, as mentioned before, this result can be used also for cases in which the dilution factor W_p has a nongeometrical interpretation. When these two equations are inserted into Eq. (3.29), the brightness temperature of the maser radiation, T_m, can be solved as a function of the pump brightness temperature, T_p. The solution can be derived analytically if both temperatures are in their respective Rayleigh-Jeans domains. It reads

$$T_m = \eta W_p \left[\frac{4\pi}{\Delta\Omega_m}\right]\left[\frac{\Delta\nu_p/\nu_p}{\Delta\nu_m/\nu_m}\right]\left[\frac{\nu_p}{\nu_m}\right]^2 T_p . \tag{3.32}$$

If the dilution factor W_p is due to geometrical effects, the equation becomes

$$T_m = \eta \left[\frac{\Delta\Omega_p}{\Delta\Omega_m}\right]\left[\frac{\nu_p}{\nu_m}\right]^2 T_p , \tag{3.33}$$

where $\Delta\Omega_p$ is the solid angle which the pump source subtends at the maser location, and where the same fractional bandwidth is assumed for the maser and pump radiation, which is usually the case. Note also that this equation demonstrates again that the power output of a saturated maser depends only on the pump properties and that it is independent of the input source radiation.

The high brightness temperatures that can be achieved by maser radiation are due mainly to the factor $(\nu_p/\nu_m)^2$, which reflects the difference in photon phase space density between the pump and maser frequencies. The inverted population enables the maser to shift efficiently high-frequency photons to low frequency, where their number is much higher than that allowed in thermal equilibrium. This point is further illustrated in Fig. 5, which assumes a $T_p = 500$ K radiative pump involving the OH rotation transition at a wavelength of 35 μ (Fig. 1). Assuming that the pump photons are converted to 18-cm maser photons with 50% efficiency, we find the resulting brightness temperature to be in excess of 10^9 K.

The upper bound on brightness temperature in Eq. (3.32) is of course the same as Eq. (3.29), since a blackbody provides the maximum number of pump photons possible. On the face of it, it would seem advantageous to pump with cycles which involve high-frequency transitions, such as vibration or electronic excitations for OH, since these can lead to higher maser intensities. However, this would also require strong sources of radiation at those frequencies, which are often not available. In addition, the inherent efficiency of such pumps, as measured by the parameter η, is usually low for the conditions which prevail in astronomical masers.

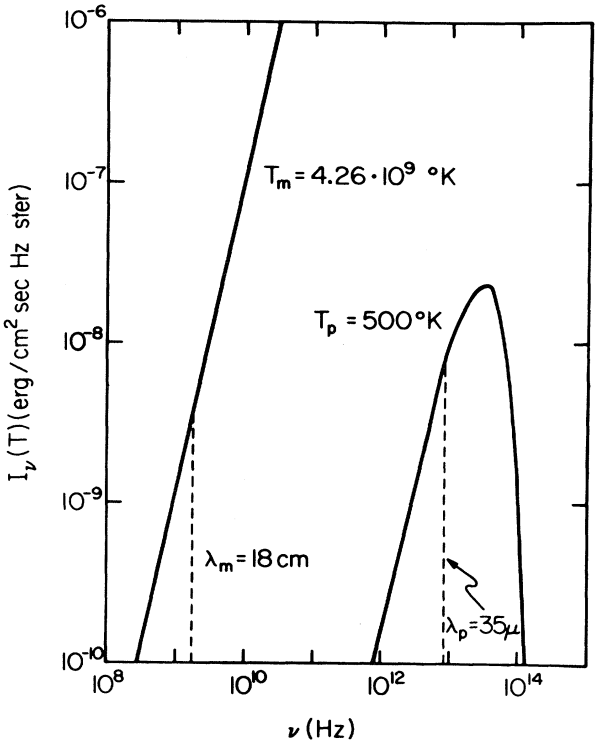

FIG. 5. The solid lines are plots of the Planck distribution function $B_\nu(T)$ at temperatures T_m and T_p, as marked on the curves. The temperature T_m was determined from the condition $B_{\nu_m}(T_m) = \frac{1}{2} B_{\nu_p}(T_p)$.

Finally, although the number of photons is roughly conserved in the pumping cycle, their energy is degraded by ν_p/ν_m. The same factor which produces the spectacular brightness temperature also ensures that the energy flux carried in the maser lines is never of importance in the energy balance of the maser source. Detection instruments are of course counting photons, though, and without the maser effect the line emission of many sources would be undetectable.

C. Collisional pumps

When the pumping is due to collisional excitations, the pump rate, as is evident from Eq. (3.10), is

$$P = C_p n_0 , \tag{3.34}$$

where C_p is the average rate for collisional excitation from the reservoir to the maser levels. The total maser

photon emission rate is then

$$N_m = \eta C_p n_0 \Delta V \ . \tag{3.35}$$

This relation does not constrain the pump in any obvious manner similar to Eq. (3.29) for a radiative pump.

Collisional pumps are constrained, however, because collision rates which are too high would thermalize the level populations and quench the maser altogether. One thermalization effect of this kind was already encountered during the general discussion of Sec. II.C. If the collision rate across the maser levels is denoted by C_m, it follows from Eq. (2.55) that the maser can operate only so long as

$$C_m < \Gamma \frac{kT}{h\nu_m} \ . \tag{3.36}$$

If the loss rate Γ is due to radiative processes, this equation puts a limit on the density of colliding particles which is allowed in the source, leading to a limit on the number of maser photons N_m which can be produced.

In addition to this general limit on the density, each pumping scheme usually provides additional, more specific, constraints on the collision rates. Using again the SiO for illustration, if the vibrational transitions are excited by collisions, then their decays (the loss process) must be radiative, because otherwise a Boltzmann distribution would be established. Hence

$$C(v=1\rightarrow 0) \lesssim \beta A (v=1\rightarrow 0) \ . \tag{3.37}$$

This constraint is similar to the previous one [Eq. (3.36)]. The question of which one is tighter depends on the relative strength of collision rates for vibration and rotation excitations for SiO. Similar arguments can be constructed for inversion schemes based on collisions for essentially every maser.

Since the loss mechanism of a collisionally excited SiO maser must be radiative decays, each pumping event would lead to the production of a photon at the frequency of the vibration transition. The number of photons emitted by the source in the vibration lines must therefore be at least as high as the number of maser photons it emits. Similar reasoning applies for other masers, and collisional pumps essentially always lead to the production of photons in certain high-frequency lines with a rate which must exceed that for the maser photons.

From the experimental point of view, this constraint is not as powerful as the seemingly similar Eq. (3.29) for radiative pumps. The reason is that the number of radiative pump photons emitted by the source is not hard to determine, since they are part of a continuum spectrum which can be studied relatively easily. In contrast, the photons produced during a collisional pumping cycle are emitted only in relatively narrow lines, which are usually hard to detect. Hence, the derivation of useful observational constraints on collisional pumps usually requires a detailed study of the particular pump model and the calculation of some observable continuum radiation associated with it. It is almost always possible to identify such radiation, because collisional pumping requires some minimum temperature for the colliding particles and hence a heat source which must be radiating at least a certain amount of energy. However, the relation between the characteristic temperature of such a heat source and the kinetic temperature of the colliding particles in the maser region always involves some model calculations, and the constraints on collisional pumps are not as straightforward as those on radiative ones.

A question frequently asked is which type of pumping, collisional or radiative, can in principle lead to the production of more maser photons in a given source. The above discussion makes it clear that this question can be answered only in a very qualitative way through order-of-magnitude estimates. Taking the SiO as an example again, the starting point would be Eq. (3.37). Using detailed balance, we can relate the collision rate for vibrational decay to that for excitation, which is just the collisional pump rate C_p. For an efficient pump, this should be as high as possible and obey, in order of magnitude, the equality sign in Eq. (3.37), which leads to

$$C_p \simeq \beta_p A_p \exp(-h\nu_p/kT) \ . \tag{3.38}$$

But $\beta_p A_p$ is directly related to the radiative pump rate in the $v=0\rightarrow 1$ transition [Eq. (3.23)]. For an optically thick source which surrounds the radiation pump source, the total collisional and radiative pump rates are then related via

$$P_t(\text{coll}) \simeq P_t(\text{rad}) \exp(-h\nu_p/kT) \times [\exp(h\nu_p/kT_r) - 1] \ . \tag{3.39}$$

The kinetic temperature, T, and the radiative pump temperature, T_r, are expected to be similar, in order of magnitude, and large enough to enable efficient excitations. It then follows that, in order of magnitude, the total pump rates for collisional and radiative excitations are expected to be roughly the same.

If both pumps were leading to a saturated maser action with the same efficiency η, they would then produce roughly the same number of maser photons. However, a radiative excitation of the $v=0\rightarrow 1$ transition corresponds to a radiative pump which involves the same transition for pump and loss and would never lead to inversion, as discussed above (Sec. III.B). Inversion is possible if the pump excitation is through a higher vibration level, for instance (Kwan and Scoville, 1974).

Apart from the obvious point that inversion is sometimes possible through pumping in a certain transition with one mechanism but not the other, the discussion illustrates that the total possible pump rate in that transition would be approximately the same for collisions and radiation if the overall conditions are close to thermal equilibrium in the pump lines. Although widely different rates can be achieved in some specially concocted models, it can be argued that the conditions in most strong maser sources should be such that they are indeed close to thermal equilibrium in the pump lines. This is because the largest maser fluxes are obtained when the pump, either radiative or collisional, is operating as close to thermalization as possible. For a radiative pump, a

strong maser requires W to be as close to 1 as possible, whereas for a collisional pump, the collision rates would have to be as large as possible. In either case, increasing the pump intensity a bit further would thermalize the levels, as discussed above.

D. Chemical pumps

An obvious explanation for any inversion effect is that the molecules production mechanism somehow favors the upper level of the maser transition. This idea may seem particularly attractive for OH, since the laboratory production of this molecule in flames does indeed lead to inverted populations in excited electronic levels which could conceivably be transferred to the ground state.

It is quite easy to show, however, that this idea cannot work for astronomical masers in steady state (Litvak, 1969a). The argument is as follows. At steady state, the OH production and destruction rates are equal. For the temperatures which can exist in the maser sources, the most likely destruction mechanisms are reactions with species other than H or H_2, even though they are the most abundant, because of the high energy barriers for hydrogen exchange. Taking into account the low relative abundance of any other element, it follows immediately that the rates for OH excitations through collisions with hydrogen are much larger than those for its destruction in interaction with any other species. The rates for collisional pumps, therefore, always dominate those for chemical pumps for masers in steady state.

Inversion through chemical pumping can thus be of significance only as a very short, transient effect. This holds true for other molecules, as well, and for this reason chemical pumps did not gain much popularity and will not be discussed here any further.

IV. OH

The OH molecule is somewhat unique in that a maser effect occurs in its ground state which contains most of the population and which is therefore also the maser reservoir. It is then not immediately obvious what the pump and loss mechanisms are. This is not a real difficulty, though.

To demonstrate how this is handled, consider for illustration purposes a model in which the four ground-state levels interact with just one other level and in which the rate for such population exchange is denoted by $R_{i,0}(i=1,4)$. In steady state, the population of the external level, n_0, is then given by

$$n_0=\frac{1}{R_0}\sum_{i=1}^{4} n_i R_{i,0} , \tag{4.1}$$

where

$$R_0=\sum_{i=1}^{4} R_{0,i}$$

is the total decay rate (inverse lifetime) of the external level.

The terms in the rate equation for the ith ground-state level corresponding to interaction with the external level are

$$\frac{dn_i}{dt}=-n_i R_{i,0}+n_0 R_{0,i} . \tag{4.2}$$

We can now identify the pump and loss rates for each ground-state level as follows:

$$\Gamma_i=R_{i,0} \tag{4.3}$$

and

$$P_i=n_0 R_{0,i}=P_{\rm tot} T_{0,i} ,$$

where

$$P_{\rm tot}=\sum_{i=1}^{4} n_i R_{i,0}$$

is the total rate for pumping OH molecules out of the ground state, and

$$T_{0,i}=R_{0,i}/R_0$$

is the probability that a molecular will choose the ith level during its cascade back to the ground state.

The expression for the rate $P_{\rm tot}$ can be further simplified because in lowest order, the ground-state sublevels are roughly equally populated and contain the bulk of the OH density. Hence

$$P_{\rm tot}\simeq N_{\rm OH}\sum_{i=1}^{4} R_{i,0}\Big/\sum_{i=1}^{4} g_i . \tag{4.4}$$

The results of this simple model are easy to generalize. The loss rate for each ground-state level is its rate of excitation to the higher levels, and its pump rate, per OH molecule, is the total rate of pumping out of the ground state times the probability that that particular level will be chosen in the cascade back to the ground state.

A. Satellite-line inversions

The OH satellite lines are almost never in thermal equilibrium. In most extended sources, the main lines appear in either weak emission or absorption, and the ratio of these two lines is in agreement with local thermodynamic equilibrium. Even in these cases, however, the satellite lines almost always exhibit an anomalous pattern. One of the lines usually appears in absorption and the other in emission. Even the line which behaves similarly to the main lines shows a strength of emission, or depth of absorption, exceeding that which would be expected from the main lines' behavior. The excitation temperature which is inferred for the emitting satellite line, is either negative or, within the errors, positive and extremely high.

A glance at the OH energy-level diagram (Fig. 1) reveals that such a pattern could be explained with a mechanism that transfers molecules between levels with different total angular momentum F, but not among the

two halves of the Λ doublet. A transfer of molecules from $F=1$ to $F=2$, say, within the same Λ-doublet component (same parity), would lead to an inversion of one satellite line (the 1720-MHz line) and "anti-inversion" of the other (the 1612 line), while the main lines remained in equilibrium. The opposite transfer, from $F=2$ to $F=1$, would reverse the roles of the two satellite lines, without affecting the main lines. A mechanism for such population exchange, capable of operation in a wide range of parameters, would provide an adequate explanation for the observations.

It is evident that the sought-after mechanism can involve only pumping of the OH rotation levels because vibration and electronic excitations require radiative fields which are not available in general. Among the rotation levels, only those which connect directly to the ground state need be considered, because the rates for double excitations can always be safely neglected. This is the reason for the selection of the levels in Fig. 1.

Assuming that most of the OH molecules are in the ground state, where they are distributed roughly equally among all magnetic sublevels, which is always approximately correct, the optical depth, τ_{IR}, of the strongest transition to the first rotational level [$^2\Pi_{3/2}(J=\frac{5}{2})$; see Fig. 1] is

$$\tau_{IR}=4\cdot10^{-14}\frac{N_{OH}l}{\Delta v}\ , \tag{4.5}$$

where Δv is the linewidth in km/sec, $N_{OH}(\text{cm}^{-3})$ is the OH density, and l(cm) is the source dimension. Similarly, the optical depth of the 1667 line, the strongest ground-state transition, is given by

$$\tau(1667)=4.5\cdot10^{-15}\frac{N_{OH}l}{\Delta v}\frac{1}{T_x(1667)}\ . \tag{4.6}$$

Since the excitation temperature, $T_x(1667)$, is typically in the range of $\sim5-10$ K, the rotation transition optical depth is about $50-100$ times larger than that for a ground-state transition, which has to be at least $\sim10^{-1}-10^{-2}$ in a detectable source. The transitions to the first rotational excited state are therefore usually optically thick.

The OH molecules are pumped by collisions or radiation from the ground state to the rotation levels depicted in Fig. 1 and then cascade back. It is easy to show, and the numerical calculations confirm, that the most important step in the pump cycle is the cascade back to the ground state, which determines the probability T_{ij} [Eq. (4.3)]. Consider a cascade route whose last step is the decay $^2\Pi_{3/2}(J=\frac{5}{2})\rightarrow{}^2\Pi_{3/2}(J=\frac{3}{2})$. This transition is optically thick, as demonstrated, and the line strengths for its various branches therefore drop out of the calculation, as shown above (Sec. III.A). We need to consider, then, only the number of transitions to any given level. The $F=2$ levels of the $^2\Pi_{3/2}(J=\frac{5}{2})$ state can decay to either of the ground-state levels, but the $F=3$ levels decay only to $F=2$, because a decay to $F=1$ would involve a forbidden $\Delta F=2$ transition. A pumping cycle which ends with this step is then effectively transferring population from each ground-state $F=1$ level to $F=2$ of the same parity, thus producing the pattern of 1720 inversion and 1612 "anti-inversion" with the main lines in equilibrium.

Precisely the opposite would happen during a cascade from $^2\Pi_{1/2}(J=\frac{1}{2})$ to the ground state, because in this case the transition from $F=0$ to $F=2$ is forbidden. A pumping cycle which ends with this transition would therefore lead to the opposite effect of 1612 inversion and 1720 "anti-inversion." Rotational excitations can therefore explain the observed patterns of satellite line anomalies in a most natural way.

When the two inversion mechanisms compete with each other, namely, when the cascades from $^2\Pi_{3/2}(J=\frac{5}{2})$ and from $^2\Pi_{1/2}(J=\frac{1}{2})$ are equally important, the 1612 inversion always wins. The explanation of this effect requires a more detailed study of the cascades to the ground state but is also easy to understand on more general grounds. The 1612 inversion involves the overpopulation of the three sublevels of $F=1$, which is easier than that of the five sublevels of $F=2$, required for the 1720 inversion. The population per sublevel is of course the relevant one for the optical depths [Eq. (1.20)]. The production of a 1720 maser requires, then, that the inversion effect based on cascades from $^2\Pi_{1/2}(J=\frac{1}{2})$ be suppressed. This can happen in either of the situations discussed below.

The OH column density may be small enough that the cascade transitions from $^2\Pi_{1/2}(J=\frac{1}{2})$ are optically thin. It is easy to show that in this case the cascade simply preserves the original population distribution and that the 1612 inversion mechanism is not activated. If at the same time the cascade transitions from $^2\Pi_{3/2}(J=\frac{5}{2})$ are optically thick, an inversion of the 1720 line would be achieved. A situation in which one cascade is optically thin while the other is thick is possible because the line strengths for transitions which cross between the rotation ladders are about an order of magnitude smaller than those for interladder transitions.[4] But such a situation, obviously, is possible only in sources where the OH column density is not too high. The ground-state optical depths are therefore also limited to small values, and such sources can produce only weak 1720 masers, with gains smaller than 1.

The other possibility is to suppress the excitation of all the states other than the $^2\Pi_{3/2}(J=\frac{5}{2})$. This can be done with collisions, if the kinetic temperature is not too high, and is practically impossible in the case of radiative pumping.

We can conclude that (Elitzur, 1976a): (1) at small OH column densities, such that $^2\Pi_{1/2}(J=\frac{1}{2})\rightarrow{}^2\Pi_{3/2}(J=\frac{3}{2})$ is optically thin, the 1720 line is inverted

[4]The OH molecule is intermediate between Hund's cases a and b. Its rotation ladders are actually mixtures of pure $^2\Pi_{1/2}$ and $^2\Pi_{3/2}$ states. The notation used for the physical levels stands for the dominant state in each mixture. Transitions between the pure states are forbidden.

no matter which is the dominant pump excitation, and (2) the only way to achieve strong 1720 masers, i.e., $\tau > 1$, is with collisions at relatively low ($\lesssim 200$ K) temperatures; in other words, the only strong satellite-line maser that can be produced by radiative pumping is the 1612.

The results of a detailed calculation for a collisional pumping model are plotted in Fig. 6. In addition to demonstrating the basic inversion effect, they also display some of the general results of maser theory discussed in the previous sections. The saturation of the 1720 maser gain at a value which is determined by the pump properties, through the temperature of the colliding particles, is evident. In addition, the collisions thermalize the rotational transition, and the inversion turns off at large OH column densities, because the optical depth of the cascade transition becomes large enough that the radiative decays are slowed down by the trapped radiation and their rates become smaller than those of the competing collisions. Higher temperatures lead to larger collision rates, which is why the thermalization occurs earlier.

A problem which plagues most pump mechanisms employing collisional excitations is a poor knowledge of the detailed cross sections. The calculations displayed in Fig. 6 used hard-sphere cross sections with equal strength for decays to any magnetic sublevel. The resulting 1720 inversion is little affected by moderate deviations from this simple geometric cross section so long as these deviations do not introduce some peculiar selection rules. A study by Kosloff, Kafri, and Levine (1977) shows that hard-sphere collisions are actually a reasonable first-order approximation for OH rotational excitations. A series of detailed calculations by Green and collaborators (e.g., Green and Thaddeus, 1976) shows that geometric cross sections, without any particular selection rules, are in fact always the most plausible *a priori* guess for collisions between neutral molecules.

The predictions of the scheme described above for satellite-line anomalies were confronted with observations by Haynes and Caswell (1977), who found satisfactory agreement. There is a preponderance of the 1720 line enhancement among the weak sources, and all the sources which emit strongly in this line are believed to be very dense, so that the excitations in them are indeed dominated by collisions. Guibert, Elitzur, and Rieu (1978) performed some detailed calculations along the lines described above and compared the results with OH observations for a number of sources. The source parameters which were obtained from this comparison were in good agreement with those inferred from other types of observations.

The strongest argument in favor of any particular pumping mechanism is usually obtained when the source displays time variations such that the maser and the potential pump are varying together. Since there is no general class of objects which are strong emitters in the 1720 line, unlike the 1612, such regular time variations are not observed for it. Fortunately, however, a unique flare in the star V1057 Cyg was accompanied by a strong burst of 1720 maser emission (Lo and Bechis, 1973). The scheme described here would require such a maser to be pumped by collisions. An analysis of the observations of the burst's decline in various wavelengths, corresponding to all potential radiative pumps, showed indeed that none of them varied by a sufficient amount to explain the decay of the maser output, thus eliminating radiative excitations as the pump mechanism. At the same time, it was possible to construct a model based on collisional pumping which did reproduce adequately the time variation of the maser radiation (Elitzur, 1976b). This source therefore provides strong support for the satellite-lines anomalies scheme described above. The source actually underwent another 1720 burst a few years later (Winnberg *et al.* 1981), but this one was much weaker than the first and was not studied in any other radiation wavelength.

The greatest success of the satellite-lines scheme, and where these ideas were in fact first formulated (Elitzur, Goldreich, and Scoville, 1976), is the 1612 emission from late-type stars. Since these stars comprise the most studied and best understood family of astronomical masers, they are described separately in the next section (IV.B).

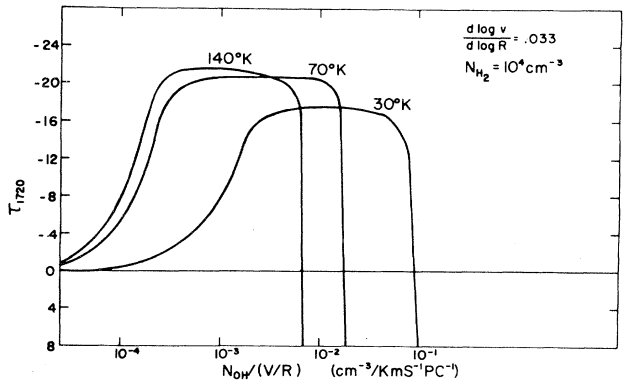

FIG. 6. Optical depths of the 1720-MHz line as a function of OH column density in a region of operation of a collisional pump. The colliding particles (H_2 molecules) have a density of 10^4 cm^{-3} and a temperature which is marked on the curves (from Elitzur, 1976a).

B. Late-type stars

These are stars at the late stages of stellar evolution (spectral type later than M5). They are very big, with radii of about 10^{13} cm for a "giant," to more than 10^{14} cm for a "supergiant." A giant star placed at the center of the solar system would fill it up to the orbit of the Earth, a supergiant would reach beyond the orbit of Saturn, almost to Uranus. Their energy output ranges from $10^4 L_\odot$[5] for a typical giant to $\gtrsim 10^5 L_\odot$ for a super giant. Their outer layers are very cool, with surface tem-

[5]The solar luminosity, $L_\odot$, is $3.86 \cdot 10^{33}$ erg/sec.

peratures of about 2000–2500 K, and are therefore rich in molecules. Since their radiation is concentrated at the red end of the spectrum, they are known as red giants and red supergiants. These stars are losing mass through a process whose details are still under debate. At a distance which is probably only about two stellar radii or less, the out-flowing material is forming dust particles. At that point the radiation pressure on the dust grains becomes the dominant force on the expanding envelope and the material is accelerated outward, with the dust carrying the gas particles through the effect of collisions. The dust grains absorb the stellar radiation and reemit it at their own temperature, established through the equilibrium between absorption and reemission. This temperature decreases as the material flows away from the central star. The dust reradiation occurs at infrared wavelengths and sometimes exceeds the directly observed radiation from the star.

The luminosities of all red giants and supergiants display time variations. The types of variability range from the smooth, periodic variations, with time scale of ~ 1 yr, of the regular ("Mira-type") variables, to the erratic variations, on a time scale of a few months, of the irregular variables, usually supergiants.

These stars provide the only class of astronomical masers, so far, with radiation in all three known strong maser molecules. The SiO maser radiation emanates from excited vibrational states (Fig. 3). The state with $v=3$, for instance, is 3650 cm^{-1} (or equivalently 5251 K) above the ground state. A substantial population, required for a strong maser gain, in this state can therefore be maintained only very close to the star. The SiO masers are thus probing the immediate stellar vicinity.

The H_2O and OH masers, on the other hand, emanate from transitions in the ground vibration states of these molecules and do not require such extreme conditions for pumping. The water-maser transition is between two rotation levels which lie about 447 cm^{-1} (or 643 K) above the ground state (Fig. 2). Hence, substantial temperatures (for the colliding particles or the radiation field—whichever is the relevant inverting agent) are required for the pumping, though not as high as for SiO. The water masers may therefore occur at distances of up to about 10^{15} cm from the central star.

The strong OH maser radiation emanates from the ground rotation state at wavelengths of about 18 cm (or 0.08 K). The ground-state levels themselves are therefore excited everywhere, but the inversion requires cycling of the molecules through the excited rotation states, which are 120–411 K above the ground state. The excitation energies of these levels are lower than those of the water masering levels. In addition, there is a need only for cycling molecules through them and not for a population buildup. OH masers can therefore be inverted at yet lower temperatures, and may be located even further away from the star, up to distances of a few times 10^{16} cm.

There is a number of additional arguments which support locating the SiO masers in the immediate vicinity of the star, the H_2O masers in the expanding shell at distances of $\lesssim 10^{15}$ cm, and the OH masers in the outer parts of the shell, at $\sim 10^{16}$ cm (e.g., Elitzur, 1981). The strongest argument is probably provided by the results of interferometric observations. A particularly elegant experiment was performed recently for the OH and H_2O masers around the star VX Sgr by Moran *et al.* (1981). By assuming a simple radius-velocity relation for the outflowing material they were able to construct the full three-dimensional structure of the maser spots. The H_2O is clustered within 2.10^{15} cm and the OH within $4.7 \cdot 10^{16}$ cm (Fig. 7). Another method for measuring the OH shell size, which yields similar results, is discussed below.

Maser radiation is therefore probing all regions of interest around late-type stars, from the outer parts of their atmospheres (SiO) to the expanding shells (H_2O and OH), and has become an invaluable tool in studies of them.

OH maser radiation from late-type stars has now been detected in all of the ground-state lines except the 1720 MHz. The stars which are stronger in the main lines are called "type I" and those with stronger 1612 emission are "type II." All are sometimes referred to as "OH/IR stars." The type II stars have been studied more thoroughly so far and will be discussed first.

1. Type II OH/IR stars—1612 MHz masers

The 1612 maser line from late-type stars has a characteristic profile of two strong spikes separated in velocity by about 20–50 km/sec. An often discussed prototype is the star IRC + 10011 (Fig. 8). This characteristic profile can be easily explained as resulting from the need for velocity coherence for the masering molecules (see II.A). In a radially expanding shell, the velocities of molecules

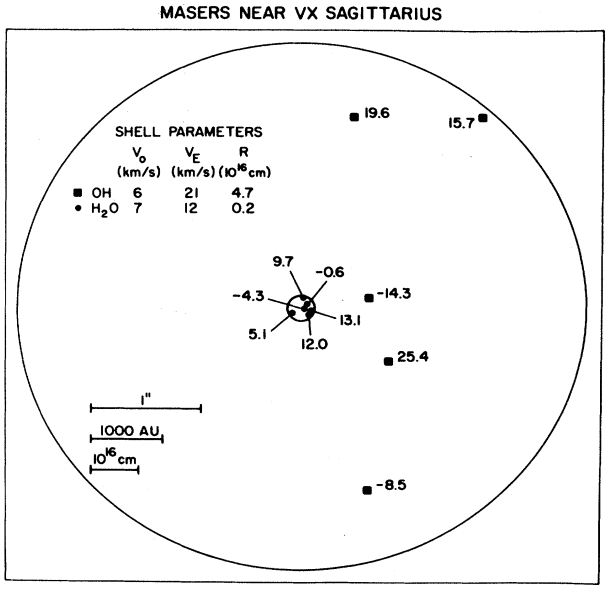

FIG. 7. The distribution of OH 1612-MHz and H_2O maser spots around the star VX Sgr (from Moran *et al.*, 1981).

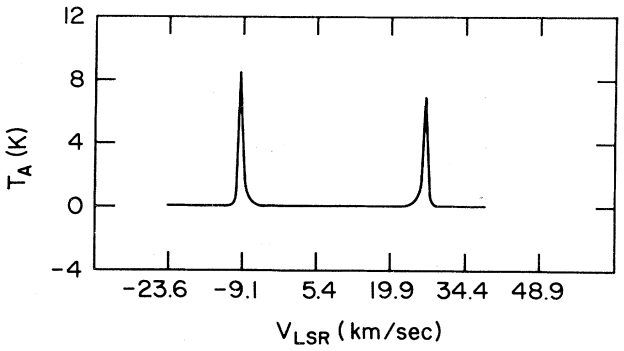

FIG. 8. The OH 1612-MHz spectrum of the star IRC + 10011 on August 9–10, 1978 (courtesy P. R. Jewell).

on different radius vectors are pointing at different directions, and strong amplification is therefore possible only along radial directions. Each segment of the shell can then emit strong maser radiation only along the radius vector, toward and away from the center. An observer at an arbitrary location will therefore detect radiation from only two regions which correspond to the intersection of the shell with his line of sight. The "blue-" and "red-" shifted components, corresponding to the front and back of the shell, respectively, should therefore be separated by twice the shell velocity of expansion, and their midpoint should correspond to the stellar velocity.

A strong and direct confirmation for the "front-back" model was provided recently in an experiment by Booth *et al.* (1981), who performed precise interferometric observations of various intervals in the spectrum of a type II star. They found that the radiation of the two spikes emanates from compact, well-defined regions which are coincident in position, along the line of sight. In contrast, the radiation from the inner shoulders of the spikes covers a larger, circular region, as expected from an expanding shell (Fig. 9). Another experiment which provides support for the front-back model is mentioned below.

In spite of its success in explaining the observations, an ideal, radial flow cannot always be the full, correct description. The observations of VX Sgr (Fig. 7) show deviations from the ideal front-back structure. Deviations from pure spherical symmetry are to be expected for a number of reasons, including irregularities in the basic outflow and interaction of the outflowing material with the ambient gas.

The monitoring project of Harvey *et al.* (1974) demonstrated that the maser radiation follows the stellar IR variation in regular variables with almost no phase delay. This has essentially established IR radiation as the inverting pump mechanism. Collisions are eliminated because their presumed variations, due to temperature fluctuations with the heating rate, would involve a larger phase lag than measured. Estimating the various excitation rates in the shell at the expected OH location, $\sim 10^{16}$ cm, it is found that the rotational excitations by the dust IR radiation are indeed the dominant ones. From the above discussion it then follows that the 1612 line, but never the 1720, is expected to be strongly inverted, provided the cascades from $^2\Pi_{1/2}(J=\frac{1}{2})$ are optically thick.

The optical depth, τ_w, of the weakest among these transitions (the $F=1\rightarrow 1$) can be written in the following form, which eliminates density in favor of the mass loss rate,

$$\tau_w = 2.1 \frac{\dot{M}_6}{R_{16} V_{10}} , \tag{4.7}$$

where $\dot{M}_6$ (in units of $10^6 M_\odot$/yr) is the mass loss rate, R_{16} (in 10^{16} cm) is the radius of the OH masering region, and V_{10} (in 10 km/sec) its expansion velocity. It follows that $\tau_w > 1$ requires mass loss rates in the range of $\sim 10^{-5} - 10^{-6} M_\odot$/yr. These are very large rates by stellar evolutionary standards and cannot last for more than $\sim 10^4 - 10^5$ yr, a relatively short phase in the late evolutionary stages of these stars (the mass of these stars is at most a few $M_\odot$). The infrared observations support the large mass loss rates derived for the type II OH/IR stage from maser theory. This stage has not yet been identified by stellar evolution theory, nor have the theories yet been able to produce such huge mass loss rates in calculations (Elitzur, 1981).

The maser radiation is expected to follow the temporal variations of the stellar luminosity—the ultimate source of the pumping radiation. The OH radiation from the entire shell is therefore expected to vary in unison. For an external observer, however, the maser photons in the red-shifted component are arriving from the back of the shell and are delayed in comparison with the blue-shifted photons by their travel time across the shell. The light curves of the "red" and "blue" components should thus be displaced, with the red trailing the blue by the shell crossing time. For IRC + 10011 the detailed theoretical model of Elitzur, Goldreich, and Scoville (1976) based on the shell chemistry calculations of Goldreich and Scoville (1976) led to a shell size estimate of $\sim 3 \cdot 10^{16}$ cm or a phase lag of about 23 days.

The first attempt at measuring the phase lag was by Schultz, Sherwood, and Winnberg (1978). Due to the lack of sufficient data their results were somewhat inconclusive. Jewell *et al.* (1979) undertook the same experiment with more sampling. They were able to demonstrate with certainty that the red component was indeed lagging behind the blue and finally determined the phase lag for IRC + 10011 as 25 ± 5 days (Jewell *et al.*, 1980). This gives strong support to the front-back idea and to the theoretical maser model, in addition to providing the best estimate of the shell size. Some other preliminary phase-lag determinations were reported recently by Herman and Habing (1981) and all of them lead to shell sizes of a few 10^{16} cm.

The model calculations of Elitzur, Goldreich, and Scoville (1976) led to the conclusion that the type II OH/IR masers are saturated and that it takes about four IR pump photons to produce one maser photon. In a recent experiment, Werner *et al.* (1980) performed simultaneous

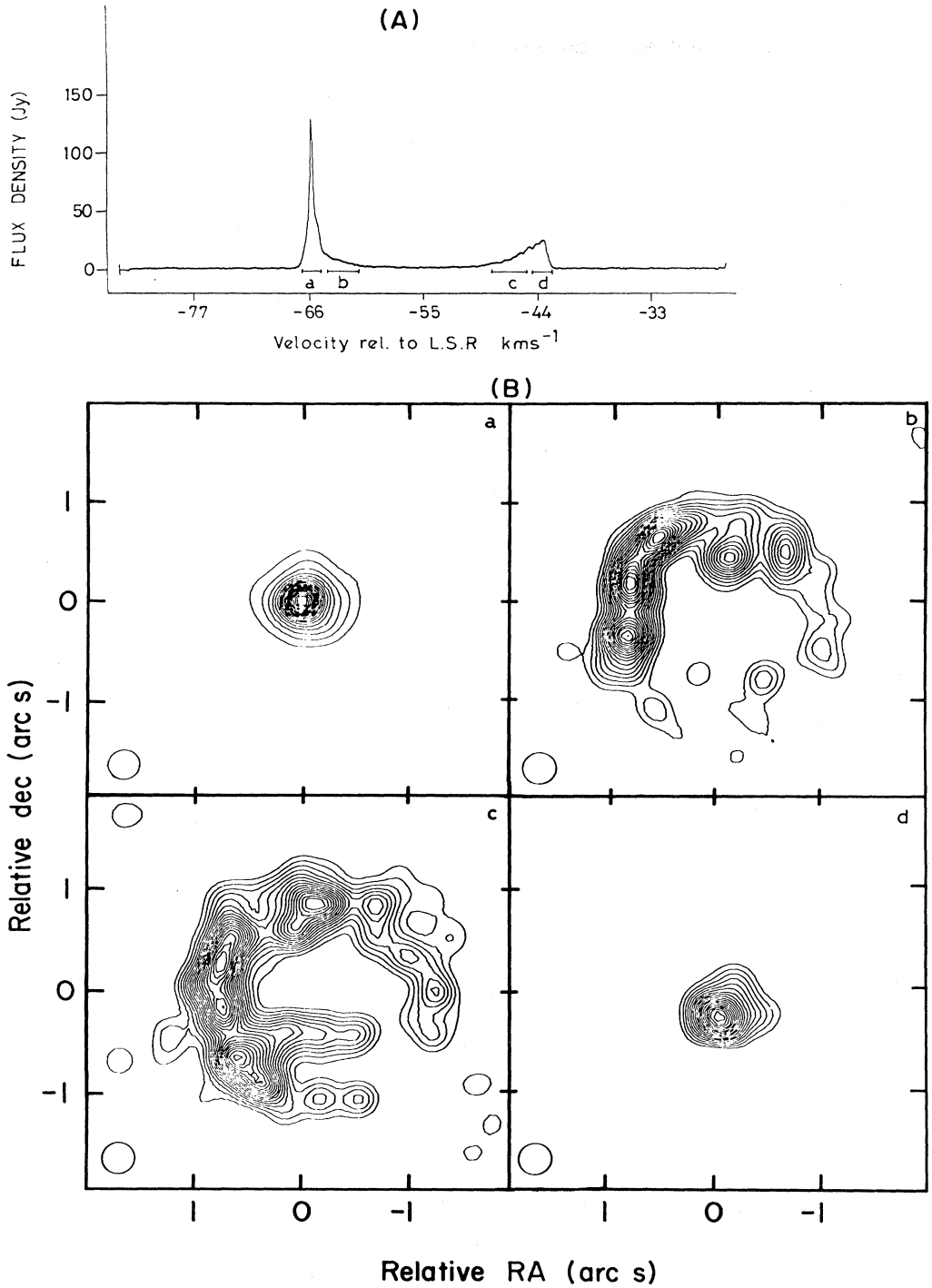

FIG. 9. Spectrum and interferometric measurements of the 1612-MHz maser emission from the star OH127.8-0.0. (A) shows the spectrum; the bracketed velocity intervals a—d represent those ranges of velocity for which spatial distribution of the maser emission has been determined. The maps are shown in (B) and plot the spatial distribution of OH emission in each velocity interval. In each map the contour interval is 5% of the peak emission in that velocity interval and the lowest contour is ~10% (from Booth *et al.*, 1981).

IR and radio observations of five 1612 emitters and showed that the ratio of pump-to-maser photons is indeed about 4:1, in agreement with the calculations. They also showed that the ratio remained approximately constant with time, unaffected by temporal variations, as expected for a saturated maser.

The experiments of Jewell *et al.* and of Werner *et al.* provide strong support for the theory of the 1612 maser

stars. In fact, one would not expect *a priori* such close agreement in details from an idealized, simplified model calculation. It is conceivable that for strongly saturated masers, the details of the model are not too important and that all models with the same essential ingredients lead to the same results. The case for strong saturation for these masers is supported by the results of the above-mentioned experiments, as well as by the failure to detect very small-scale structure in the maser components. Unlike a saturated maser, an unsaturated one would amplify the radiation of a background source (Sec. II.C). In the case of type II OH/IR stars the maser would then amplify the radio component of the stellar radiation, leading to a bright, compact maser spot in the blue-shifted component. The observations of Reid and Muhelman (1975) show that such a compact structure, in fact, does not exist.

2. Type I OH/IR stars—main-line masers

In contrast with the satellite lines, the OH main lines appear as strong masers in HII/OH regions [next section (IV.C)] and OH/IR stars but are otherwise in thermal equilibrium in most other sources. A glance at the OH energy-level diagram (Fig. 1) reveals that the difference between satellite and main-line behavior should not come as a surprise. Satellite lines connect levels with different F values, which couple to different levels in the excited rotation states, due to the dipole selection rules. This effect leads to a difference in the molecular flow patterns. It is then easy to transfer molecules between the various ground-state F levels, as described above. Satellite-line inversion is achieved even though the population of the ground-state upper half may be equal to (or even smaller than) that of the lower half.

Main-line anomalies, on the other hand, are much more selective phenomena and are much more difficult to explain. The energy-level diagram is symmetric with regard to the two halves of the Λ doublet, and for every cycle that carries molecules from negative to positive parity there is an image cycle operating in the opposite direction with similar efficiency. Main-line inversions therefore require some parity selectivity in the pump process itself.

As in the case of the 1612 line, the main-line inversion in late-type stars has to result from rotational excitations by IR radiation, the dominant excitation process in the shell. The Λ-doublet separation in each excited rotation state is larger than that in the ground state. The transitions from the ground state to the upper half of a rotation state therefore have higher frequencies than the transitions to the lower half. Rotational excitations with a radiation field $\tilde{I}(\nu)$ [Eqs. (3.10) and (3.11)] increasing with frequency would then pump more molecules into the upper Λ-doublet component of each rotation level. The cascades to the ground state, proceeding preferentially within the rotation ladders, preserve the sense of upper or lower half (Fig. 1) and would transfer the population inversion to the ground state.

A monotonically increasing function $\tilde{I}(\nu)$ is a possibility when the IR radiation is emitted by optically thin dust, since the Planck function is then multiplied by the dust's optical depth $\tau_d(\nu)$ [Eq. (3.15)], which is given by

$$\tau_d(\nu)=\text{const}\cdot\nu^p\ , \tag{4.8}$$

with $p\sim 2$ in the relevant wavelength region (Aanestad, 1975). The resulting $\tilde{I}_d(\nu)$ is then increasing with ν for dust which is warm enough.

A detailed calculation (Elitzur, 1978) showed that pumping by warm ($T_d \gtrsim 200$ K), optically thin dust does indeed produce main-line inversions which can explain the observed features in late-type stars. In this model, the main lines are inverted together in a region which is inside the 1612 emitting region; and the OH optical depths cannot be too large, since the inversion would then shift to the 1612 line. The main lines are also not as strongly saturated.

The model calculations were repeated by Bujarrabal, Destombes, *et al.* (1980) with an improved set of Einstein A coefficients. The new coefficients introduced more asymmetries, which enhanced the pumping efficiency, to inversions of about 1%, but did not modify the overall conclusions. The same researchers have since attempted to incorporate in the calculations nonlocal effects due to line overlaps caused by velocity gradients in the shell (Bujarrabal, Guibert *et al.*, 1980).

The idea of line overlap, like so many others in the theory of astronomical masers, was first introduced by Litvak (1969b). It is based on the observation that various transitions between a given pair of rotation states are sometimes close enough in frequency that Doppler shifts in the source can bring them to a common overlap. A photon emitted in some part of the shell in one line can then be absorbed somewhere else in the other one. The incorporation of this effect in the calculations is a complicated affair because of its highly nonlocal nature. Bujarrabal, Guibert, *et al.* (1980) attempted to bypass these difficulties with a local approximation based on the escape probability. There appear to be some difficulties with their calculation, which is still rather complicated. Whatever problems there may be, their calculations lead to yet higher efficiencies in inverting the main lines, but the results and basic pump requirements are otherwise similar to the previous calculations. Observationally it does not seem that this extra efficiency is really required. The maser output in the calculations of Bajarrabal, Destombes, *et al.* (1980), which incorporated the new A coefficients without line overlaps, appears to be sufficient.

The overall agreement between theory and observations is good (Elitzur, 1981). In particular, the models imply that every OH/IR star should exhibit main-line emission from the inner parts of its OH shell, but only those with very high mass loss rates will emit also in the 1612 line. This is in agreement with observations. Type I and type II OH/IR stars are similar in their main-line properties, but the latter emit also 1612-MHz maser radiation, which the former lack.

C. HII/OH regions

The strongest OH masers, and the first ones discovered, are main-line masers in regions which are also the sites of active star formation. Early speculations centered on the idea that each maser spot is perhaps a protostellar object. This can be easily dismissed, because the observed dimensions of the spots (which are comparable to the spots' separations) are so small that their inferred masses are many orders of magnitude less than typical stellar masses.

Cook (1966a) was the first to conjecture that these masers may be located at the surface of HII regions, which are abundant in sites of star formation. These are regions of ionized hydrogen[6] surrounding bright, young stars (spectral type O and B) with the ionization caused by the hot stellar radiation. Baldwin, Harris, and Ryle (1973) suggested that radio observations of W3(OH), one of the strongest maser sources, were indeed consistent with the idea that the maser spots are located on the circumference of the associated HII region. This suggestion received support from the observations of Goss, Lockhart, and Fomalont (1975) and was finally verified with certainty by the observations of Forster, Welch, and Wright (1977).

A parallel observational effort was mounted by a group of European radio astronomers who were trying to identify the underlying HII regions at selected maser sources. As a result of this effort, they were able to conclude that one could always find an HII region coincident in location with the maser source and that the associated HII regions were always compact, with dimensions of less than about $5\cdot10^{17}$ cm (Habing *et al.*, 1974). The association with HII regions thus seems established; hence the name HII/OH regions.

The energy released in the ionization process heats up HII regions to temperatures of about 10^4 K, and their pressure is therefore much higher than that of the surrounding medium, leading to their expansion with velocities of order 10 km/sec. It then takes only $\sim10^4$ yr for an HII region to expand to dimensions of $\sim10^{17}$ cm. The OH maser is therefore a short-lived and early phase in the evolution of dense HII regions.

The calculations of HII region development show that the expansion leads to the formation of a shock front at the edge of the ionized material [a comprehensive description can be found in Spitzer (1978)]. By following the chemical development of the material behind the shock front, Elitzur and de Jong (1978) were able to show that there is an enhancement in the OH abundance at the high-density postshock region, which they therefore suggested as the location of the maser action. When the HII region expands further, the shock slows down, the OH abundance decreases, and the maser turns off.

Reid *et al.* (1980) suggested instead that the masers are located at infalling material that the central star is still accreting from the surrounding cloud. Their argument is based on a velocity difference between the OH maser spots and the centroid of the recombination line profile in W3(OH). The main problem with this suggestion is that, in order to achieve the proposed infall velocity (taking account of the angle with the line of sight) at the measured radius, the mass of the central star has to be larger than what is inferred from its ionizing flux and even larger than what is allowed for a star on the main sequence. In addition, Norris and Booth (1981) claim that the infall idea is in conflict with their recent observations of W3(OH).

A possible fault in the argument of Reid *et al.* is the velocity determination of the central star. The recombination lines in W3(OH) are especially broad (full width of $\sim$50 km/sec), and it is not clear that their center can be determined with the degree of accuracy needed to establish the velocity difference (of a few km/sec) with the OH maser spots. The question of the symmetry of the lines at the required accuracy is particularly difficult. If, for instance, the ionized gas is expanding in a nonhomogeneous region with different densities in its front and back, the line profiles would not be perfectly symmetric and their apparent center would be different from the true stellar velocity. It is not clear how to resolve the issue for W3(OH) and it is conceivable that the only way to proceed is with a statistical approach. A survey of HII/OH regions can determine whether there is a preponderance of infall velocities for the maser spots in these sources.

Whatever the outcome of the infall issue may be, it is evident that the OH maser spots are probing density enhancements in the immediate surroundings of compact HII regions. The structure of these regions is certainly more complex than the simple picture described by Spitzer (1978) which Elitzur and de Jong have employed. The fact that W3(OH) is indeed not a smooth, spherically symmetric region but rather complex and clumpy is evident from the recent radio observations of Dreher and Welch (1981). Recent calculations which account for the interaction of the star with the material out of which it had condensed already lead to an involved structure (Yorke and Krugel, 1977). The picture gets even more complicated when the cloud finite dimensions and various velocity configurations for the star and cloud are included (e.g., Tenorio Tagle, Yorke, and Bodenheimer, 1979). A shock front, leading to a density increase and an enhancement of the OH abundance, may occur at various positions. The OH masers are presumably probing such fronts and, together with other observations, can be used as a tool to study their complex structure.

The understanding of the pump mechanism for these sources is even less satisfactory. Reid *et al.* (1980) suggested that the maser spots are amplifying the radio emission from the HII region, trying to explain in this way the fact that their power output is much higher than that of OH/IR stars, where the background radio radiation is weaker. However, a saturated maser is not amplifying any input radiation but is instead converting

[6]Hence the name HII, since each H atom gives rise to two particles.

pump events to maser photons, as shown above (Sec. II.C), and HII/OH masers are most certainly strongly saturated. Indeed, Norris and Booth (1981) noticed that the maser spots in W3(OH) which are not projected against the background radio radiation of the HII region are as strong as those which are, as expected from a saturated maser. The larger maser power output in HII/OH regions therefore simply implies higher pump rates. But the total energy output of a strong HII/OH source is about $10^5 L_\odot$, not really larger than that of a luminous supergiant. These sources are somehow more efficient in utilizing their luminosity to produce maser radiation.

Radiative pumps based on electronic or vibrational excitations can be dismissed immediately, because the number of photons emitted from these sources in the UV and near-IR wavelengths is much smaller than the number of maser photons. The mid-IR observations of Wynn-Williams, Becklin, and Neugebauer (1972) led to a similar conclusion with regard to rotational excitations. These observations were performed in the $10-20$ μ wavelength region and had to be extrapolated to the relevant region, longward of 35 μ, a procedure that was accomplished by adopting a reasonable temperature ($\sim 100-150$ K) for the emitting dust.

It therefore appeared that radiative pumps could be dismissed on observational grounds alone. This was in agreement with the results of the model of Elitzur and de Jong, which showed that collisional excitation rates were larger than any radiative rates, and they therefore suggested that the masers are pumped by collisions.

The situation has changed recently because of far-infrared observations in the $40-150$ μ region, with the NASA Kuiper Airborne Observatory, by Thronson and Harper (1979). These observations show that instead of turning down at $\sim 30-50$ μ, as expected from ~ 150 K dust, the IR radiation keeps increasing and peaks only at the long wavelength of ~ 100 μ, corresponding to a dust temperature of only $\sim 30-50$ K. The number of photons emitted in the rotational transitions wavelengths ($35-120$ μ) is therefore larger than the number of maser photons, making radiative pumping a possibility, in principle. The problem is that such low dust temperatures are possible only far away from the HII region, because close to it, at the presumed location of the OH masers, the stellar radiation would heat the dust to at least 100 K. Indeed, the Thronson and Harper observations show that the cool, long-wavelength radiation is coming from rather extended regions. In the case of W3(OH) the dimensions of the emitting region are in excess of 10^{17} cm, in agreement with the expected value for the dust temperature at such a distance. The OH maser is therefore located inside the far-IR emitting region; and the geometrical dilution then reduces the number of available photons to a value which is below the number of maser photons, eliminating far-IR as a viable pump mechanism.

It is evident, though, that this argument involves some assumptions about the source structure and is therefore not as powerful as the straightforward elimination of near-IR and UV radiation as possible pumps for which the strictly observed photon number is smaller than the number of maser photons. Better understanding of the far-IR emission from HII/OH regions is required before its viability as a pump can be decided with certainty. The significance of this point is obvious. If radiative pumps could be eliminated altogether, efforts could be concentrated on finding the right collisional process.

At this stage, however, far-IR radiation has to be considered still a possible candidate, together with collisions, for the pump mechanism. It is not clear, though, how to achieve an efficient main-line inversion with either of these pumps. The only detailed calculation of a radiative pump which claims a sufficient efficiency for main-line inversion is by Lucas (1980). It is based on the line overlap formalism mentioned in the previous section. In addition to being subject to the same uncertainties mentioned above, it also requires some *ad hoc* assumptions about the velocity field in the maser region.

The status of collisional pumps is not much more encouraging at this point. A main-line inversion can result from a rotational excitation, followed by radiative decay, if the collision cross sections prefer the excitation of one-half of the Λ-doublet components. If only the ${}^2\Pi_{3/2}(J=\frac{5}{2})$ is significantly excited, for instance, a preferential excitation of the upper half would lead to an inversion during the decay back to the ground state. Kaplan and Shapiro (1979) and Dixon and Field (1980) claim, on the basis of detailed cross-section calculations, that this is in fact the case. A difficulty with this result is that the main lines would then be inverted with great ease in a variety of situations and main-line anomalies would be as prevalent as those of satellite lines. This is contrary to observations. In addition, even if the cross sections were leading to a selective excitation, the density and OH abundance in HII/OH regions appear to be so high that the optical depths of the rotation transitions would probably be very large. Radiative trapping in these lines then slows the decays to such an extent that they would thermalize and would not lead to inversions, irrespective of the form of their collision cross sections.

It seems that the only collisional pump which stands a chance of explaining the large maser flux in HII/OH regions would involve inversion resulting from the direct population exchange between the maser levels. It is a question of general interest whether an inversion is possible at all in a two-level model dominated by collisions. The discussion of Sec. II.A shows that this is, of course, impossible when the colliding particles obey the Maxwell-Boltzmann distribution. Are there any velocity distributions which can in principle lead to inversion? For isotropic distributions, the rate constant K ($=\langle \sigma v \rangle$) is determined by a one-dimensional integral over kinetic energies. Using invariance under time reversal it is easy to show (Elitzur, 1979) that the rates for excitations and deexcitations are related through

$$K_{12}=K_{21}-\frac{h\nu_{21}}{kT}H\ , \tag{4.9}$$

where

$$H = g_2 \int_0^\infty \left[\frac{2}{m\varepsilon}\right]^{1/2} f(\varepsilon)\frac{\partial}{\partial\varepsilon}(\varepsilon\sigma_{21}(\varepsilon))d\varepsilon ,$$

and $f(\varepsilon)$ is the normalized energy distribution of the colliding particles. An inversion would require H to be negative. This is possible only if there is an energy region where the cross section is decreasing faster than $1/\varepsilon$ and if the distribution function $f(\varepsilon)$ is peaked in that region.

Although the cross sections for Λ-doublet excitations of OH in H_2 collisions are not known, it appears that inversion by an isotropic nonthermal distribution is very unlikely (Elitzur, 1979).

The other possible deviation from a thermal distribution is a nonisotropic velocity distribution—a particle stream, for instance. In this case one can define a specific direction which can serve as an axis of alignment for the magnetic sublevels. The cross section between specific sublevels is not related in any particular way to the reversed process; detailed balance applies only to the overall cross section, summed over sublevels (e.g., Sakurai, 1964, p. 90).[7] The cross section for excitation in a given transition may even be larger than the one for deexcitation.

Johnston (1967) was the first to notice that the cross section in Born approximation for dipole excitation of a $\Delta m = 0$ transition in an electron collision is larger than the one for deexcitation. Collisions with an electron stream could therefore produce OH main-line inversions. The resulting maser radiation could essentially be arbitrarily strong, because there is no thermodynamic limit on the density of the colliding particles. The problem is that since electrons are so light, they have thermal velocities of ~40 km/sec, and it seems unlikely that such a large streaming velocity for electron motions relative to the bulk of the material could be sustained.

Ionic thermal velocities, on the other hand, are much smaller, and an ion drift of only ~1 km/sec would constitute a "stream." Such a drift can result from the effect of a magnetic field anchored to the surrounding medium and compressed by the shock at the regions where the shock propagates perpendicular to the field lines. If the streaming ions in this region of "ambipolar diffusion" were capable of inverting the OH, the magnetic field and polarization would become an integral part of the inversion process itself, which seems like a desirable effect (Elitzur, 1979). Unfortunately, the Born approximation does not hold for OH excitations in ion collisions (Elitzur, 1977; Bouloy and Omont, 1977, 1979), and the detailed cross sections are not known.

The main encouraging point about collisional excitations in particle streams is the apparent ease with which the large maser outputs could be produced. Evans *et al.* (1979) calculated what would be the maser output in a number of HII/OH regions they observed if ion streams were capable of inversion with an efficiency determined from the analytic expression for electron collisions. The resulting ion collision asymmetry, of only about 10^{-4}, suffices to produce a maser photon flux always as large as observed, using the observed densities. It is evident, though, that there is little hope of performing a believable cross-section calculation with the accuracy to test an effect at the level of one part in 10^4.

The status of collisional pumps is therefore not encouraging because of lack of precise knowledge of cross sections. As mentioned above, the status of IR pumping is not that much better. There is therefore the need to try to search for more clues from observations about the nature of the pump. One possibility, for instance, is to find out whether the OH output is correlated with the far-IR flux, through an experiment similar to the one by Werner *et al.* (1980) for OH/IR stars. A linear correlation would provide an argument, although not a conclusive one, in favor of IR pumping. The lack of such a correlation would be a conclusive argument against IR pumps.

The output of a collisionally pumped and saturated maser, on the other hand, is proportional to the density squared [Eq. (3.35)], because of its dependence on the collision rate, as well as the density of the maser molecule. The density inside the HII region can be determined from its continuum emission and can be used to provide a reasonable estimate for the density in the maser region. A correlation between the maser output and the density squared, if found, would provide a strong argument for collisional pumping. The lack of any correlation with density, coupled with a correlation with IR flux, would pretty much eliminate collisional pumping in favor of radiative pumping.

Another possible line of investigation is discussed in the next section. (IV.D)

D. ^{18}OH

The terrestrial abundance of ^{18}O is about 490 times lower than that of the major isotope, ^{16}O. The energy-level structure of the ^{18}OH is identical to that of ^{16}OH, but the frequencies are modified because of the different molecular mass. The ground-state frequencies are 1637 and 1639 MHz for the main lines, and 1584 and 1693 MHz for the satellite lines. Three of those, with the exception of the 1693-MHz line, which is close to a frequency used for a satellite communication, were discovered in absorption in a number of sources. The inferred abundance ratio of ^{18}OH to ^{16}OH is enhanced by about a factor of 2, compared with the ^{18}O/^{16}O terrestrial ratio, in all the sources but one, where the ratio is suppressed (Whiteoak and Gardner, 1978). The question of the interstellar chemistry of ^{18}OH has not been tackled yet.

The smaller ^{18}OH abundance can provide a test for various inversion schemes. For instance, in Sec. IV.A it was shown that rotation transitions to the first excited

[7] It had already been noticed by Boltzmann that spherical objects' cross sections obey a detailed balance relation, whereas those of nonspherical ones do not.

state have optical depths which are about two orders of magnitude larger than those of the ground state. For ground-state optical depths of order unity for ^{16}OH, the rotation transition of ^{18}OH would then be optically thick. One would then expect to detect satellite-line anomalies in ^{18}OH, and this is indeed the case, according to the recent observations of Williams and Gardner (1981) toward the source Sgr B2.

In contrast, the crucial parameter for the 1612 inversion is the optical depth τ_w [Eq. (4.7)], which is larger than unity for ^{16}OH only for mass loss rates of OH/IR stars in excess of $10^{-5}-10^{-6}M_\odot$/yr. The operation of the equivalent 1584-MHz maser of ^{18}OH would therefore require mass loss rates in excess of at least $10^{-4}M_\odot$/yr, which is pretty much out of the question. Indeed, Caswell and Haynes (1981) performed observations of four 1612 emitters in the 1584 line and found no emission, as expected. This is a nontrivial negative result. A radiatively pumped saturated 1584 maser, if produced, would emit as strongly as the 1612 maser, because the number of maser photons would then be determined by the number of pump photons and would be independent of densities.

An attempt at detecting ^{18}OH main-line emission from HII/OH regions was made by Wilson and Barrett (1970). It led to negative results, showing that the possible emission is at most 2000 and 3000 times weaker than the corresponding ^{16}OH emission in W3(OH) and W49, respectively.

In the absence of detailed reliable calculations, one can only speculate about the implications of this result for collisional and far-IR line overlap pumps. The output of a collisionally pumped and saturated maser is proportional to the density of the maser molecules [Eq. (3.35)]. If the ^{18}OH density were about two orders of magnitude less than ^{16}OH, its maser output would have been detected by Wilson and Barrett under such circumstances. However, such a decrease in density could also lower the maser emission below the saturation flux, leading to a further exponential decrease of the maser output. Most calculations show that the masers saturate at gains of order 10. If a main line of ^{18}OH had an optical depth of ~ 10, the ^{16}OH column density would be in excess of $\sim 10^{19}$ cm^{-2}, which is larger than any reasonable estimate. It does not seem likely, then, that the ^{18}OH could ever by saturated.

This argument could be countered, leading to a possible elimination of collisional pumps, if the HII/OH regions which emit the ^{16}OH maser radiation spanned a density range of two orders of magnitude, in which case the most rarefied ones would have ^{16}OH densities comparable with the ^{18}OH densities of the densest. This does not seem to be the case (Evans *et al.*, 1979). Note that the output of a collisionally pumped maser is proportional to the density squared, as mentioned in the previous section (IV.C).

If the masers were radiatively pumped and always saturated, their output would be proportional to the pump rate and independent of density, and the ^{18}OH maser would be as strong as the ^{16}OH. But the fact that the ^{18}OH maser probably cannot saturate can be used to argue that their emission should be undetectable for a radiative pump, too. Another possible excuse for line overlap pumps, that the overlap mechanism cannot work for ^{18}OH because of the different frequencies, can actually be dismissed. The recent tabulation of Beaudet and Poynter (1978) shows that the ^{16}OH frequencies are close enough to those of ^{18}OH that if a frequency overlap occurred in one it should occur in the other, as well.

It is evident that at this stage it is impossible to draw conclusions and eliminate either of the pumps on the strength of the ^{18}OH negative result. It is also evident that this is a line of investigation worth pursuing for both theory and observations.

E. Comets

As a comet is approaching the sun, the increased heating rate causes evaporation of surface ices. Photodissociation of the water vapors produces OH, which is pumped by the solar UV radiation, leading to nonthermal population distribution. This results in a very weak maser effect, because of the relatively small dimensions, with limited gain. It is, however, one of the most beautiful examples of the operation of basic laws of physics in an astronomical object and is therefore described here briefly.

When excitations by UV radiation are dominant, molecules are pumped from the ground state to the first excited electronic state (a $^2\Sigma^+_{1/2}$ level about 33 000 cm^{-1} above the ground state) and cascade rapidly back. There are six allowed excitation transitions. The pump cycles initiated with some of those lead to inversion of the ground-state Λ doublet, whereas some lead to anti-inversion. When the excitations are by featureless, "white" radiation, the overall pumping effect is to cause anti-inversion (Litvak *et al.*, 1966). The solar spectrum, however, is anything but featureless. The radiation is propagating through the cool outer layers of the solar atmosphere, leading to the production of an enormous number of Fraunhofer absorption lines. Naturally, with so many absorption features, some have frequencies close to those of the UV pump lines of OH.

The comet heliocentric velocity is varying according to Kepler's second law (equal areal velocities). The Fraunhofer line frequencies are therefore varying in the comet frame as it moves around the Sun. At a certain point, an absorption line would overlap a pump line, one which leads to anti-inversion, say. At that point, the pump rate becomes deficient in an anti-inversion cycle, and the net effect of the pump is inversion. At a later stage an absorption line may coincide with an inversion causing pump line, and the population would switch to anti-inversion. The OH radiation would therefore oscillate between emission and absorption.

This elegant effect was not only observed, but, what is even more impressive, was actually predicted. Following

the publication of results of radio observations of comet Kouhoutek in December 1973, Mies (1974) pointed out that the OH absorption was in agreement with UV-pump calculations and that in January 1974 the absorption should have switched to emission. He was unaware that observations were, in fact, already made by a group of radio astronomers at Meudon (France), who, on their part, were unaware of his calculations. They detected the changeover from absorption to emission and figured out, independently, the correct explanation (Biraud *et al.*, 1974). The oscillations between emission and absorption have been detected by now in a number of comets, and the observed pattern is in satisfactory agreement with the UV-pump predictions.

V. SiO

A. The pump mechanism

With the possible exception of a source in the Orion nebula, all SiO masers occur in red giants and supergiants. The fact that a substantial population can be maintained at $v=3$, so high above the ground state (Fig. 3), implies a close proximity to the star, as mentioned above, irrespective of the nature of the pump. This conclusion is corroborated by interferometric observations which show that for the star R Cas, with a stellar radius of about $3\cdot10^{13}$ cm, the SiO maser spots are clustered within $\sim4\cdot10^{13}$ cm around the star (Moran *et al.*, 1979).

The operation of a strong maser in $v=2$ and 3 essentially eliminates radiative excitations as a viable pump mechanism. Only $\Delta v=1$ transitions are allowed for a pure harmonic oscillator, and the line strengths therefore decrease very fast with Δv. A substantial radiative pumping of $v=3$, and also $v=2$, is then possible only through successive $\Delta v=1$ excitations, which are entirely negligible, apart from an insignificant region right at the edge of the photosphere. It has also already been shown (Sec. III) that $\Delta v=1$ radiative excitations correspond to a radiative pump with the same reservoir levels for the pump and loss processes and can never lead to inversion.

Radiative pumping could work in principle, though, for the $v=1$ maser. In that case, the pumping would have to be indirect, through a higher vibration state. Indeed, Kwan and Scoville (1974) have shown that a radiative excitation of the $v=2$ state followed by the double cascade $v=2\rightarrow1\rightarrow0$ would invert the low-lying rotation levels of $v=1$, provided the $v=1\rightarrow0$ transitions were optically thick, while $v=2\rightarrow1$ were optically thin. If both were thick (or thin), the selection rules for populating the rotation levels of $v=1$ would be the same as for emptying them and no inversion could be produced. It follows that this mechanism could not apply to the higher vibration states in the same location, in principle, since a $v=2$ maser, produced through a $v=3\rightarrow2\rightarrow1$ cascade, would require the $v=2\rightarrow1$ transitions to be optically thick, while for a $v=1$ maser action they would have to be thin. However, the line profiles of masers in different vibration states show similarity which is sometimes displayed even in minute details, implying that they are inverted in the same region by the same pump mechanism. Recent simultaneous interferometric measurements of various vibration states confirm that they are located in the same region (Reid and Moran, 1981). This then helps to eliminate radiative pumps altogether, a conclusion which is supported by recent observations which show a definite lack of correlation between time variability of SiO and optical output in certain sources (Clark, Troland, and Johnson, 1981).

Large Δv excitations are rather strong in collisions with neutral molecules [e.g., Procaccia and Levine (1975) and references therein], and collisions can therefore adequately pump all the masers. Strong pumping would require, though, high temperatures and large densities. If the late-type star's atmospheres were quiescent, the required conditions would occur only in a limited region, too small to provide enough maser output to explain the observations. The reason is that in steady-state atmospheres, the temperature and density profiles follow the barometric equation of exponential decrease with a scale height that is much smaller than the stellar dimensions. The atmospheres of these stars, however, are anything but quiescent. Strong motions, which are discussed below, manifest themselves in the behavior of many atomic and molecular lines, including those of the SiO masers themselves. The violent activity in the outer layers of these stars can transport hot and dense material to distances which would be impossible under static conditions. The atmospheric activity is therefore an integral part of the maser action.

For a collisional pump to work, the vibration decays must be radiative, because otherwise the level populations would thermalize with the kinetic temperature. This restricts the maser action to regions with densities which are $\lesssim10^{12}$ cm^{-3}. Denote by $n^{(0)}$ the population of the $v=J=0$ state. With collisional excitations and radiative decays for the vibration transitions, the population per magnetic sublevel of the (v,J) state, $n_J^{(v)}$, is given by

$$n_J^{(v)}=n^{(0)}P_J^{(v)}(T)/\Gamma_J^{(v)}\,, \tag{5.1}$$

where the pump rate is

$$P_J^{(v)}(T)=\sum_{J_i}\frac{(2J_i+1)}{(2J+1)}\exp(-E_{J_i}/kT)\,C(T;0,J_i;v,J)\,, \tag{5.2}$$

$C(T;0,J_i;v,J)$ being the collision rate for the $(0,J_i)\rightarrow(v,J)$ vibration-rotation transition and $\Gamma_J^{(v)}$ the radiative decay rate.

The discussion of OH collisional pumps in the previous section (IV) would lead one to believe that this problem is hopeless, because it requires detailed knowledge of precise vibration-rotation cross sections. Fortunately, this is not the case. The SiO rotation constant is only 1.05 K, whereas the atmospheric temperatures are about 2000 K and the relation

$$E_{J_i}\ll kT \tag{5.3}$$

is obeyed by all the relevant rotation states. The ex-

ponential terms in the definition of $P_J^{(v)}(T)$ [Eq. (5.2)] can then be approximated by unity and the sum becomes one over cross sections. Equation (5.3) also ensures that the molecular rotational motions are in the domain of applicability of the sudden approximation. The cross sections for vibration-rotation transitions therefore obey (e.g., Parker and Pack, 1978)

$$\sigma(v_1,J_1;v_2,J_2)=(2J_2+1)\sum_l \begin{bmatrix} J_1 & l & J_2 \\ 0 & 0 & 0 \end{bmatrix}^2 \times\sigma(v_1,0;v_2,l) \ . \quad (5.4)$$

The summation in Eq. (5.2) becomes one over vector coupling coefficients and can be performed in closed form. The result is

$$P_J^{(v)}(T)=C_{0,v}(T) \ , \quad (5.5)$$

where $C_{0,v}(T)$ is the collisional $0\rightarrow v$ vibration rate. We have therefore obtained an expression for the collisional pump rate into any (v,J) level (with small J) without the need for a detailed knowledge of vibration-rotation cross sections (Watson, Elitzur, and Bieniek, 1980).

The expression obtained in Eq. (5.5) shows that the pump rates $P_J^{(v)}(T)$ are J independent, for the low values of J. A J dependence for $n_J^{(v)}$ can therefore result only from a possible J dependence of the radiative decay rate $\Gamma_J^{(v)}$. In the optically thin case, this is equal to the Einstein-A, which is also J independent to a high degree of accuracy. Hence, a J dependence can be obtained only when the $\Delta v=1$ SiO vibration transitions are optically thick and $\Gamma_J^{(v)}(=A/\tau)$ is given by

$$\Gamma_J^{(v)}=\Gamma_0^{(v)}2/(2J+1), \quad J\geq 1 \quad (5.6)$$

(Kwan and Scoville, 1974). Because of photon trapping, the decay rates are decreasing with J and

$$n_J^{(v)}=(J+\tfrac{1}{2})n_0^{(v)}, \quad J\geq 1 \ , \quad (5.7)$$

which exhibits the sought-after inversion effect. Since the SiO vibration transitions are optically thick in the stellar atmospheres, this calculation is applicable there and demonstrates the viability of inversion by collisional pumping (Elitzur, 1980).

The analysis here was presented as an illustration of the basic inversion effect and neglected any process other than collisional excitations and radiative decays across the vibration transitions. The inclusion of exchanges among the maser levels themselves is quite straightforward in the approximation in which all other rotation levels (in the same vibration state) are neglected, which is reasonable for a lowest-order estimate of the maser effect. The resulting photon production rate for a saturated maser between the J and $J-1$ rotation levels of $v=1$ is

$$\phi_{J,J-1}^{(1)}=n^{(0)}C_{0,1}\frac{1-2J^2(2J+1)\dfrac{C_{J,J-1}^{(v)}}{\Gamma_0^{(1)}}\dfrac{B}{kT}}{1+J\dfrac{C_{1,0}}{\Gamma_0^{(1)}}} \ , \quad (5.8)$$

where $B(=1.05$ K) is the SiO rotation constant. The expression demonstrates the turn-off of the maser effect for large J, in agreement with observations. The denominator increases due to the fact that the radiative decays slow down, for large J, because of photon trapping [Eq. (5.6)]. The vibration decays then become collision dominated and the inversion disappears. The expression in the numerator reflects the thermalizing effect of collisions across the maser levels themselves (as in Secs. II.C and III.C). This term is missing in the original derivation of Elitzur (1980), which neglected effects of order $\Delta E_J/kT$, but should be included in view of the fact that cross sections for rotation transitions are a few hundred times stronger than those for vibrations (Bieniek and Green, 1981). This thermalization, in fact, is probably more important in explaining the maser turn-off at larger J's.

B. Atmospheric motions

The close proximity to the star implies that only the front half of the SiO masering region is visible. However, strong maser spikes appear in velocities which are both blue shifted and red shifted from the velocity of the central star. The only possibility, then, is that the SiO maser-emitting material is taking part in large mass motions which are directed both toward and away from the star (Elitzur, 1980). Similar upward and downward motions are detected also in many optical and IR observations of various other atomic and molecular species. As mentioned above, such activity is essential for the SiO maser operation.

A number of theoretical models propose to explain the atmospheric activity in late-type stars. The one which seems most in line with SiO maser observations, although proposed entirely independently, is Schwarzschild's (1975) idea of convection in very large cells. The ordered convective motions would then lead to the velocity coherence which is required for the large maser gains. The largest velocity variations are expected to occur along the direction of the streaming, so material which moves in a direction perpendicular to the line of sight will provide the best velocity coherence for a maser photon propagating outward. Maser spikes should therefore appear mainly at the stellar velocity and should not deviate from it in general by more than the convection velocity—a few km/sec. Both conclusions are in agreement with observations. The convective cells' lifetime is estimated at about 150–200 days, so maser spikes should appear and disappear with this time scale. Indeed, the SiO profiles are unique in sometimes displaying a complete changeover with such a time scale, in spite of a great stability in their day-to-day behavior. From the collision rate and the observed maser output one can get an estimate of the size of the convection cells ($\sim 10^{13}$ cm), which is in agreement with the independent estimate of Schwarzschild. In addition, supergiants are expected to have more convection cells than giants; indeed, they appear to have more maser spikes.

Another property which is explained naturally with the convective cells is the recently discovered linear polarization (Troland *et al.* 1979), which requires some inherent nonsphericity for the masering material. This also provides a unique opportunity for studying motions on the stellar surface, because a variation of the polarization position angle corresponds to rotation of some specific axis. Such rotation has now been detected (Clark, Troland, and Johnson, 1981).

As mentioned above, convection is not the only idea which was proposed for atmospheric activity in late-type stars. Another model, which is by now well supported by observations, is the passage of periodic shock waves, powered by the pulsating atmosphere (Willson and Hill, 1979). The effect of these shocks on the SiO masers appears to have been detected recently (Clark, Troland, and Johnson, 1981).

It is evident that the atmospheres of late-type stars are rather complex. The difficult task of integrating the various motions into a grand overall atmospheric model has not yet even been attempted. SiO maser observations provide important clues for this hard problem.

As a final point it should be mentioned that although convective cells apparently provide an excellent explanation for the location of the maser material, the motions themselves could originate from something other than classical convection. The characteristics used here were that the maser radiation is emitted from well-defined cells with properties—size, internal velocity, and hence typical lifetime—which are similar to typical atmospheric scales. This fits the scales of convective motions rather well, but it is conceivable that the motions on the surface of these stars may be somewhat different than standard convection.

VI. H_2O

Radio emission from the water vapor (H_2O) molecule provides the most intense maser emission of any astronomical source. The observations of W49 by Burke *et al.* (1970) and Johnston *et al.* (1971) revealed brightness temperatures in excess of 10^{14} K. The H_2O maser radiation from this source is so intense that it emits almost a full solar luminosity in a single line with a bandwidth of only 50 kHz. H_2O maser emission was detected in late-type stars and in regions of active star formation. Unlike the case of OH emission from such regions, where an association with compact HII regions was established, the water emission was not yet clearly linked with any other class of astronomical objects.

In spite of the spectacular intensities of H_2O masers, theoretical work on the topic was rather minimal and very few detailed calculations have been performed so far. The main reason, undoubtedly, has to do with the great complexity of the H_2O spectrum (Fig. 2). The H_2O molecule is an asymmetric rotor and the levels are labeled $J_{K_+K_-}$, where J is the total angular momentum and K_+ and K_- are its projections on two molecular axes (Townes and Schawlow, 1956). The states in which both K quantum numbers are even or odd are para states (with statistical weight of 1) and those in which they have a different evenness are ortho (with statistical weight of 3). Ortho and para states do not couple for the particular rotation constants of the H_2O molecule. Other than that, the selection rules for the K numbers are rather involved, whereas those for J follow the usual dipole rules ($\Delta J = 0, \pm 1$).

A close examination of the energy-level diagram and routes of allowed transitions shows that for each J manifold, the lowest level should carry the bulk of the population, and was therefore called a "backbone" level by de Jong (1973). This result is particularly transparent when the optical depths are large, because the backbone levels can exchange population only among themselves, except for the two weak transitions 4_{14}-3_{21} and 6_{16}-5_{23}, whereas the other levels have a large number of cascade transitions open to them. The molecules would therefore tend to pile up along the backbone ladder and any transition between these and other levels would have inverted populations. The 6_{16}-5_{23} transition, where the two levels are accidentally near coincidence, is of course the strong maser transition. The 4_{14}-3_{21} transition, at 380 GHz, was detected recently using NASA's Kuiper Airborne Observatory (Phillips, Kwan, and Huggins, 1980). The detected radiation is consistent with weak maser action, as predicted by de Jong (1973).

Although the qualitative results of de Jong's detailed calculation are in agreement with observations, their details cannot carry much weight. The excitation mechanism was collisions with H_2 molecules for which variations on dipole collisions were used. Dipole collisions describe adequately electron excitations, but they cannot be considered a reasonable first-order guess for collisions with neutral molecules, as mentioned above. In fact, the cross sections for H_2O excitations in collisions with H_2 were calculated recently (Green, 1980), and the tabulated results do not even remotely resemble anything which can be deduced from dipole collisions. Unfortunately, the few detailed calculations performed since de Jong's original work used the same cross sections (e.g., Deguchi, 1977). A calculation using the cross sections of Green is long overdue.

The qualitative discussion and the potential detection of weak maser radiation in 4_{14}-3_{21} show that an inversion is apparently easy to achieve, irrespective of the excitation mechanism of the H_2O molecule. The question of the pump therefore has to be decided on the basis of the conditions at the given source. A detailed model calculation of the H_2O emission from late-type stars should not be too difficult in view of the good success of OH calculations, which indicate that the basic conditions in the expanding shell are probably reasonably well understood. A successful calculation for these masers would provide some further insight into the question of H_2O inversion mechanisms and may help in tackling the much more difficult—and interesting—problem of H_2O masers in regions of star formation.

The H_2O masers in regions of star formation are the least understood maser sources. No clear-cut association between the maser spots and any other family of objects has yet been established. In addition, the emission from the strongest sources is so intense that it is not even clear what type of mechanism could be pumping them. Every pump mechanism proposed so far leads also to production of IR radiation in amounts which exceed the observed values (Forster, Welch, and Wright, 1977). The difficulty in achieving the highest observed brightness temperatures is best illustrated using the result for a radiatively pumped saturated maser [Eqs. (3.32) and (3.33)]

$$T_m = \eta W_p \left[\frac{4\pi}{\Delta\Omega_m} \right] \left[\frac{\nu_p}{\nu_m} \right]^2 T_p \ , \tag{6.1}$$

assuming the same fractional bandwidth for the maser and pump radiations. The pump temperature T_p cannot be more than a few times 10^3 K, at most. Its enhancement through the effect of the beaming factor $4\pi/\Delta\Omega_m$ is probably approximately offset by the combined effects of the pump efficiency η and the dilution factor W_p. The main enhancement has to come from the phase-space factor $(\nu_p/\nu_m)^2$. Even pumping through high vibrational transitions at $\sim 2\ \mu$, which are, in fact, deficient in photons, would produce an enhancement factor of only $\lesssim 10^8$, much too small to account for $T_m \sim 10^{15}$ K.

Collisional pumps could not overcome this difficulty if the source was at a state of approximate overall thermal equilibrium, since in that case the excitation rates for collisions and radiation would be expected to be roughly similar (Sec. III.C).

It therefore appears that the pump cycle must involve a situation which is inherently different from thermodynamic equilibrium. One possible mechanism which falls under this category was proposed by Goldreich and Kwan (1974b), and it involves the presence of dust in the source, at a temperature different from that of the gas. If the dust was the species which determined the optical depths of the entire system in certain pump lines, the thermodynamic constraints could be eased, because the dust could be "hiding" some of the photons. Such a situation would have to be time varying, however, because the dust and gas temperatures would equilibrate in steady state. Tarter and Welch (1979) suggested that the surface of collision between a small, dense cloud and either an HII region or another cloud could be locally heated and provide, for a short time, the conditions for operation of such a pump. No detailed calculation of such a model has yet been presented, however.

The great difficulties with understanding the strong water masers are evident. At the same time, these may be the most interesting masers from the astronomical point of view. They are closely related to some stages, as yet not understood, of star formation and as such may carry the most important information of any maser source.

VII. CONCLUDING REMARKS

Maser radiation provides useful information about the conditions at the emitting source. This is best demonstrated in the case of late-type stars where the upper atmosphere is probed by the SiO maser emission and the expanding shell is mapped by the H_2O and OH masers. The maser radiation in itself carries only a limited amount of information and has to be supplemented by other information to become a valuable tool. This may well be the reason why the level of understanding of masers in late-type stars is so much higher than that of masers in regions of star formation. It may simply reflect our overall poor understanding of the star formation process, which is generally regarded as the most important outstanding problem in galactic astronomy. In spite of the great difficulties associated with them, studies of the masers in regions of star formation are probably the most important task of maser research in the near future.

Portions of this review were written at CSIRO, Epping, Australia, during a visit sponsored by the Distinguished Visitors Program at the Division of Radiophysics. I would like to thank all members of the division, especially R. N. Manchester and J. L. Caswell, for the warm hospitality and stimulating discussions enjoyed during the visit. In addition, I thank P. Goldreich, J. Moran, T. Troland, J. Welch, and B. Zuckerman for their very useful comments on the manuscript. This work was supported in part by NSF grant AST-8023712 to the University of Kentucky.

REFERENCES

Aannestad, P. A., 1975, Astrophys. J. **200**, 30.
Baldwin, J. E., C. S. Harris, and M. Ryle, 1973, Nature (London) **241**, 38.
Barrett, A. H., and A. E. E. Rogers, 1966, Nature (London) **210**, 188.
Beaudet, R. A., and R. L. Poynter, 1978, J. Phys. Chem. Ref. Data **7**, 311.
Bieniek, R. J., and S. Green, 1981, Chem. Phys. Lett. **84**, 380.
Biraud, F., G. Bourgois, J. Crovisier, R. Fillit, E. Gerard, and I. Kazes, 1974, Astron. Astrophys. **34**, 163.
Booth, R. S., A. J. Kus, R. P. Norris, and N. D. Porter, 1981, Nature (London) **290**, 382.
Bouloy, D., and A. Omont, 1977, Astron. Astrophys. **61**, 405.
Bouloy, D., and A. Omont, 1979, Astron. Astrophys. Suppl. Ser. **38**, 101.
Bujarrabal, V., J. L. Destombes, J. Guibert, C. Marliere-Demuynck, N. Q. Rieu, and A. Omont, 1980, Astron. Astrophys. **81**, 1.
Bujarrabal, V., J. Guibert, N. Q. Rieu, and A. Omont, 1980, Astron. Astrophys. **84**, 311.
Burdyuzha, V. V., and D. A. Varshalovich, 1972, Astron. Zh. **49**, 727 [Sov. Astron. AJ **16**, 597 (1973)].
Burke, B. F., D. C. Papa, G. D. Papadopoulos, P. R. Schwartz, S. H. Knowles, W. T. Sullivan, M. L. Meeks, and J. M. Moran, 1970, Astrophys. J. Lett. **160**, L63.
Castor, J. I., 1970, Mon. Not. R. Astron. Soc. **149**, 111.

Caswell, J. L., and R. F. Haynes, 1981, CSIRO preprint RPP 2517.
Cheung, A. C., D. M. Rank, C. H. Townes, D. D. Thornton, and W. J. Welch, 1969, Nature (London) **221**, 626.
Clark, F. O., T. H. Troland, and D. R. Johnson, 1981, Univ. of Kentucky preprint.
Cook, A. H., 1966a, Nature (London) **210**, 611.
Cook, A. H., 1966b, Nature (London) **211**, 503.
Deguchi, S., 1977, Publ. Astron. Soc. Jpn. **29**, 669.
Dixon, R. N., and D. Field, 1980, in *Interstellar Molecules,* edited by B. H. Andrew, International Astronomical Union Symposium 87 (Reidel, Dordrecht), p. 583.
Dreher, J. W., and W. J. Welch, 1981, Astrophys. J. **245**, 857.
Elitzur, M., 1976a, Astrophys. J. **203**, 124.
Elitzur, M., 1976b, Astron. Astrophys. **52**, 213.
Elitzur, M., 1977, Astron. Astrophys. **57**, 179.
Elitzur, M., 1978, Astron. Astrophys. **62**, 305.
Elitzur, M., 1979, Astron. Astrophys. **73**, 322.
Elitzur, M., 1980, Astrophys. J. **240**, 553.
Elitzur, M., 1981, in *Physical Processes in Red Giants,* edited by I. Iben and A. Renzini (Reidel, Dordrecht), p. 363.
Elitzur, M., and T. de Jong, 1978, Astron. Astrophys. **67**, 323.
Elitzur, M., P. Goldreich, and N. Scoville, 1976, Astrophys. J. **205**, 384.
Evans, N. J., R. E. Hills, O. E. H. Rydbeck, and E. Kollberg, 1972, Phys. Rev. A **6**, 1643.
Evans, N. J., S. Beckwith, R. L. Brown, and W. Gilmore, 1979, Astrophys. J. **227**, 450.
Forster, J. R., W. J. Welch, and M. C. H. Wright, 1977, Astrophys. J. Lett. **215**, L121.
Goldreich, P., and D. A. Keely, 1972, Astrophys. J. **174**, 517.
Goldreich, P., D. A. Keely, and J. Y. Kwan, 1973a, Astrophys. J. **179**, 111.
Goldreich, P., D. A. Keely, and J. Y. Kwan, 1973b, Astrophys. J. **182**, 55.
Goldreich, P., and J. Kwan, 1974a, Astrophys. J. **190**, 27.
Goldreich, P., and J. Kwan, 1974b, Astrophys. J. **191**, 93.
Goldreich, P., and N. Scoville, 1976, Astrophys. J. **205**, 144.
Goss, W. M., I. A. Lockhart, and E. B. Fomalont, 1975, Astron. Astrophys. **40**, 439.
Green S., 1980, Astrophys. J. Suppl. **42**, 103.
Green, S., and P. Thaddeus, 1976, Astrophys. J. **205**, 766.
Guilbert, J., M. Elitzur, and N. Q. Rieu, 1978, Astron. Astrophys. **66**, 395.
Habing, H. J., W. M. Goss, H. E. Matthews, and A. Winnberg, 1974, Astron. Astrophys. **35**, 1.
Harvey, P. M., K. P. Bechis, W. J. Wilson, and J. A. Ball, 1974, Astrophys. J. Suppl. **27**, 331.
Haynes, R. F., and J. L. Caswell, 1977, Mon. Not. R. Astron. Soc. **178**, 219.
Herman, J., and H. Habing, 1981, in *Physical Processes in Red Giants,* edited by I. Iben and A. Renzini (Reidel, Dordrecht), p. 383.
Jaynes, E. T., and F. W. Cummings, 1963, Proc. IEEE **51**, 89.
Jewell, P. R., M. Elitzur, J. C. Webber, and L. E. Snyder, 1979, Astrophys. J. Suppl. **41**, 191.
Jewell, P. R., J. C. Webber, and L. E. Snyder, 1980, Astrophys. J. Lett. **242**, L29.
Johnston, I. D., 1967, Astrophys. J. **150**, 33.
Johnston, K. J., S. H. Knowles, W. T. Sullivan, J. M. Moran, B. F. Burke, K. Y. Lo, D. C. Papa, G. D. Papadopoulos, P. R. Schwartz, C. A. Knight, I. I. Shapiro, and W. J. Welch, 1971, Astrophys. J. Lett. **166**, L21.
de Jong, T., 1973, Astron. Astrophys. **26**, 297.
Kaplan, H., and M. Shapiro, 1979, Astrophys. J. Lett. **229**, L91.
Kosloff, R., A. Kafri, and R. D. Levine, 1977, Astrophys. J. **215**, 497.
Kwan, J., and N. Scoville, 1974, Astrophys. J. Lett. **194**, L97.
Lamb, W. E., 1964, Phys. Rev. **134**, 1429A.
Litvak, M. M., 1969a, Science **165**, 855.
Litvak, M. M., 1969b, Astrophys. J. **156**, 471.
Litvak, M. M., 1970, Phys. Rev. A **2**, 2107.
Litvak, M. M., 1973, Astrophys. J. **182**, 711.
Litvak, M. M., A. L. McWhorter, M. L. Meeks, and H. J. Zeiger, 1966, Phys. Rev. Lett. **17**, 821.
Lo, K. Y., and K. D. Bechis, 1973, Astrophys. J. Lett. **185**, L71.
Lucas, R., 1980, Astron. Astrophys. **84**, 36.
Mies, F. H., 1974, Astrophys. J. Lett. **191**, L145.
Moran, J. M., 1981, Bull. Am. Astron. Soc. **13**, 508.
Moran, J. M., B. F. Burke, A. H. Barrett, A. E. E. Rogers, J. C. Carter, J. A. Ball, and D. D. Cudaback, 1968, Astrophys. J. Lett. **152**, L98.
Moran, J. M., J. A. Ball, C. R. Predmore, A. P. Lane, G. R. Huguenin, M. J. Reid, and S. S. Hansen, 1979, Astrophys. J. Lett. **231**, L67.
Moran, J. M., S. Lichten, M. Reid, R. Huguenin, and R. Predmore, 1981, in preparation.
Norris, R. P., and R. S. Booth, 1981, Mon. Not. R. Astron. Soc. **195**, 213.
Parker, G. A., and R. T. Pack, 1978, J. Chem. Phys. **68**, 1585.
Phillips, T. G., J. Kwan, and P. J. Huggins, 1980, in *Interstellar Molecules,* edited by B. H. Andrew, International Astronomical Union Symposium 87 (Reidel, Dordrecht), p. 21.
Procaccia, I., and R. D. Levine, 1975, J. Chem. Phys. **63**, 4261.
Rieu, N. Q., A. Winnberg, J. Guibert, J. R. D. Lepine, L. E. B. Johansson, and W. M. Goss, 1976, Astron. Astrophys. **46**, 413.
Rosen, R. A., 1974, Astrophys. J. Lett. **190**, L73.
Reid, M. J., and D. O. Muhelman, 1975, Astrophys. J. Lett. **196**, L35.
Reid, M. J., A. D. Maschick, B. F. Burke, J. M. Moran, K. J. Johnston, and G. W. Swenson, 1980, Astrophys. J. **239**, 89.
Reid, M. J., and J. M. Moran, 1981, Ann. Rev. Astron. Astrophys. **19**, 231.
Sakurai, J. J., 1964, *Invariance Principles and Elementary Particles* (Princeton University, Princeton).
Schultz, G. V., Sherwood, W. A., and A. Winnberg, 1978, Astron. Astrophys. **63**, L5.
Schwarzchild, M., 1975, Astrophys. J. **195**, 137.
Serber, R., and C. H. Townes, 1960, in *Quantum Electronics,* edited by C. H. Townes (Columbia University, New York), p. 233.
Shklovskii, I. S., 1969, Astron. Zh. **46**, 3 [Sov. Astron. AJ **13**, 1 (1969)].
Snyder, L. E., and D. Buhl, 1974, Astrophys. J. Lett. **189**, L31.
Sobolev, V. V., 1958, "Problems of the physics of stars with bright spectral lines," in *Theoretical Astrophysics,* edited by V. A. Ambartsumyan, translated by J. B. Sykes (Pergamon, London), Chap. 29, p. 497.
Spitzer, L., 1978, *Physical Processes in the Interstellar Medium* (Wiley, New York).
Tarter, J. C., and W. J. Welch, 1979, Bull, Am. Astron. Soc. **11**, 689.
Tenorio Tagle, G., H. W. Yorke, and P. Bodenheimer, 1979, Astron. Astrophys. **80**, 110.
Thronson, A. H., and D. A. Harper, 1979, Astrophys. J. **230**,

133.
Townes, C. H., and A. L. Schawlow, 1956, *Microwave Spectroscopy* (McGraw-Hill, New York).
Troland, T. H., C. Heiles, D. R. Johnson, and F. O. Clark, 1979, Astrophys. J. **232**, 143.
Varshalovich, D. A., and V. V. Burdyuzha, 1975, Astron. Zh. **52**, 1178 [Sov. Astron. AJ **19**, 705 (1976)].
Watson, W. D., M. Elitzur, and R. J. Bieniek, 1980, Astrophys. J. **240**, 547.
Weinreb, S., A. H. Barrett, M. L. Meeks, and J. C. Henry, 1963, Nature **200**, 829.
Werner, M. W., S. Beckwith, I. Gately, K. Sellgreen, G. Berriman, and D. L. Whiting, 1980, Astrophys. J. **239**, 540.
Whiteoak, J. B., and F. F. Gardner, 1978, Mon. Not. R. Astron. Soc. **183**, 67p.
Williams, D. R., and F. F. Gardner, 1981, CSIRO preprint.
Willson, L. A., and S. J. Hill, 1979, Astrophys. J. **228**, 854.
Wilson, W. J., and A. H. Barrett, 1970, Astrophys. Lett. **6**, 231.
Winnberg, A., D. A. Graham, and C. M. Walmsley, 1981, Astron. Astrophys. **93**, 79.
Wynn-Williams, C. G., E. E. Becklin, and G. Neugebauer, 1972, Mon. Not. R. Astron. Soc. **160**, 1.
Yorke, H. W., and E. Krugel, 1977, Astron. Astrophys. **54**, 183.