THE ASTROPHYSICAL JOURNAL, 494:218–235, 1998 February 10


# DOPPLER SHIFT ASYMMETRY IN HIGH-VELOCITY MASER EMISSION FROM SHOCKS IN CIRCUMNUCLEAR DISKS

EYAL MAOZ[1,2] AND CHRISTOPHER F. MCKEE[1,3]
*Received 1997 March 5; accepted 1997 September 18*

## ABSTRACT

The rapidly rotating, masing circumnuclear disk in the central subparsec region of the galaxy NGC 4258 is remarkably circular and Keplerian, yet a striking asymmetry appears in the maser spectrum: the redshifted, high-velocity sources are much more numerous and significantly more intense than the blueshifted ones. A similar strong asymmetry also appears in the recently discovered, masing, circumnuclear disks in NGC 1068 and NGC 4945, thus suggesting it may be a general phenomenon.

We show that the observed Doppler shift asymmetry can naturally arise due to spiral shocks in circumnuclear disks. We argue that population inversion can largely be quenched in these systems because of IR photon trapping, and that the high-velocity maser emission originates within thin slabs of postshock gas, where the physical conditions are conducive to maser action. The high-velocity masers with the longest gain paths appear where the line of sight is tangent to shock fronts. Since the spirals have a trailing geometry due to the action of differential rotation, the locations of the masers make the blueshifted radiation travel through a column of noninverted gas that maintains close velocity coherence with the maser source, where absorption occurs. The resulting asymmetry in the high-velocity maser spectrum, where the redshifted emission appears systematically stronger, is independent of the existence of a warp in the disk or the azimuthal direction to the observer, and is insensitive to small distortions in the velocity field in the disk.

The high velocities of these features reflect the rotational velocities in the disk and have nothing to do with the shock speed. The low-velocity emission arises within a narrow annulus near the inner edge of the disk, where direct irradiation by a central source may provide the energy that ultimately powers these masers. In NGC 4258—currently the most well-defined masing disk—the proposed scenario can also account for the intriguing clustering of the high-velocity maser spots in distinct clumps, the restricted radial distribution of the low-velocity sources, and the dip in the maser spectrum at the systemic velocity of the disk. In this case, we infer a molecular density of $\sim 10^9$ cm$^{-3}$, a disk mass of $\sim 10^4$ $M_\odot$, and a mass accretion rate of order $\sim 7 \times 10^{-3}$ $M_\odot$ yr$^{-1}$, which is consistent with an advection-dominated accretion flow. These results differ significantly from those of the Neufeld and Maloney model ($\simeq 10^{7.5}$ cm$^{-3}$, $\sim 100$ $M_\odot$, and $\sim 7 \times 10^{-5}\alpha$ $M_\odot$ yr$^{-1}$, respectively).

The predicted maser luminosities of the blueshifted and redshifted, high-velocity features in NGC 4258 are consistent with the observations, both in the case of C-type (MHD) shocks and dissociative J-type shocks, where the shock speed is about 20 km s$^{-1}$. The high-velocity features arise nearly along a diameter through the disk that makes an angle of about 2° with the midline. It does not introduce any noticeable deviation from a Keplerian rotation curve (the velocity gradient across the shock is always perpendicular to the line of sight at a maser location). The corrections to the previously derived black hole mass and galaxy distance are negligible. Predictions include slow systematic drifts in the velocity and position of all the high-velocity features, a systematic displacement in the locations of the high-velocity maser sources from the disk midline, and the existence of circumnuclear disks that are delineated only by high-velocity maser emission.

*Subject headings:* accretion, accretion disks — galaxies: individual (NGC 4258, NGC 1068, NGC 4945) — galaxies: nuclei — masers — shock waves

## 1. INTRODUCTION

Very long baseline interferometry (VLBI) observations by Miyoshi et al. (1995) have revealed that the water maser emission at the center of the galaxy NGC 4258 arises in a rapidly rotating molecular disk ($\simeq 900$ km s$^{-1}$ at $\simeq 0.2$ pc), viewed nearly edge-on, surrounding a massive black hole (Maoz 1995a). The high-velocity maser sources have line-of-sight velocities that trace a Keplerian rotation curve to a remarkably high precision. The fact that the redshifted and blueshifted high-velocity maser emission fit into the same Keplerian relation indicates that the disk is very nearly circular. Yet, a striking asymmetry appears in the maser spectrum: the redshifted high-velocity features on the receding side of the disk are much more numerous and significantly more intense than the blueshifted ones on the approaching side of the disk, having a total luminosity ratio of $\simeq 23$ (Nakai, Inoue, & Miyoshi 1993; Miyoshi et al. 1995).

Evidence for masing circumnuclear disks has also been found in other galaxies. The most compelling case is the rotating structure in NGC 1068 (Gallimore et al. 1996; Greenhill et al. 1996). It is delineated by high-velocity, $H_2O$ maser emission ($\simeq 300$ km s$^{-1}$ at $\simeq 0.7$ pc) where, again, the redshifted features are much stronger and more numerous than the blueshifted ones, being even stronger than the fea-

[1] Astronomy Department, University of California, Berkeley, Berkeley, CA 94720.
[2] Miller Fellow.
[3] Physics Department, University of California, Berkeley, Berkeley, CA 94720.

tures with the systemic velocity of the galaxy (hereafter, the "systemic" features). A similar situation appears at the center of NGC 4945, where the water maser emission consists of systemic features and features shifted roughly symmetrically, in a nearly planar structure, by about $\pm$(100–150) km s$^{-1}$ (Greenhill, Moran, & Herrnstein 1997). A disk origin of the maser emission is indicated also by the position-velocity gradient detected in the systemic emission (Greenhill 1996) and by the fact that the line connecting the systemic and redshifted features is perpendicular to the axis of the observed ionization cone (Nakai 1989). The disk radius is about $\approx$0.3 pc, and the binding mass is $\sim 10^6 M_\odot$. The redshifted, high-velocity features are stronger than the systemic features and significantly stronger than the blueshifted features (Greenhill et al. 1997). Evidence for a rotating disk structure ($\simeq$100 km s$^{-1}$ at $r \simeq 40$ pc), delineated by OH maser emission, has also been found in the galaxy III Zw 35, which is located at a distance of $\simeq$100 Mpc (Montgomery & Cohen 1992). Again, the redshifted maser features are stronger. Wilson, Braatz, & Henkel (1995) found that the systemic maser emission from the nucleus of the galaxy NGC 2639 exhibits a velocity drift similar to that observed in NGC 4258, which suggests that it too may arise due to centripetal acceleration in a rotating disk. High-velocity maser features, though, are yet to be found in NGC 2639. At present there is no evidence for a masing circumnuclear disk where the blueshifted, high-velocity emission is stronger than the redshifted emission. This suggests that the relatively enhanced, redshifted, high-velocity maser emission may be a general phenomenon in circumnuclear disks.

High-velocity maser emission, by which we mean emission displaced by $\gtrsim$100 km s$^{-1}$ from the systemic velocity of the galaxy, need not necessarily always arise in a rotating disk. It may arise in outflows from galactic nuclei, in which case the blueshifted features are likely to appear stronger because the radially approaching side of the flow may amplify a background continuum source at the center of expansion, as found in a case of interstellar masers (Gwinn 1994). This might be the case in NGC 3079, for example, where the maser emission features cover a velocity range that lies completely below the systemic velocity of the galaxy (Baan & Haschick 1996). High-velocity maser emission may arise along radio jets, possibly at shock interfaces between the jet and the dense near-nucleus gas, as found along the jet of NGC 1068 (Gallimore et al. 1996). If the maser emission originates in the vicinity of a massive black hole, whether in a disk, an outflow, or jets, then identifying features as "systemic" or "satellite" can be confused by the black hole motion within the galaxy nucleus: massive black holes may have substantial peculiar velocities (see, e.g., Miller & Smith 1992 and references therein) that would shift the maser spectrum in either direction by up to the velocity dispersion in the galaxy nucleus. Therefore, evidence for a disk origin of maser emission must rely on both the kinematic and spatial distribution of the maser sources.

Water maser emission has been also detected at the centers of 15 other galaxies (Braatz, Wilson, & Henkel 1996 and references therein). The origin of the maser emission in these galaxies is currently unknown, but it is reasonable to expect a disk origin in at least a few of the cases. Thus, it is interesting to see whether the redshifted features in these maser spectra tend, in a statistical sense, to be stronger relative to the blueshifted ones. Indeed, the strongest maser features in each spectrum are redshifted by 45–200 km s$^{-1}$ relative to the galaxy systemic velocity in 10 cases (Mrk 1, NGC 1052, NGC 1386, Mrk 1210, IC 2560, Circinus, ESO 103-G35, TXFS 2226-184, IC 1481, and M51), blueshifted by 85–150 km s$^{-1}$ in three cases (NGC 3079, NGC 5506, and NGC 253), and consistent with the systemic velocity of the galaxy in NGC 2639 and NGC 5347. This is compatible with the suggestion that the Doppler shift asymmetry is a general phenomenon in circumnuclear disks, yet the above finding should not be regarded as conclusive evidence because (1) the origin of the maser emission in these galaxies has not yet been determined, (2) the asymmetry is not always very pronounced, (3) the relative strength of spectral features may vary in time, and (4) the velocities of some of the features are not very "high."

Herrnstein, Greenhill, & Moran (1996) proposed that the Doppler shift asymmetry in NGC 4258 may be attributed to the warp in the disk. They suggested that free-free absorption in the upper atmosphere of the disk attenuates the maser radiation only from the blueshifted side of the disk because of the orientation of the warp with respect to the line of sight. Although plausible for an isolated case, such an explanation relies on a coincidentally special direction to the observer and is therefore improbable for a class of circumnuclear disks, particularly in light of the absence of a clear counterexample (a disk with stronger blueshifted emission). Conceivable explanations that rely on maser amplification of nonuniform continuum source at the background of the disk or on anisotropic irradiation of the disk by the central source are improbable for the same reasons.

We show that the observed Doppler shift asymmetry—in which redshifted high-velocity maser emission appears stronger than the blueshifted emission—can naturally arise due to shocks in circumnuclear disks, independent of the azimuthal direction to the observer. In § 2 we argue that circumnuclear disks could be largely nonmasing due to IR photon trapping, meaning that maser radiation can be absorbed in regions in the disk that maintain close velocity coherence with a maser source. In § 3 we summarize results of previous investigations that have shown that the conditions for maser action are favorable behind shock fronts and argue that spiral shocks may form in circumnuclear disks for the same reasons they do in other astrophysical disk systems. In § 4 we describe the geometry of the masing regions behind the spiral shock fronts and derive the expected isotropic luminosities of redshifted, high-velocity features. In § 5 we show that the trailing geometry of the shock fronts gives rise to excess absorption of the blueshifted, high-velocity maser emission, thus producing the Doppler shift asymmetry. We describe the geometry of the absorbing regions and calculate the optical depth for absorption. In § 6 we discuss, in detail, the remarkably well-defined disk in NGC 4258, where our model can naturally explain the intriguing spectrum and spatial distribution of the high-velocity maser emission, as well as the general properties of the low-velocity emission. We discuss predictions and compare our model with that of Neufeld & Maloney (1995). In § 7 we discuss the implications of our results to other circumnuclear disks in general, and to NGC 1068 in particular. In § 8 we conclude and summarize the main points of this investigation.

## 2. A CONDITION FOR INVERTED LEVEL POPULATIONS

Maser radiation passing though molecular gas that is at rest relative to the maser source is either amplified, when

the level populations of the masing transition are inverted, or absorbed, when the level populations are not inverted. Inversion can be maintained by a radiative or collisional pumping mechanism, but the strong extragalactic masers are unlikely to be pumped radiatively, since radiative pumping from continuous-spectrum sources requires pumping efficiency of $\sim 100\%$, which is implausible (Goldreich & Kwan 1974; Downes 1983; Genzel 1986; Reid & Moran 1988). The most likely pump for extragalactic, 22 GHz water masers is collisional.

Inversion can be achieved only when either the pump or loss rates for the two levels are different. In collisional pumping of $H_2O$, inversion arises owing to there being more spontaneous decays from high-lying states, populated by collisions, down to the upper maser level than there are down to the lower maser level (De Jong 1973; Strelnitskii 1971; Goldreich & Kwan 1974). However, at densities greater than $\sim 10^{12}$ cm$^{-3}$, collisions reshuffle the $H_2O$ level populations into a thermal distribution with the kinetic temperature of the gas (Hollenbach, Elitzur, & McKee 1998; hereafter HEM98). Inversion cannot be maintained at such high densities, except in cases of inherently non-LTE processes such as collisional pumping by charged particles that are much hotter than neutrals.

The collisional pump in $H_2O$ is characterized by the emission of infrared line photons owing to spontaneous decays from high-lying states to the upper maser level and to spontaneous decays from the lower maser level. It is essential that these infrared photons not be reabsorbed by the molecules if inversion is to be maintained. The IR photons could be removed from the medium either by escaping from the masing region, or they may be destroyed by being absorbed by cold dust within the masing region. Thus, the level populations could not be inverted in regions where the minimal distance to the boundary of the local velocity-coherent region is above some critical value—a scale-length that decreases with increasing density and water abundance and increases with increasing abundance of cold dust.

The global geometry plays a crucial role in determining whether the level populations are locally inverted or not, and it strongly affects the local maser efficiency. Detailed numerical calculations of the equilibrium level populations in collisional pumping schemes (Elitzur, Hollenbach, & McKee 1989, hereafter EHM89; HEM98) show that the characteristics of $H_2O$ maser inversion can be summarized quite accurately using the scaling variable, $\xi$, effectively the maser emission measure, defined by

$$\xi \equiv \frac{x_{-4}(\mathrm{H_2O}) n_{\mathrm{H},9}^2 \, d_{13}}{\Delta v_5}, \tag{1}$$

where $d_{13} \equiv d/10^{13}$ cm is the optical-depth length scale that controls photon escape, $n_{\mathrm{H},9} \equiv n_{\mathrm{H}}/(10^9 \text{ cm}^{-3})$ is the hydrogen nuclei density, $x_{-4}(\mathrm{H_2O}) \equiv [n(\mathrm{H_2O})/n_{\mathrm{H}}]/10^{-4}$ is the water abundance, and $\Delta v_5 \equiv \Delta v/(10^5 \text{ cm s}^{-1})$ is the maser line width (typically between 1 and a few km s$^{-1}$). These calculations assume static media where the only velocity gradients present are due to thermal velocities and possibly to small-scale chaotic motions (turbulence) in the gas. Neufeld & Melnick (1991) performed similar calculations at the presence of large-scale velocity gradients (LVG) and found that in the LVG limit (where $|dv_z/dz|$ across the masing region is much larger than $\Delta v/d$), the local equilibrium level populations are generally a function of the variable $\xi'$, defined as

$$\xi' \equiv \frac{x_{-4}(\mathrm{H_2O}) n_{\mathrm{H},9}^2}{(dv_z/dz)_{-8}}. \tag{2}$$

This approach enables us to treat also cases of media in which the velocity gradient is negligible by using the approximate scaling, $\xi' \simeq \xi/9$ (Neufeld & Melnick 1991).

The pump efficiency, $\eta \equiv \delta n/n$, which is positive when the level populations are inverted, falls monotonically with increasing $\xi$ in the standard maser parameter space, $0.01 \lesssim \xi \lesssim 10$, and drops abruptly at larger values of $\xi$ as the maser levels rapidly approach thermalization due to radiative trapping in the transitions that connect them to other levels. Population inversion disappears altogether ($\eta$ becomes negative) when $\xi$ is above some critical value, $\xi_{\mathrm{crit}}$, which depends primarily on the gas temperature (Neufeld & Melnick 1990; HEM98). The transition point occurs, for example, at $\xi_{\mathrm{crit}} \simeq 200$ at 400 K, and drops to about 40 at 100 K (again, notice that $\xi'_{\mathrm{crit}} \simeq \xi_{\mathrm{crit}}/9$). Evidently, in circumnuclear molecular disks, $\xi$ may exceed $\xi_{\mathrm{crit}}$ throughout most of the disk volume: molecular densities of $10^9$ cm$^{-3}$ are not outrageously high for these systems (notice that $n_{\mathrm{H},9} = 2n_9[\mathrm{H_2}]$ for molecular gas); the scale height of the gas distribution in the already discovered masing disks (§§ 6 and 7) is a few orders of magnitude larger than $10^{13}$ cm; and finally, the water abundance may be a few times higher than the interstellar medium value ($\sim 10^{-4}$) both because of the high metallicity in the galactic nucleus as well as the fact that copious amounts of water may be produced by X-ray irradiation of the disk (Neufeld, Maloney, & Conger 1994) and by shocks in the disk (§ 3). In such disks, the level populations are generally not inverted, except near the inner edge of the disk, which is irradiated directly by X-rays from a central nuclear source, and within two, very thin slabs in the upper atmosphere of the disk, where the IR photons can escape in a short distance vertically.

The effect of cold dust on the radiative transfer would allow larger values of $\xi$ before quenching takes place (Deguchi 1981; Collison & Watson 1995). Absorption of IR photons by the dust provides an additional heat sink for collisional pumps, but it requires that the dust be cooler than the gas. The higher the temperature difference, the larger is $\xi_{\mathrm{crit}}$ above which inversion is quenched. We shall assume that the dust and gas are nearly in thermal equilibrium, so the presence of dust would not increase $\xi_{\mathrm{crit}}$ enough to prevent quenching of the inversion in most of the disk volume.

## 3. MASER EMISSION FROM SHOCKS IN CIRCUMNUCLEAR DISKS

### 3.1. *Maser Action behind Shock Fronts*

A shock origin of interstellar $H_2O$ masers has been suggested by many authors (Litvak 1969; Strelnitskii & Sunyaev 1973; Shmeld, Strelnitskii, & Muzylev 1976; Elitzur 1979; Strelnitskii 1984; Tartar & Welch 1986; Hollenbach, McKee, & Chernoff 1987; EHM89; Melnick et al. 1993; Kaufman & Neufeld 1996; HEM98). It has been successful in explaining the observed, highly supersonic, radial and proper motion velocities of interstellar water maser features, as well as their high luminosities and small dimensions. Detailed modeling revealed that the physical

and chemical conditions in postshock regions are favorable for the generation of luminous, 22 GHz water maser emission by collisional pumping. The energy to pump the masers is naturally provided by the dissipation of the relative kinetic energies of the shocked and unshocked gas. In dissociative, J-type shocks, chemical reactions in the warm, dense postshock gas lead to the production of copious amounts of $H_2O$ (Elitzur 1979; McKee & Hollenbach 1980; EHM89; Neufeld & Hollenbach 1994). High abundances of warm $H_2O$ are also produced behind nondissociative, MHD (C-type) shocks ($10 \lesssim v_s \lesssim 40$–$50$ km s$^{-1}$) that occur in media of low fractional ionization at the presence of transverse magnetic field (Draine 1980; Draine, Roberge, & Dalgarno 1983; Chernoff, McKee, & Hollenbach 1982). The possibility of maser emission from C-type shocks was first recognized by EHM89 and later confirmed by detailed investigations (Melnick et al. 1993; Kaufman & Neufeld 1996). In either type of shock, maser action occurs within a column density of $\sim 10^{22}$–$10^{23}$ cm$^{-2}$ behind the shock front, in a slab of thickness $\sim 10^{13}$–$10^{14}/n_{\rm H,9}$ cm, where the maser emission measure is typically $\xi \sim 0.1$–$10$. The sheet-like geometry of the shocks allows long gain paths for amplification along rays in the shock plane. The IR photons produced in the pump cycle easily escape from the masing slab through the shock front due to the velocity difference between the postshock and preshock gas, thus avoiding quenching of the maser *regardless* of the dimension of the surrounding gas distribution. For all these reasons, shocks provide an ideal location for the generation of strong water masers.

It is important to note that maser action can take place behind shock fronts only when the preshock density does not exceed $\sim 10^{10}$ cm$^{-3}$, depending exactly on the shock type and strength, since otherwise the postshock density may reach $\approx 10^{12}$ cm$^{-3}$, at which $H_2O$ masers are quenched due to collisions (HEM98). The maximum allowed preshock density is higher in the case of C-type shocks since, although the total compression in a C-type shock is $\sim 10$–$20$, the hottest part of the shock occurs where the compression in the neutral component is less than a factor of 2 (Kaufman & Neufeld 1996).

### 3.2. *On the Presence of Spiral Shocks*

In circular disks, the relative speed between orbits separated by a distance larger than the disk thickness is supersonic, so even small nonaxisymmetric disturbances may lead to the formation of nonaxisymmetric shocks, which shear into a trailing spiral geometry because of the action of differential rotation. The existence and importance of spiral shocks were first recognized in the context of galactic disks, where shocks were found to develop naturally, indeed necessarily, in the presence of spiral activity (Fujimoto 1966; Roberts 1969, 1970; Shu, Milione, & Roberts 1973; Woodward 1973; Roberts, Roberts, & Shu 1975). Shocks along the trailing, slowly rotating spiral waves were suggested to be the mechanism that triggers gravitational collapse and rapid star formation, explaining the observed tendency of young star associations and brilliant H II regions to lie along the inner sides of gaseous spiral arms (see, e.g., Roberts 1967; Westerhout 1968a, 1968b). The formation of spiral shocks and their role in the outward transport of angular momentum have also been discussed and simulated in the context of protostellar disks and accretion disks (see, e.g., Sawada, Matsuda, & Hachisu 1986; Matsuda et al. 1987; Spruit 1987; Spruit et al. 1987; Morfill, Spruit, & Levy 1993).

It is only reasonable to expect that spiral shocks may form in circumnuclear disks as well. In fact, spiral shocks in active galactic nucleus disks have been suggested to explain the observed variability of double-peaked broad emission lines seen in some active galactic nuclei (Chakrabarti & Wiita 1994). The theory of spiral shock generation is quite complex, and current understanding of it may not be complete. It is clear that shock formation requires that the disk be irritated, even by weak disturbances, of which there is no shortage at the bottom of a galactic potential well. There are two basic mechanisms for spiral shock formation that have been thoroughly investigated in the contexts of accretion disks and spiral galaxies. The first is the generation of self-sustained spiral shocks (Lynden-Bell 1974; Donner 1979; Spruit 1987). This process requires a weak source of nonaxisymmetric disturbances in the outer parts of a disk. As a perturbation propagates inward, it is sheared by the differential rotation into a trailing spiral pattern, and its amplitude grows. The wave eventually steepens into a shock, with a further increase in amplitude until further steepening is balanced by dissipation in the shock (Spruit et al. 1987). From there inward, the strength of the shock is independent of the initial amplitude of the perturbation. Though shocks in this scenario are set up by some disturbance in the outer parts of the disk, they extend all the way in *without* further external forcing. The generation and support of such self-sustained, spiral-shaped shock waves has been observed in simulations (Sawada et al. 1986, 1987).

The second type of mechanism for generating spiral shocks relies on the self-gravity of a disk that maintains spiral density waves, as in the case of spiral galaxies. First, we should point out the distinction made between "grand-design" spiral galaxies (e.g., M51 and M81) and flocculent spirals (e.g., Sc galaxies). Grand-design spirals are believed to have a well-defined pattern speed, so material in the disk is orbiting faster than the waves inside corotation and more slowly than the waves outside corotation. The shocks occur as the orbiting gas slams into the spiral waves, which rotate at a different speed (Fujimoto 1966; Roberts 1969; Shu et al. 1972; Shu et al. 1973). Flocculent spirals consist of bits and pieces of spiral arms whose origin lie with the local gravitational instability (Goldreich & Lynden-Bell 1965; Julian & Toomre 1966; Toomre 1981). The constantly forming and dying spiral waves are triggered by imperfections in the disk, such as lumps that are formed accidently. The spiral activity is confined to the range of radii where the disk is on the border of stability, namely, where the local value of the Toomre Q parameter is of order unity. Each spiral wave rotates with the angular speed of its "irritator," so disk particles inside "corotation" overtake the wave because of the differential rotation, while those orbiting at larger radii are overtaken by the wave. Shocks may form along those portions of a wave where the local relative speed between the gas and the wave is sufficiently high, either inside or outside corotation.

To conclude, spiral shocks in circumnuclear disks may arise either due to steepening of self-sustained waves, the slamming of gas into spiral density waves (either of grand-design or flocculent type), or perhaps due to some as yet unknown mechanism. It is clear, though, that spiral shock formation depends, in one way or another, on the presence of disturbances, either local or at the outer edge of the disk,

of which there is no shortage at the bottom of a galactic potential well. Conceivable sources for disturbances include irregularities in the mass distribution in the disk and in the mass accretion rate at the disk edge, turbulence due to convection (Spruit et al. 1987), magnetic fields, flyby of globular clusters or chunks of giant molecular clouds, and self-gravitating clumps forming in the disk. The "outer edge" of the disk means, in our context, the region outside (but not too far outside) the observable disk. Obviously, the disk may extend much beyond the masing part, being undetectable in regions where shocks do not occur or where the physical conditions are unfavorable for maser action. In the case of NGC 4258, which we discuss in detail in § 6, the existence of a massive object orbiting outside the outer edge of the masing disk has been recently proposed to explain the origin of the warp in the disk (Papaloizou, Terquem, & Lin 1998). If such an object does exist, it might be the source of the nonaxisymmetric perturbation that triggers spiral shock formation in this system.

## 4. MASER EMISSION FROM SPIRAL SHOCKS

### 4.1. *General Geometry*

Let us consider a largely nonmasing circumnuclear disk that contains spiral shocks. The shock fronts can be described, to zeroth-order approximation, as bent sheets that are as wide as the disk thickness and perpendicular to the disk plane. As elaborated by Roberts (1969), the gas passes through successive shocks along its orbit: if the angular speed of the spiral waves is lower than that of the disk material, the gas enters each shock front from its concave side as preshock gas and emerges on its convex side as postshock gas (the other way around if a wave overruns the gas); the streamlines nearly repeat themselves through every intershock interval, being nearly concentric because there must be a small radial transfer of mass due to energy dissipation in the shocks. The shock speed is the flow velocity normal to the spiral in a frame corotating with the wave. Denoting the angle between the tangent to the spiral and the circumferential direction by $\theta_p$ (hereafter, the pitch angle), the shock speed is

$$v_s \simeq r\,|\,\omega - \Omega\,|\,\sin\theta_p\,, \tag{3}$$

where $\omega$ is the angular velocity of the unperturbed flow at radius $r$, and $\Omega$ is the angular velocity of the spiral wave.

Assuming the shock speed is high enough above the sound speed and the turbulent speed in the preshock gas to allow efficient IR photon escape through the shock front, the level populations of the $5_{23}$–$6_{16}$ water maser transition are inverted within thin, spiral-shaped slabs of thickness $d \sim (10^{13}$–$10^{14})/n_{\rm H,9}$ cm behind the shock fronts (§ 3.1). The general geometry is illustrated in Figure 1*a*, where each spiral region represents a masing slab behind a shock front, and the horizontal line is the midline—the diameter through the disk that is perpendicular to the observer's line of sight. The maser intensity depends on the path length through a velocity-coherent, inverted region: the mean intensity scales exponentially with column density when the maser is unsaturated and linearly when it is saturated. Thus, in the case of a curved slab, the most luminous, high-velocity masers would appear where the line of sight is tangent to an arc. The narrow, dark, rectangular region in Figure 1*a*, located where the line of sight is tangent to one of the masing slabs, represents such a strong maser. Since the

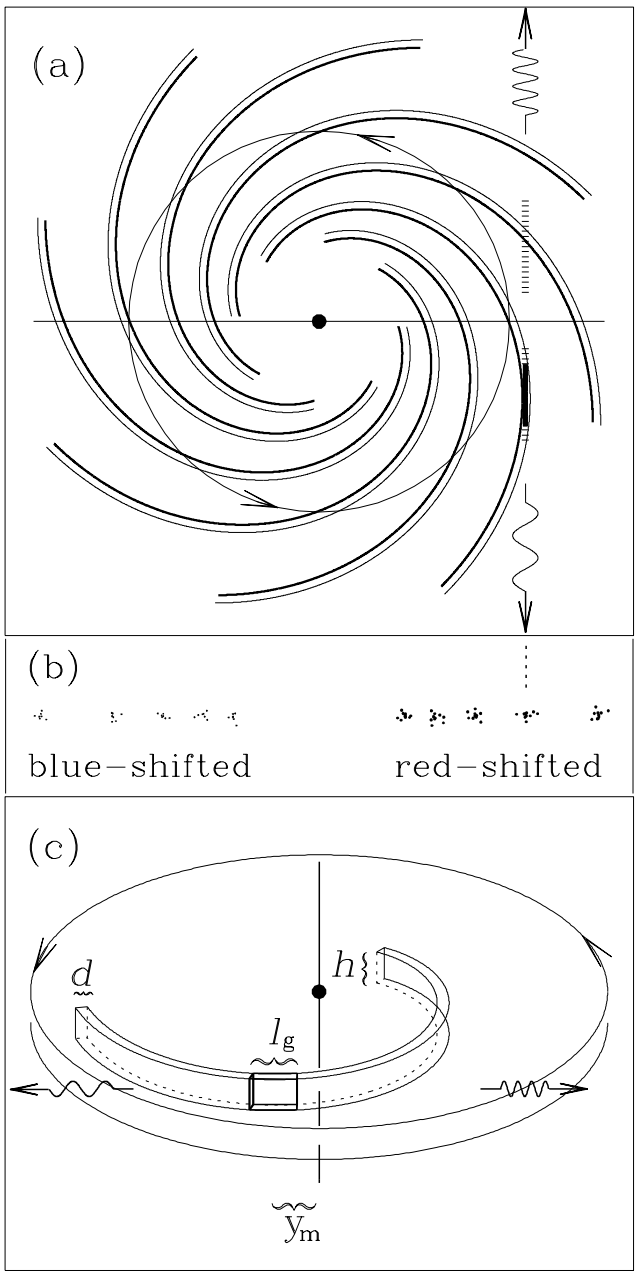


FIG. 1.—Illustration of the general geometry: (*a*) Spiral regions represent the thin masing slabs behind shock fronts, the horizontal line is the midline—the diameter through the disk perpendicular to the line of sight and the dark rectangular region—located where the line of sight is tangent to a spiral represents a high-velocity maser (§ 4.1). Since the disk is largely not masing, photons from that maser are absorbed in the shaded regions, where close velocity coherence is maintained with the maser source. The Doppler shift asymmetry arises owing to the trailing geometry of the spirals, which gives rise to excess absorption of the blueshifted radiation (§ 5.1). (*b*) Expected image of the maser emission as viewed by a nearly edge-on observer. The high-velocity maser features break into maser spots due to microturbulence and density inhomogeneities in the emitting region. The spots should appear clustered in distinct clumps rather than being uniformly or randomly distributed. The clumps may appear either narrow or broad, depending on whether the slabs are heavily corrugated and thereby effectively thickened. The positions of all the high-velocity maser features are predicted to slowly drift outward due to the rotation of the spiral waves (§ 6.3). (*c*) The high-velocity maser emission originates in a box-shaped region of thickness $d \sim (10^{13}$–$10^{14})/n_{\rm H,9}$ cm, height $h$, and length $l_g$, which is determined by $d$ and the local radius of curvature of the spiral. It is located at a distance $|\,y_m\,| \simeq r \sin\theta_p$ from the midline, where $\theta_p$ is the pitch angle of the spiral (§ 4.1).

maser amplifies only photons produced by spontaneous decays in the maser itself, there would be no systematic difference between the intensity of radiation emanating from the two ends of the maser (the relativistic beaming effect is negligible at the rotation speeds under consideration). The maser photons will appear redshifted to a distant observer in one direction and blueshifted to an observer in the diametrically opposed direction.

An immediate consequence of this scenario is that, regardless of the azimuthal direction to the observer, the disk should appear to be delineated by clumps of high-velocity maser spots rather than by uniformly or randomly distributed spots (maser features may each consist of several maser "spots" due to microturbulence or inhomogeneities in the emitting region). It should be made clear that the observed high-velocities of these masers reflect the rotational velocities in the disk and have nothing to do with the shock speed.

### 4.2. *Maser Luminosities*

The maser luminosity depends on the structure of the shocked region and thus on the shock type. At shock velocities $\lesssim 50$ km s$^{-1}$, shocks might be either of J- or C-type, depending upon such poorly known parameters such as the preshock magnetic field and the length scale of the ion-neutral coupling. We shall now estimate the expected isotropic luminosity of a high-velocity maser feature in either type of shock. We denote the disk radius at the maser location (where the line of sight is tangent to a shock front) by $r$ and the local thickness of the disk by $h$.

#### 4.2.1. *C-type Shocks*

Kaufman & Neufeld (1996; hereafter KN96) constructed a series of models of dense, slow, nondissociative, C-type (magnetohydrodynamic) shocks and computed the expected water maser emission. The maser luminosity produced per unit area of the shock front is $\mathscr{L}_{\rm prod} = \epsilon n_{\rm H} \mu_{\rm H} v_s^3/2$, where $\epsilon$ is the efficiency with which shock energy is converted into water maser emission, and $\mu_{\rm H} = 2.34 \times 10^{-24}$ g is the mass of gas per hydrogen nucleus. The integration of $\mathscr{L}_{\rm prod}$ over the surface area of the spiral shock is not an observable quantity, since measurement of it requires integration over all lines of sight. However, provided our line of sight is typical, this integral is directly related to the isotropic luminosity of the maser, $L_{\rm iso}$, which is observable ($L_{\rm iso} = 4\pi F_{\rm obs} D^2$, where $F_{\rm obs}$ is the observed flux from the maser, and $D$ is the galaxy distance). Since we are interested in the isotropic luminosity of a maser at a given radius, $r$, we may imagine that the segment of the spiral shock that emits maser photons in our direction is a part of a circular masing slab of radius $r$, width $h$, and maser luminosity per unit area given by $\mathscr{L}_{\rm prod}$. Denoting the half-beaming angle in the direction normal to the disk plane by $\theta_b$, the maser radiation from the ring is concentrated into a solid angle $\Delta\Omega = 4\pi \sin\theta_b$ (the maser radiation can be detected only by observers that view the disk with inclination larger than $90^\circ - \theta_b$). Keeping in mind that the observer's lines of sight are tangent to the circular slab at two points, which would result in two, equally bright maser features, the isotropic luminosity of a single high-velocity feature at a radius $r$ is $L_{\rm iso} = \frac{1}{2}(2\pi r h \mathscr{L}_{\rm prod})(4\pi/\Delta\Omega)$.

The fraction, $\epsilon_{\rm sat}$, of the mechanical energy in the shock that emerges in a water maser transition under the conditions where the maser action is saturated has been calculated by KN96, assuming cosmic-ray ionization and a preshock magnetic field of magnitude $B = 32 n_{\rm H,9}^{1/2}$ mG parallel to the shock plane (such scaling relation corresponds to a constant Alfvén speed of about 2 km s$^{-1}$). For example, assuming a shock speed of 20 km s$^{-1}$, molecular hydrogen density of $10^9$ cm$^{-3}$, and transverse magnetic field given by the above scaling relation, KN96 obtained $\epsilon_{\rm sat} \simeq 10^{-5.5}$, in which case the isotropic luminosity of a high-velocity maser feature would be

$$L_{\rm iso} \simeq 2.7 r_{18} h_{16} n_9({\rm H}_2) \times \left(\frac{\epsilon_{\rm sat}}{10^{-5.5}}\right)\left(\frac{v_s}{20\ {\rm km\ s^{-1}}}\right)^3\left(\frac{\theta_b}{10^\circ}\right)^{-1} L_\odot\ , \quad (4)$$

where $r_{18} \equiv r/10^{18}$ cm, and $h_{16} \equiv h/10^{16}$ cm.

We should emphasize the limitations of the above estimate for the expected maser luminosity: (1) the fractional ionization may be much higher in the nuclear environment, depending on the amount of X-ray irradiation and the vertical column density of gas in the disk; (2) assuming other quantities being equal, the energy conversion efficiency in the case of unsaturated masers, $\epsilon$, is probably different than $\epsilon_{\rm sat}$; (3) the efficiency, $\epsilon_{\rm sat}$, is not an independent variable but depends on the magnetic field, gas density, and the shock speed; and (4) the assumed scaling of the transverse magnetic field with gas density means that only a part of the parameter space relevant to circumnuclear conditions has been explored (it is possible that the Alfvén velocity in circumnuclear disks may be much different than $\simeq 2$ km s$^{-1}$). In NGC 4258, for example, there is an upper limit to the Alfvén velocity in the masing part of the disk of about 10 km s$^{-1}$, based on the upper limit to the disk thickness (§ 6.1.2). The strength of the magnetic field in the masing region is currently unknown, but it may be possible to measure it in the future if Zeeman splitting is detected in the maser spectrum. It is not obvious, at present, how variations in the preshock magnetic field, ionization, and grain size distribution will affect the emerging maser luminosity, so extending the study of KN96 to include other regions of parameter space is most desirable. We can only point out that the preshock magnetic field and the grain size distribution control the thickness and temperature of the shocked layer behind the C-shock front. A higher magnetic field would result in a shocked region that is thinner and hotter. On the other hand, the ion-neutral coupling length depends on the grain size distribution, and it increases if small ($\lesssim 4$ Å) grains are absent, leading to a thicker but cooler shocked region. If the fractional ionization is very high, then the shocks would be of J-type rather than C-type. In the case of J-type shocks, we take a different approach for estimating the expected water maser luminosity, as discussed below.

#### 4.2.2. *J-type Shocks*

In a J-type shock in molecular gas, the temperature just behind the shock front is $\approx 5300\ (v_s/10\ {\rm km\ s^{-1}})^2$ K—much higher than that in C-type shocks with the same shock speed—and it rapidly falls with postshock column density (McKee & Hollenbach 1980; Neufeld & Hollenbach 1994). A J-type shock may be either dissociative or nondissociative. At shock velocities smaller than $\sim 15$ km s$^{-1}$, J-type shocks dissociate only a small fraction of molecules (Hollenbach & McKee 1980; Neufeld & Hollenbach 1994). As demonstrated by Neufeld & Hollenbach (1994, Fig. 7) for the case of a preshock density of $10^9$ cm$^{-3}$ and $v_s = 10$

km $s^{-1}$ (nondissociative shock), maser action is collisionally quenched at postshock column densities $\gtrsim 10^{20}$ $cm^{-2}$ owing to the high compression factor. Consequently, the volume of the masing region is much smaller than that in a C-type shock, so we expect the maser luminosity to be significantly lower than that estimated in § 4.2.1.

A dissociative J-type shock, on the other hand, can produce high intrinsic maser luminosity (EHM89). While the temperature profile just behind the shock front is similar to that in a nondissociative shock, the heat of $H_2$ reformation on dust grains provides a substantial energy source that maintains a large column of $\sim$400 K gas, where the chemistry drives a considerable fraction of the oxygen not in CO to form $H_2O$. We now estimate the expected isotropic luminosity of a high-velocity maser source, assuming a dissociative J-type shock.

Denoting the maximum rate at which maser photons are produced per unit volume by $\Phi_{\rm sat}$, the emergent luminosity per unit surface area of the shock front is given by

$$\mathscr{L}_{\rm prod} \equiv h\nu \int \Phi_{\rm sat}\, dz\ , \tag{5}$$

where $z$ is the distance behind the shock front, and the integration is performed over the postshock region where the level populations of the maser transition are inverted. The maser photon production rate, $\Phi_{\rm sat}$, depends on the gas kinetic temperature, density, water abundance, and the velocity gradients in the masing region. It can be expressed as

$$\Phi_{\rm sat} = q\eta n_{\rm H}^2\, x({\rm H_2O})\ , \tag{6}$$

where $q$ is the pump rate coefficient, which depends primarily on $T$ and is almost independent of density. For the water 22 GHz transition, $q$ is well fit by (HEM98)

$$q \simeq 2.5 \times 10^{-11} \exp\,(-644/T)\ {\rm cm^3\ s^{-1}}\ . \tag{7}$$

We may approximate the integral in equation (5) by $\Phi_{\rm sat}(z_{\rm p})z_{\rm p}$, where $z_{\rm p}$ is the distance behind the shock front where $z\Phi_{\rm sat}$ peaks. Multiplying and dividing equation (5) by $\Delta v_5$—the maser line width in kilometers per second—and substituting equations (6) and (1), we obtain

$$\mathscr{L}_{\rm prod} \approx 10^{27} h\nu \Delta v_5\, q(T)\eta(\xi_p, T)\xi_p\ {\rm ergs\ cm^{-2}\ s^{-1}}\ , \tag{8}$$

where $\xi_p$ and $T$ are the maser emission measure and temperature at a distance, $z_{\rm p}$, behind the shock front. The pump efficiency, $\eta(\xi, T)$, was calculated by HEM98 for $120 \le T \le 600$ K: it falls monotonically with increasing $\xi$, and can be fit, for $\xi \lesssim 10$, by (HEM98)

$$\eta \equiv \frac{\delta n}{n} \simeq \frac{1}{20\xi^{1/2} + 0.084 n_{\rm H,9}^{1.3}} - \frac{1.05}{T}\ , \tag{9}$$

which is almost independent of temperature and density. At higher values of $\xi$, the pump efficiency drops much faster; the deviation from the above fit begins at a $\xi$ that increases with increasing temperature. Since the temperature in the $H_2$ reformation plateau is nearly constant (EHM89), the maximum of $q\eta\xi$ and that of $\eta\xi$ occurs nearly at the same $\xi_p$. Using the numerical results of HEM98, we find that the peak value of $\xi\eta(\xi)$ increases monotonically from 0.1 at 200 K (where $\xi_p \simeq 20.5$ and $\eta[\xi_p] \simeq 5.0 \times 10^{-3}$) to 0.17 at 600 K (where $\xi_p \simeq 46$ and $\eta[\xi_p] \simeq 3.7 \times 10^{-3}$). Combining these results with equations (8) and (7), we obtain

$$\mathscr{L}_{\rm prod} \approx (0.04|_{200\ {\rm K}} - 0.21|_{600\ {\rm K}})\Delta v_5\ {\rm ergs\ cm^{-2}\ s^{-1}}\ , \tag{10}$$

where the numerical coefficients (0.04 at 200 K and 0.21 at 600 K) correspond to the temperature range of 200–600 K explored by HEM98. Converting the intrinsic luminosity to isotropic luminosity as in § 4.2.1, we obtain

$$L_{\rm iso} \simeq (1.9|_{200\ {\rm K}} - 9.8|_{600\ {\rm K}}) r_{18}\, h_{16}\, \Delta v_5 \left(\frac{\theta_b}{10^\circ}\right)^{-1} L_\odot\ , \tag{11}$$

where the temperature in the $H_2$ reformation plateau, which determines the numerical coefficient in equation (11), was estimated by EHM89 to scale with preshock gas density and shock speed roughly as

$$T_{\rm plateau} \simeq 660 n_9^{2/9}({\rm H_2}) \left(\frac{v_s}{20\ {\rm km\ s^{-1}}}\right)^{2/9} \left(\frac{\Delta v_5}{2}\right)^{-2/9} {\rm K}\ . \tag{12}$$

The above derivation assumes a negligible large-scale velocity gradient over the masing region, which is a reasonable approximation in the case of J-type shocks where, unlike in C-type shocks, the velocity change occurs in the shock front.

### 4.3. *Structure of the High-Velocity Masers*

The high-velocity maser emission originates in box-shaped regions of thickness $d$, height $h$, and length $l_g \simeq (8\mathscr{R}d)^{1/2}$, where $\mathscr{R}$ is the local radius of curvature of a spiral (Fig. 1*b*). Any monotonically increasing function, $\varphi(r)$, describes a spiral with a local pitch angle given by $\cot\theta_p = r\,|d\varphi/dr|$, but since the global form of $\varphi(r)$ will be clearly irrelevant to our discussion, we shall consider for simplicity only logarithmic spirals, $\varphi = \cot\theta_p \ln r$, where the pitch angle is constant and the local radius of curvature is $\mathscr{R} = r/\cos\theta_p$. Thus, the geometric length of the maser, $l_{\rm g}$, defined as the path length through the masing slab, is given by

$$\frac{l_g}{r} \simeq 10^{-2}\left(\frac{d_{13}}{r_{18}}\right)^{1/2} (\cos\theta_p)^{-1/2}\ , \tag{13}$$

where $(\cos\theta_p)^{1/2}$ is a factor of nearly unity, unless the pitch angle is implausibly large. The effective length of the maser, which is also constrained by the condition of velocity coherence, could be smaller than $l_{\rm g}$, depending on the velocity field in the disk. Thus, we now address the issue of velocity coherence along the line of sight through a high-velocity maser.

Let us denote the location of the maser in the disk plane by $(R, y_{\rm m})$, where $y_{\rm m}$ is the distance of the maser center from the midline (see Fig. 1*b*), and $R$ is the projected distance of the line of sight from the center of rotation. Assuming a nearly circular, Keplerian velocity field at the maser vicinity, the line-of-sight velocity along the line of sight through the maser, in the rest frame of the disk, is $v_{\rm los} = v_{\rm c}(R)[1 + (y^2/R^2)]^{-3/4}$, where $v_{\rm c}(R)$ is the circular velocity at radius $R$. Velocity coherence is maintained along the entire geometric length of the maser if the peculiar line-of-sight velocity along the maser does not exceed the maser full line width, $\Delta v$. Since the gradient of the line-of-sight velocity increases with increasing distance from the midline, a sufficient condition for velocity coherence is $\delta v_{\rm pec} \lesssim \Delta v/2$, where

$\delta v_{\rm pec} \equiv |\, v_{\rm los}(R, y_m) - v_{\rm los}(R, y_m - l_g/2)\,|$, and $(R, y_{\rm m} - l_{\rm g}/2)$ is the location of the maser end furthest away from the midline. Assuming a logarithmic spiral, which means that $y_{\rm m} = R \tan \theta_p$, expanding $\delta v_{\rm pec}$ up to second order in $\delta y$ (the first order vanishes at the midline) and using equation (13), the condition for maintaining close velocity coherence along the entire geometric length of the maser reads

$$v_{\rm los,8}\left[\frac{\sin\theta_p}{\sin 4^\circ}\sqrt{\frac{d_{13}}{R_{18}}} + 0.03\left(\frac{d_{13}}{R_{18}}\right)\right] \lesssim \Delta v_5\ , \qquad (14)$$

where $R_{18} \equiv R/10^{18}$ cm, and $v_{\rm los,8} \equiv v_{\rm los}/1000$ km s$^{-1}$. If the above condition does not hold, the effective length of the maser is smaller than $l_{\rm g}$. Note that the velocity gradient in the postshock gas associated with the shock itself would not introduce much of a gradient in the line-of-sight velocity, since the two directions are perpendicular to each other.

However, the maser radiation is unlikely to be emitted in a single beam of vertical half beaming angle, $\theta_b \simeq \tan^{-1}(h/l_g)$, for several reasons: the thin masing slab ($d \ll h, l_{\rm g}$) is likely to break up, owing to microturbulence, into several filaments, each giving rise to a single maser spot with typical diameter comparable to the slab thickness; the shock front may be corrugated (in two dimensions), which would fragment the filamentary masers into shorter pieces, depending on the typical amplitude and wavelength of the corrugation, and would distort the shape of the "masing box"; and the vertically stratified structure of the disk may also affect the beaming pattern, especially when the disk thickness is much larger then its scale height. Hence, we leave $\theta_b$ as a free parameter, with the only obvious constraint that $\theta_b$ must exceed $90^\circ - \theta_{\rm inc}$ for the maser radiation to be detected. The horizontal extent of each high-velocity maser clump will be at least as large as the uncertainty in the maser positions, which is typically $\simeq 0.1$ mas and possibly broader than that if the slabs are heavily corrugated and thereby effectively thickened (see, e.g., Wardle 1990).

We shall continue to adhere to the working hypothesis that the geometry of the emitting region behind the shock fronts is of spiral slabs, as wide as the disk thickness and perpendicular to the disk plane, while bearing in mind that it is only a zeroth-order approximation for a possibly much more detailed and complex structure.

## 5. THE DOPPLER SHIFT ASYMMETRY

The maser radiation travels through some part of the largely nonmasing disk along its way to the observer. Since the interaction of maser photons with water molecules is a line process, the noninverted gas is transparent to the maser radiation, except in regions in the disk that maintain close velocity coherence with the maser source, where maser photons are absorbed.

### 5.1. *The Absorbing Region*

In a circular disk, $v_{\rm los}$ along any line of sight is symmetric about the midline: it peaks at $(R, 0)$ and drops monotonically with increasing $y$. This means that in a largely noninverted disk, maser radiation originating on one side of the midline at $(R, y_{\rm m})$, which goes through the disk on the other side of the midline, will be absorbed at $(R, -y_{\rm m})$. The location of this absorbing region is shown in Figure 1*a* as the shaded area above the midline. Since the spiral-shaped regions of masing gas have a *trailing* geometry due to the action of differential rotation, the high-velocity masers are always located in front of the midline in the receding side of the disk and behind the midline in the approaching side of the disk (see Fig. 1*a*). The Doppler shift asymmetry arises because the blueshifted maser photons are subject to absorption as they pass through the velocity-coherent region on the other side of the midline, while redshifted photons never cross such a region, since $v_{\rm los}$ in the disk declines monotonically along their path to the observer.

The redshifted and blueshifted beams may also be slightly attenuated just outside the maser, where noninverted gas may maintain close enough velocity coherence with the maser source. Such regions are shown in Figure 1*a* as the short, shaded areas on both sides of the maser. First, we shall discuss the excess absorption of the blueshifted maser beam, which is suggested to account for the Doppler shift asymmetry in the high-velocity maser spectrum, and then the possible, relatively weaker absorption of redshifted radiation just outside the maser.

The first step is to derive the path length that maser photons travel through an absorbing region. Let us denote the height above the midplane of the disk by $z$, where $-h_{\rm abs}/2 \le z \le h_{\rm abs}/2$, $h_{\rm abs}$ is the disk thickness of the absorbing gas, and $z = h_{\rm abs}/2$ is the height of the "upper" surface of a disk, which is directly observable when $\theta_{\rm inc} \neq 90^\circ$. Note that $h_{\rm abs}$ may, in principle, be larger than $h$—the disk thickness of the masing gas behind shock fronts—owing to, for example, the vertically stratified structure of the disk, which may lead to more favorable conditions for maser action closer to the disk midplane. If a disk is viewed exactly edge-on, all *detectable* rays from a high-velocity maser will travel the same path length through the disk. However, if the disk is not viewed edge-on, the rays connecting the maser and the observer are inclined to the disk plane, so photons leaving the maser at different heights in the disk would travel different path lengths through the disk: those originating with $z = h_{\rm abs}/2$ would not pass at all through noninverted gas, while those leaving the maser at $z = -h_{\rm abs}/2$ would travel the largest path length through the disk.

Denoting the difference in the total path length of blueshifted and redshifted maser photons, emanating from the same maser at the same $z$, through velocity-coherent absorbing regions by $l_{\rm ex}$, and the location of the maser center in the disk plane by $(R, y_m)$ as in § 4.3, the excess absorption of the blueshifted radiation occurs within a region of length

$$l_{\rm ex}(z) \equiv \int_{-\infty}^{y_m - (l_g/2)} \Gamma\, dy - \int_{y_m + (l_g/2)}^{+\infty} \Gamma\, dy \quad -\frac{h}{2} \le z \le \frac{h}{2}\ , \qquad (15)$$

where

$$\Gamma(z) \equiv H\left[\frac{\Delta v}{2} - |\, v_{\rm los}(R, y_m) - v_{\rm los}(R, y)\,|\right] H \times \left[\frac{(h_{\rm abs}/2) - z}{|\, y - y_m\,| - l_g/2} - \tan(90^\circ - \theta_{\rm inc})\right], \qquad (16)$$

and $H(x)$ is the Heaviside step function, which equals unity when $x > 0$ and vanishes otherwise. The first term in equation (16) reflects the condition of velocity coherence, where $\Delta v$ is the maser line width. The second term ensures that the integration is carried out only along the path where

the photons travel through the disk. As will be demonstrated in § 6.1.3, it is possible that blueshifted rays may pass only through part of the absorbing region on the other side of the midline, or even miss it altogether.

Redshifted maser radiation may also be absorbed just outside the maser, where noninverted gas may maintain close enough velocity coherence with the maser source. This region is shown in Figure 1*a* as the short, shaded area below the maser. The length of the absorbing column is given by the first term in equation (15). An estimate for the length of the absorbing column on the redshifted side of the maser is given by $l_{\rm abs,red} = \max\,[l_{v,\rm red} - (l_g/2), 0]$, where $l_{v,\rm red}$ is determined by the condition of velocity coherence $|v_{\rm los}(R, y_{\rm m}) - v_{\rm los}(R, y_{\rm m} - l_{v,\rm red})| = \Delta v/2$, and $l_g$ is given by equation (13). Clearly, $l_{\rm abs,red}$ vanishes when the noninverted gas just outside the maser does not maintain close enough velocity coherence with the maser source. Approximating $l_{v,\rm red}$ by $(\Delta v/2)(dv_{\rm los}/dy)^{-1}$, assuming Keplerian rotation as in § 4.3, and substituting $y = R\,\tan\,\theta_p$ assuming a logarithmic spiral, we obtain

$$l_{\rm abs,red} \simeq \min\left[\frac{2\Delta v R}{3v_{\rm los}\,\sin\,(2\theta_p)} - \frac{l_g}{2}\,,\,0\right]. \qquad (17)$$

This length scale would typically be much smaller than the path length for absorption of the blueshifted beam (e.g., see § 6.1.3). We also note that the effective absorption coefficient in this region is smaller than that along the path of the blueshifted beam, since the absorbing region on the other side of the midline is in velocity coherence with the center of the maser line, by definition, while the absorbing region just outside the maser is in coherence with the wing of the maser line profile.

The velocity-coherent region on the other side of the midline may include a small masing section of another spiral slab, as illustrated in Figure 1*a*. We shall ignore this situation since, if it occurs, the length of that masing section will typically be only a few times larger than the thickness of the very thin slabs, so it would not reduce much the length of the absorbing region. It may only marginally increase the effective length of the maser, causing a negligible displacement in the location of the maser core.

### 5.2. *Optical Depth For Absorption*

The absorption in the noninverted gas depends on the level populations of the maser transition. Trapping of the IR photons, which quenches the inversion, would bring the level populations close to their LTE values. However, if the intensity of maser radiation passing through an absorbing region is very high, it may perturb the level populations, as the radiation "tries to invert" the levels. This effect is analogous to the process of maser saturation in which stimulated emission starts to diminish the population inversion and so reduces the magnitude of the (negative) absorption coefficient. Thus, we shall consider the cases of saturated and unsaturated absorption of maser radiation separately.

#### 5.2.1. *Unsaturated Absorption*

Following Hollenbach & McKee (1979), let $n_l$ and $n_u$ be the populations in the lower and upper levels of the transition of interest, $g_l$ and $g_u$ the respective statistical weights, $A_{ul}$ the Einstein coefficient, $\lambda_{\mu\rm m} \equiv \lambda/1\ \mu$m the wavelength, $\sigma_5$ the line-of-sight velocity dispersion in units of $10^5$ cm s$^{-1}$, and $N$ the column density of absorbing molecules. The optical depth at line center is

$$\tau_{ul} = 2.2 \times 10^{-19} N\left(\frac{n_l}{n}\right)\left(\frac{g_u}{g_l}\,\frac{A_{ul}\,\lambda_{\mu\rm m}^3}{\sigma_5}\right)\left(1 - \frac{n_u\,g_l}{n_l\,g_u}\right), \qquad (18)$$

where $(n_l/n)$ is the fraction of molecules in the lower state, and the third term is the correction for stimulated emission. In the case of the $6_{16} \rightarrow 5_{23}$ rotational transition in water vapor at 22.23508 GHz, $\lambda \simeq 1.35$ cm, $A = 1.9 \times 10^{-9}$ s$^{-1}$, and the rightmost term in equation (18) equals $1 - e^{-(\Delta E/KT)} \simeq 1.07/T$ for $T \gg 1$ K. Given the energy levels of $H_2O$, kindly provided by D. Neufeld, we calculated the partition function of water vapor in LTE. It is well fit by $Z \simeq 8.79T_2^{3/2}$ over the range 50 K $\lesssim T \lesssim$ 500 K to an accuracy of $\lesssim$2%, where $T_2 \equiv T/100$ K. The fraction of molecules in a given state to all ortho-water molecules is $(2J+1)e^{-(E/KT)}/Z$. Taking an ortho-to-para ratio of 3, we obtain $n_{5_{23}}/n_{\rm H_2O} \simeq 0.94T_2^{-3/2}\,\exp\,(-6.42/T_2)$, where $n_{\rm H_2O}$ is the density of all water molecules. The line-of-sight velocity dispersion due to thermal broadening alone is $\sigma_5 \simeq 0.13(T/A_{\rm H_2O})^{1/2}$, where $A_{\rm H_2O}$ is the atomic weight of water, so in order to take the effect of turbulence into account, we multiply it by the factor max (1, $\mathcal{M}$), where $\mathcal{M}$ is the Mach number of the mean turbulent velocity. Denoting the column density of water molecules by $N_{20}({\rm H_2O}) \equiv N({\rm H_2O})/10^{20}$ cm$^{-2}$, equation (18) yields

$$\tau(5_{23} - 6_{16}) \simeq 4.5N_{20}({\rm H_2O})\left[\frac{f(T)}{f(100\ {\rm K})}\right]\min\,[1, \mathcal{M}^{-1}]\,, \qquad (19)$$

where $f(T) \equiv T^{-3}\,\exp\,(-642/T)$. The function $f(T)$ increases by a factor of $\simeq$3 from 100 to 200 K and then slowly drops with further increase in temperature. For $T < 100$ K, it approaches an exponential decline with decreasing temperature. We discuss the possible value of $\mathcal{M}$ for the case of NGC 4258 in § 6.1.3.

In a stationary medium, the line opacity is generally calculated by averaging the optical depth, which depends on wavelength, over the line profile. However, in the case of absorption on the other side of the midline, the optical depth is essentially constant across the line profile, being equal to its peak value at line center. The reason for that is the smooth change of $v_{\rm los}$ along the line of sight, which guarantees that line center and off–line center photons will be in coherence with absorbing columns of gas of essentially equal length—columns that largely overlap but are slightly displaced relative to each other along the line of sight. Thus, the optical depth at line center (eq. [19]) provides an adequate approximation for the line opacity. We note that the above argument also means that mild, large-scale deviations from the assumed velocity field in the disk would only displace the location of the absorbing column of gas on the other side of the midline but would not affect the optical depth much.

In our case of maser emission from spiral shocks, we are interested in the excess absorption of the blueshifted beams, which depends on the excess path length of the blueshifted radiation through the absorbing region, $l_{\rm ex}$, given by equation (15). However, as discussed in § 5.1, photons leaving the maser at different heights in the disk travel different path lengths through the disk, such that $l_{\rm ex}(z)$ increases with decreasing $z$, where $-h/2 \le z \le h/2$. Instead of integrating the amount of absorbed flux over $z$, which would require detailed modeling of the stratified structure of the disk, we shall adopt $\bar{l}_{\rm ex} \equiv l(z = h/4)$ as an estimate for

the effective path length for excess absorption of the blueshifted beams. This is not a precise estimate, but it would allow us to demonstrate that substantial absorption may still occur, even for maser photons originating halfway up in the atmosphere of the disk (§ 6.1.3). With this estimate, the optical depth for excess absorption of the blueshifted radiation, $\tau_{ex}$, is given by equation (19) after substituting $N(H_2O) \simeq n^2 x(H_2O)\bar{l}_{ex}$.

#### 5.2.2. *Saturated Absorption*

The rate of absorption per unit volume is limited by saturation effects to $-\Phi_{sat}$, where $\Phi_{sat}$ is given by Neufeld & Melnick (1991) in their eq. [16] and is negative when the level populations are not inverted. Neufeld & Melnick (1991) demonstrated that the level populations of the water maser transition in the LVG limit (§ 2) are a function only of $\xi'$ and $T$ if the relevant nonmasing transitions are optically thick. They defined a maser rate coefficient $Q \equiv 10^{-14} Q_{-14}$ cm$^3$ s$^{-1}$ such that the volume emissivity is given by

$$\Phi_{sat} = Q(\xi', T) n(H_2O) n = Q_{-14} n_{H,9}^2 x_{-4}(H_2O) \text{ photons s}^{-1} \text{ cm}^{-3} . \quad (20)$$

For example, $Q_{-14}(\xi', 100\text{ K})$ for the water maser transition becomes negative at $\xi' \sim 5$ and asymptotically approaches a value of $-0.0185$ with increasing $\xi'$, and $Q_{-14}(\xi', 400\text{ K})$ becomes negative at $\xi' \sim 24$ and asymptotically approaches a value of $-0.324$ (D. Neufeld 1997, private communication). These values are smaller in magnitude than the corresponding values in masing gas, yet we shall show that absorption is substantial owing to the typically larger relative volume of absorbing gas.

Equation (20) gives the maximum rate of absorption of water maser photons per unit volume as a function of gas temperature, density, water abundance, and the amount of IR photon trapping (through the variable $\xi'$). Thus, in the case of saturated absorption, the maximum fraction of photons that can be absorbed is

$$f_{sat} = \frac{-\int \Phi_{sat}\, dl}{\mathscr{F}} , \quad (21)$$

where $\mathscr{F}$ is the incident photon flux on the absorbing region, and the integration is carried out over the velocity-coherent path length of the beam through noninverted gas. However, notice that although the integral in equation (21) can be arbitrarily large, $f_{sat}$ obviously cannot exceed unity since there is an upper limit on the true fraction of absorbed flux, $f_{max}$, given by

$$f_{max} = 1 - e^{-\tau} . \quad (22)$$

What happens is that when saturation effects are important, the flux decreases linearly until (if the absorbing slab is large enough) the flux falls to the level where saturation effects become unimportant, and from there, the flux decreases exponentially. It is exactly equivalent to the case of amplification in saturated masers, except in reverse. Let us consider the case in which the maser intensity remains high enough to significantly perturb the level populations along the entire path length of the beam. The fraction of absorbed photons, $f_{sat}$, is the lowest possible, since the intensity never falls exponentially within the absorbing region. In order to estimate $f_{sat}$, it is useful to notice that the excess absorption of blueshifted maser beams occurs within a spiral-shaped absorbing region located parallel to, but just outside of, the convex side of the masing spiral slab. The thickness of that absorbing spiral region—defined by velocity coherence—is about $\bar{l}_{ex} \sin \theta_p$, assuming the pitch angle is not very large. Since we are interested in the effect of absorption on the *isotropic* luminosity of a blueshifted, high-velocity feature, we may imagine that the absorption of the isotropic luminosity occurs within a spherical shell of thickness $\bar{l}_{ex} \sin \theta_p$ and a radius slightly larger than the disk radius at the maser location, $r$. Denoting the isotropic luminosity at the absence of absorption by $L_{iso}$, which should be comparable to that of redshifted, high-velocity features, we obtain

$$f_{sat} = \frac{(4\pi r^2 \bar{l}_{ex} \sin \theta_p)(-\Phi_{sat} h\nu)}{L_{iso}} , \quad (23)$$

where $\Phi_{sat}$ is given by equation (20).

Note that the above expression may exceed $f_{max}$ or even unity, which means that a transition from saturated to unsaturated absorption occurs somewhere along the path of the beam through the absorbing region. The higher the $f_{sat}$ above unity, the earlier along the path of the beam the transition occurs and the larger is the path length of exponential attenuation. The fraction of absorbed photons in such intermediate cases is clearly high, but deriving the exact value requires a detailed modeling of the effect of the maser radiation on the level populations as a function of maser intensity near the transition point from saturated to unsaturated absorption, which is beyond the scope of this paper.

## 6. THE MASING DISK IN NGC 4258

### 6.1. *High-Velocity Features*

#### 6.1.1. *Spatial Distribution and the Keplerian Rotation*

The masing circumnuclear disk in NGC 4258 is remarkably well defined: it is delineated over the range of radii 0.16–0.25 pc by high-velocity water maser sources with line-of-sight velocities (770–1080 km s$^{-1}$) that trace a Keplerian rotation curve around a black hole of $3.6 \times 10^7\ M_\odot$ (Miyoshi et al. 1995; Nakai et al. 1993). The disk is very thin ($h \lesssim 0.003$ pc), slightly warped (Herrnstein et al. 1996), and viewed nearly edge-on ($\theta_{inc} \simeq 83°$).

The high-velocity maser spots are clustered into several distinct clumps that are almost equally spaced apart from each other (Miyoshi et al. 1995). Nakai et al. (1995) reported on two additional high-velocity features in the maser spectrum that were too weak to appear in the VLBI image. Assuming that these features also fit in the Keplerian relation, their inferred locations were found to obey the regularity in the interclump spacing ($\simeq 0.75$ mas on both sides of the disk), making it even more pronounced. The possibility that the apparent clustering of the maser sources is simply due to randomly distributed gas clumps (or small star-forming regions) in the disk is implausible, not only because of the suspicious regularity in their spatial distribution, but also because they all fit a single Keplerian rotation curve, meaning that they *all* arise near a single diameter through the disk to within a few degrees, given the accuracy of the Keplerian fit.

Maoz (1995b) proposed that the clumps of maser spots are located at the intersections of spiral waves with the line that supports the longest coherent gain paths for amplification in our direction (the midline), in a largely masing disk,

and are thus spaced apart at the characteristic crest-to-crest radial distance between the waves. In the present investigation, we suggest that the disk in NGC 4258 is largely not masing and that it contains spiral shocks rather than just spiral density waves. In this scenario, high-velocity maser radiation arises where the line of sight is tangent to a spiral shock, which naturally accounts for the high-velocity maser spots being clumped rather than uniformly or randomly distributed.

We emphasize that the global pattern of spiral shocks in the disk of NGC 4258 cannot be uniquely determined from observations at present (but see the prediction in § 6.3), since it depends on the number of spiral waves, the pitch angle, the total azimuthal angle subtended by each spiral, and on whether the spirals are continuous or have gaps. It is interesting to notice, though, that assuming the simplest spiral pattern of $m$ equally spaced, winding spirals with a small, constant pitch angle, $r(\varphi) = \exp(\varphi \tan \theta_p)$, the radial crest-to-crest distance between the shocks is $\Delta r \simeq 0.1 r(\theta_{p,\deg}/m)$, where $\theta_{p,\deg} \equiv \theta_p/1^\circ$. This is consistent with the observed mean interclump separation of $\Delta r/r \simeq 0.1$ if $\theta_{p,\deg} \simeq m$. Yet, one should bear in mind that these are not the only possible spiral patterns that can reproduce the spatial distribution of the maser clumps and that in any case, the Doppler shift asymmetry is independent of the global spiral pattern.

The most remarkable finding about the disk in NGC 4258 is its nearly Keplerian rotation curve, which indicates that the high-velocity masers arise along a diameter through the disk. If the disk was largely masing, that diameter would be the midline—the major axis of the projected disk (see, e.g., Elmegreen & Morris 1979; Ponomarev, Smith, & Strelnitski 1994). In the shock-origin scenario proposed in the present investigation, the high-velocity masers are located at a distance $|y_m| \simeq r \sin \theta_p$ from the midline, where $r$ is the disk radius at the position of the maser center. If the local pitch angle does not vary much throughout the spiral pattern, all the high-velocity masers will appear along a diameter through the disk that makes an angle $\theta_p$ with the midline and thus traces a (slightly different) Keplerian rotation curve. This means that the true disk radius, circular velocity, and binding mass are higher by a factor $(\cos \theta_p)^{-1}$ than the values derived directly from the projected radius and line-of-sight velocities. This introduces a negligible correction in the case of NGC 4258, where the typical pitch angle is probably about 2° (see § 6.1.2). In any case, the predicted off-midline locations of the masers are consistent with the observational constraint that the high-velocity sources must lie within 6° of the midline, based on the derived upper limit for their centripetal acceleration (Greenhill et al. 1995). The shocks themselves would not introduce substantial deviations from a Keplerian rotation curve because the velocity jump occurs along the normal to the shock fronts, which are perpendicular to the line of sight at the positions of the masers.

As pointed out in § 2, even in a largely noninverted disk, the level populations may be inverted within two thin slabs in the upper atmosphere of the disk, where the IR photons can escape in a short distance vertically. The thickness of these slabs would be similar to that of the postshock masing regions ($\sim 10^{13}$–$10^{14}/n_{\rm H,9}$ cm), which is much smaller than the disk thickness. Although amplification along rays within these slabs may produce strong masers, these masers are unlikely to be detected because their length, defined by velocity coherence, is so much larger than the thickness of these slabs that the vertical beaming angle would be very small. Such masers could be detected only in regions in the disk where the locally tangent plane is viewed edge-on. It is improbable that the observed high-velocity maser emission in NGC 4258 arises in these slabs, since it would require wiggles or wrinkles in the disk that happen to bring the extremely narrow beams into sight only along a single diameter through the disk (as indicated by the Keplerian rotation), on both sides of the disk, and only along that line.

#### 6.1.2. *Shock Speed and Luminosity of the Redshifted Features*

The shock speed must be higher than the sound speed in the gas, $c_s$. Given the condition of hydrostatic equilibrium, $H/r \simeq c_s/v_c$, where $H \lesssim h/2$ is the disk scale height, and the observed upper limit to the thickness of the masing disk, $h \lesssim 10^{16}$ cm (Miyoshi et al. 1995), the upper limit to the sound speed at the mean radius of the disk ($r = 0.2$ pc, $v_c = 900$ km s$^{-1}$) is $c_s \lesssim 15$ km s$^{-1}$. It should be emphasized that magnetic pressure may play an important role in the vertical support of the disk, in which case the sound speed could be much lower than 15 km s$^{-1}$. This may well be the situation in the disk of NGC 4258, where the typical maser line width is only a few kilometers per second. Thus, a reasonable lower limit to the required shock speed in the disk is $\approx 10$ km s$^{-1}$.

The shocks might be either J- or C-type, depending upon such poorly known parameters as the preshock magnetic field and the fractional ionization. In the case of J-type shocks, significant maser luminosity is produced only if a large fraction of the molecules are dissociated (§ 4.2.2), which requires $v_s \gtrsim 20$ km s$^{-1}$. Thus, we shall assume that the shock speed in the masing disk of NGC 4258 is at least 20 km s$^{-1}$, in which case the conditions for maser action are favorable behind shock fronts in either type of shock (§ 4).

The shock speed and pitch angle of a spiral are locally related through the pattern speed and rotational velocity in the disk (eq. [3]). Assuming the angular speed of the spiral waves is of order one-half the disk rotation speed, as expected in cases where the spirals originate in disturbances at the outer edge of the disk (§ 3), we obtain

$$\theta_p \approx 2.5^\circ \left(\frac{v_s}{20 \text{ km s}^{-1}}\right)\left(\frac{v_c}{900 \text{ km s}^{-1}}\right)^{-1} . \quad (24)$$

Thus, given the uncertainties in the angular speed of the spiral waves and in the shock speed, we expect in NGC 4258 a typical pitch angle of order 2°, assuming a shock speed of about 20 km s$^{-1}$. We note that for such a pitch angle and rotational velocity, the inequality expressed in equation [14] is satisfied, which means that close velocity coherence is maintained along the entire path length of the line of sight through the masing slabs.

Substituting $h \lesssim 10^{16}$ cm and $\theta_b \gtrsim 90 - \theta_{\rm inc} = 7^\circ$ (Miyoshi et al. 1995) in equations (4) and (11), the expected isotropic luminosity of a redshifted, high-velocity feature at the mean radius of the disk is, in the case of C-type shocks,

$$L_{\rm iso} \lesssim 2.3 n_9({\rm H}_2)\left(\frac{\epsilon_{\rm sat}}{10^{-5.5}}\right)\left(\frac{v_s}{20 \text{ km s}^{-1}}\right)^3 L_\odot , \quad (25)$$

and in the case of slow, dissociative, J-type shocks,

$$L_{\rm iso} \lesssim 8.4|_{600\,\rm K} \Delta v_5 \, L_\odot , \quad (26)$$

where $\Delta v_5$ is typically between 1 and a few. The numerical coefficient in equation (26) is for a temperature of 600 K in

the $H_2$ reformation plateau (the approximate dependence of $T_{\rm plateau}$ on gas density and shock speed is given by eq. [12]). The observed, mean isotropic luminosity of the redshifted, high-velocity features in NGC 4258 is about 4 $L_\odot$ (Nakai et al. 1995), which can be easily reproduced in the case of dissociative, J-type shocks (eq. [26]) and is marginally consistent with C-type shocks, given the uncertainties in the magnetic field, fractional ionization, estimated energy conversion efficiency, and shock speed.

#### 6.1.3. *Absorption of the Blueshifted Radiation*

The asymmetry in the high-velocity maser spectrum of NGC 4258 is striking: the blueshifted features are systematically weaker than the redshifted ones by about an order of magnitude (Nakai et al. 1993; Miyoshi et al. 1995). We now check whether such substantial absorption is consistent with our model.

First, we calculate numerically the velocity-coherent path length for excess absorption of the blueshifted maser radiation, $l_{\rm ex}$, at the mean radius of the masing disk (eq. [15]). The result is presented in Figure 2 as a function of pitch angle for three different choices of $z$—the height in the disk atmosphere where the photons emanate from the maser. This calculation assumes a disk thickness of $h_{\rm abs} = 0.003$ pc for the absorbing gas, $d = 10^{13}$ cm and consequently $l_{\rm g} \simeq 8 \times 10^{15}$ cm (eq. [13]), and maser line width of 2 km s$^{-1}$ ($\Delta v_5$ is typically between 1 and a few). The three curves in Figure 2 show that maser photons emerging from the maser closer to the upper surface of the disk pass through a smaller column of absorbing gas. The solid line, which corresponds to maser photons emanating from the maser at the midplane of the disk, gives the mean path length for excess absorption of blueshifted photons.

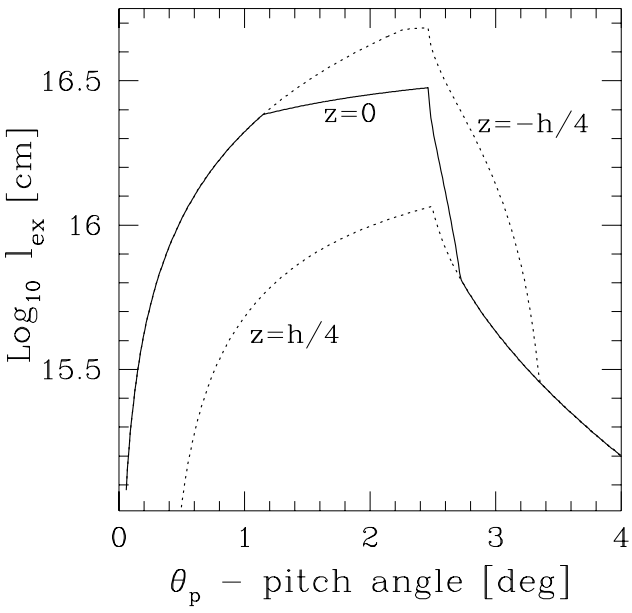


FIG. 2.—Velocity-coherent path length for excess absorption of blueshifted, high-velocity radiation, $l_{\rm ex}(z)$, at the mean radius of the disk in NGC 4258, as a function of pitch angle (eq. [15]). It assumes a disk thickness of $10^{16}$ cm and maser line width of 2 km s$^{-1}$ (see text). The three curves correspond to maser radiation emanating from the maser at the midplane of the disk ($z = 0$) and at halfway up (and down) in the disk atmosphere ($z = \pm h/4$). The sharp decline at $\theta_p \gtrsim 3°$ occurs because the line of sight, which is inclined to the disk plane by 7°, does not pass through the absorbing region on the other side of the midline. The estimated pitch angle in NGC 4258 is of order 2° (§ 6.1.2).

The sharp peak in these curves corresponds to the situation in which the maser is so close to the midline that the velocity-coherent regions on both sides of the midline touch and connect into one. The steep falloff at larger pitch angles is because the maser is too far from the absorbing region on the other side of the midline, so the line of sight through the maser, which is inclined to the disk plane by $\simeq 7°$, does not pass through the absorbing region. The effect of larger maser line width or larger disk thickness would be to broaden the range of $\theta_p$ where $l_{\rm ex}(z)$ is substantial.

The amount of absorption strongly depends on the gas temperature, which is quite uncertain. The fact that the disk is slightly warped and hence obliquely irradiated by the central source provides a lower limit to the mean gas temperature in the disk. The X-ray luminosity of the nucleus in the 2–10 keV range is $L_x \simeq 4 \times 10^{40}$ ergs s$^{-1}$ (Makishima et al. 1994), and the warping angle of the disk is $\theta_w \lesssim 20°$ (Herrnstein et al. 1996). For direct attenuated radiation, about one-third of the incident radiation is absorbed (Cunningham 1976), so the black body temperature of the disk due to the X-ray irradiation is given by $\sigma T_{\rm irr}^4 \approx \frac{1}{3} L \sin \theta_w/(4\pi r^2)$, which yields $T_{\rm irr} = 60$ K $(\sin \theta_w/\sin 20°)^{1/4}$ at the mean radius of the disk. The mean gas temperature is likely to be somewhat higher than that because there are other possible heat sources, such as direct irradiation by Compton-heated corona and viscous energy dissipation in the disk. Moreover, if the absorbing column is not too distant from the shock front, which happens when the pitch angle is small, the gas in the absorbing region may not have completely cooled down to its preshock temperature (notice that in Fig. 1*a*, $\theta_p = 20°$ for a clearer illustration, while we expect $\theta_p \approx 2°$ in the case of NGC 4258).

Replacing $N(H_2O)$ in equation (19) by $n(H_2)x(H_2O)\bar{l}_{\rm ex}$ and substituting $\bar{l}_{\rm ex} = 10^{16}$cm (defined in § 5.2.1), the optical depth for excess absorption of the blueshifted radiation in the unsaturated case (§ 5.2.1) is

$$\tau_{\rm ex} \approx 32 n_9(H_2) x_{-4}(H_2O) \left[\frac{f(T)}{f(100\ {\rm K})}\right] \min\,[1, \mathscr{M}^{-1}]\ , \quad (27)$$

where the mean Mach number of the turbulent velocities, $\mathscr{M}$, is probably of the order of a few and is certainly limited by the condition of hydrostatic equilibrium to $\mathscr{M} < (v_c\, h/2r)/(0.04 T^{1/2}\ {\rm km\ s^{-1}}) \simeq 18(T/100\ {\rm K})^{-1/2}$. Again, notice that if magnetic pressure plays an important role in the vertical support of the disk, then $\mathscr{M}$ could be much lower than the above limit. We also note that the gas density in the absorbing region could be higher than the preshock (or mean) density in the disk since, when the pitch angle is small, the absorbing region is not very far behind the shock front (relative to the intershock spacing). Thus, we conclude from equation (27) that the optical depth for unsaturated absorption in the disk of NGC 4258 can be easily high enough to attenuate the intensity of the blueshifted beams by roughly an order of magnitude, as observed. We now examine whether substantial attenuation may occur also in the case of saturated absorption (§ 5.2.2).

The rate of absorption per unit volume, in the limit where saturation effects are important, depends on the amount of IR photon trapping through the variable $\xi$, defined in equation (1). Assuming preshock gas density of $n(H_2) \sim 10^9$ cm$^{-3}$, disk thickness of $h \sim 10^{16}$ cm, and expected water abundance in the galactic nucleus of 1 to a few times $10^{-4}$, we obtain $\xi \sim 10^3$ throughout most of the disk volume

(notice that $n_{\rm H,9} = 2n_9[{\rm H_2}]$). Assuming that the velocity gradient in the absorbing region is small, which is an adequate assumption in the case of J-type shocks, we can use the results derived in the LVG limit (D. Neufeld 1997, private communication), where $\xi' \simeq \xi/9 \sim 100$. Substituting $Q_{-14}(100\ {\rm K}) \simeq 0.02$ and $Q_{-14}(400\ {\rm K}) \simeq 0.22$ for $\xi' \sim 100$ (D. Neufeld 1997, private communication), equation (20) gives

$$\Phi_{\rm sat} \simeq (0.08|_{100\ {\rm K}} - 0.8|_{400\ {\rm K}}) n_9^2({\rm H_2O}) x_{-4}({\rm H_2O}) \quad \text{photons s}^{-1}\ \text{cm}^{-3}\ . \quad (28)$$

Substituting $\bar{l}_{\rm ex} = 10^{16}$ cm in equation (23), we obtain

$$f_{\rm sat} \simeq (1.1|_{100\ {\rm K}} - 11|_{400\ {\rm K}}) \left(\frac{\bar{l}_{\rm ex}}{10^{16}\ {\rm cm}}\right)\left(\frac{\theta_p}{2^\circ}\right)\left(\frac{L_{\rm iso}}{4\ L_\odot}\right)^{-1}, \quad (29)$$

which is above unity; as discussed in § 5.2.2, this indicates that the absorption has become unsaturated (note that we have assumed that the typical isotropic luminosity of the blueshifted, high-velocity features would have been comparable to that observed for the redshifted ones, $\simeq 4\ L_\odot$, in the absence of absorption.) Therefore, we conclude that our model is consistent with the intensity of the blueshifted, high-velocity maser beams in NGC 4258 being significantly lower than that of the redshifted ones.

We note that the path length for possible absorption of the redshifted beams in NGC 4258, given by equation (17) and evaluated at the mean radius of the masing disk, is $\lesssim 3 \times 10^{15}$ cm, which is about an order of magnitude smaller than the path length for absorption of the blueshifted beams. Combined with the fact that this relatively small absorbing column is in velocity coherence with the wing of the maser line rather than with the line-center (§ 5.1), we conclude that any possible attenuation of the redshifted, high-velocity beams in NGC 4258 is negligible.

### 6.2. *The Low-Velocity Features*

The strongest features in the maser spectrum of NGC 4258 arise near the line of sight to the center of rotation, and have low velocities with respect to the systemic velocity of the disk (Miyoshi et al. 1995). If the gas in the disk is largely not inverted, as we argue in the present investigation, what is the origin of the low-velocity maser emission? It is unlikely to arise in the masing spiral slabs because these are very thin ($\sim 10^{13}$–$10^{14}/n_{\rm H,9}$ cm) and are nearly perpendicular to the lines of sight under consideration. Before addressing that question, we point out two intriguing details in the spatial distribution and spectrum of the low-velocity maser emission.

Close velocity coherence is maintained along the entire length of the minor axis of the projected disk. Yet all the observed low-velocity maser sources are confined to a narrow annulus of radius 0.13 pc, inner to the region where high-velocity sources are detected (0.16–0.25 pc). The width of that annulus is $\delta r \lesssim 0.013$ pc, as indicated by the dispersion in the centripetal acceleration of the maser spots (Greenhill et al. 1995). It is difficult to understand the restricted spatial distribution of the low-velocity masers if the entire disk is masing. Miyoshi et al. (1995) suggested that it may result from the amplification of a continuum source of finite extent at the center of rotation, which appears—due to the disk inclination—only at the background of the inner part of the disk (see illustration in Fig. 20 of Nakai et al. 1995). Although the low-velocity masers may well amplify a central continuum source, the suggested explanation for their restricted distribution is somewhat improbable because it requires precise fine tuning: the effective size of the background continuum source must coincide with the projected distance of the inner edge of the disk to the center, to within $\simeq 0.013 \cot \theta_{\rm inc} = 0.0016$ pc. Also, if the entire disk is masing, it is unclear why no high-velocity features were found at 0.13 pc $\le r <$ 0.16 pc on either side of the disk.

In a largely masing disk, the low-velocity maser emission should generally peak at the systemic velocity of the disk and decline with increasing relative velocity. The reason for that is twofold: the gradient of the line-of-sight velocity vanishes along the minor axis of the projected disk, or in the case of an edge-on disk, along the line of sight to the center, and it increases with increasing impact parameter, $b$. Thus, the velocity-coherent gain path for amplification in our direction drops with increasing $b$, which should result in a diminishing $F_\nu$ with increasing $v_{\rm los}$. In addition, the surface brightness profile of a background continuum source is likely to be centrally concentrated, thus making the expected peak at $b = 0$ even more pronounced. However, in NGC 4258, instead of a peak there is a striking *dip* in the maser spectrum at the systemic velocity of the disk, which splits the low-velocity complex into two distinct components. The flux density at the systemic velocity is weaker even than that of the redshifted, high-velocity features (see, e.g., Greenhill et al. 1995). The dip is unlikely to simply be just a fluctuation because of its prominence, its location at the systemic velocity of the disk, and because it persists in spite of the variability of the entire low-velocity complex (Nakai et al. 1995). Watson & Wallin (1994) suggested that absorption in the disk may play an important role in splitting the low-velocity maser emission complex in NGC 4258 at the systemic velocity.

We suggest that the low-velocity maser emission in NGC 4258 arises in a ring of masing gas at the inner edge of the disk, where the direct X-ray irradiation by the central source may provide the energy that ultimately powers the maser (see, e.g., Neufeld, Maloney, & Conger 1994). We now argue that the existence of a narrow, masing annulus at the inner edge of the largely nonmasing disk can naturally explain the intriguing findings discussed above.

Low-velocity maser radiation from the inner edge of the disk is subject to absorption as it passes through the largely nonmasing disk. The length of the absorbing column of gas in the foreground of the low-velocity masers is maximal at the projected minor axis of the disk and drops rapidly with increasing impact parameter, $b$. The absorbing column extends from the outer boundary of the masing annulus, along the line of sight in our direction, out to the point where $v_{\rm los}$ in the disk deviates from that of the maser source by the maser line width ($\Delta v$). Denoting the radius and circular velocity of the masing annulus by $r'$ and $v_c'$, respectively, it is simple geometry to derive the length of the velocity-coherent absorbing column, $l'(b)$ or $l'(v_{\rm los})$. In the case of an edge-on disk with a Keplerian rotation curve, we obtain $l'(v_{\rm los}) \simeq (r'/4)(\Delta v/|v_{\rm los}|)$ for $\Delta v \lesssim |v_{\rm los}| \lesssim v_c'/2$ (notice that it is independent of the circular velocity at the maser). When $|v_{\rm los}| < \Delta v$, $l'(v_{\rm los} \to 0) \to (r_{\rm max} - r')$, where $r_{\rm max}$ is the maximum radius of the disk. However, as discussed in § 5.1, in a disk that is not viewed exactly edge-on, there is an

upper limit to the path length of any line of sight through the disk, which is determined by the disk thickness, $h_{abs}$, and inclination, $\theta_{inc}$. Thus, maser photons emanating in our direction from a low-velocity maser spot at a height $z$ in the disk atmosphere (see definition in § 5.1) pass through an absorbing column of length

$$l'(v_{los}) \simeq \min\left[\frac{r'\,\Delta v}{4\,|\,v_{los}\,|}, \left(\frac{h_{abs}}{2} - z\right)\tan\theta_{inc}, r_{max} - r'\right]. \quad (30)$$

Figure 3 presents $l'(v_{los})$ in the case of NGC 4258, where $r' = 0.13$ pc, assuming $\Delta v = 2$ km s$^{-1}$, $h_{abs} = 0.003$ pc, and $\theta_{inc} = 83°$. The absorption of low-velocity maser emission is maximal around the systemic velocity of the disk and drops rapidly with increasing relative velocity. We suggest that this may be the reason for the prominent dip in the low-velocity maser spectrum near the systemic velocity of the disk.

The reason for the concentration of the low-velocity sources within a narrow annulus at the inner edge of the disk is now also obvious. We have to explain, though, the absence of high-velocity emission at the intersections of the masing annulus and the midline. The reason for that is twofold: the velocity-coherent path length in the masing annulus for amplification in the tangential direction is $l_{tan} \simeq 0.004$ pc, derived from the condition $|v_{los}(r', 0) - v_{los}(r', l_{tan})| = \Delta v/2$ as in § 4.3, while the path length for amplification in the radial direction, which is the width of the annulus, is $l_{rad} = \delta r \lesssim 0.013$ pc, as estimated from the dispersion in the centripetal acceleration of the maser sources. Thus, $l_{rad}$ may well be larger than $l_{tan}$, in which case the velocity-coherent regions within the masing annulus are elongated in the *radial* directions. The radially propagating beams will then win in the competition for available pump photons over beams in the locally perpendicular direction (the direction tangent to the annulus). Moreover, the annulus of inverted gas is clearly illuminated by a central source of continuum radiation—presumably the bisymmetric nuclear jet that has been recently detected on a milliarcsecond scale in the radio continuum (Herrnstein et al. 1997) and on much larger scales in Hα (Ford et al. 1986; Cecil, Wilson, & Tully 1992), the radio continuum (van der Kruit, Oort, & Mathewson 1972), and in the X-ray (Cecil, Wilson, & De Pree 1995). Further evidence for amplification of a central source comes from the absence of low-velocity masers along the minor axis of the projected disk on the *rear* side of the disk, in contrast to the strong low-velocity emission observed along the same axis on the front side of the disk, although the two regions equally maintain velocity coherence. The central continuum source provides ample seed photons to be amplified in the *radial* directions, while beams in the locally perpendicular directions only amplify photons generated due to spontaneous decays in the masing region. For these two reasons, the maser radiation within the masing annulus at the inner edge of the disk may be strongly beamed in the radial directions, so it would not be detectable where lines of sight are tangent to the ring.

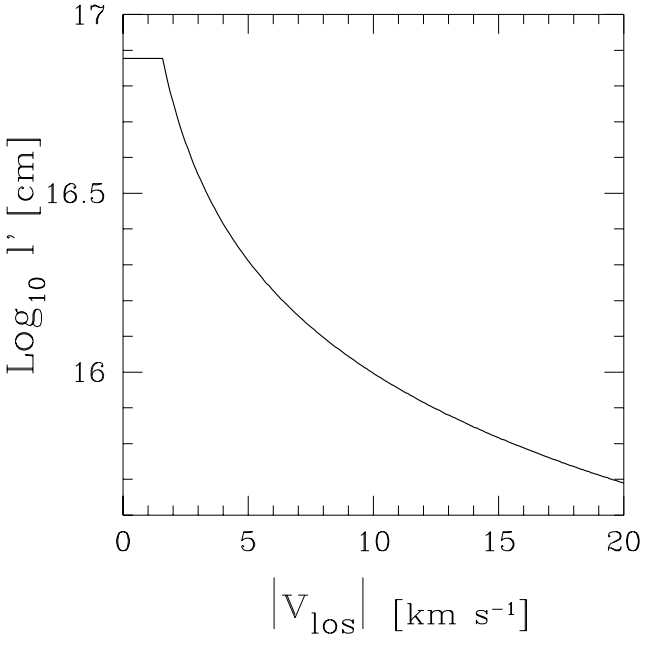


FIG. 3.—Length of the absorbing column of gas in the foreground of the low-velocity masers in NGC 4258, as a function of $v_{los}$, assuming maser line width of 2 km s$^{-1}$, disk thickness of 0.003 pc, and inclination $\theta_{inc} = 83°$ (eq. [30]). The absorption of low-velocity maser emission, which drops rapidly with increasing relative velocity, may explain the prominent dip in the maser spectrum at the systemic velocity of the disk (§ 6.2).

### 6.3. *Predictions*

The shock-origin scenario has two unique observational consequences. The high-velocity masers arise where the line of sight is tangent to the spirals, so the slow rotation of the waves must result in a continuous, outward drift in the positions of all the high-velocity maser clumps. Assuming a rotating logarithmic spiral, $r(\varphi, t) = \bar{r}\exp[\tan\theta_p(\varphi + \Omega t)]$, where $\Omega$ is the angular velocity of the waves at the mean radius of the disk, $\bar{r}$, the proper motion of the high-velocity masers is given by $\Omega\tan\theta_p\bar{r}/D$, where $D$ is the galaxy distance. Assuming that $\Omega$ is about one-half the angular velocity of the disk, we predict a systematic proper motion of $\approx 0.6(\theta_p/2.5°)$ μas yr$^{-1}$. The slow increase in the distance of the high-velocity clumps from the center of rotation must be accompanied by a slow decrease in their line-of-sight velocities at a rate $dv/dt = \frac{1}{2}v\Omega\tan\theta_p$, which yields $\simeq 0.05(\theta_p/2.5°)$ km s$^{-1}$ yr$^{-1}$. The outermost maser clump will eventually disappear, and a new high-velocity clump will appear at the inner edge of the disk, sometime within the next $\approx 1200(\theta_p/2.5°)$ yr, given the observed mean separation between the clumps ($\simeq 0.75$ mas). At present, the predicted drift rates in position and velocity are too small to be detected for individual maser spots. However, since the drifts are systematic, it may be possible to test their existence in a statistical analysis of the ensemble of high-velocity maser spots.

The second prediction is a systematic offset in the locations of the high-velocity maser sources in the disk from the midline. The redshifted features should arise in front of the midline, while the blueshifted features should arise behind the midline. The predicted displacement from the midline is comparable to the typical pitch angle of the spiral shocks and thus can be quite small. Indeed, in the case of NGC 4258, the currently most well-defined masing disk, the derived locations of the blueshifted high-velocity features are clearly slightly behind the midline, while those of the redshifted ones are marginally in front of the midline. This finding seems to be insensitive to the exact model of the warp (J. Herrnstein 1997, private communication).

Assuming a standard "alpha disk", where $\dot{M} \simeq 3\pi\nu\Sigma$ and $\nu = \alpha h^2 \omega$ is the kinematic viscosity, we obtain for our model $\dot{M} \simeq 7 \times 10^{-3}(\alpha/0.1)h_{16}^3 n_9(\mathrm{H}_2)\ M_\odot\ \mathrm{yr}^{-1}$. However, it is possible to estimate directly the mass accretion rate due to energy dissipation in the shocks: the energy dissipation rate per unit length of the shock front is $\dot{E}\,dl = \Sigma v_s^3\,dl/2$; the total length of the shock front within an annulus of width $dr$ is $m\,dr/\sin\theta_p$, assuming the presence of $m$ "grand-design" spirals with pitch angle $\theta_p$; the kinematic viscosity is $\nu = (2/3)rv_r$, where $v_r \approx r/t_r$ is the radial velocity of the gas; and $t_r \approx E/\dot{E}$ is the radial drift timescale. We obtain

$$\dot{M} \sim 7 \times 10^{-3} m \left(\frac{v_s}{20\ \mathrm{km\ s}^{-1}}\right)^2 n_9(\mathrm{H}_2) h_{16}\ M_\odot\ \mathrm{yr}^{-1}\ . \quad (31)$$

We emphasize that equation (31) gives only a rough estimate for the mass accretion rate due to energy dissipation in the shocks, since the spiral pattern cannot be uniquely determined (see § 6.1.1). Yet this result may suggest an advection-dominated accretion flow in NGC 4258, as proposed by Lasota et al. (1996) to explain the observed spectrum from the galaxy nucleus. They derived a mass accretion rate of $\dot{M} \sim 1.3 \times 10^{-3}(\alpha/0.1)\ M_\odot\ \mathrm{yr}^{-1}$, which is consistent with equation (31) if $\alpha$ in their model is of order 0.5. We note, though, that Lasota et al. derived the *current* mass accretion rate onto the black hole, while the above derivation gives $\dot{M}$ at a radius of about 0.2 pc ($\sim 10^5$ Schwarzschild radii), which need not necessarily be the same (it is possible that nucleus of NGC 4258 is relatively quiescent at the present time).

### 6.4. *Comparison with the Model of Neufeld and Maloney*

Neufeld & Maloney (1995; hereafter NM95) proposed that the location and extent of the masing region in NGC 4258 is determined by the X-ray irradiation of the warped disk by the central source. According to their model, the circumnuclear disk is completely atomic outside some critical radius, where the ionization parameter exceeds a critical value. Inside that radius, the disk becomes increasingly molecular, where warm molecular gas near the disk midplane is "sandwiched" by hot atomic gas in the disk atmosphere. The level populations of the water maser transition are inverted throughout the warm molecular gas, so the high-velocity masers arise along the midline of the disk that supports long coherent paths for amplification in our direction. Assuming that the observed outer radius of the masing disk (0.25 pc) represents the critical radius beyond which molecules may not exist, NM95 predict molecular hydrogen density of $3 \times 10^7\ \mathrm{cm}^{-3}$ in the masing disk, and total disk mass of $\sim 100\ M_\odot$. The inner edge of the masing region is suggested to represent the point at which the warp flattens out and the disk can no longer be heated by obliquely incident X-rays, so the temperature of the molecular gas at small radii is too low to produce significant water maser emission.

In the model presented in the present paper, the molecular density in the masing disk is $\sim 10^9\ \mathrm{cm}^{-3}$, which implies a disk mass of $\sim 10^4\ M_\odot$ (notice that the disk mass could be higher if it is dominated by stars or any other form of compact objects). Such disk mass is consistent with the model of Papaloizou et al. (1998) for the origin of the warp in the masing disk of NGC 4258. The gas is molecular throughout the region under consideration due to the larger shielding column density of hydrogen nuclei, except perhaps high in the disk atmosphere. Thus, in contrast to the NM95 model, the maser locations are insensitive to possible variations in the X-ray irradiating flux, which may be tested by observations if X-ray variability is detected. The level populations of the water maser transition are largely not inverted owing to IR photon trapping. High-velocity maser emission arises behind spiral shocks in the disk, where the conditions are conducive for maser action. The observed high-velocity maser features originate where the line of sight is tangent to spiral waves, where the path length through velocity-coherent, inverted gas is the largest. The high-velocity masers appear nearly along a diameter through the disk which, in the case of NGC 4258, makes an angle of about 2° with the midline (it does not introduce a noticeable deviation from Keplerian rotation).

The NM95 model accounts for the observed extent of the masing disk, while in the present model, high-velocity features are detectable in the region in the disk that maintain spiral activity, where the strength of spiral shocks, combined with the local gas density and water abundance, provide favorable conditions for maser action. In contrast to NM95, our model may naturally account for the clustering of the high-velocity maser spots into distinct clumps, the prominent dip in the maser spectrum at the systemic velocity of the disk, and the striking asymmetry in the high-velocity maser spectrum of NGC 4258 in particular, and that of other masing disks in general.

We should note that the Doppler shift asymmetry in the spectrum of NGC 4258 may not pose a problem for the NM95 model, as Herrnstein et al. (1996) pointed out that free-free absorption in the upper atmosphere of the disk may attenuate predominantly the blueshifted maser beams due to the presumed geometry of the warp (§ 1). This effect relies on a coincidentally special orientation of the warp with respect to the line of sight; it is unlikely to account for the Doppler shift asymmetry for circumnuclear disks as a class, and, in any case, it is not in conflict with the model presented in the present investigation.

Finally, the model of NM95 yields a very low mass accretion rate of $7 \times 10^{-5}\alpha\ M_\odot\ \mathrm{yr}^{-1}$, where $\alpha < 1$ is the dimensionless viscosity parameter introduced by Shakura & Sunyaev (1973). The present model predicts an accretion rate of order $0.007\ M_\odot\ \mathrm{yr}^{-1}$, which is consistent with the prediction of the advection-dominated accretion flow model for this system (§ 6.3).

## 7. OTHER CIRCUMNUCLEAR DISKS

We have proposed that high-velocity water maser emission in NGC 4258 arises behind shock fronts in a largely noninverted disk, and suggested that that model may also account for the Doppler shift asymmetry in other circumnuclear disks. We should emphasize that water maser emission may also arise in circumnuclear disks under different circumstances, such as within some range of radii in a disk where the gas is largely inverted (not enough IR photon trapping) or preferentially behind shock fronts in a largely inverted disk that maintains spiral activity.

The prevalence of each situation is unknown, as it depends on the combined probability distribution of many characteristics of circumnuclear disks and of their environment, as well as on biases in the selection of systems to be searched for high-velocity maser features. We argue that if the Doppler shift asymmetry continues to be common in an increasingly larger number of identified circumnuclear disks, then the scenario proposed in the present paper is the

most prevalent circumstance for the generation of detectable high-velocity maser emission in circumnuclear disks. We now discuss some implications of our model for circumnuclear disks in general and elaborate on the case of NGC 1068 in particular.

High-velocity maser emission need not necessarily arise in every circumnuclear disk that exhibits low-velocity emission, even if the disk is observed nearly edge-on. This may happen, among other reasons, simply because there are no spiral shocks in the disk. The low-velocity features may arise in a masing annulus at the inner edge of the disk that is directly irradiated by X-rays from a central source, as we suggest happens in NGC 4258 (§ 6.2). There is already one case—NGC 2639—where a disk origin of the low-velocity maser emission is indicated by its velocity drift (Wilson et al. 1995), possibly caused by centripetal acceleration as in NGC 4258, but a search for "satellite," high-velocity maser features did not discover any.

High-velocity maser emission that originates behind spiral shock fronts need not necessarily exhibit the Doppler shift asymmetry: the optical depth for absorption of the blueshifted beams may be too small for producing noticeable asymmetry in the maser spectrum; the blueshifted beams may not pass through the velocity-coherent regions on the other side of the midline, where absorption occurs, if the combination of disk thickness, inclination, and pitch angle of the spirals satisfy the approximate condition $h/r < 2 \tan \theta_p \cot \theta_{\rm inc}$ (see eq. [16]); and the disk may be largely inverted if the gas density and/or disk thickness are sufficiently small, in which case there would be no absorption. As mentioned above, it is unclear how common these possible situations are.

In cases where the high-velocity maser spectrum from a circumnuclear disk exhibits the Doppler shift asymmetry, we predict that there should generally be substantial absorption of maser emission near the systemic velocity of the disk. We argued that it may account for the dip in the low-velocity maser complex in NGC 4258 (§ 6.2), and we suggest that it may also explain why the systemic maser emission in NGC 1068 and NGC 4945 is weaker than the redshifted emission, in spite of probably being boosted due to illumination by a central background continuum source.

According to our model, the energy that ultimately powers the high-velocity masers is provided by energy dissipation in shocks, while the low-velocity masers probably arise near the inner edge of the disk that is irradiated directly by a central source. Since epochs of spiral shock activity in a disk may not necessarily overlap with epochs of substantial nuclear activity, we predict that high-velocity maser emission could, in principle, appear in the absence of detectable low-velocity emission. This situation may be detected in galactic nuclei that do not contain, at the present epoch, an apparent central source that heats up the disk and produces background radio continuum. Obviously, it requires the existence of a molecular, circumnuclear disk that is viewed nearly edge-on and that contains spiral shocks.

We predict that high-velocity maser spots in circumnuclear disks that exhibit the Doppler shift asymmetry will appear clustered in clumps rather than being uniformly or randomly distributed, since they arise where the lines of sight are tangent to the spiral shocks (turbulent motions in the gas presumably cause the clumps to break up into individual spots). The clumps may be either narrow or broad, depending on whether the shock fronts are smooth or heavily corrugated. The number of high-velocity maser clumps depends on the spiral pattern and could be as low as one.

The second most compelling case of a rotating structure, delineated by water maser emission, is found in the nucleus of NGC 1068 (Gallimore et al. 1996; Greenhill et al. 1996). The maser emission extends about $\pm 300$ km s$^{-1}$ from the galactic systemic velocity, tracing sub-Keplerian differential rotation around a central mass of $\sim 10^7\ M_\odot$ that lies within a radius of about 0.7 pc. The Doppler shift asymmetry is as striking as in NGC 4258, except that in NGC 1068 the redshifted, high-velocity maser emission is stronger than even the low-velocity emission complex at the systemic velocity of the galaxy. The VLBI observations by Greenhill et al. (1996) currently provide the most precise image of the disk but only of its receding side. It is delineated by five dominant clumps of maser emission, distributed in a nearly planar configuration on the sky: one at the systemic velocity of the galaxy and the others at relatively redshifted velocities. The weaker, blueshifted, high-velocity features have been recently imaged by Greenhill and coworkers (L. J. Greenhill 1997, private communication), but the results are not yet available.

The disk in NGC 1068 is not nearly as well defined as that in NGC 4258, and in fact, at present, it is not yet clear whether the observed structure is a disk or a thick torus. Greenhill et al. (1996) proposed that the observed emission traces part of the limb of an edge-on rotating torus with an opening angle of about 90°, rather than the midplane of a disk as suggested by Gallimore et al. (1996). The reason for that is twofold: the radio jet axis is misaligned from the nearly planar distribution of the redshifted, high-velocity sources by about 45° rather than $\sim 90°$. In addition, the location of the blueshifted emission, observed with lower resolution by Gallimore et al. (1996), is more consistent with being symmetric about the jet axis with respect to the redshifted emission than with all of the features being nearly aligned in a plane.

Yet there are some difficulties with this model, as pointed out by Greenhill et al. (1996). The condition of hydrostatic equilibrium implies gas motions of order 100 km s$^{-1}$ which, if thermal, correspond to temperatures at which molecular gas could not survive. If the maser emission arises in cold gas clouds with highly supersonic velocities, it is unclear why the high-velocity maser spots appear clustered in a few distinct "superclumps." It is also difficult to understand the absence of high-velocity maser emission from the southern half of the torus and near its midplane. Greenhill et al. (1996) suggested that it may be related to the absence of strong southern jet emission. But if the irradiation of the inner surface of the torus is dominated by the jet, which is likely to produce also radio continuum emission, it is unclear why we do not detect systematic emission superimposed along some length of the jet or at least along its northern part. These problems do not arise if the rotating structure is a relatively thin disk, which cannot be ruled out at present because the uncertainties in the position of the weak, blueshifted masers are quite large (Gallimore et al. 1996). The axis of the radio jet may not be perpendicular to the masing disk, and its direction may be determined by the accretion disk on smaller scales, which could have a different orientation, or by the equatorial plane of a misaligned, spinning black hole. VLBI observations of the blueshifted

emission would probably allow one to distinguish between the torus and disk hypotheses.

We suggest that, as in NGC 4258, the high-velocity maser emission in NGC 1068 arises owing to spiral shocks in a largely nonmasing disk, which accounts for the similar Doppler shift asymmetry in the maser spectrum. The low-velocity masers may arise within a thin layer at the inner edge of the rotating structure that is irradiated directly by the central source and amplify a central background continuum source. Assuming that the disk is relatively thin, the shock-origin scenario described in the present paper can account for the clustering of the high-velocity maser spots in NGC 1068 into four clumps rather than being randomly distributed. However, unlike in NGC 4258, the shocks in NGC 1068 are probably heavily corrugated, as is perhaps indicated by the larger extent of the clumps relative to the mean interclump separation. The rotational speed in NGC 1068 is about 3 times lower than that in NGC 4258, so assuming the shock speed is at least as high as in NGC 4258, it is reasonable to expect the typical pitch angle of the spirals in NGC 1068 to be $\gtrsim 6°$ (eqs. [3] and [24]). Higher mean pitch angles allow larger possible variations in the local pitch angle throughout the spiral pattern. Thus, the higher the pitch angle, the more unlikely it is that the maser clumps would all lie exactly along a single diameter through the disk. This may contribute to the observed deviations from a Keplerian rotation curve in NGC 1068, in addition to the probable effect of the star distribution on the gravitational potential (the volume enclosed within the radius of the masing disk in NGC 1068 is a few hundred times larger than the corresponding volume in NGC 4258).

The origin of the maser emission in NGC 1068 and the role of absorption in producing the Doppler shift asymmetry in this system will be examined in more detail when VLBI images of the receding side of the disk become available and the nature of the rotating structure is revealed.

## 8. DISCUSSION

We propose a new explanation for the origin of water maser emission in circumnuclear disks, where the high-velocity maser features arise behind spiral shock fronts in a largely noninverted disk. The model has considerable explanatory power as well as predictive power (§ 6.3).

The main results of this investigation are summarized in the abstract, so we will avoid redundancy here . We summarize only the basic assumptions of the model, which are as follows.

1. The molecular density in the disk should not exceed a few times $10^9$ cm$^{-3}$, depending exactly on the shock type and strength, in order not to collisionally quench inversion in the postshock gas (§ 3.1), and the density should also not be much lower than a few times $10^8$ cm$^{-3}$, depending exactly on the dimensions of the disk and the water abundance, in order for the gas in the disk to be largely quenched due to IR photon trapping (§ 2).

2. Spiral shocks may form in circumnuclear disks for the same reasons they do in disk galaxies, accretion disks, and in protostellar disks, which is basically due to the presence of nonaxisymmetric disturbances, of which there is no shortage at the bottom of a galactic potential well (§ 3.2).

3. The shock speed needs to be high enough above the sound speed and the turbulent velocity in order to allow efficient escape of IR photons from the warm, postshock region through the shock front (§ 4.1).

4. The pitch angle of the spiral should not be too high, depending exactly on the disk thickness and inclination, in order for the blueshifted maser beams to pass through the corresponding velocity-coherent regions on the other side of the midline, where absorption occurs (§§ 5.1 and 7).

5. The gas and dust are nearly in thermal equilibrium, so reprocessing of the IR photons (produced in the pump cycle) by the dust does not significantly affect the level populations of the water maser transition as to become inverted.

These assumptions should be satisfied under a wide variety of conditions in circumnuclear disks, and as a result we believe that the model proposed here provides a robust explanation for masers in these systems.

Special thanks to D. Neufeld for providing us with the energy levels of water and his unpublished results for the maser rate coefficient $Q$ at very high $\xi$ values, as well as for useful discussions. We also thank M. Elitzur for comments and discussions, D. Hollenbach for discussions, and S. Shore for useful comments. The research of C. F. M. is supported in part by National Science Foundation grant AST95-30480.

## REFERENCES

Baan, W. A., & Haschick A. 1996, ApJ, 473, 269
Braatz, J. A., Wilson, A. S., & Henkel, C. 1996, ApJS, 106, 51
Cecil, G., Wilson, A. S., & De Pree, C. 1995, ApJ, 440, 181
Chakrabarti, S. K., & Wiita, P. J. 1994, ApJ, 434, 518
Cecil, G., Wilson, A. S., & Tully, R. B. 1992, ApJ, 390, 365
Chernoff, D. F., McKee, C. F., & Hollenbach, D. J. 1982, ApJ, 259, L97
Collison, A. J., & Watson, W. D. 1995, ApJ, 452, 103
Cunningham, C. 1976, ApJ, 208, 534
Deguchi, S. 1981, ApJ, 249, 145
De Jong, T. 1973, A&A, 26, 297
Donnor, K. J. 1979, Ph.D. thesis, Cambridge University
Downes, D. 1983, in The Birth and Infancy of Stars, ed. R. Lucas, A. Omont, & R. Stora (Amsterdam: North Holland), 560
Draine, B. T., Roberge, W. G., & Dalgarno, A. 1983, ApJ, 264, 485
Draine, B. T. 1980, ApJ, 241, 1021
Elitzur, M. 1979, ApJ, 229, 560
Elitzur, M., Hollenbach, D. J., & McKee, C. F. 1989, ApJ, 346, 983 (EHM89)
Elmegreen, B. G., & Morris, M. 1979, ApJ, 229, 593
Ford, H. C., et al. 1986, ApJ, 311, L7
Fujimoto, M. 1966, in IAU Symp. 29, Non-Stable Phenomena in Galaxies, ed. V. Ambartsumian (Yerevan, SSR: Armenian Acad. Sci.), 453
Gallimore, J. F., Baum, S. A., O'Dea, C. P., Brinks, E., & Pedlar, A. 1996, ApJ, 462, 740
Genzel, R. 1986, in Maser, Molecules and Mass Outflows in Star Forming Regions, ed. A. D. Haschick (Westford, MA: Haystack Observatory), 233
Goldreich, P., & Kwan, J. 1974, ApJ, 191, 93
Goldreich, P., & Lynden-Bell, D. 1964, MNRAS, 130, 125
Greenhill, L. J. 1996, to appear in IAU Colloq. 159, Emission Lines in Active Galaxies, ed. B. M. Peterson, F. Z. Cheng, & A. Wilson (San Francisco: ASP)
Greenhill, L. J., Henkel, C., Becker, R., Wilson, T. L., & Wouterloot, J. G. A. 1995, A&A, 304, 21
Greenhill, L. J., Gwinn, C. R., Antonucci, R., & Barvainis, R. 1996, ApJ, 472, L21
Greenhill, L. J., Moran, J. M., & Herrnstein, J. R. 1997, ApJ, 481, L23
Gwinn, C. R. 1994, ApJ, 429, 241
Herrnstein, J., Greenhill, L., & Moran, J. 1996, ApJ, 468, L17
Herrnstein, J., et al. 1997, ApJ, 475, L17
Hollenbach, D. J., & McKee, C. F. 1979, ApJS, 41, 555
———. 1980, ApJ, 241, L47
Hollenbach, D. J., Elitzur, M., & McKee, C. F. 1998, in preparation (HEM98)
Hollenbach, D. J., McKee, C. F., & Chernoff, D. 1987, in Star Forming Regions, ed. M. Peimbert & J. Jaguka (Dordrecht: Reidel), 344
Julian, W. H., & Toomre, A. 1966, ApJ, 146, 810
Kaufman, M., & Neufeld, D. 1996, ApJ, 456, 611

Lasota, J. P., Abramowicz, M. A., Chen, X., Krolik, J., Narayan, R., & Yi, I. 1996, ApJ, 462, 142
Litvak, M. M. 1969, Science, 165, 855
Lynden-Bell, D. 1974, In Galaxies and Relativistic Astrophysics, ed. B. Barbanis & J. D. Hadjidemetriou (Berlin: Springer), 224
Makishima, K., et al. 1994, PASJ, 46, L77
Maoz, E. 1995a, ApJ, 447, L91
———. 1995b, ApJ, 455, L131
Matsuda, T., Inoue, M., Sawada, K., Shima, E., & Wakamatsu, K. 1987, MNRAS, 229, 295
McKee, C. F., & Hollenbach, D. J. 1980, ARA&A, 18, 219
Melnick, G. J., Menten, K. M., Phillips, T. G., & Hunter, T. 1993, ApJ, 416, L37
Miller, R. H., & Smith, B. F. 1992, ApJ, 393, 508
Miyoshi, M., Moran, J., Herrenstein, J., Greenhill, L., Nakai, N., Diamond, P., & Makoto, I. 1995, Nature, 373, 127
Montgomery, A. S., & Cohen, R. J. 1992, MNRAS, 254, 23p
Morfill, G., Spruit, H., & Levy, E. H. 1993, In Protostars and Planets III, ed. E. H. Levy & J. I. Lunine (Tuscon: Univ. of Arizona Press), 950
Nakai, N. 1989, PASJ, 41, 1107
Nakai, N., Inoue, M., & Miyoshi, M. 1993, Nature, 361, 45
Nakai, N., et al. 1995, PASJ, 47, 771
Neufeld, D. A., & Hollenbach, D. A. 1994, ApJ, 428, 170
Neufeld, D. A., & Maloney, P. R. 1995, ApJ, 447, L17
Neufeld, D. A., Maloney, P. R., & Conger, S. 1994, ApJ, 436, L127
Neufeld, D. A., & Melnick, G. 1990, ApJ, 352, L9
———. 1991, ApJ, 368, 215
Papaloizou, C. B., Terquem, C., & Lin, D. N. C. 1998, ApJ, in press
Ponomarev, V. O., Smith, H. A., & Strelnitski, V. S. 1994, ApJ, 424, 976
Reid, M. J., & Moran, J. M. 1988, in Galactic and Extragalactic Radio Astronomy, ed. G. L. Verschuur & K. I. Kellermann (Berlin: Springer), 255
Roberts, M. S. 1967, PASP, 69, 59
Roberts, W. W. 1969, ApJ, 158, 123
———. 1970, IAU Symp. 38, The Spiral Structure of Our Galaxy, ed. W. Becker & G. Contopoulos (Dordrecht: Reidel), 415
Roberts, W. W., Roberts, M. S., & Shu, F. H. 1975, ApJ, 196, 381
Sawada, K., Matsuda, T., & Hachisu, I. 1986, MNRAS, 219, 75
Sawada, K., Matsuda, T., Inoue, M., & Hachisu, I. 1987, MNRAS, 224, 307
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337
Shmeld, I. K., Strelnitskii, V. S., & Muzylev, V. V. 1976, Sov. Astron., 20, 4111
Shu, F. H., Milione, V., Gebel, W., Yuan, C., Goldsmith, D. W., & Roberts, W. W. 1972, ApJ, 173, 557
Shu, F. H., Milione, V., & Roberts, W. W. 1973, ApJ, 183, 819
Spruit, H. C. 1987, A&A, 184, 173
Spruit, H. C., Matsuda, T., Inoue, M., & Sawada, K. 1987, MNRAS, 229, 517
Strelnitskii, V. S. 1971, Astron. Tsirk., 609, 1
———. 1984, MNRAS, 207, 339
Strelnitskii, V. S., & Sunyaev, R. A. 1973, Sov. Astron., 16, 579
Tartar, J. C., & Welch, W. J. 1986, ApJ, 305, 467
Toomre, A. 1981, In The Structure and Evolution of Normal Galaxies, ed. S. M. Fall & D. Lynden-Bell (Cambridge: Cambridge Univ. Press), 111
van der Kruit, P. C., Oort, J. H., & Mathewson, D. S. 1972, A&A, 21, 169
Wardle, M. 1990, MNRAS, 246, 98
Watson, W. D., & Wallin, B. K. 1994, ApJ, 432, L35
Westerhout, G. 1968a, Bull. Astron. Inst. Neth., 14, 215
———. 1968b, Bull. Astron. Inst. Neth., 14, 261
Wilson, A. S., Braatz, J. A., & Henkel, C. 1995, ApJ, 455, L127
Woodward, P. R. 1973, Ph.D. thesis, University of California, Berkeley