THE ASTROPHYSICAL JOURNAL, 428:693–712, 1994 June 20

DETERMINING STRUCTURE IN MOLECULAR CLOUDS

JONATHAN P. WILLIAMS[1]
Radio Astronomy Laboratory, University of California, Berkeley, CA 94720
AND
EUGÈNE J. DE GEUS[2] AND LEO BLITZ[3]
Astronomy Department, University of Maryland, College Park, MD 20742
Received 1993 September 7; accepted 1993 December 16

ABSTRACT

We describe an automatic, objective routine for analyzing the clumpy structure in a spectral line position-position-velocity data cube. The algorithm works by first contouring the data at a multiple of the rms noise of the observations, then searches for peaks of emission which locate the clumps, and then follows them down to lower intensities. No a priori clump profile is assumed. By creating simulated data, we test the performance of the algorithm and show that a contour map most accurately depicts internal structure at a contouring interval equal to twice the rms noise of the map. Blending of clump emission leads to small errrors in mass and size determinations and in severe cases can result in a number of clumps being misidentified as a single unit, flattening the measured clump mass spectrum.

The algorithm is applied to two real data sets as an example of its use. The Rosette molecular cloud is a "typical" star-forming cloud, but in the Maddalena molecular cloud high-mass star formation is completely absent. Comparison of the two clump lists generated by the algorithm show that on a one-to-one basis the clumps in the star-forming cloud have higher peak temperatures, higher average densities, and are more gravitationally bound than in the non–star-forming cloud. Collective properties of the clumps, such as temperature–size–line-width–mass relations appear very similar, however. Contrary to the initial results reported in a previous paper (Williams & Blitz 1993), we find that the current, more thoroughly tested analysis finds no significant difference in the clump mass spectrum of the two clouds.

Subject headings: ISM: clouds — ISM: individual (Rosette Nebula, Maddalena Cloud) — ISM: structure — methods: analytical — stars: formation

1. INTRODUCTION

Within the solar circle, half the mass of the interstellar medium is in the form of molecular clouds (Dame 1993), ranging from high-latitude clouds with masses of a few times 10 $M_\odot$ and sizes of a few parsecs to giant molecular clouds (GMCs) with masses as large as several times 10^6 $M_\odot$ and sizes greater than 100 pc. Star formation has been observed in each type of molecular cloud, from the smallest (Magnani, Blitz, & Mundy 1985) to the largest (e.g., Myers et al. 1986). Because the stars form from the clouds, understanding their formation and evolution explains the earliest stages in the process of star formation. The molecular clouds are always observed to be highly fragmented (e.g., Blitz & Shu 1980 and references therein). Thus the ability to quantify the fragmentation or clumpiness of clouds may make it possible to learn how cloud structure is related to cloud formation and evolution.

Molecular clouds have been observed at optical, infrared, and millimeter wavelengths by means of, respectively, the stellar extinction due to associated dust (e.g., Dickman 1978), the thermal emission of this dust (e.g., Cox, Deharveng, & Leene 1991), and the rotational and vibrational spectral lines of the molecules themselves (e.g., Turner & Ziurys 1988). Radio spectral line mapping has the advantage over optical and infrared of providing high velocity resolution which not only gives dynamical information about the cloud but also separates out components along the same line of sight, creating a "data cube" with intensity a function of two positional and one velocity coordinate.

The most commonly used tracer of molecular gas is the CO molecule because of its high abundance. However, the volume-averaged density of a molecular cloud is typically $\bar{n} \sim 100$ cm^{-3}, an order of magnitude less than that needed to excite CO emission, suggesting that the clouds consist of density enhancements with a surface filling fraction of unity but a volume filling fraction of about 10% (Blitz & Shu 1980). Higher resolution mapping, which became possible with larger telescopes and lower noise receivers, showed that these density enhancements form discrete, localized *clumps* of emission (e.g., Blitz & Stark 1986). The measurements of the far-infrared C II line by Stutzki et al. (1988), which traces the presence of ionizing photons, provide further independent evidence for clumpiness: ionizing radiation penetrates far deeper into the cloud than would be expected in a uniform medium. *All* molecular clouds that have been observed at high resolution show a clumpy internal structure, and it is through an understanding of this structure that questions about cloud evolution and star formation can be addressed. For example, are evolutionary differences (i.e., star forming or not) between clouds reflected in the distribution of clumps? Even in a star-forming cloud, there are clumps that do not form stars: what are the conditions of these clumps, and will they *eventually* form stars, or are they simply inert gas? What is the effect of stars on clumps—can star formation be induced in clumps by nearby stars? Do the

[1] Also Astronomy Department, University of Maryland; e-mail address: williams@astron.berkeley.edu.
[2] Present address: Astronomy Department 105-24, Caltech, Pasadena, CA 91125; edg@astro.caltech.edu.
[3] E-mail: blitz@astro.umd.edu.

clumps themselves possess internal structure? Can we see the actual process of clump formation?

There have been several observations of different molecular clouds with the purpose of observing and describing the clumpy internal structure (Lada, Bally, & Stark 1991; Loren 1989; Bally et al. 1987; Carr 1987; Stutzki & Güsten 1990; Herbertz, Ungerechts, & Winnewisser 1991; see Blitz 1993 for a review). ^{13}CO is the most commonly used tracer for such observations because it has reasonably strong lines without the strong saturation effects associated with CO along most lines of sight. However, other optically thin lines can be and have been used. Because these lines directly trace the column density rather than the temperature, clump peaks are readily apparent as emission maxima in the data cube. The emission associated with each peak determines the clump size and mass, and as long as individual clumps are sufficiently well separated, it is possible to catalog clumps and clump quantities by eye—the method used for the majority of cases above. In regions of high clump density, blending of emisson makes estimates of the relative contributions of each clump more difficult to calculate accurately. Eye analyses generally split up the blended emission by an intuitive extrapolation of the size of the contributing clumps, thereby imposing a subjective slant on the data analysis. Moreover, as such data sets become increasingly large and more common because of faster telescope systems, manual analyses simply become unfeasible. An automatic structure analysis routine can handle large data cubes efficiently and objectively, making possible uniform comparisons of large numbers of clouds—an important step to understanding cloud structure and evolution. The first such routine, hereafter referred to as Gaussclumps, was developed by Stutzki & Güsten (1990), who fitted the data with a sum of triaxial Gaussians. The algorithm that we have developed is closely based on eye-analysis techniques, and simply automates the process.

This paper describes, tests, and applies this new algorithm, for analyzing the clumpy structure of spectral line data cubes. We describe how the algorithm, henceforth called Clumpfind, works, and compare it with the different approach of Gaussclumps in § 2. It is tested on simulated data, and its performance is evaluated, in § 3: the results of the simulations determine optical values of user-entered variables and demonstrate the effect of blending. In § 4 we apply the algorithm, comparing the clump properties in a star-forming GMC and a non–star-forming GMC. We conclude in § 5.

2. THE CLUMP-FINDING ALGORITHM

2.1. *Appearance of the Data*

The internal structure of molecular clouds observed in millimeter-wave spectral lines is most readily seen in contour maps. Clump analysis by eye, or by the algorithm we describe here, work not on the actual continuous intensity map itself but on its representation as a set of discrete contours. The appearance of such a map is that of a set of peaks surrounded by gradually decreasing contours (see, e.g., Fig. 8). For the case of an optically thin line, intensity peaks correspond to density maxima in the cloud which are the clump peaks themselves. It is therefore an easy matter in principle to *count* the clumps (it is shown in § 3.2 that this consideration alone determines how best to contour the data). However, to go beyond simply counting the numbers of clumps and actually describing their physical properties requires following the clump profiles to lower levels. In the case where the clumps are well separated, each peak can be readily and unambiguously followed down to the noise level of the map.

As the clump-to-clump separation decreases, however, blending of clump emission will occur above the noise level, corresponding to contours that surround more than one peak. Each peak still corresponds to a clump, but now the task of following the clump profile out to lower intensities requires dividing up these shared contours. It is this step that is the most subjective in producing a list of clump properties by eye. The Rosette molecular cloud is an example where there clearly is some blending of emission but it is sufficiently mild that clump decomposition by eye is still possible (Blitz & Stark 1986).

When there is strong blending so that there are many peaks sharing contours at high intensities, it becomes more difficult for the eye to divide up the blended emission accurately and correctly describe the clumps. Part of the problem is the difficulty with visualization of the full three-dimensional structure using only two-dimensional slices. A computer routine has the immediate advantage of not being hindered by a three-dimensional array and also of applying the same rules used in the low-level blending case in a uniform, unbiased way both from clump to clump within a single data cube and from one data cube to another.

Ultimately there is the case where clumps are so close together in projected position and radial velocity that they appear as only a single peak. In this case the observations cannot distinguish between the different clumps, and errors will inevitably result, *no matter what method of clump analysis is used.* It has already been demonstrated in the case of CO surveys of the Galaxy (which are severely blended because of the low contrast of CO emission and velocity crowding at the tangent points) that cloud decomposition can be very misleading (Adler & Roberts 1992).

The computer algorithm that we have developed was motivated by how the eye decomposes the maps into clumps and mimics what an infinitely patient observer would do in the same manner that we have heuristically described above. However, the nature of the observations—finite resolution and instrumental noise—make additional considerations necessary.

2.2. *Description of the Algorithm*

Spectral line mapping of a molecular cloud produces a data cube or a set of pixels in position and velocity measuring the emission, generally in temperature units. Clumpfind directly reads the data cube into a three-dimensional array and, just as in producing contour maps, requires a contouring interval with which to "view" the data. Linear contours are assumed, $T = \Delta T, 2\Delta T, 3\Delta T, \ldots$, where ΔT is a user-defined parameter whose value we discuss in § 3.2.

Clumpfind traces structures by connecting pixels, at each contour level, that are within one resolution element of each other. Generally, the resolution is equal to the beam size in position (or the width of one spectrometer channel for the velocity axis). For example, if the mapping is made at full beam spacing, then one resolution element is just the distance from one pixel to its immediate neighbor, but is twice this for fully sampled (half-beam) observations. However, if the observations are undersampled, then structures smaller than the map grid spacing cannot be detected, and the effective resolution is again equal to the interpixel distance.

Typically, the contour levels are only 1 or 2 pixels wide (i.e.,

down the steepest gradient of a clump profile each or every other pixel is at a different contour level). The task of connecting neighboring pixels within the same contour level and of then following structures from one contour level to another therefore requires careful consideration. We introduce the parameter N_{axis}, which limits a neighboring pixel to be 1 resolution element away in at most $N_{axis} = 1, 2, 3$ axes (either position or velocity). Figure 1 demonstrates how the neighborhood of a single pixel depends on the value of N_{axis}. It is assumed, for diagrammatical ease, that the resolution is equal to the interpixel distance (if the resolution is greater, then there are more pixels along each axis of the cube). If $N_{axis} = 1$, only pixels that share the faces of a cube are considered to be neighbors (each pixel has 6 neighbors). If $N_{axis} = 2$, all pixels along the edges, but not including the vertices, of a cube are considered to be neighbors (each pixel has 18 neighbors). If $N_{axis} = 3$, all pixels in the cube are considered to be neighbors (each pixel has 26 neighbors). It is shown in § 3.2 that the value of N_{axis} is critical to the accuracy of the clump decomposition.

The two parameters, ΔT and N_{axis}, dictate, respectively, how the data are contoured and how the pixels in the data cube are connected. Clumpfind now works on the full three-dimensional data cube in exactly the same way as the eye analyzes a two-dimensional contour map as described in the previous subsection. Clumps are located by their peaks, which are local maxima in the map. A local maximum is an isolated (three-dimensional) contour, i.e., at set of pixels at one contour level that is not connected to other pixels at higher levels.

Clumpfind works from the highest contour levels to the lowest. At each level, isolated contours are identified and labeled as new clumps. The other contours at this level are extensions of previously defined clumps. Contours that surround just one peak are simply assigned to the corresponding clump, but blended contours that surround more than one peak must be split up using a "friends-of-friends" algorithm: the unblended part of each contributing clump (defined at higher contour levels) possesses an immediate neighborhood defined by the parameter N_{axis} that extends into the shared contour level. Similarly, the pixels in these neighborhoods have neighbors that further extend into this contour level, and so on. Eventually, all the pixels in the blended contour level are connected via their "friends" to a clump to which they are thereby assigned. There may be a small number of pixels "in the middle" that end up being connected to more than one clump in this way. These are assigned to the clump whose peak is closer. The process is repeated at the next lower contour level, including any newly merged clumps. As we have emphasized, an important part of the clump decomposition lies in the handling of the blended emission. We test the performance of this method using simulated data in § 3, from which we find that, despite the simplicity of this approach, clump parameters are accurately determined.

The input to Clumpfind is a data cube of intensity in position and velocity. The output is a companion data cube of clump assignments in position and velocity. Therefore, by combining the two data cubes, each position corresponds to an intensity and a particular clump. Conversely, by searching the second data cube for any one clump, its extent and profile are easily followed: this way different clump properties can be quickly calculated (Appendix A), individual clumps can be plotted in relation to the rest of the data cube, and so on.

As a visual example we describe how Clumpfind would work on a two-dimensional contour map in Figure 2*a*. Starting from the highest level, it would define clump A. At the next lower contour level, clump A extends and clump B appears. Clump C appears at the next level and A and B merge: the surrounding contour is split up between the two using the friends-of-friends method. At the next (second lowest) level, clump D appears and A and C merge. The shared contour surrounding A, B, and C is split up between them. Finally, A, B, C, and D all merge at the lowest contour level. The final result is four clumps defined as in Figure 2*b*.

2.3. *Variations on the Method*

A possible problem with the approach described above is that each pixel in a blended region is either assigned completely to one clump or completely to another when in fact the measured intensity is actually a sum of emission from each of the contributing clumps. A more sophisticated treatment would be to permit each pixel to be assigned fractionally to more than one clump. We have, in fact, developed such an algorithm with fractional assignments based on extrapolating the unmerged parts of clumps into blended regions, but found its performance to be no better than the simple single-assignment approach. Indeed, the simulations, which we describe in § 3, demonstrate that this kind of analysis is unnecessary, since the simple friends-of-friends approach works extremely well. The reason that blending can be treated in such an unsophisticated fashion, and also that previous clump decompositions by eye proved so successful, is that the degree of blending of the clumps that are seen in the data is typically very small. Appendix B shows, for the case of two Gaussian clumps blended together, that if more than just $\sim 10\%$ of the flux of one is blended with the other, then the individual peaks of the two clumps will not be resolved and they will not be identified as distinct. Therefore, if the peaks of

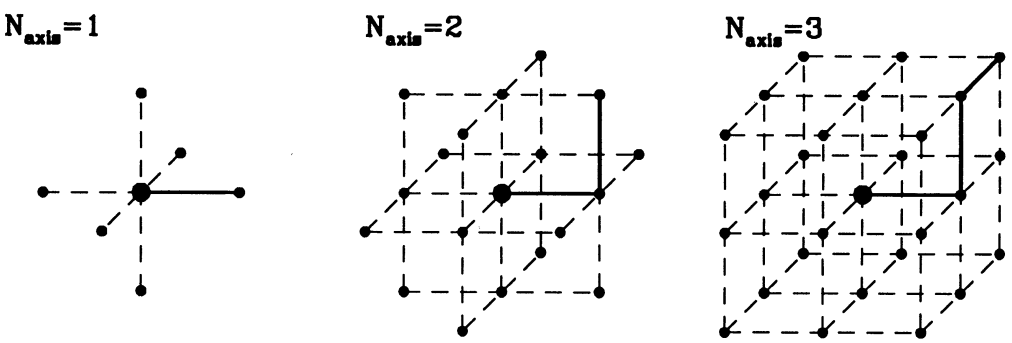

FIG. 1.—Diagram of neighboring pixels for different values of N_{axis}. A neighbor is 1 resolution element away in $N_{axis} = 1, 2, 3$ axes (either position or velocity).

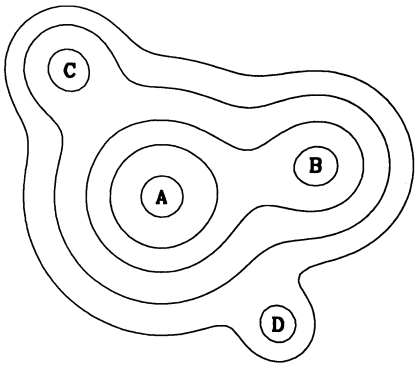

FIG. 2*a*

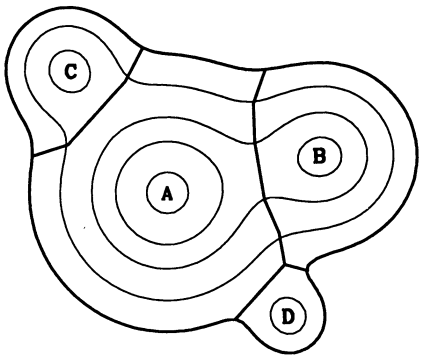

FIG. 2*b*

FIG. 2.—(*a*) Sample contour map to demonstrate how Clumpfind works. (*b*) The clumps found by Clumpfind.

the clumps *can* be distinguished, more than 90% of their flux is contained within the unblended contours and is therefore easily assigned. These numbers are less extreme if large numbers of clumps are all blended together, and also may be different for non-Gaussian profiles. For moderate blending, if the peaks of each clump remain detectable, the flux contained in the blended contours will probably still be relatively small.

On the other hand, it is clear that clumps can readily merge and appear as a single peak which cannot be deciphered using *any* algorithm that identifies clumps as local maxima in the data cube. Either a completely different method of clump deconvolution must be employed (which attempts to determine the constituent components of the blend, and therefore necessarily assumes a particular clump profile), or the cloud should be reobserved in a different molecular line that samples higher densities for which the blending may be less severe.

2.4. *Comparison with Gaussclumps*

The identification of clumps and clump properties is affected by an observer's philosophy about how to separate blended emision into discrete units. There exists in the literature one other automated routine for clump deconvolution: Gaussclumps (Stutzki & Güsten 1990), which fits the data in a generalized least-squares sense as a sum of three-dimensional Gaussian clumps. The assumptions of this method and those of Clumpfind are quite different, and we discuss them here.

The basic assumption of Clumpfind is that local maxima in the data are the peaks of clumps, and thus that the number of peaks in the data cube is equal to the number of clumps in the cloud. The corresponding assumption of Gaussclumps is that the clump profiles are triaxial Gaussians. There are both advantages and disadvantages to either of these differing philosophies. By fitting the data points, Gaussclumps works directly on the continuous intensity distribution. Clumpfind, on the other hand, contours the data into a finite number of intensity steps. For a single, isolated peak, Clumpfind will always find one clump, however peculiar the profile may be. Gaussclumps, on the other hand, may find many clumps (each fitted clump having 11 free parameters) associated with this peak, depending on the profile's degree of departure from a single tri-axial Gaussian. For the case of several clumps blended together, if the peak of any one clump is not visible in the data, that clump will not be found by Clumpfind. To deconvolve such a blend requires fitting the peaks that can be seen, and then fitting the residuals of these fits. The fitting, however, requires a particular clump profile or shape which must be assumed a priori, as in Gaussclumps.

In practice, the real differences between these two approaches are small. Observationally, the large-scale three-dimensional clumpy structure seen in a data cube consists of clumps that are fairly circular and are not strongly blended together. (Elongated structures, e.g., wisps or filaments, are often seen at smaller scales, but even then typically break up into multiple peaks.) Under these conditions, (and since clump profiles are not greatly different from triaxial Gaussians), Clumpfind and Gaussclumps both agree well on the size and location of the largest clumps. We have run some of the simulations described in the following section on Gaussclumps in addition to Clumpfind and find that, whereas Clumpfind misses low-mass clumps that lie below the lowest contour, causing the mass spectrum to flatten at the low-mass end, there is a tendency for Gaussclumps to find enhanced numbers of low-mass clumps from the residuals of the fits to more massive clumps, particularly in the case where the simulated clump profile is not Gaussian in position. This has the opposite effect of steepening the mass spectrum at the low-mass end. At intermediate to high masses, we find that both algorithms generally perform well, except in cases of very high blending, when the both fail. We discuss these results for Clumpfind in more detail in the following section.

3. TESTING WITH SIMULATIONS

The complexity of the three-dimensional structure in a data cube makes it important to check the performance of Clumpfind in realistic cases and to assess its limitations in terms of blending, signal-to-noise ratio, and resolution. This is best done by creating simulated data and modifying the input conditions in a controlled way. In doing so, we have found more general results about analyses of this type of structure that go beyond the application of this particular algorithm.

3.1. *Creating Simulated Data Cubes*

To create a simulated data set, we first need to create a simulated clump. This requires specifying the mass, temperature profile, size, and line width. Since the integrated temperature is proportional to the mass, only three of these are independent. If there exists a relation between mass, size, and line width (e.g., Larson 1981), we need only to specify two: the temperature profile and mass. Observations of an isolated clump generally exhibit a Gaussian velocity profile, which we adopt for the clumps in the simulations. Similarly, we have

chosen a Gaussian for the spatial properties of a clump; the simulated clumps are therefore triaxial Gaussians. For existing data sets, the widest clumps are usually only 5–6 pixels wide, so that the differences with other profiles (e.g., an inverse power law) are small. We have created simulations with different clump profiles and find that the performance of Clumpfind is not significantly affected.

With the adoption of a triaxial Gaussian profile, the clumps are then completely specified by their mass, with the size and line width determined via mass–size–line-width relations. We have attempted to use the observations to guide us, but relations that have appeared in the literature cannot be exactly followed, since, as we will show here, they are inconsistent with each other.

For an optically thin tracer all the clump emission is observed, and the mass is directly proportional to the integrated intensity. For any reasonable clump profile with a central peak, dimensional arguments then imply

$$M \propto T_{\rm peak} \Delta R^2 \Delta V ,$$

where $T_{\rm peak}$, ΔR, and ΔV are the peak temperature, clump size, and (FWHM) line width, respectively (see Appendix A).

Beginning with Larson (1981), relations between size and line width have been both observationally measured and theoretically (Myers & Goodman 1988) deduced. Larson's size–line-width relation, $\Delta R \sim \Delta V^2$, together with his density-size relation, $\bar{n} \sim \Delta R^{-1}$, imply $T_{\rm peak} \sim M^{-1/4}$, so that the more massive clumps have lower temperatures than the less massive clumps, which is not seen in any of the aforementioned studies of molecular cloud structure. The assumption of small optical depth and these mass–size–line-width relations are therefore incompatible.

In fact, the peak temperature of ^{13}CO increases from $\sim$1 K for the smallest (a few solar masses) clumps to $\sim$10 K for the largest clumps of $\sim 10^3$ solar masses (Blitz 1993 and references therein), so that $T_{\rm peak} \sim M^{0.3}$ is a better approximation. The simulations presented here use this scaling along with $\Delta V \sim M^{0.2}$ and $\Delta R \sim M^{0.25}$ (these are similar to what are actually found for two real cloud data sets in § 4). A random ellipticity is then chosen so that clumps have aspect ratios varying from 1 to 2. Changing these relations is easily done and is found to have no significant effect on the performance of the algorithm.

In these simulations, then, the clumps are completely determined from their masses. By now imposing a mass distribution, i.e., the number of clumps as a function of mass, a data cube may be created by simply combining (summing the emission from) a family of such clumps. An input list with the position, peak temperature, size, line width, and mass of each clump is created. There are two types of simulation, unblended and blended: in the first case the clumps are positioned completely separately from one another in a regular grid. In the second case the clump positions are random, with more massive clumps selected from a tighter range in position and velocity so as to simulate the situation within a GMC where the more massive clumps are confined closer to the center of the cloud and have a smaller velocity dispersion about the mean cloud velocity (Blitz 1993). Finally, Gaussian random noise is added to each pixel. Aside from the noise, all the flux in the map therefore comes from the clumps in the input list. The output of Clumpfind is a similar table of clump parameters which may be compared to the input by matching coordinates of clump peaks.

3.2. *Optimal Values of the Parameters*

By running the same simulation with different values of the two input parameters, ΔT and $N_{\rm axis}$, we can find the optimum values that maximize how well the output matches the input and are therefore most likely to give the most accurate analysis when applied to real data.

As we show here, the optimum values of the parameters can be determined without making a direct comparison of clump properties such as size, mass, etc., but simply by *counting* the clumps, i.e., how many clumps does the algorithm find that match with those of the input list of the simulation, how many does it not find, and how many clumps does it find that do not match up with any in the original input list? The detailed workings of Clumpfind (e.g., following clumps to lower intensities, dealing with merging, etc.) are not relevant to the discussion here: the number of clumps is simply determined by counting local maxima, and the results in this subsection could just as easily have been achieved by eye.

There are three errors that may result from the clump deconvolution: some clumps in the input may go undetected, clumps additional to those input may be falsely created, and some of the flux in the map may not be assigned to any clumps at all. We expect these errors to increase as the noise in the simulations is increased, and as the blending of clump emission is increased. We have created two types of simulations, one where clumps are widely separated so that there is very little blending of clump emission and a second where blending does occur. In Figures 3*a* and 3*b* we plot how each of these varies as a function of the two input parameters for a blended and an unblended simulation.

As long as there is no lower limit to the clump spectrum, there will always be some clumps below the detectability limit of a map, because of either resolution or temperature sensitivity or both. So as not to confuse chance agglomerations of noisy pixels as real structure, all clumps detected by Clumpfind must contain at least four resolution elements, with at least one in the second intensity range or higher, to be considered real. Since this requires $T_{\rm peak} \geq 2\Delta T$, the detection limit changes with ΔT, so we expect Clumpfind to be unable to detect small clumps at large contour increments. However, all clumps with peaks above the second contour level should in principle be detected. The first panel in Figures 3*a* and 3*b* plots the percentage of clumps that Clumpfind does not detect, but should. These clumps may be missed due to noise lowering the peak temperature below detectability or blending with nearby clumps that can "hide" the peak (Appendix B).

For the unblended case this percentage is very small for all values of $\Delta T/T_{\rm rms}$. There is a slight increase as $\Delta T/T_{\rm rms}$ decreases below 2, since it increases the likelihood that noise can lower a pixel by a complete contour level and decrease detectable clumps' peaks to below detectability. The percentage actually dips below zero because the minimum detectable clump mass can be less than that expected by setting $T = 2\Delta T$ in the mass-temperature relation. The Gaussian clump profiles in the simulations possess extended wings that collectively increase the peaks of all clumps above that predicted by this relation, with the effect that supposedly undetectable clumps at a given ΔT may in fact extend above the second contour level and be identified by the algorithm. The result is that Clumpfind, instead of missing clumps, actually appears to find more than it should!

For the blended case the percentages are higher overall than would be expected, and along with the increase at $\Delta T/T_{\rm rms} =$

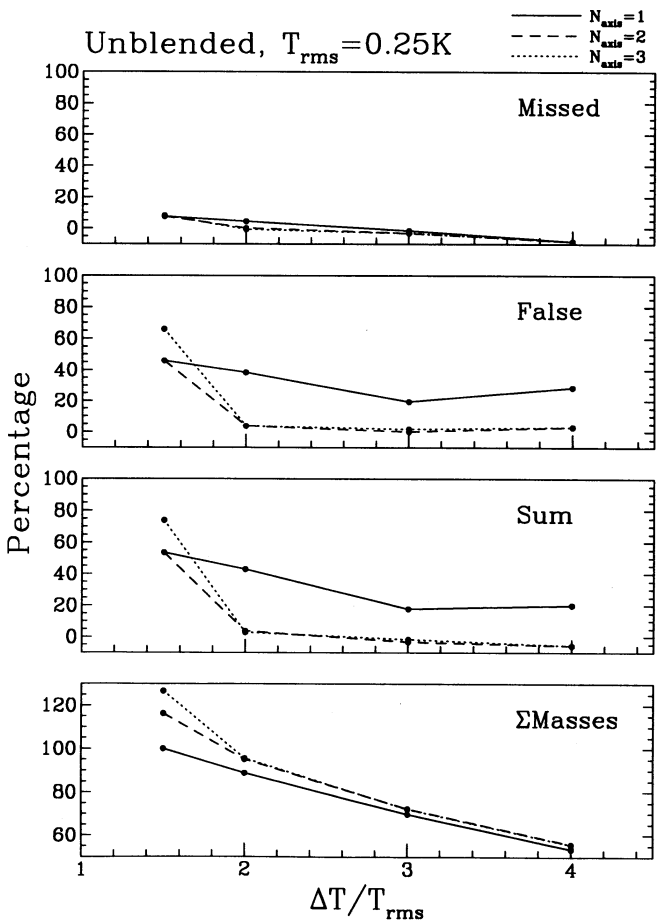

FIG. 3*a*

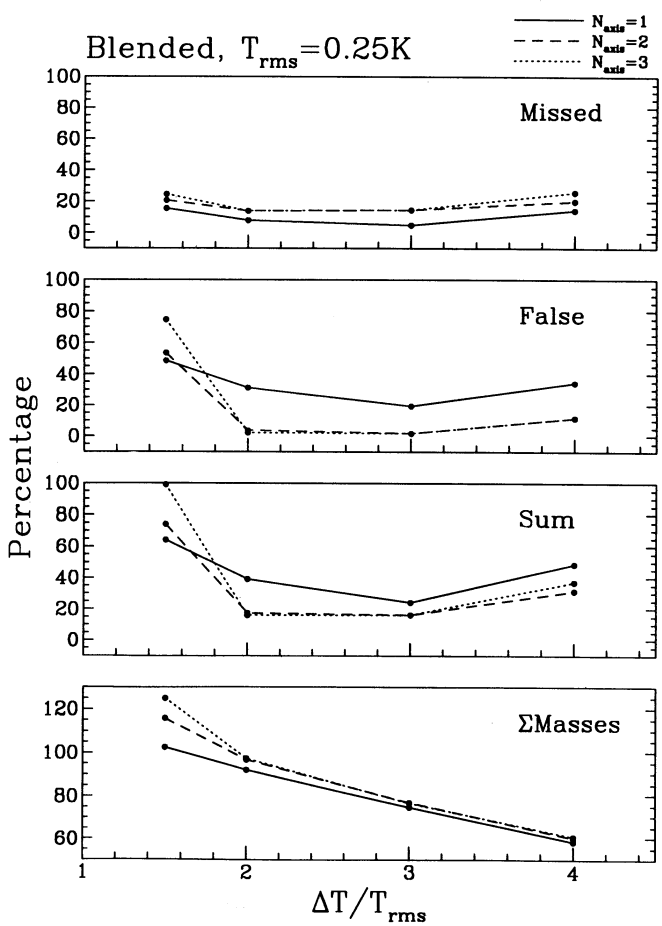

FIG. 3*b*

FIG. 3.—Errors made by Clumpfind as a function of the two user input parameters ΔT and N_{axis} for an unblended (*a*) and blended (*b*) simulation. The first panel (*top*) in each is the number that are not detected as a fraction of the number that are theoretically detectable ($T_{peak} > 2\Delta T$). The second panel is the fraction of clumps in the output that do not match up with any clumps in the input (for $N_{axis} > 1$ these are generally noise spikes). The third panel is the sum of the first two and represents the total number of errors that Clumpfind makes. In the fourth panel the mass of all the output clumps is plotted as a percentage of the sum of the input and demonstrates that increasing flux is missed as the lower threshold $T = \Delta T$ increases.

1.5 there is also a slight increase at high $\Delta T/T_{rms}$ which is due to the lack of contrast as the increments in contouring levels increase and peaks which could be distinguished at a smaller contouring interval merge. $N_{axis} = 1$ appears to find clumps more reliably than higher values of N_{axis}, but, as becomes clear from the next panel, this is due to the very large number of false clumps that are found with $N_{axis} = 1$, some of which happen to match up with an input clump.

The percentage of false clumps, or clumps that the algorithm finds but which do not match up with any clump on the input list, is graphed on the second panel of each of Figures 3*a* and 3*b*. This error is very small except for $N_{axis} = 1$ or $\Delta T/T_{rms} < 1.5$. The problem with $N_{axis} = 1$ is that each pixel is only connected to 6 others (Fig. 1), compared to 18 for $N_{axis} = 2$ and 24 for $N_{axis} = 3$. Contour levels that are typically only 1 pixel wide (cf. § 2.2) cannot be adequately followed with a limited neighborhood defined by $N_{axis} = 1$ that lacks diagonal partners, and therefore are broken up into disconnected regions which can take on the appearance to Clumpfind of a new, false clump. With unlimited memory the data cube could be smoothed to a finer grid and the problem of connecting pixels would be academic. When working directly on the raw data cube, however, it is clear from these figures that we must use $N_{axis} \geq 2$.

Even for $N_{axis} \geq 2$, however, there are large numbers of errors for $\Delta T/T_{rms} = 1.5$. It is to be expected that if the contour increment becomes too small, then the noise of the observations will appear as a number of spikes at different contour levels giving the appearance of false structure. Increasing the increment will decrease this effect but will lower the contrast (number of contours per clump) in the map. Therefore, the optimum setting for the parameter ΔT should be the smallest contour level which is not severely affected by noise. Figures 3*a* and 3*b* show that this is at $\Delta T/T_{rms} = 2$, for which there are almost no errors due to false clumps. The errors increase rapidly, however, for smaller contour increments. We conclude that the structure seen in a map that is contoured at increments less than twice the noise level is liable to be confused with noise.

The third panel is simply the sum of the two errors and demonstrates that Clumpfind makes the fewest errors when $N_{axis} = 2$, 3 and $\Delta T/T_{rms} \geq 2$. There is a slight preference for $\Delta T/T_{rms} = 2$, but it is the total flux recovered that really settles the issue. The fourth panel shows how the sum of the output clump masses as a fraction of the sum of the input clump masses decreases as the lowest level $T = \Delta T$ increases. This is simply due to truncation of the clump profiles (the increase

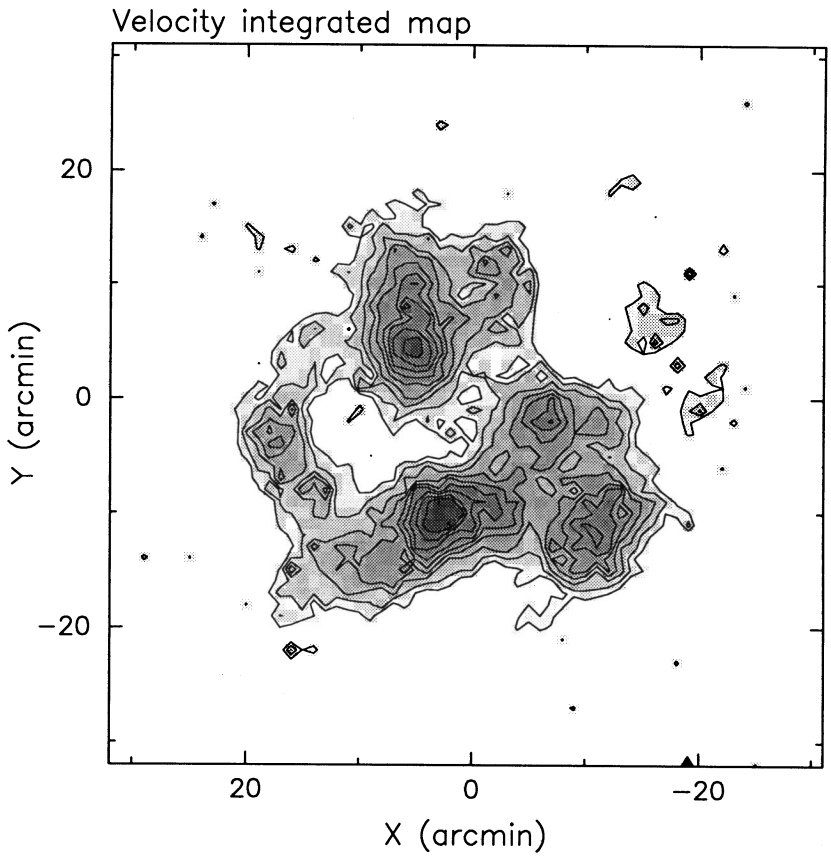

FIG. 4*a*

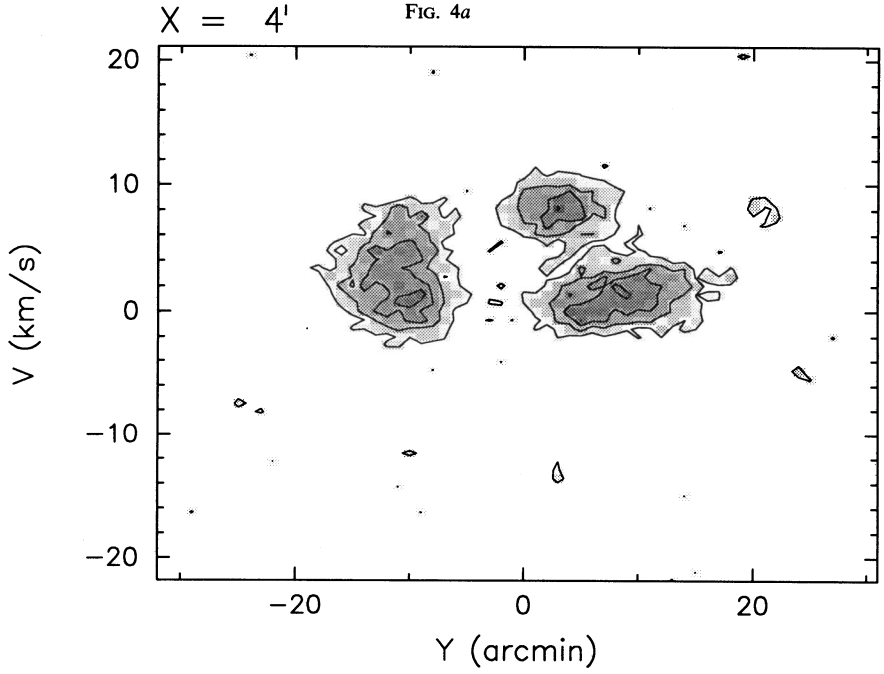

FIG. 4*b*

FIG. 4.—(*a*) Integrated velocity map and (*b*) position-velocity slice along $x = 4'$ of a simulated data cube with the parameters of Table 1. Contours start at 4 K km s^{-1} and increase linearly by 2 K km s^{-1} for the integrated map, and 1 K increasing by 0.5 K for the position velocity slice. The data cube consists of 200 clumps ranging in mass from 1 to 500 $M_\odot$ with a mass distribution $dN/dM \sim M^{-1.5}$.

above 100% for $\Delta T/T_{\rm rms} = 1.5$ is due to counting some of the noise as clumps). The contrast in the map increases, and the truncation of clump edges decreases, as the contouring interval decreases. However, contours can be contaminated by noise for a contouring interval that is too small. The optimum contour interval represents a balance between these considerations.

There are therefore two reasons to choose contour increments $\Delta T = 2T_{\rm rms}$: (1) it is the smallest contouring interval (hence lowest mass detectability and greatest contrast) which is not overwhelmed by the noise, and (2) it allows clump mass estimates to be more accurate by extending as far as possible into the wings of the emission. We emphasize again that this is a general statement about the contouring of molecular cloud observations with the intent of accurately showing structure and not about the internal workings of Clumpfind itself.

3.3. *Performance of the Algorithm*

We have demonstrated that the optimum values of the input parameters are $N_{\rm axis} = 2$ or 3 and $\Delta T = 2T_{\rm rms}$. Here we show the comparison of input and output clump lists for a typical simulation using these optimum parameter values. Two hundred clumps ranging in mass from 1 to 500 $M_\odot$ from a mass spectrum $dN/dM \sim M^{-1.5}$ were placed at a distance of 1000 pc and sampled at 1′ in position and 0.68 km s^{-1} in velocity (corresponding to full-beam sampling on a 12 m telescope and 250 kHz filter banks at the frequency of ^{13}CO). These parameters are summarized in Table 1. Once again, blended and unblended data cubes were created. An integrated velocity map and a position-velocity slice of the blended simulation (shown in Figs. 4*a* and 4*b*) compare well with existing maps of GMCs.

Figure 5*a* is the comparison between input and output for the ideal case with no blending of clump emission, and zero noise added to the data cube. Figure 5*b* plots a similar comparison for a blended, noisy simulation corresponding to the maps in Figures 4*a* and 4*b*. There are six panels, four of which plot input on the *x*-axis against output on the *y*-axis and two of which plot the mass spectrum and a least-squares fit for the input and output clump lists separately. Clumps whose peak positions agree (within 1 resolution element in each axis) can be directly compared and are plotted as filled circles. Clumps which do not match up are plotted as *crosses*, either in a line along the *x*-axis for those which are in the input clump list but have no matching output ("missing" clumps), or along the *y*-axis for those in the output clump list with no corresponding input partner ("false" clumps).

TABLE 1

SIMULATION PARAMETERS

Parameter	Value or Specification
Distance	1000 pc
Resolution	1′, 0.68 km s^{-1}
Number of clumps	200
Mass range	1–500 $M_\odot$
Mass spectrum	$dN/dM \propto M^{-1.5}$
Clump profiles	Triaxial Gaussian
Scaling relations:	
Temperature	$T_{\rm peak} \propto M^{0.3}$
Size	$\Delta R \propto M^{0.25}$
Line width	$\Delta v \propto M^{0.2}$

As expected, the match is excellent in the noiseless, unblended case (Fig. 5*a*). In principle, since $T_{\rm rms} = 0$, the contouring interval can be arbitrarily small. However, in practice, for very small contour intervals, each data point is at a separate contour level, and the algorithm breaks down. We set contour levels at $\Delta T = 0.5$ K for uniformity with the noisy simulations that we discuss next. Sizes, line widths, and the lower ends of the masses are underestimated due to the truncation of the clump profiles at the lowest working level in Clumpfind, $T = \Delta T$. Thus, it is the comparison of continuous input clump parameters with measurements from a discrete data cube that is responsible for the small deviations from a straight line and serves as a benchmark to compare with the effect on the output of adding noise and allowing blending. Mass errors are greater for smaller, lower mass clumps for which the truncation is relatively more severe. Sizes and line widths, which vary less from clump to clump (see § 3.1) are more uniformly affected), however. There are no clumps in the output that do not match up with the input (i.e., there are no crosses along the *y*-axis) but Clumpfind fails to detect clumps with $T_{\rm peak} < 2\Delta T = 1$ K (i.e, crosses along the *x*-axis). This is simply due to the requirement that a clump contain at least 1 pixel in the second working level or higher. The same temperature sensitivity limit will apply to real data and can be translated into a mass-sensitivity limit if a mass-temperature relation is known.

Care must be taken when applying the mass-sensitivity limit to statistics, since noise can make clumps that lie right on this limit undetectable. The more relevant mass is the *completeness* mass above which *all* clumps are found. Clumps on the detection limit have only 1 resolution element (pixel) above $2\Delta T$, which may be dropped below this minimal level by noise. Clumps with more than 1 pixel above the second contour level are much less likely to be "lost in the noise" in this way. A heuristic criterion for completeness then is that there be more than 1 pixel with $T > 2\Delta T$—a resolution-dependent criterion—which approaches the basic mass-sensitivity limit in the case of infinitely good resolution (more pixels per clump). For the simulations, which are made to resemble real data, this completeness mass was of order 3–4 times greater than the theoretical mass sensitivity.

As a more realistic example, Figure 5*b* is the comparison between input and output for a simulation with blending of clump emission and noise with an rms of 0.25 K per pixel (so that contours are set to $\Delta T = 2T_{\rm rms} = 0.5$ K). The comparison is, naturally, not as good as the preious ideal case, but is nevertheless generally within a factor of 2 for sizes, line widths, and masses. The combination of noise which tends to increase the maximum, and blending (adding together) of clump emission, explain why the output peak temperature is found to be greater than the input. The effect of truncation, which lowered the output (measured) sizes, line widths, and masses below the input in the noiseless, unblended simulations, is no longer as apparent. This may be due to the parts of the extended wings of clumps that would lie just below the lowest contour, $T = \Delta T = 2T_{\rm rms}$, with no blending, being boosted into this contour because of blending. Also, if the wings are quite extended, so that there are more pixels that lie just below the lowest contour than just above it in the absence of noise, then noise will also tend to increase the overall number of pixels that are found in the lowest contour. Since Clumpfind defines clumps out to this lowest contour level, size, line widths, and masses will tend to increase because of these effects.

There are now several clumps along the *x*-axis that are unde-

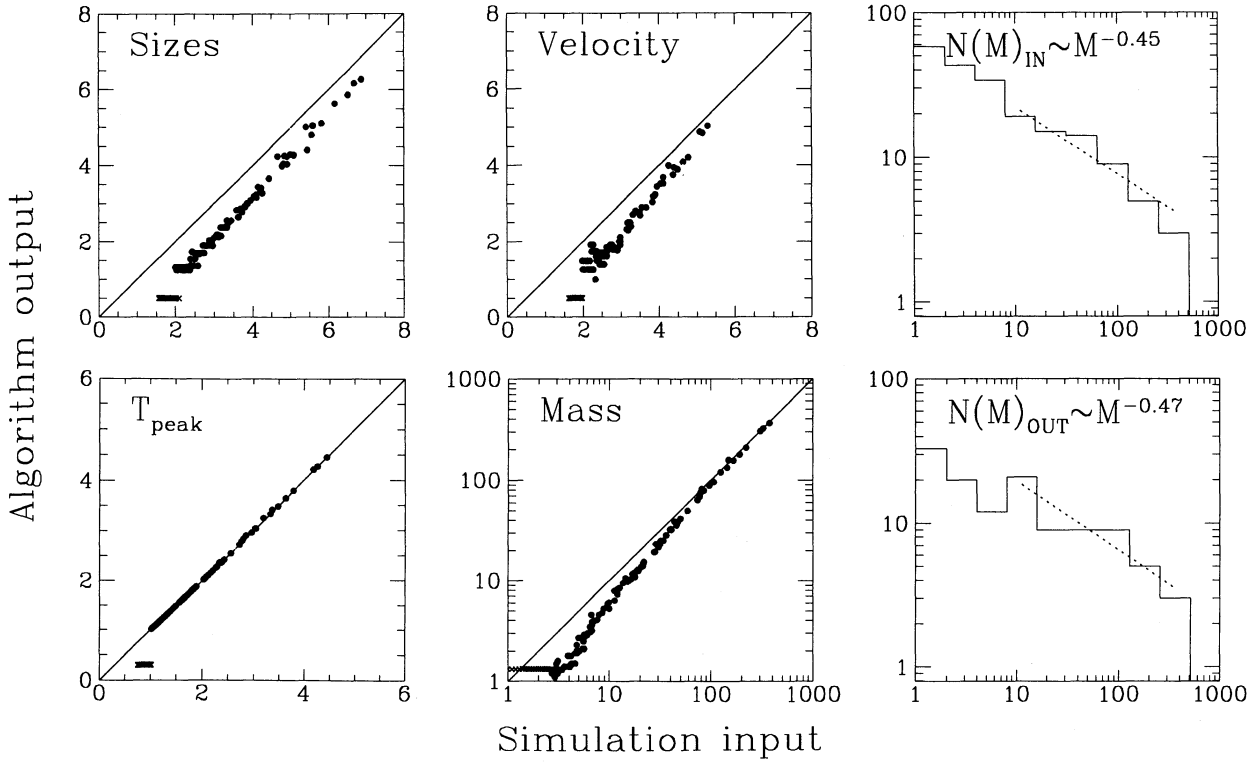

FIG. 5a

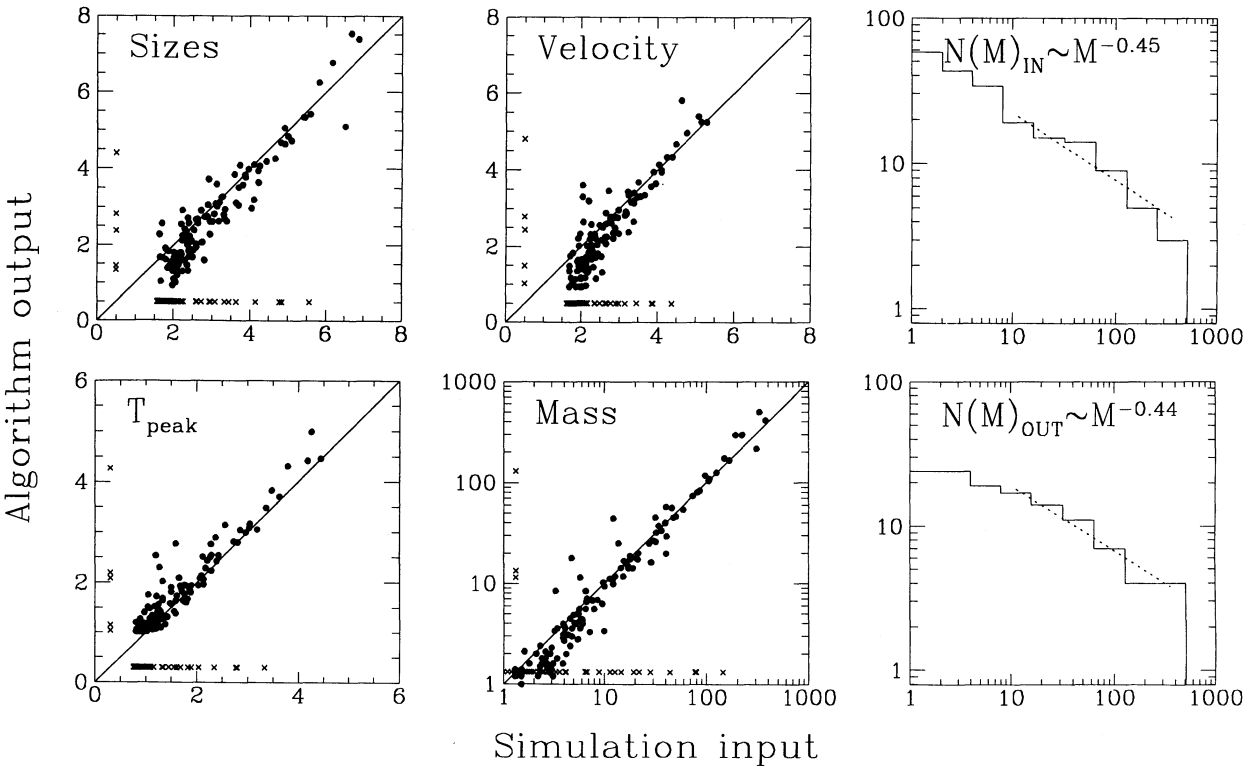

FIG. 5b

FIG. 5.—(*a*) Output clump properties as measured by Clumpfind against the input values. Clumps that are in the input but not matched to any in the output appear as *crosses* along the *x*-axis ("missing" clumps). There are such large numbers of small clumps that are not detected that the crosses are in fact indistinguishable in this figure. Crosses along the *y*-axis are clumps that are in the output but have no counterpart in the input ("flase" clumps). The two panels on the far right are the clump mass distributions of the input and output with a least-squares fit to the power-law exponent indicated. Here the simulation consists of completely separated clumps (no blending), and no noise has been added to the data cube. Clumps below the detection limit $T = 2\Delta T$ are not detected, and masses, sizes, and line widths are underestimated because of truncation of the clump profile at $T = \Delta T$. (*b*) As in (*a*), but for a simulation with blending and noise ($T_{rms} = 0.25$ K) (see Figs. 4*a* and 4*b*). Noise tends to increase the measured peak temperatures above input and acts in the opposite sense of the effect of truncation by increasing masses, sizes, and line widths. The increased scatter in the size and mass relations is also due to the noise, but the number of missing and false clumps—crosses along each axis—increases because of blending.

tected yet are above the detection limit $T_{peak} = 2\Delta T = 1$ K. These are the missing clumps discussed in the previous subsection and result from close merging of two or more clumps. The noise in the map limits the contouring and therefore the temperature contrast of the map. Similarly, the resolution with which the cloud was observed also limits the ability to determine structure. Therefore, although these missing clumps represent errors in the analysis, they are unavoidable errors: the clumps are not resolved. However, for clumps that can be resolved, Figure 5*b* shows that clump properties are measured accurately despite (low-level) blending and noise.

There are five clumps along the *y*-axis which are found in the output but which do not match up with any in the input. Since $N_{axis} > 1$, these "false" clumps are not due to disconnected contour levels, nor are they likely to be noise effects, since $\Delta T/T_{rms} = 2$. Rather, they are the results of merging of more than one clump into one large apparent clump whose peak does not match with any of the original. As we discuss in the next section, blending inevitably causes loss of information about the input clump properties.

In summary, Clumpfield can detect and accurately measure individual clump properties in realistic simulations (i.e., including noise and blending). The optimum values of the user input parameters are $\Delta T = 2T_{rms}$ and $N_{axis} = 2$ or 3, for which only a listing of clumps with peak temperatures $T_{peak} \geq 2\Delta T = 4T_{rms}$ can be expected to be detected.

3.4. *Determination of the Mass Spectrum*

Besides measuring individual clump parameters, the collective properties of the clump ensemble are also of interest particularly for understanding the formation and evolution of clouds. One such quantity is the mass spectrum or distribution of numbers of clumps with mass. The way in which the mass of a cloud arranges itself should be a basic signature of its formation and evolution. All the previous analyses of clumpy structure in molecular clouds discussed in the Introduction have found strikingly similar distributions $dN/dM \sim M^{-1.5}$ (Blitz 1993). It is not understood why these distribtions should be the same from cloud to cloud.

The determination of the mass distribution is not straightforward: clumps must first be accurately identified and their masses measured over at least two orders of magnitude. The numbers are binned by mass and generally fitted to some functional form (most often a power law). Since statistical errors are larger for small numbers, it is also clearly an advantage to map enough of the cloud, and at sufficient sensitivity to be able to detect large numbers of clumps to better determine the mass distribution. We ran a simple test of randomly selecting N clumps with masses from 1 to 1000 $M_\odot$ from an infinite pool with mass distribution $dN/dM \sim M^{-1.5}$; we then binned by mass and made a least-squares fit to find the power-law exponent of the mass spectrum for this subsample. Figure 6 plots how the measured slope and formal least-squares error vary as a function of the size N of the subsample. We ran 100 of these "simulations" for each N, and plot the mean measured slope and the mean error in the slope determination as a function of N, with error bars indicating the dispersion over the 100 samples. Errors are very large (with very large swings from sample to sample) for $N < 50$ clumps, but even at large N the error remains large, and there appears to be an effective limit to the minimal formal least-squares error of ~ 0.2. The measured slope is systematically underestimated for all N, although the effect is small for $N > 100$. This is due to the small numbers of

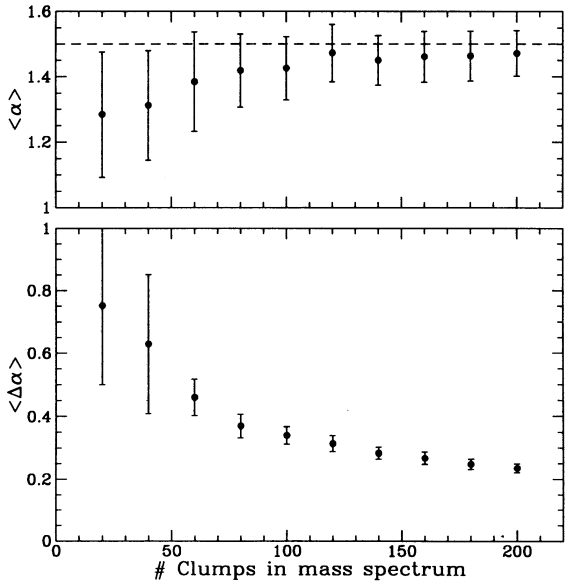

FIG. 6.—Results of a least-squares fit to a sample of N clumps drawn from an infinite pool with mass distribution $dN/dM \sim M^{-1.5}$. One hundred simulations are run for each value of N. The top panel graphs the averaged measured slope; the lower panel, the average formal least-squares error. Error bars in each panel plot the dispersion over the 100 simulations. The slope is systematically underestimated, and the error decreases very slowly with increasing N. Typically $\gtrsim 100$ clumps are needed for a good estimation of the mass distribution (given sufficient mass coverage).

clumps in the higher mass ranges which do not constrain the upper end of the mass distribution well, resulting in a fitted slope that tends to be flatter than the original. We conclude that, because of the need to sample the small numbers of high mass clumps adequately, typically ~ 100 clumps should be used in determining the mass distribution.

Aside from the statistics, there are three more possible errors that can affect this measurement: not measuring clump masses accurately, missing clumps because of blending, and including false clumps because of noise. For clumps that match up in the input and output lists Clumpfind measures the masses very well. There is a small percentage of false clumps that are found, but the main error comes from the clumps that are hidden in the blending. This problem worsens as the blending becomes more severe, and eventually clump deconvolution completely fails—blending irretrievably causes loss of information about the underlying distribution. Again, only by running simulations can we evaluate the effect of blending on the measurement of a mass spectrum.

We ran 50 simulations, similar to those described in § 3.3, using the same clump list (200 clumps with masses from 1 to 500 $M_\odot$ from a mass spectrum $dN/dM \sim M^{-1.5}$). Gaussian random noise, $\Delta T/T_{rms} = 0.25$ K, was added each time (Clumpfind parameters were fixed at $\Delta T = 0.5$ K, $N_{axis} = 3$), and the clump positions (and therefore the blending) were randomly changed. The average least-squares fit to the output mass spectrum was $\alpha_{out} = -1.45 \pm 0.08$ (the least-squares error in each individual fit was typically ~ 0.25), and the range varied from -1.26 to -1.58, showing that there can be very large swings depending on the degree of blending and the way the observations fit into the (fixed) mass bins. Inspection of the output distributions showed that although clumps with masses at the theoretical sensitivity limit $T_{peak} = 2\Delta T$ were detected, the distribution was not necessarily *complete* down to this

mass, i.e., not *all* the clumps at this mass were detected, because of noise effects on the peaks of these clumps, as discussed in the previous subsection.

Similar tests showed that the effect of additional noise, while maintaining $\Delta T/T_{\rm rms} = 2$, has no significant effect on the slope until ΔT becomes so large that contrast in the map decreases to the point where blending of clump emission becomes severe: merging of many clumps into one observed clump most affects the comparison of measured slope with input. The blending of clumps in a data cube will increase (intrinsically) if the number density of clumps is higher or (apparently) if the resolution is poorer. By running the following experiment, we find that the effect on the measured slope is the same. Two sets of simulations with the same input clump list, $dN/dM \sim M^{-1.5}$, as the earlier simulations were created: the first set was created in the exact same manner as previous simulations, but the data were smoothed to half the resolution in each axis before applying Clumpfind. The second set consisted of the same clumps, but forced into a data cube of half the size in each axis (doubling the number density of clumps). The data cube in both sets of simulations was therefore the same size and equal to half that of the previous simulations. Approximately 100 clumps were found in each data set. Power-law fits to the output mass distributions averaged over 10 simulations each were $\alpha_{\rm out} = -1.32 \pm 0.12$ for the first set and $\alpha_{\rm out} = -1.30 \pm 0.10$ for the second set. We conclude that resolution effects on the mass spectrum are indistinguishable from intrinsic blending effects. Comparisons of mass distributions between clouds must therefore be made at similar linear resolutions. However, even then they may still be affected by intrinsic differences in blending, and clump number densities (see the following section) should be calculated. In strongly blended regions, a higher density tracer may be more appropriate.

4. APPLICATION TO TWO MOLECULAR CLOUD DATA CUBES

The major advantage of an automated routine is that, both across an individual data cube and from one cube to another, the analysis is uniform (with the caveat about differences in the intrinsic blending of emission discussed in the last section). Hence comparisons between different regions of a data cube or between two different data cubes are far more reliable than they would be if done by eye (especially when comparing analyses done by two different observers). In this section we apply Clumpfind to real data by comparing two clouds apparently in different evolutionary states, concentrating on the application of the algorithm rather than giving a detailed scientific discussion, which we defer to a later paper (Williams & Blitz 1994).

All stars form in molecular clouds, and, conversely, all GMCs that have been studied in detail are associated with signs of active star formation. The one known exception is a large, cold cloud discovered in a CO survey by Maddalena & Thaddeus (1985) that lacks any clear optical signs of star formation. The Maddalena molecular cloud, or MMC, is also not obviously noticeable in the *IRAS* survey: its integrated luminosity is less 5000 $L_\odot$ at 100 μm (Blitz 1993 and references therein), corresponding to a star formation efficiency less than 0.1%. The preponderance of star-forming clouds in the solar neighborhood (Dame et al. 1986)—where even such cold clouds should be apparent—suggests that all large molecular clouds will eventually harbor star formation, and that the MMC is simply a young cloud in which star formation has not yet begun.

On the other hand, the Rosette molecular cloud, or RMC, is a more typical star-forming cloud surrounded by H II regions and supernova remnants and also containing strong embedded infrared sources (Blitz & Thaddeus 1980; Cox et al. 1991). Blitz & Stark (1986) mapped the RMC in ^{13}CO at a resolution of 1.'7, and recently Williams & Blitz (1993) mapped two regions of the MMC in ^{13}CO at 1', corresponding to the same linear resolution of 0.7 pc for adopted kinematic distances of 1600 and 2500 pc to the RMC and the MMC, respectively. The area mapped in each cloud is similar, and the two data sets are well suited for a comparative study using Clumpfind.

This comparison was in fact made by Williams & Blitz (1993), although only in order to measure the mass spectrum of each cloud. We repeat the analysis here to show the comparison of other clump quantities, and also as a recalculation of the mass spectra, previously made with an older, less well tested version of the clump algorithm will an effective value of $N_{\rm axis} = 1$, which we have since learned is prone to error.

For the analysis presented here, we used parameters $N_{\rm axis} = 3$ and $\Delta T/T_{\rm rms} = 2$, corresponding to $\Delta T = 0.5$ K for the RMC and $\Delta T = 0.25$ K for the MMC. Any clumps that were too close to the edges of the data cube were not used in the analysis, since a significant fraction of their emission is not mapped. The final clump lists consist of 83 clumps for the RMC and a total of 78 for the two observed regions in the MMC (see Tables 2, 3A, and 3B, respectively). Sizes, ΔR, and line widths, ΔV, have been calculated (and corrected for instrumental resolution) in the manner described in Appendix A. We assume that the ^{13}CO line is optically thin throughout, so that the mass, $M_{\rm LTE}$, is proportional to the integrated temperature. The size and line width combine to give the virial mass $M_{\rm vir}$ (Appendix A); for virialized clumps $M_{\rm LTE} \geq M_{\rm vir}$. Clump peak temperatures, sizes, line widths, and virial parameter $M_{\rm LTE}/M_{\rm vir}$ are plotted against $M_{\rm LTE}$ in Figure 7. Both clouds are shown on the same plot for comparison.

It is immediately apparent that there are simple relations between these independently measured clump properties. Peak temperature, size, line width, and virial parameter all vary as a positive power of the clump mass $M_{\rm LTE}$ (as noted by, e.g., Larson 1981 and Bertoldi & McKee 1991). The relations are remarkably similar for each cloud. Least-squares fits have been calculated separately for each cloud and have been drawn on the plots to further emphasize that, despite the differences in cloud star formation history, the clump ensembles in each cloud obey the same scaling laws.

The plots with the most scatter are the line-width–mass and virial parameter–mass relations. The second follows from the first, since $M_{\rm vir} \propto \Delta R\, \Delta V^2$. The scatter is large because the resolution in velocity (0.68 km s^{-1}) is poor: clumps are only a few pixels wide along the velocity axis, making it difficult to measure the dispersion accurately. Observations of the MMC were made simultaneously with 100 kHz filter banks corresponding to a velocity resolution of 0.26 km s^{-1}, but then the noise, and hence mass detectability ($M_{\rm LTE} \sim T_{\rm peak}^3$), would be higher. We estimate that only about half the number of clumps would then have been found, which would be inadequate for a reliable statistical analysis. A secondary factor is the small range in clump line width from the most massive clumps to the least massive. For the clumps that were observed in each cloud, the integrated temperature (proportional to mass) varies by more than two orders of magnitude, but the line width varies by only about a factor of 6.

In contrast to the similarities in the slope of the relations

TABLE 2

CLUMPS IN THE ROSETTE MOLECULAR CLOUD

Clump	l_{peak}	b_{peak}	v_{peak}	T_{peak}	ΔR[a]	Δv[b]	M_{LTE}[c]	M_{vir}[d]
1.....	207.000	−1.823	15.6	8.6	2.36	1.95	1183	1126
2.....	207.100	−1.848	10.8	6.3	2.82	1.73	1030	1055
3.....	207.550	−1.723	12.9	6.1	4.21	2.28	1878	2749
4.....	207.250	−1.823	12.9	6.0	3.06	2.30	1618	2033
5.....	207.275	−2.148	15.6	6.0	2.79	1.33	742	623
6.....	207.125	−1.898	12.9	5.5	1.90	1.64	753	644
7.....	207.075	−1.873	12.9	5.0	1.98	2.00	493	1003
8.....	206.850	−2.373	14.2	5.1	2.21	1.89	595	995
9.....	206.825	−1.998	16.3	5.2	1.94	2.03	592	1013
10.....	206.775	−1.948	15.6	5.1	1.33	1.50	296	378
11.....	207.400	−1.948	16.3	5.3	2.31	1.39	386	561
12.....	207.150	−1.798	12.2	4.5	1.50	1.50	318	423
13.....	206.750	−1.923	14.2	4.9	0.73	1.29	88	153
14.....	207.100	−1.873	16.3	4.5	1.95	1.47	327	533
15.....	207.775	−1.773	11.5	4.3	3.77	1.51	1546	1088
16.....	207.350	−1.898	12.2	4.1	2.96	1.85	793	1280
17.....	206.775	−1.773	12.9	4.0	1.40	2.07	237	757
18.....	206.875	−1.898	14.9	4.5	1.99	2.24	457	1255
19.....	207.700	−1.923	14.9	3.9	3.60	1.92	1397	1681
20.....	207.900	−1.798	10.8	3.1	1.93	1.18	251	338
21.....	207.350	−1.423	12.2	3.0	2.64	1.92	621	1227
22.....	206.925	−1.598	14.2	3.2	0.57	1.11	62	89
23.....	207.600	−1.898	15.6	3.1	1.52	1.89	148	685
24.....	207.325	−2.573	16.3	3.0	0.78	1.18	75	137
25.....	207.250	−2.523	16.3	3.3	1.24	0.96	126	143
26.....	207.100	−1.648	9.5	2.7	0.79	1.11	49	122
27.....	207.600	−1.948	11.5	2.9	1.85	1.58	254	581
28.....	207.675	−1.923	11.5	2.5	2.09	1.12	291	327
29.....	207.225	−1.573	11.5	2.8	2.23	2.48	335	1725
30.....	207.250	−1.423	12.9	2.7	1.48	2.67	210	1329
31.....	206.800	−2.523	13.6	2.8	1.42	0.96	142	167
32.....	207.650	−1.573	14.9	2.9	1.89	1.43	255	487
33.....	206.925	−1.648	14.9	2.9	1.31	1.56	120	402
34.....	207.450	−1.248	4.7	2.2	3.08	0.94	382	341
35.....	207.500	−2.048	10.2	2.2	3.01	2.54	592	2446
36.....	207.925	−2.573	11.5	2.0	0.92	1.51	65	262
37.....	206.775	−2.498	11.5	2.2	1.13	1.04	61	153
38.....	207.750	−2.273	12.2	2.3	0.90	0.82	47	76
39.....	206.900	−2.248	12.9	2.0	1.14	1.30	52	244
40.....	206.825	−2.073	14.2	2.1	1.49	2.08	85	815
41.....	207.375	−1.273	14.2	2.3	1.54	1.26	105	307
42.....	207.450	−1.273	13.6	2.3	0.71	0.90	30	73
43.....	207.800	−1.773	14.9	2.0	2.17	1.25	199	430
44.....	207.325	−2.298	16.3	2.1	0.72	1.06	39	102
45.....	207.250	−2.448	17.0	2.1	0.83	0.68	48	48
46.....	207.500	−1.773	5.4	1.6	1.09	1.74	34	418
47.....	207.475	−1.723	5.4	1.9	1.08	1.13	65	174
48.....	207.500	−1.273	5.4	1.8	0.74	0.68	30	43
49.....	207.275	−1.198	8.8	1.7	1.70	3.32	115	2355
50.....	207.275	−1.723	10.2	1.5	1.02	1.71	54	376
51.....	207.275	−1.223	10.2	1.8	0.74	1.10	52	113
52.....	207.300	−1.073	10.2	1.5	1.02	1.60	56	330
53.....	206.975	−2.498	11.5	1.6	1.88	1.20	145	342
54.....	207.500	−1.873	10.8	1.5	0.76	2.89	34	802
55.....	207.850	−1.598	10.8	1.9	2.25	2.56	191	1858
56.....	207.750	−2.198	11.5	1.7	1.09	1.91	54	498
57.....	208.225	−1.698	11.5	1.5	0.52	1.13	16	113
58.....	207.725	−1.523	12.2	1.9	2.38	3.96	193	4689
59.....	207.825	−2.073	12.9	1.7	0.46	0.68	18	27
61.....	207.425	−1.523	12.9	1.7	2.03	2.99	143	2288
62.....	207.500	−1.348	12.9	1.6	1.40	1.78	59	560
63.....	206.925	−2.423	14.2	1.5	0.76	1.10	36	116
64.....	207.825	−1.698	14.9	1.7	1.34	1.26	51	270
65.....	207.900	−2.048	15.6	1.9	1.55	1.04	94	210
66.....	207.575	−2.048	15.6	1.5	0.65	0.68	21	38
67.....	207.300	−1.923	16.3	1.6	0.52	0.68	18	30
68.....	206.950	−2.073	19.0	1.8	1.45	3.17	77	1833
69.....	207.500	−1.373	5.4	1.1	1.31	0.68	36	77
70.....	207.425	−1.373	6.8	1.4	1.71	1.41	88	430
72.....	207.325	−2.048	8.1	1.4	2.00	1.16	130	337
73.....	207.475	−1.873	6.8	1.3	1.11	1.12	36	176

TABLE 2—*Continued*

Clump	l_{peak}	b_{peak}	v_{peak}	T_{peak}	ΔR[a]	Δv[b]	M_{LTE}[c]	M_{vir}[d]
74.....	207.400	−1.548	7.4	1.0	1.43	2.29	53	941
75.....	207.300	−1.273	8.8	1.3	0.78	2.17	23	464
76.....	207.350	−1.673	9.5	1.5	0.76	2.14	24	441
77.....	208.150	−2.073	10.2	1.0	0.58	1.24	10	113
78.....	208.175	−1.973	10.2	1.3	1.57	0.68	49	92
80.....	207.075	−2.498	11.5	1.4	0.87	0.88	31	84
82.....	208.150	−1.698	10.8	1.2	0.56	2.21	14	342
83.....	207.475	−1.873	12.2	1.2	0.56	2.95	17	609
84.....	207.900	−1.573	12.2	1.1	0.58	1.55	13	174
85.....	207.300	−2.098	12.9	1.3	0.78	0.68	20	46
86.....	207.675	−2.073	12.9	1.1	0.69	0.93	13	76
87.....	208.100	−1.723	12.9	1.1	0.70	0.90	18	71
89.....	207.075	−2.423	14.9	1.1	0.70	0.68	13	41
90.....	207.450	−1.523	15.6	1.2	0.56	0.68	12	32
92.....	208.000	−1.573	15.6	1.0	1.13	2.26	27	724
94.....	206.950	−2.298	17.0	1.3	1.11	0.68	35	65
95.....	207.500	−1.798	17.6	1.2	0.88	1.16	13	151

[a] Equivalent circular radius in pc, corrected for beam size as described in Appendix A.
[b] FWHM in km s^{-1}, corrected for beam size as described in Appendix A.
[c] Clump masses in $M_\odot$, derived from the integrated ^{13}CO luminosity assuming LTE, $\tau_{13} < 0.5$, $N_{H_2}/N_{^{13}CO} = 4.8 \times 10^5$, and $T_{ex} = 20$ K. Includes an additional factor of 1.4 for helium.
[d] Virial mass in $M_\odot$, assuming an inverse-square power-law density profile (see Appendix A).

TABLE 3A

CLUMPS IN THE MADDALENA MOLECULAR CLOUD: REGION 1 ($\alpha = 6^h46^m50^s$, $\delta = -4°31'14''$)

Clump	$\Delta\alpha_{peak}$[a]	$\Delta\delta_{peak}$[b]	v_{peak}	T_{peak}	ΔR[c]	Δv[d]	M_{LTE}[e]	M_{vir}[f]
1.....	−5	−2	22.6	3.4	2.99	2.63	590	2593
2.....	−2	2	23.9	3.2	4.21	2.27	872	2739
3.....	−9	−4	23.3	2.8	4.74	3.51	1056	7347
4.....	−2	−1	22.6	2.7	3.58	3.01	610	4084
5.....	−11	2	22.6	2.2	2.97	2.64	302	2610
6.....	−17	−3	23.9	2.0	2.12	2.08	169	1153
7.....	−14	−5	23.9	1.8	1.21	1.85	53	523
8.....	−13	−3	23.9	1.8	1.52	1.86	91	663
9.....	−10	−8	21.9	1.5	1.74	2.32	71	1175
10.....	−17	−10	23.9	1.7	2.87	1.91	201	1320
11.....	−17	−7	23.9	1.6	2.35	2.70	136	2168
12.....	20	2	19.2	1.4	1.74	2.14	67	1004
13.....	20	4	19.2	1.3	1.93	2.81	52	1916
14.....	20	0	27.3	1.4	3.28	2.60	217	2799
15.....	32	2	28.0	1.3	1.89	1.89	75	844
16.....	27	6	18.5	1.2	2.22	2.44	82	1660
17.....	34	5	19.9	1.1	1.61	1.29	27	336
18.....	35	6	21.9	1.1	2.06	1.86	56	900
19.....	8	−4	23.3	1.2	1.49	1.53	37	440
20.....	2	8	23.9	1.2	2.06	2.16	110	1211
21.....	−23	3	24.6	1.0	1.45	1.77	40	572
22.....	6	7	25.3	1.0	1.81	2.34	76	1246
23.....	29	2	27.3	1.0	2.15	2.45	74	1628
24.....	26	9	19.9	0.8	1.21	3.98	20	2409
25.....	27	1	20.5	0.8	1.35	3.76	20	2395
26.....	28	3	21.2	0.9	0.96	1.76	15	374
27.....	−20	−1	23.3	0.9	1.99	2.07	67	1073
28.....	8	5	23.3	0.9	2.34	3.05	82	2735
29.....	−20	4	22.6	0.8	1.21	0.91	20	128
30.....	−24	−2	24.6	0.9	1.56	2.72	59	1455
31.....	−13	6	23.3	0.8	1.28	1.90	34	584
32.....	40	9	23.3	0.9	0.48	1.01	6	63
33.....	−23	7	23.9	0.9	2.23	2.53	107	1790
34.....	2	−5	24.6	0.9	1.27	1.77	28	502
35.....	3	−3	24.6	0.8	0.46	1.45	4	122
36.....	−23	5	24.6	0.9	0.97	1.59	17	307
37.....	35	−12	26.7	0.8	1.05	1.99	13	527
38.....	23	8	27.3	0.8	1.27	2.07	27	688
39.....	37	−10	27.3	0.8	2.00	3.12	43	2448
40.....	32	−2	27.3	0.9	1.20	2.13	17	687
41.....	26	4	28.0	0.9	1.67	1.48	57	458
42.....	−20	7	14.4	0.6	0.72	0.68	4	42

TABLE 3A—*Continued*

Clump	$\Delta\alpha_{peak}$[a]	$\Delta\delta_{peak}$[b]	v_{peak}	T_{peak}	ΔR[c]	Δv[d]	M_{LTE}[e]	M_{vir}[f]
43.....	25	−2	19.2	0.6	1.59	2.18	19	952
44.....	18	−1	20.5	0.6	1.97	3.26	32	2643
45.....	−14	9	21.2	0.7	1.29	3.99	23	2580
47.....	20	7	23.3	0.6	1.69	4.28	21	3891
48.....	2	−12	23.9	0.5	1.60	2.32	27	1083
50.....	16	9	23.9	0.7	0.81	1.40	7	199
51.....	16	7	24.6	0.5	0.84	3.01	10	956
52.....	6	−8	25.3	0.6	1.36	2.15	18	789
53.....	14	8	25.3	0.5	0.73	0.95	4	84
56.....	−17	−9	27.3	0.6	1.23	0.68	10	72
58.....	−25	−5	29.4	0.5	1.49	0.77	11	112

[a] Offset in arcminutes from $\alpha = 6^h46^m50^s$.
[b] Offset in arcminutes from $\delta = -4°31'14''$.
[c] Equivalent circular radius in pc, as in Table 2.
[d] FWHM in km s^{-1}, as in Table 2.
[e] Clump masses in $M_\odot$, as in Table 2, but assuming $T_{ex} = 10$ K.
[f] Virial mass in $M_\odot$, as in Table 2.

TABLE 3B

CLUMPS IN THE MADDALENA MOLECULAR CLOUD: REGION 2 ($\alpha = 6^h43^m39^s$, $\delta = -3°30'23''$)

Clump	$\Delta\alpha_{peak}$[a]	$\Delta\delta_{peak}$[b]	v_{peak}	T_{peak}	ΔR[c]	Δv[d]	M_{LTE}[e]	M_{vir}[f]
1.....	−6	1	22.6	1.6	3.25	3.15	236	4062
2.....	4	−8	24.6	1.7	3.78	2.57	356	3149
3.....	−15	−1	22.6	1.3	2.36	2.65	139	2086
4.....	−1	1	23.9	1.3	3.44	2.85	258	3506
5.....	4	−11	24.6	1.3	2.36	2.64	138	2072
6.....	4	−5	24.6	1.3	2.10	1.43	90	543
7.....	8	−4	25.3	1.5	3.85	3.39	408	5562
8.....	−1	−12	23.3	1.1	1.90	1.98	59	933
9.....	−4	5	24.6	1.1	2.18	1.93	78	1027
10.....	0	−8	25.3	1.1	3.13	3.47	219	4763
11.....	−10	−3	25.3	1.0	1.26	1.36	22	294
12.....	8	2	26.0	1.1	1.32	0.93	31	144
13.....	−6	−4	22.6	0.8	1.47	1.09	26	219
14.....	−7	−2	23.3	0.8	0.88	0.68	10	52
15.....	14	−12	23.9	0.9	1.51	1.66	37	526
16.....	−5	−12	23.9	0.9	1.51	1.94	28	715
17.....	3	−15	24.6	0.9	1.27	2.35	23	881
18.....	12	−1	24.6	0.8	2.31	2.68	43	2090
19.....	−6	−4	25.3	0.9	1.20	2.73	17	1127
20.....	7	4	25.3	0.8	1.57	1.76	30	616
21.....	11	1	26.0	0.8	1.14	1.21	14	209
22.....	0	−13	26.7	0.8	0.68	1.69	7	243
23.....	11	−7	20.5	0.6	1.54	2.03	16	794
24.....	15	−11	21.9	0.5	0.73	0.68	4	43
25.....	−13	−13	22.6	0.6	1.88	4.86	31	5604
26.....	8	0	22.6	0.6	1.37	2.42	13	1006
27.....	−15	−6	23.3	0.5	0.44	1.00	3	56
28.....	−10	4	24.6	0.6	0.82	1.90	10	371
29.....	−15	−9	24.6	0.7	1.53	2.99	28	1719
30.....	15	0	23.9	0.5	0.60	2.80	8	597
31.....	8	3	23.9	0.6	0.81	1.08	5	119
32.....	−12	4	23.9	0.6	0.83	1.72	6	309
34.....	−11	−9	25.3	0.5	1.10	2.01	14	556
36.....	6	6	26.0	0.6	0.41	1.35	4	93
37.....	−11	1	26.0	0.6	1.09	1.27	8	221
38.....	−9	3	26.0	0.5	0.73	0.68	4	42
41.....	−9	−4	27.3	0.5	0.73	0.83	3	64
42.....	−10	5	27.3	0.5	0.73	1.12	4	116
43.....	2	6	29.4	0.6	1.01	1.01	6	128

[a] Offset in arcminutes from $\alpha = 6^h43^m39^s$.
[b] Offset in arcminutes from $\delta = -3°30'23''$.
[c] Equivalent circular radius in pc, as in Table 2.
[d] FWHM in km s^{-1}, as in Table 2.
[e] Clump masses in $M_\odot$, as in Table 2, but assuming $T_{ex} = 10$ K.
[f] Virial mass in $M_\odot$, as in Table 2.

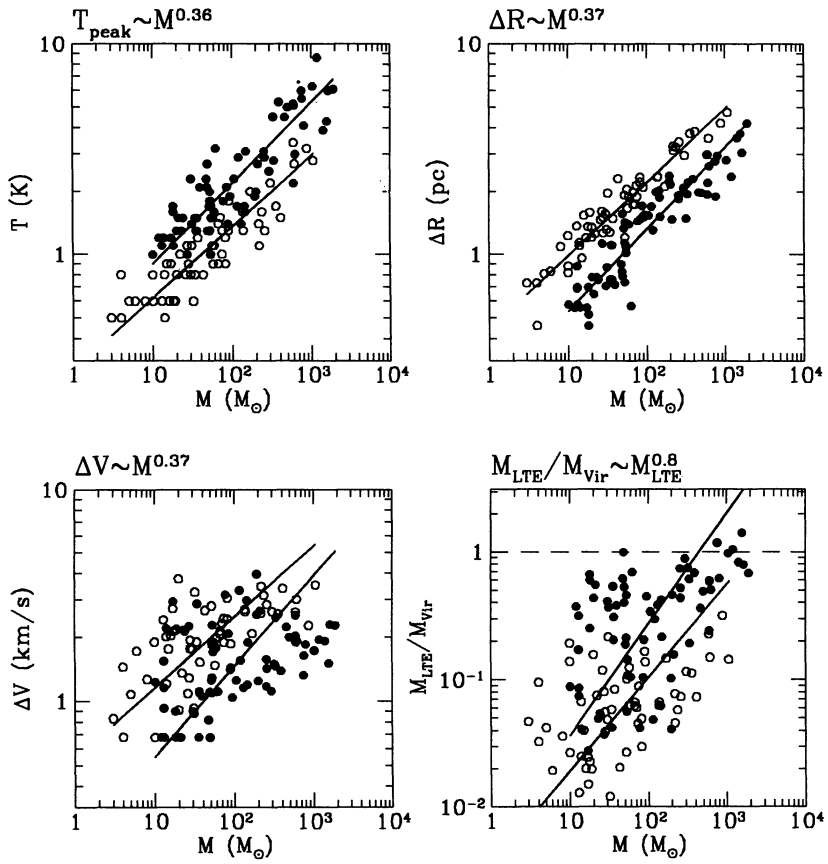

FIG. 7.—Relations between temperature, size, line width, and mass for clumps in the RMC (*filled circles*) and the MMC (*open circles*). A least-squares fit to the clumps in each cloud is indicated on each plot. The slope, equal to the power-law exponent, is also displayed at the top of each plot (averaged over the two clouds). The slopes, or forms of the relationships, between clump peak temperature, size, line width, and virial parameter are remarkably similar between the two clouds, but there are clear individual differences indicated by the offsets of one cloud from another. Clumps in the star-forming RMC are hotter, smaller (and hence denser), possess a smaller line width, are are more bound that clumps of the same mass in the MMC. These differences cannot all be simultaneously resolved by assuming that the distances to the clouds are in error.

between basic clump quantities and mass, there are clear differences in the individual clump properties between the two clouds (the offset of one cloud from another in each plot). Clumps in the star-forming RMC are hotter, smaller, possess a smaller line width, and are more bound than clumps of the same mass in the MMC.

Could any or all of these differences be attributed to errors in the assumed distances to the clouds? The distance to the RMC, d_{RMC}, is fairly well determined (see Blitz 1978 and references therein), but the distance to the MMC, d_{MMC}, is less well known, perhaps accurate only to 50%. The line-width difference cannot be explained in this way, since the measurement of the velocity dispersion is distance-independent. Differences in clump peak temperatures, sizes, and virial parameter could each conceivably be explained by distance errors, but not all simultaneously. The measured clump peak temperature is a resolution smeared average and therefore gives a lower limit to the intrinsic peak intensity. The assumed distances imply that the linear resolution of the observations is about the same for each cloud, therefore the difference in peak temperatures might be smaller if the actual linear resolution, were greater in the RMC than in the MMC, which corresponds to d_{RMC} being overestimated and/or d_{MMC} being underestimated. The clump size directly scales with the assumed distance to each cloud, so the observed difference in clump sizes could be explained by d_{RMC} being underestimated and/or d_{MMC} being overestimated. Finally, the virial parameter M_{LTE}/M_{vir} varies as (distance)^2/(distance), i.e., linearly with distance, which is the same scaling as clump size. However, the observed difference between the two clouds is in the opposite sense: to take away the difference in virial parameter between the two clouds requires d_{RMC} to be overestimated and/or d_{MMC} to be underestimated. Therefore, although each difference could, in principle, be due to errors in the distance determination to one or both of the clouds, they cannot *all* be explained simultaneously this way, and we conclude that there are real, fundamental differences in individual clump properties between the two clouds.

For clumps of the same mass, greater clump sizes in the MMC than in the RMC imply lower average densities. Observations of clumps in each cloud in higher transition CO lines and higher dipole moment molecules confirm this result independently (Williams & Blitz 1994). If the MMC is truly at an earlier evolutionary state than the RMC, then theories of clump and cloud formation and evolution must explain not only how these scaling relations arise but also how they are preserved as the clumps become denser, hotter, more bound, and ultimately star-forming.

Using the output (clump assignments) data cube, we plot a

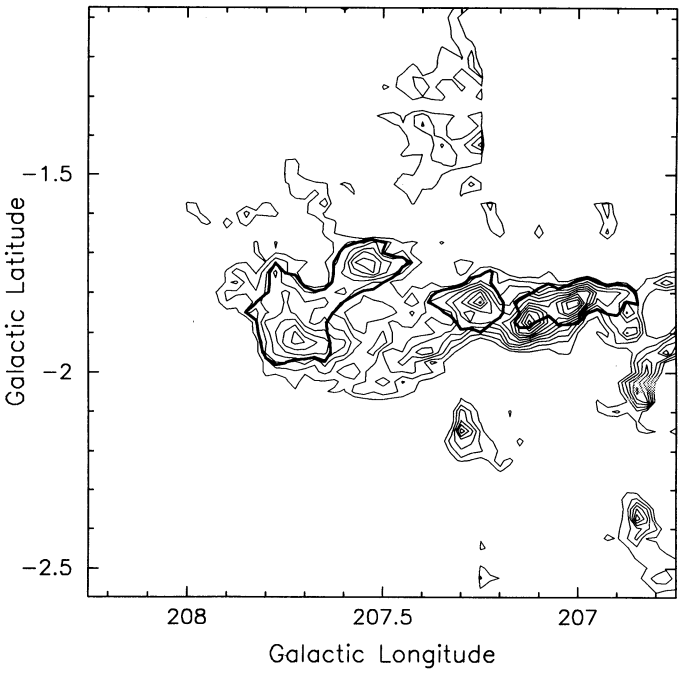

FIG. 8.—Outline of the five most massive clumps in the RMC ($M > 1000\ M_\odot$) against an integrated velocity map of the entire data cube. The velocity range is $v = 0$–20 km s^{-1}, and contours are at intervals of 2 K km s^{-1} beginning at 4 K km s^{-1}. Their proximity to the midplane of the cloud along a line of constant Galactic latitude is evidence of dynamical evolution.

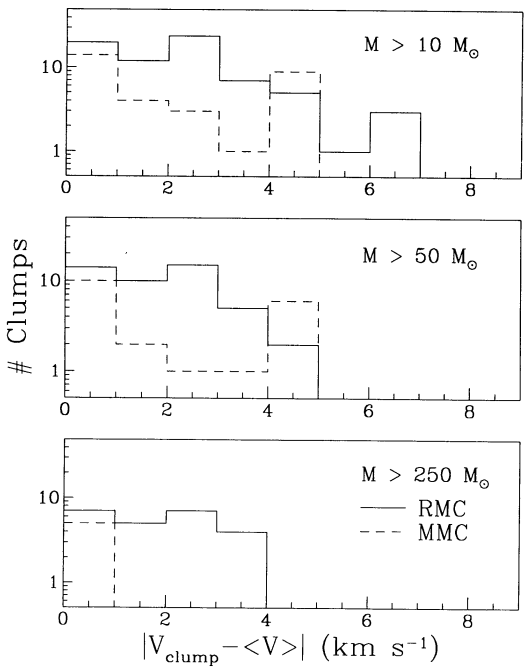

FIG. 9.—Histograms of clump centroid velocity about the mean cloud velocity for different mass ranges. The solid line is the distribution for clumps in the RMC, and the dashed line is for the MMC. The dispersion about the mean decreases with increasing mass in both clouds. However, although the dispersion is very similar for the small to intermediate mass clumps, $M = 10$–$250\ M_\odot$ (*top two panels*), in the MMC, it decreases more uniformly with mass in the RMC—a possible dynamical indication of evolutionary differences.

single contour at the lowest level for the six most massive clumps ($M > 1000\ M_\odot$) over an integrated velocity map of the RMC (Fig. 8). That they all lie along the midplane of the cloud is evidence that dynamical evolution of the system of clumps has taken place (Blitz 1987, 1993). A similar plot for the MMC is not as useful, and may in fact be misleading because only a small fraction of the entire cloud was mapped and the spatial sampling is insufficient for such a comparison.

Although the observed clump velocities in the MMC do not cover the entire cloud velocity range, an analysis of the velocity distribution has more validity than an equivalent positional analysis because the range about the mean velocity for the observed ensemble is reasonably large. In Figure 9 we plot histograms of the velocity distribution about the mean of the clump ensemble for the RMC observations and for the larger of the two mapped regions in the MMC. It is clear that, in each cloud, the more massive clumps are more clustered about the mean than the less massive clumps. However, there is little difference in the distribution of small-mass clumps, $M = 10$–$50\ M_\odot$, and intermediate masses, $M = 50$–$250\ M_\odot$, in the MMC, although the dispersion about the mean decreases uniformly for the same mass ranges in the RMC. The dispersion about the mean strongly decreases at high masses, $M > 250\ M_\odot$, in the MMC, and it also decreases in the RMC, but there may be a large uncertainty because of the small number of clumps in the MMC with these masses. The difference between the two clouds at small to intermediate clump masses may be a dynamical indication of the evolutionary difference between the two.

As long as a representative region of the cloud is mapped, complete spatial and velocity coverage is not necessary to determine the mass distribution of clumps, and a comparison of the RMC and the MMC is valid (Fig. 10). Logarithmic bins

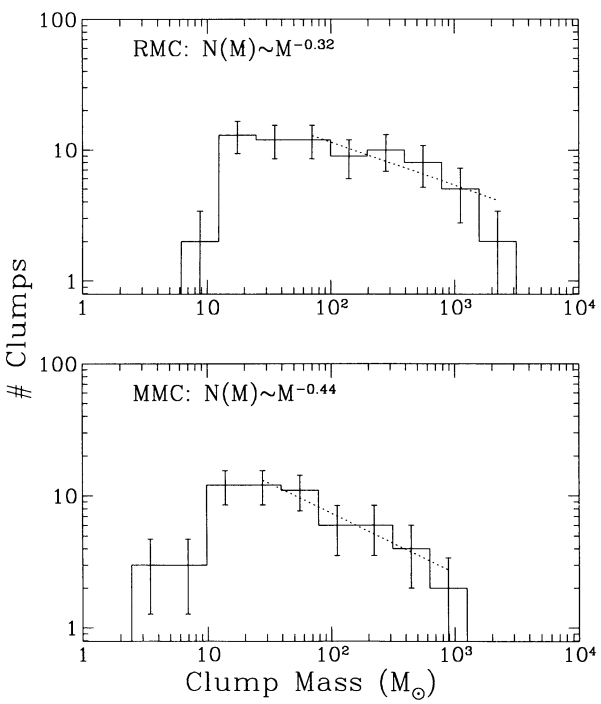

FIG. 10.—Mass spectra of, and least-squares power-law fits to, the RMC and the MMC. The error bars are equal to the square root of the number of clumps in each mass bin and are used as weights in the fitting process. The power law is shown only over those bins used in the fit.

of size $\log_{10} 2$ are chosen, ending at the highest mass clump in each cloud. $N(M)$ is the number of clumps within the bin at mass M and decreases with increasing mass for each cloud. A least-squares fit to each implies power laws, which seem to be a reasonable description of the data. We choose to plot $N(M)$ versus M rather than the more commonly quoted exponent of dN/dM versus M, because it makes differences in slope easier to see. Since $dN/dM \sim N(M)/M$, the exponent of dN/dM versus M is 1 less than for $N(M)$ versus M. We find that $N(M) \sim M^{-0.44}$ for the MMC and $N(M) \sim M^{-0.32}$ for the RMC. We note that both are close to the value of -0.5 (-1.5 for dN/dM) found for other clouds.

The temperature-sensitivity limit is $T_{\rm peak} = 2\,\Delta T = 0.5$ K for the MMC and 1 K for the RMC. Figure 7 shows that this corresponds roughly to clumps of masses $\sim 3\ M_\odot$ (MCC) and $\sim 10\ M_\odot$ (RMC. The mass spectra are therefore probably complete at a mass $\gtrsim 10\ M_\odot$ (MMC) and $\gtrsim 30\ M_\odot$ (RMC), so the power-law fits only include bins with masses greater than this. The surveyed areas of the two clouds are very similar: 1100 pc^2 for the MMC and 900 pc^2 for the RMC. The cloud extent along the line of sight is unknown, but since the projected cloud size of the MMC is larger than that of the RMC, it might be expected that the extent is greater for the former, so that the actual volume of the MMC data cube is probably a little larger than that of the RMC data cube. Approximately equal numbers of clumps were found in both data sets, so the clump number density is about the same between the two clouds—perhaps somewhat smaller in the MMC, but the difference is small enough that the mass spectrum comparison is probably not biased by blending differences.

The similar values of the exponents, both similar to the seemingly ubiquitous -1.5, disagree with the conclusions of Williams & Blitz (1993), which used an older, less well tested version of the same clump-finding method. The old method used a different criterion for connecting pixels, without the $N_{\rm axis}$ parameter as we have defined it here. However, a neighborhood of a pixel in the old method is the same as in the new with an effective $N_{\rm axis} = 1$. As we showed in § 3.2, setting $N_{\rm axis} = 1$ creates large numbers of false clumps because the contours cannot be followed adequately. The present analysis, which has been comprehensively tested with simulations such as those we have presented here, shows no evidence for a difference in the mass distribution between the two clouds. We conclude that, if the two clouds are at different evolutionary states and representative of each state, then it appears that the mass spectrum of a GMC is determined at its formation and does not significantly evolve over the lifetime of a GMC. In addition, it appears that other collective properties of the clouds, such as mass–temperature–size–line-width relations, are also very similar. On the other hand, the largest difference between the two clouds appears to be the degree of virialization of the clumps. This deviation from gravitational boundedness may, in fact, turn out to be a useful indicator of the evolutionary status of a GMC.

5. CONCLUSIONS

We have developed an automatic procedure for finding clumps or discrete density enhancements in three-dimensional (primarily molecular spectral line) data cubes. The algorithm, Clumpfind, operates on the data in much the same way as the eye would analyze a set of maps, but with the advantage of visualizing the full position-position-velocity nature of the data at once in a uniform manner from region to region across any one data cube and also between different data cubes. The algorithm's performance has been evaluated by creating simulated data sets consisting of a family of model clumps to which noise is added and which are either allowed to blend together or not.

Clumpfind works by first contouring the data. Noise can look like cloud structure if contour increments are too small. Simple consideration of the numbers of apparent clumps shows that contour levels must be at least equal to twice the rms noise level to depict the cloud structure accurately. This condition imposes a certain maximum contrast on the map which determines the minimum detectable clump mass and the ability to distinguish closely merged clumps. This is analogous to the beam size of the telescope determining the minimum size scale of structure that can be observed.

Clumpfind deals with blended emission (or shared contours) by breaking them up and assigning them to contributing clumps using a simple "friends-of-friends" method. That this method is adequate for the case of relatively mild blending, when clump peaks can be separately resolved, is shown both theoretically in Appendix B and practically by the results of the simulations. The algorithm inevitably fails when blending is so strong that clumps cannot be resolved as distinct. The effect of such severe blending, whether caused by inadequate resolution or low signal-to-noise ratio, or intrinsic to the cloud structure, is to count many clumps as one large object, flattening the measured mass spectrum.

As an example of its use, Clumpfind has been applied to two real data sets. The information contained in each clump deconvolution, and the comparison between the two, demonstrate the power of this type of structure analysis. The collection of large data cubes is becoming more common with the advent of faster telescope systems: we therefore expect this routine, or similar versions, to become an important part of molecular cloud studies. Clumpfind is available as a task in the MIRIAD framework, the data analysis package for the Berkeley-Illinois-Maryland Millimeter Array (BIMA) at Hat Creek.

We would like to thank the referee, J. Stutzki, for his careful reading of the manuscript and insightful comments regarding this work. This work was partially supported by NSF grant AST 89-18912. E. J. D. also acknowledges financial support from NASA grant NAG 5-1736. The simulations were run on the computing facility NCSA C3880, with support from the NSF under grant ASC 9217384. The observations of the MMC were made at the NRAO 12 m telescope, operated by Associated Universities, Inc., under cooperative agreement with the NSF.

APPENDIX A

DETERMINATION OF CLUMP SIZES AND MASSES

Clumpfind finds clumps in the data set, but does not itself determine their properties. Other programs read in the two data cubes—the raw data set of temperature as a function of position and Clumpfind output clump assignment cube—to determine the position and temperature of each pixel of each clump and then calculate the desired clump statistics. Fundamental quantities are the peak temperature, peak position, and size. The size is calculated as

$$\sigma_x = \left[\frac{\sum x^2 t}{\sum t} - \left(\frac{\sum xt}{\sum t}\right)^2\right]^{1/2}, \tag{A1}$$

where x can be any coordinate, either position or velocity; $t = t(x)$ is the temperature at this position; and the summation is over all pixels assigned to the clump. For an assumed Gaussian profile—most appropriate for the velocity axis—this translates into a diameter at half-peak temperature equal to 2.355 σ_x.

An alternative definition of clump size is the effective circular radius,

$$\Delta R = (A/\pi)^{1/2}, \tag{A2}$$

where A is the projected area of the clump on the sky. This is a better estimator to use in calculating the virial mass because it takes into account the full clump extent, which is precisely what is needed for the calculation of the potential energy term,

$$W = -\frac{3}{5}\alpha_{\rm vir}\frac{GM^2}{\Delta R}. \tag{A3}$$

The parameter $\alpha_{\rm vir}$ allows for the effect of nonuniform densities in the calculation of the gravitational potential energy; $\alpha_{\rm vir}$ is unity for a uniform density profile and 5/3 for an inverse-square profile (the value used in § 4). On the other hand, the kinetic energy $T = (3/2)M\sigma_v^2$ requires the one-dimensional velocity dispersion σ_v rather than a measure of the maximum velocity range of the clump. The virial condition $2T + W = 0$ then determines the virial mass,

$$M_{\rm vir} = 5\Delta R\sigma_v^2/\alpha_{\rm vir} G. \tag{A4}$$

The other basic clump statistic is its mass, or integrated intensity:

$$M \propto (\sum t)\Delta x^2 \Delta v, \tag{A5}$$

where the summation is again over all pixels in the clump, Δx is the pixel size (grid spacing of the map), and Δv is 1 spectrometer channel velocity width. The constant of proportionality depends on the abundance of the molecular tracer, line optical depth, excitation temperature, etc. (e.g., Martin & Barrett 1978).

Since there is a spectrum of clumps, with increasing numbers at small masses and sizes, there will be many clumps at or near the resolution limit. It is therefore important that dispersions be corrected for the beam size ($\sigma_{\rm obs}$ = beam FWHM/2.355) and velocity resolution, $\sigma_{\rm obs} = \Delta v$(spectrometer), of the observations,

$$\sigma' = (\sigma^2 - \sigma_{\rm obs}^2)^{1/2}. \tag{A6}$$

Similarly, the circular radius should be corrected for the (Gaussian) beam size of a point source with the same $T_{\rm peak}$ as the clump, computed at the farthest extent at which the projected area is measured, $T = \Delta T$:

$$\Delta R' = \left[\Delta R^2 - \left\{\frac{b}{2.355}\left[2\ln\left(\frac{T_{\rm peak}}{\Delta T}\right)\right]^{1/2}\right\}^2\right]^{1/2}, \tag{A7}$$

where b is the beam FWHM is the same units as ΔR. These corrected forms for size (equivalent circular radius for position, and 2.355 times the dispersion for line width) were used in § 4. We also note that the measured peak intensity of a clump is a resolution smeared average and therefore a lower limit to the intrinsic peak. However, integral quantities, such as clump mass, are independent of resolution.

Other quantities, such as clump pressure, density, etc., are generally a combination of sizes, masses, and line widths. More complicated statistics, e.g., higher order clump moments, may be simply calculated by integrating over the clump profile in the same manner.

APPENDIX B

MINIMUM SEPARATION FOR RESOLUTION OF TWO CLUMPS

In this appendix we calculate the criterion for the resolution of two clumps. Gaussian profiles are assumed, and we work only in one dimension, since the axes of the data cube can be considered independently. In a real data cube two clumps would be unresolved if the criterion were not met in all three axes simultaneously.

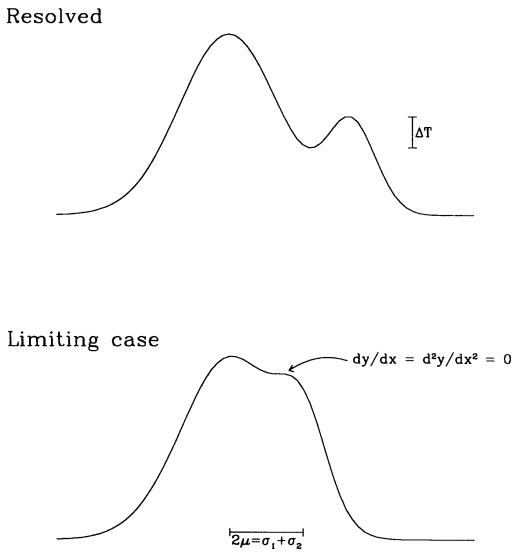

FIG. 11.—Two Gaussians separated by $x = 2\mu$. *Top:* resolved; *bottom:* limiting criterion for resolution.

We start with the case of two identical Gaussians separated by $x = 2\mu$,

$$y(x) = T_{\rm peak}\, e^{-x^2/2\sigma^2} + T_{\rm peak}\, e^{-(x-2\mu)^2/2\sigma^2} . \tag{B1}$$

Widely spaced, the two Gaussians are clearly distinguishable, but as the separation 2μ decreases, the two peaks will begin to merge into one. At this point the clumps are no longer resolvable. To be resolved into two clumps implies that there are two maxima, i.e., there is a minimum between the two (Fig. 11). Since the Gaussians are identical, this minimum must occur at the midpoint between the two, and therefore the criterion for resolution is

$$\left.\frac{d^2y}{dx^2}\right|_{x=\mu} > 0 . \tag{B2}$$

This reduces to

$$\mu > \sigma , \tag{B3}$$

that is, the separation between the two must be greater than twice their standard deviation, or almost equal to their FWHM. This is a basic limit, independent of resolution or signal-to-noise limitations. For the observational case of finite signal-to-noise, the situation is even worse: the minimum between the two clumps must be sufficiently strong so as not to be confused with random noise fluctuations. For instance, Clumpfind works by contouring the data and searches for isolated peaks. For clumps to be resolved, minima must be at a lower contour level, so the condition for resolution is

$$y(x = 0) - y(x = \mu) > \Delta T , \tag{B4}$$

where ΔT is the contour interval optimally set at twice the rms noise (see § 3.2). For typical large-scale maps of molecular clouds, the ratio of clump peak temperature to noise ranges from $S \sim 2$–3 for small clumps near the detection limit to perhaps $S \gtrsim 20$ for the most massive clumps. Then $\Delta T = 2T_{\rm rms} = 2T_{\rm peak}/S$, and the above criterion becomes

$$e^{-\mu^2/2\sigma^2} < \frac{1}{2} - \frac{1}{S} , \tag{B5}$$

ignoring a small fourth-order term from the contribution of the second Gaussian to the peak of the first. Therefore, resolution requires a separation between the two clumps of

$$2\mu > 2\sigma\left[2\ln\left(\frac{2S}{S-2}\right)\right]^{1/2} , \tag{B6}$$

which is 3.1 σ, 2.7 σ, and 2.5 σ for $S = 5$, 10, and 20, respectively. Clumpfind would assign the emission out to the midpoint ($x = \mu$) to each clump, and in doing so would miss, respectively, 6%, 9%, and 11% of the flux of each clump. Thus, the errors introduced by such a rudimentary treatment of the shared contours is small.

The situation becomes more serious if several clumps are blended together, because then up to ~10% of the flux of the central clump may be blended with each clump that it is blended with. For maps with good signal-to-noise this may amount to more than 50% in the worst cases. A more sophisticated treatment is then necessary.

What about two clumps of different sizes and peak temperatures? It is possible to imagine the case where a very narrow spike could in principle be resolved arbitrarily close to the center of a Gaussian. However, lower mass clumps have lower peak

temperatures along with their smaller sizes, in fact the scalings with mass are very similar (Fig. 7), so to a first approximation $T_{\rm peak} \propto \sigma$. Then

$$y(x) \propto \sigma_1 e^{-x^2/2\sigma_1^2} + \sigma_2 e^{-(x-2\mu)^2/2\sigma_2^2} . \tag{B7}$$

The limiting case between resolvability and nonresolvability is when there is an inflection point (Fig. 11). That is, for some x_0, $0 < x_0 < 2\mu$,

$$\frac{dy}{dx} = \frac{d^2y}{dx^2} = 0 \qquad \text{at} \qquad x = x_0 . \tag{B8}$$

One can quickly verify that the solution is $x_0 = \sigma_1$ and $2\mu = \sigma_1 + \sigma_2$. Again, this is a basic limit independent of resolution or temperature-sensitivity considerations. Such observational considerations will increase the separation required to resolve the clumps, implying that the contribution of one clump to another is smaller still, as we calculated analytically for the case of two equal clumps above. By generalizing that result, therefore, we conclude that splitting up the merged contours, in the manner of the friends-of-friends algorithm, makes relatively minor errors for the case of moderate blending.

REFERENCES

Adler, D. S., & Roberts, W. W. 1992, ApJ, 384, 95
Bally, J., Langer, W. D., Stark, A. A., & Wilson, R. W. 1987, ApJ, 312, L45
Bertoldi, F., & McKee, C. F. 1991, ApJ, 395, 140
Blitz, L. 1978, Ph.D. thesis, Columbia Univ.
———. 1987, in Physical Processes in Interstellar Clouds, ed. G. E. Morfill & M. Scholer (Dordrecht: Kluwer), 35
———. 1993, in Protostars and Planets III, ed. E. H. Levy & J. I. Lunine (Tucson: Univ. Arizona Press), 125
Blitz, L., & Shu, F. H. 1980, ApJ, 238, 148
Blitz, L., & Stark, A. A. 1986, ApJ, 300, L89
Blitz, L., & Thaddeus, P. 1980, ApJ, 241, 676
Carr, J. S. 1987, ApJ, 323, 170
Cox, P., Deharveng, L., & Leene, A. 1991, A&A, 230, 171
Dame, T. M. 1993, in Back to the Galaxy, ed. S. S. Holt & F. Verter (New York: AIP), 267
Dame, T. M., Elmegreen, B. G., Cohen, R. S., & Thaddeus, P. 1986, ApJ, 305, 892
Dickman, R. L. 1978, AJ, 83, 363
Herbertz, R., Ungerechts, H., & Winnewisser, G. 1991, A&A, 249, 483
Lada, E. A., Bally, J., & Stark, A. A. 1991, ApJ, 368, 432
Larson, R. 1981, MNRAS, 194, 809
Loren, R. B. 1989, ApJ, 338, 902
Maddalena, R., & Thaddeus, P. 1985, ApJ, 294, 231
Magnani, L., Blitz, L., & Mundy, L. 1985, ApJ, 295, 402
Martin, R. N., & Barrett, A. H. 1978, ApJS, 36, 1
Myers, P. C., Dame, T. M., Thaddeus, P., Cohen, R. S., Silverberg, R. F., Dwek, E., & Hauser, M. G. 1986, ApJ, 301, 398
Myers, P. C., & Goodman, A. A. 1988, ApJ, 329, 392
Stutzki, J., & Güsten, R. 1990, ApJ, 356, 513
Stutzki, J., Stacey, G. J., Genzel, R., Harris, A. I., Jaffe, D. T., & Lugten, J. B. 1988, ApJ, 379, 399
Turner, B. E., & Ziurys, L. M. 1988, in Galactic and Extragalactic Radio Astronomy, ed. G. L. Verschuur & K. I. Kellermann (Berlin: Springer), 200
Williams, J. P., & Blitz, L. 1993, ApJ, 405, L75
———. 1994, in preparation