HERSCHEL–SPIRE IMAGING SPECTROSCOPY OF MOLECULAR GAS IN M82

J. Kamenetzky1, J. Glenn1, N. Rangwala1, P. Maloney1, M. Bradford2, C. D. Wilson3, G. J. Bendo4, M. Baes5, A. Boselli6, A. Cooray7, K. G. Isaac8, V. Lebouteiller9, S. Madden9, P. Panuzzo9, M. R. P. Schirm1, L. Spinoglio10, and R. Wu9

1 Center for Astrophysics and Space Astronomy, 389-UCB, University of Colorado, Boulder, CO 80303, USA
2 NASA Jet Propulsion Laboratory, Pasadena, CA 91109, USA
3 Department of Physics & Astronomy, McMaster University, Hamilton, Ontario L8S 4M1, Canada
4 UK ALMA Regional Centre Node, Jordell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK
5 Sterrenkundig Observatorium, Universiteit Gent, Krijgslaan 281 S9, 9000 Gent, Belgium
6 Laboratoire d’Astrophysique de Marseille, UM6161 CNRS, 38 rue F. Joliot-Curie, 13388 Marseille, France
7 Department of Physics & Astronomy, University of California, Irvine, CA 92697, USA
8 ESA Astrophysics Missions Division, ESTEC, P.O. Box 299, 2200 AG Noordwijk, The Netherlands
9 CEA, Laboratoire AIM, Irfu/SAp, Orme des Merisiers, 91191 Gif-sur-Yvette, France
10 Istituto di Fisica dello Spazio Interplanetario, INAF, Via del Fosso del Cavaliere 100, 00133 Roma, Italy

Received 2012 January 13; accepted 2012 April 30; published 2012 June 14

ABSTRACT

We present new Herschel–SPIRE imaging spectroscopy (194–671 μm) of the bright starburst galaxy M82. Covering the CO ladder from J = 4 → 3 to J = 13 → 12, spectra were obtained at multiple positions for a fully sampled ∼3 × 3 arcmin map, including a longer exposure at the central position. We present measurements of 12CO, 13CO, [C i], [N ii], HCN, and HCO+ in emission, along with OH*, H2O*, and HF in absorption and H2O in both emission and absorption, with discussion. We use a radiative transfer code and Bayesian likelihood analysis to model the temperature, density, column density, and filling factor of multiple components of molecular gas traced by 12CO and 13CO, adding further evidence to the high-J lines tracing a much warmer (∼500 K), less massive component than the low-J lines. The addition of 13CO (and [C i]) is new and indicates that [C i] may be tracing different gas than 12CO. No temperature/density gradients can be inferred from the map, indicating that the single-pointing spectrum is descriptive of the bulk properties of the galaxy. At such a high temperature, cooling is dominated by molecular hydrogen. Photon-dominated region (PDR) models require higher densities than those indicated by our Bayesian likelihood analysis in order to explain the high-J CO line ratios, though cosmic-ray-enhanced PDR models can do a better job reproducing the emission at lower densities. Shocks and turbulent heating are likely required to explain the bright high-J emission.

Key words: galaxies: individual (M82) – galaxies: starburst – ISM: molecules

Online-only material: color figures, extended figure

1. INTRODUCTION

M82 is a nearly edge-on galaxy, notable for its spectacular bipolar outflow and high IR luminosity (5.6 × 10^10 L_☉; Sanders et al. 2003). Though its high inclination angle of 77° makes it difficult to determine, M82 is likely an SBc barred spiral galaxy with two trailing arms (Mayya et al. 2005). Its redshift-independent distance is about 3.4 Mpc (Dalcanton et al. 2009), and after correcting the commonly cited redshift (0.000677, de Vaucouleurs et al. 1991) with WMAP-7 parameters to the 3K cosmic microwave background (CMB) reference frame, we find a redshift of 0.000939. Given this distance, we assume a conversion of 17 pc/".

Due to its proximity, M82 is an exceptionally well studied starburst galaxy. High star formation rates (9.8 M_☉ yr^{-1}, likely enhanced by interactions with M81; Yun et al. 1993) and a large gas reservoir produce bright molecular and atomic emission lines. Such lines can yield important information on the interaction between the interstellar medium (ISM) and star formation (SF) processes, such as the influence of SF on the ISM through the photon-dominated region (PDR) or other excitation, as traced by intermediate to high-J CO rotational lines.

Ground-based studies of M82 are numerous (Wild et al. 1992; Mao et al. 2000; Weiß et al. 2001, 2005; Ward et al. 2003), examining both morphology and physical conditions of the gas. High-resolution CO maps of the 1 kpc disk indicate that the emission is largely concentrated in three areas: a northeast lobe, southwest lobe, and to a lesser extent, a central region (see Figure 1 of Weiß et al. 2001). The two lobes are separated dynamically, as can be seen in position–velocity diagrams (Figure 3 of Weiß et al. 2001). Outside of the disk, molecular gas emission is also detected in the halo/outflow (Taylor et al. 2001; Walter et al. 2002).

In addition to examining the morphology of molecular gas, CO emission lines can be used to determine the physical conditions of the molecular gas in galaxies. Previously, due to terrestrial atmospheric opacity, only the first few lines in the CO emission ladder could be studied. The first studies of higher-J lines (described below) have indicated that they can trace components of gas separate from those measurable with low-J lines. Many of the most interesting questions about galaxy formation, evolution, and SF concern the balance of different energy sources, i.e., what role might cosmic rays, ultraviolet light from stars, X-rays from powerful active galactic nuclei (AGNs), or turbulent motion play in the SF history of various galaxies? In what way does SF influence the molecular gas and vice versa? Estimating the influence of these various energy sources, however, often depends on knowing the physical
conditions of the gas. We therefore model physical conditions of these high-J lines first in order to inform our later discussion on energy sources. Other molecules are also useful in this study; in a ground-based survey of 18 different molecular species, Aladro et al. (2011) also found that some molecules trace different temperature components than others and that the different chemical abundances in M82 and NGC 253 may indicate different evolutionary stages of starbursts.

The Herschel Space Observatory (Pilbratt et al. 2010) is the unique facility that can measure the submillimeter properties of nearby galaxies in a frequency range that cannot be observed from the ground. As one of the brightest extragalactic submillimeter sources, M82 has been studied extensively with Herschel. For example, the imaging photometer of the Spectral and Photometric Imaging Receiver (SPIRE; Griffin et al. 2010) has been used to study the cool dust of M82, revealing wind/halo temperatures that decrease with distance from the center with warmer starburst-like filaments between dust spurs (Roussel et al. 2010). The tidal interaction with M81 was likely very effective in removing cold interstellar dust from the disk; more than two-thirds of the extraplanar dust follows the tidal streams. Panuzzo et al. (2010) used the SPIRE Fourier-Transform Spectrometer (FTS) to study a single spectrum of the 12CO emission from $J = 4 \rightarrow 3$ to $J = 13 \rightarrow 12$ to find that these higher-$J$ CO lines likely trace a $\sim 500$ K gas component not seen in the $\sim 30$ K component that can be observed from ground-based studies. Also on board Herschel is the Heterodyne Instrument for the Far Infrared (HIFI; de Graauw et al. 2010), which consists of a set of 7 heterodyne receivers with resolutions of 125 kHz to 1 MHz for electronically tunable frequency coverage of $2 \times 4$ GHz; it covers 480–1910 GHz. HIFI found ionized water absorption from diffuse gas (Weiβ et al. 2010) and high-$J$ transitions of the CO ladder. These CO transitions indicated a combination of one low and two high density gas components via comparison to PDR models (Lommen et al. 2010).

We confirm the presence of multiple molecular hydrogen thermal components in M82 by performing a more in-depth modeling analysis on a deeper data set as part of the Herschel Very Nearby Galaxies Survey. We add to existing data by using the SPIRE FTS mapping mode, providing spectroscopic imaging of a region approximately $3' \times 3'$, which helps us confirm our source–beam coupling corrections. We also present a deep pointed spectrum (64 scans versus 10 scans in Panuzzo et al. 2010) in order to detect fainter lines, such as 13CO, H2O, OH*, HF, and more.

We add depth to the analysis by modeling both the cool and warm components of molecular gas, and simultaneously accounting for 12CO, 13CO, and [C i]. We also use [C i] emission as a separate estimate of the total hydrogen mass and other absorption lines for column density estimates. We first analyze the CO excitation using likelihood analysis to determine the physical conditions and then compare to possible energy sources. Our likelihoods test the uniqueness and uncertainty in the conditions, as has also been done in Naylors et al. (2010), Kamenetzky et al. (2011), Scott et al. (2011), Bradford et al. (2009), Panuzzo et al. (2010), and Rangwala et al. (2011).

Our observations are described in Section 2. We describe the Bayesian likelihood analysis used to find the best physical properties of the molecular gas in Section 3 and present the results in Sections 4.1 and 4.2. In the remainder of Section 4, we discuss absorption results that are new to this study, possible excitation mechanisms of the warm gas, and comparisons to other galaxies. Conclusions are presented in Section 5.

2. OBSERVATIONS WITH SPIRE

2.1. The SPIRE Spectrometer

The SPIRE instrument (Griffin et al. 2010) is on board the Herschel Space Observatory (Pilbratt et al. 2010). It consists of a three-band imaging photometer (at 250, 350, and 500 μm) and an imaging FTS. We are presenting observations from the FTS, which operates in the range of 194–671 μm (447–1550 GHz). The bandwidth is split into two arrays of detectors: the Spectrometer Long Wave (SLW, 303–671 μm) and the Spectrometer Short Wave (SSW, 194–313 μm). The SPIRE spectrometer array consists of 7 (17) operational unvignetted bolometers for the SLW (SSW) detector, arranged in a hexagonal pattern. In the SLW, the beam FWHM is about 43′ at its largest, dropping to 30′ and then rising again to 35′ at higher frequency. The SSW beam is consistently around 19′.

Two SPIRE FTS observations from Operational Day 543 were utilized in this study: one long integration single pointing of 64 scans total ("deep," Observation ID 1342208389, 84 minutes (71 minute integration time), AOR “SSpec-m82-deep") and one fully sampled map ("map," Observation ID 1342208388, ~5 hr, AOR “Spec-m82"). The map observation was conducted in high-resolution (HR) mode and the deep observation was conducted in high+low-resolution (H+LR) mode. Both were processed in HR mode ($Δν \sim 1.19$ GHz) with the Herschel Interactive Processing Environment (HIPE) 7.2.0 and the version 7.0 SPIRE calibration derived from Uranus (Swinyard et al. 2010; Fulton et al. 2010).

2.2. Spectral Map Making Procedure

In mapping mode, the SPIRE detector arrays are moved around the sky to 16 different jiggle positions, creating 112 and 272 spectra of 16 scans each for SLW and SSW, respectively, covering an area of the sky approximately $3 \times 3$ arcmin. The positions of these scans on the sky are presented in Figure 1, with blue asterisks for SLW and red diamonds for SSW.

The recommended map making method bins the spectra into pixels approximately one-half the FWHM of the beam for each detector, which are about 35′ for SLW and 19′ for SSW, leading to pixel sizes of 17:5 and 9:5.11 We wrote a custom script to create the map, based largely on the NAIVEPROJECTION method described in the SPIRE documentation for HIPE. Each of the 256 scans per detector was processed individually. All scans for a given detector and jiggle position falling within a given pixel were then averaged and an error bar for each wavemeter bin average is determined as the standard deviation of the scans divided by the square root of number of scans. All detector-averaged spectra that fall into a pixel are then averaged using a weighted mean (where the weight is the inverse of the square of the error bar). We used the same 9:5 grid for both bands.

Using a 9:5 grid introduced more blank pixels in the SLW map because the SLW map is more sparsely sampled because of its larger beam, as can be seen in Figure 1. However, this enabled the comparison of the same spectral locations across both bands (where the data were available), without averaging together spatially discrepant spectra in the SSW, as would happen if pixels were made larger. We emphasize that the blank pixels are somewhat artificial; the whole galaxy has been mapped, and the pixels’ locations are simply meant to indicate the central

11 SPIRE Data Reduction Guide, http://herschel.esac.esa.int/Data_Processing.shtml, 2011 December 2.
with the signal; therefore, we first fit the strong lines (12CO, [C\text{II}], and [N\text{II}]) lines are fitted and subtracted from the spectrum, we then do a second pass to fit the weaker lines. An illustration of the difference this process can make for the $^{13}$CO $J = 5 \rightarrow 4$ line is in Figure 4.

In general, all of the lines are fit independently, with a few exceptions: the $^{12}$CO $J = 7 \rightarrow 6$ and $^{12}$CO $J = 2 \rightarrow 1$ lines are a mere 2.7 GHz away in rest frequency, and ground-state p-H$_2$O and o-H$_2$O$^+$ lines are separated by only 2 GHz. These two pairs must be fit simultaneously. Both are fit independently to supply initial guesses, which are then used to fit both lines as the sum of two sinc functions, each with their own central wavelength, width, and amplitude, but with one baseline value.

The integrated flux is simply the area under the fitted sinc function, which is proportional to the product of the amplitude and line width (converted to km s$^{-1}$). The error in the integrated flux is based on propagating the errors from the fitted parameters themselves. We note that the error estimation assumes that all wavenumber bins are independent of one another, but that is in fact not entirely true in an FT spectrometer. Though lines that are separated spectrally do not affect one another greatly (hence why we fit most lines independently), within each line fit the data points used in the 50 GHz range around the line center are not independent.

The beam size of the SPIRE spectrometer varies between the two detector arrays. In addition, it varies across the spectral range of the SLW, as described in Section 2.1, and is not strictly proportional to wavenumber due to the presence of multiple modes in the SLW detectors (Chattopadhyay et al. 2003). When examining the spectral line energy distribution (SLED), it is imperative to scale all fluxes to a single beam size. For the map observation, we first fit the spectra as they were (with no correction factor). An example of integrated flux map, prior to any convolution or beam correction, is presented in Figure 2, with all other integrated flux maps available in the online version. For the SLW detector, we present integrated flux maps using both 9.5’’5 pixels and 17.5’’ pixels.

The integrated flux maps for each line were then convolved with a kernel that matched the beams to the beam of the $^{12}$CO $J = 4 \rightarrow 3$ map, which has an FWHM of 43’’. The kernel was created using a modified version of the procedure described by Bendo et al. (2012); see also Gordon et al. (2008). However, instead of directly applying Equation (3) from Bendo et al. (2012) to the images of the beams, we applied the equation to one-dimensional slices of the beams to create the radial profile of convolution kernels, which we then used to create two-dimensional kernels. The approach worked very effectively for creating kernels to match beams observed with SSW to the $^{12}$CO $J = 4 \rightarrow 3$ map. However, in cases where we created convolution kernels for matching beams measured at two different wavelengths by the SLW array, we needed to manually edit individual values in the kernel radial profiles to produce effective two-dimensional kernels.

After the mathematical convolution of the integrated flux map with the kernel (resampled to match our map sizes), the fluxes were all converted to the units of Jy km s$^{-1}$ beam$^{-1}$, referring to the $^{12}$CO $J = 4 \rightarrow 3$ beam. The ratio of beam areas was determined empirically by convolving the kernel with the smaller (observed) beam peak normalized to 1 and determining the ratio required to produce the larger $\Omega_{\text{beam}}$ beam with the same peak (simulating the observation of a point source of 1 Jy km s$^{-1}$).

To account for blank pixels, only the portion of $\Omega_{\text{beam}}$ that encompasses data was used in the aforementioned conversion.

### Figure 1.

Spectrometer mapping point locations. SLW spectrum locations are indicated by blue asterisks, SSW by red diamonds. The pixel boundaries, spaced 9.5’’ apart, are indicated by solid black lines. The contours are also plotted in black to indicate the orientation of the (nearly edge-on) galactic disk, from the Mount Laguna 40 inch telescope (Cheng et al. 1997). Contours are in decreasing intervals of 0.2 log(maximum), i.e., $10^{0.0}$, $10^{0.2}$, $10^{0.4}$, … $10^{1.2}$. The green circle indicates the size of the $^{12}$CO $J = 4 \rightarrow 3$ 43’’ FWHM. The black × marks the position of the single-pointing (“deep”) spectrum.

(A color version of this figure is available in the online journal.)

### 2.3. Line Fitting and Convolution Procedure

The mirror scan length determines the spectral resolution of the spectrum. Because the scan length is necessarily finite, the Fourier-transformed spectrum contains ringing; therefore, the instrument’s line profile is a sinc function, as can be clearly seen in Figures 3 and 4. The spectrum also contains the underlying continuum which must be removed before fitting the lines, which we do sequentially rather than simultaneously (see exceptions below). We isolate ±25 GHz around the expected line center and mask out the ±6 GHz around the line center. We then fit the remaining signal with a second-order polynomial fit to determine the continuum shape. After subtracting this continuum fit, we then use a Levenberg–Marquardt least-squares method to fit each line as a sinc function with the following free parameters: central frequency, line width, maximum amplitude, and residual (flat) baseline value. The baseline value stays around zero because the continuum has already been subtracted. The central frequency is limited such that the line center is no greater than ±300 km s$^{-1}$ from the expected frequency given the nominal redshift of M82. For comparison, the resolution varies from 230 to 810 km s$^{-1}$, from the shorter to longer wavelength ends of the band.

For the deep spectrum, we detect weaker lines than in the map spectra. However, ringing from the strongest lines can interfere with the signal; therefore, we first fit the strong lines ($^{12}$C\text{O}, [C\text{II}], and [N\text{II}]) using the procedure outlined above and subtract their fitted line profiles from the spectrum. After all of the $^{12}$CO, [C\text{II}], and [N\text{II}] lines are fitted and subtracted from the spectrum, we then do a second pass to fit the weaker lines. An illustration of the difference this process can make for the $^{13}$CO $J = 5 \rightarrow 4$ line is in Figure 4.

In general, all of the lines are fit independently, with a few exceptions: the $^{12}$CO $J = 7 \rightarrow 6$ and $^{12}$CO $J = 2 \rightarrow 1$ lines are a mere 2.7 GHz away in rest frequency, and ground-state p-H$_2$O and o-H$_2$O$^+$ lines are separated by only 2 GHz. These two pairs must be fit simultaneously. Both are fit independently to supply initial guesses, which are then used to fit both lines as the sum of two sinc functions, each with their own central wavelength, width, and amplitude, but with one baseline value.

The integrated flux is simply the area under the fitted sinc function, which is proportional to the product of the amplitude and line width (converted to km s$^{-1}$). The error in the integrated flux is based on propagating the errors from the fitted parameters themselves. We note that the error estimation assumes that all wavenumber bins are independent of one another, but that is in fact not entirely true in an FT spectrometer. Though lines that are separated spectrally do not affect one another greatly (hence why we fit most lines independently), within each line fit the data points used in the 50 GHz range around the line center are not independent.

The beam size of the SPIRE spectrometer varies between the two detector arrays. In addition, it varies across the spectral range of the SLW, as described in Section 2.1, and is not strictly proportional to wavenumber due to the presence of multiple modes in the SLW detectors (Chattopadhyay et al. 2003). When examining the spectral line energy distribution (SLED), it is imperative to scale all fluxes to a single beam size. For the map observation, we first fit the spectra as they were (with no correction factor). An example of integrated flux map, prior to any convolution or beam correction, is presented in Figure 2, with all other integrated flux maps available in the online version. For the SLW detector, we present integrated flux maps using both 9.5’’5 pixels and 17.5’’ pixels.

The integrated flux maps for each line were then convolved with a kernel that matched the beams to the beam of the $^{12}$CO $J = 4 \rightarrow 3$ map, which has an FWHM of 43’’.

The kernel was created using a modified version of the procedure described by Bendo et al. (2012); see also Gordon et al. (2008). However, instead of directly applying Equation (3) from Bendo et al. (2012) to the images of the beams, we applied the equation to one-dimensional slices of the beams to create the radial profile of convolution kernels, which we then used to create two-dimensional kernels. The approach worked very effectively for creating kernels to match beams observed with SSW to the $^{12}$CO $J = 4 \rightarrow 3$ map. However, in cases where we created convolution kernels for matching beams measured at two different wavelengths by the SLW array, we needed to manually edit individual values in the kernel radial profiles to produce effective two-dimensional kernels.

After the mathematical convolution of the integrated flux map with the kernel (resampled to match our map sizes), the fluxes were all converted to the units of Jy km s$^{-1}$ beam$^{-1}$, referring to the $^{12}$CO $J = 4 \rightarrow 3$ beam. The ratio of beam areas was determined empirically by convolving the kernel with the smaller (observed) beam peak normalized to 1 and determining the ratio required to produce the larger $\Omega_{\text{beam}}$ beam with the same peak (simulating the observation of a point source of 1 Jy km s$^{-1}$).

To account for blank pixels, only the portion of $\Omega_{\text{kernel}}$ that encompasses data was used in the aforementioned conversion.
Figure 2. Integrated flux (top) and signal-to-noise (bottom) maps for CO $J = 4 \rightarrow 3$. Figures 2.1–2.19 are available in the online version. The top half includes no beam correction or convolution. Black corresponds to the lowest flux or zero if any fluxes are negative, at which point the color bar becomes purple. The bottom half is a map of signal to noise, though the color bar tops out at 20 in order to better illustrate which pixels are near the threshold of detectability. On the color bar, black corresponds to the lowest signal to noise or three if any pixels have S/N less than three, at which point the color bar becomes purple.

(An extended, color version of this figure is available in the online journal.)
For the deep spectrum, we used the same source–beam coupling factor \( \eta_c(\nu) \) as in Panuzzo et al. (2010), which was derived by convolving the M82 SPIRE photometer 250 \( \mu \)m map (Roussel et al. 2010) with appropriate profiles to produce the continuum light distribution seen with the FTS. The deep spectrum was multiplied by this factor before the line fitting procedure. The deep spectrum is presented in four parts in Figure 3, and the measured lines fluxes (\( \geq 3\sigma \)) are in Table 1.

The central pixel of the convolved maps offers a direct comparison to the source–beam coupling corrected deep spectrum.
In the SLW, these two SLEDS are within ±16% of one another. Later, we assume calibration error of 20%, so these differences are within those bounds. There is greater variation in the SSW band, though this is the region in which the signal-to-noise ratio (S/N) of the lines greatly drops. When 20% calibration error is included, all the line measurements have overlapping error bars with the exception of \(^{12}\)CO \(J = 13 \rightarrow 12\), where the deep spectrum measurement is more than twice that of the map. The \(^{12}\)CO \(J = 13 \rightarrow 12\) map, however, has an S/N of only 4 in the central pixel, with only 60/399 pixels having S/N greater than 3. We primarily use the deep spectrum for our likelihood analysis because of the higher S/N and access to more faint features, using two independent methods (the derived \(\eta_c\) from photometry comparisons versus map convolution) to account for source-beam coupling indicators that both methods are robust.

Though the maps do not provide adequately high spatial/spectral resolution for a detailed study of the morphology of the emission, some qualitative assessments can be made. For example, the line centroids do trace the relative redshift/ blueshift of the northeast and southwest components (\(v_{\text{hel}} \sim 300 \text{ km s}^{-1}\) and 160 km s\(^{-1}\), respectively; Loenen et al. 2010). However, we do not resolve the two separate peaks in flux. The capabilities of these maps to resolve gradients in the physical parameters modeled in this work will be discussed in Section 4.2.

### 3. BAYESIAN LIKELIHOOD ANALYSIS

We follow the method described in Ward et al. (2003) for the \(^{12}\)CO \(J = 6 \rightarrow 5\) map of M82, used frequently in the analysis of ground-based molecular observations of galaxies (e.g., the Z-spec collaboration; Naylor et al. 2010; Kamenetzky et al. 2011; Scott et al. 2011; Bradford et al. 2009), and also of the single-pointing SPIRE spectrum of M82 (Panuzzo et al. 2010) and Arp220 (Rangwala et al. 2011). The goal of our Bayesian likelihood analysis is to map the relative probabilities of physical conditions over a large parameter space; this provides a more complete statistical analysis of the physical conditions as opposed to simply finding one best-fit solution.

For each molecular species, we first calculate a grid of expected line emission temperatures for various combinations of temperature (\(T_{\text{kin}}\), density (\(n(H_2)\)), and column density per unit line width (\(N_{^{12}\text{CO}}/dv\)) using RADEX (van der Tak et al. 2007). We use the uniform sphere approximation for calculating the escape probabilities; the actual morphology in M82 shows a
more complex velocity structure, therefore this approximation is considered an average of the bulk properties of the gas (and the results are not sensitive to uniform sphere versus LVG approximation). RADEX performs statistical equilibrium calculations of the level populations, including the effects of radiative trapping, for a specified gas temperature, density, and level populations, including the effects of collisional excitation. In other words, at high temperatures, the modeled line intensities do not depend on the background radiation field. The cool component we will model is traced by low-J lines, whose background is not as affected, and so this component also finds the same results with either background. Therefore, we are presenting results using the CMB background.

In addition to 12CO, we also model 13CO and [C I] as a function of the same temperatures, densities, and column density of 12CO. For those molecular/atomic species, we also add the parameter of the relative abundance, e.g., [13CO]/[12CO] or X_{13CO}/X_{12CO}. When modeling the intensity, the column density of 13CO is simply that of 12CO times the relative abundance. [C I] is modeled with the parameter [C I]/[12CO]. Finally, we introduce one more parameter, the area fractional filling factor $\Phi_A$. The modeled emission may not entirely fill the beam, so the flux may be reduced by this factor. All model grid points are therefore multiplied by each value of $\Phi_A$. The ranges of parameters, as well as the number of grid points, are presented in Table 2.

The RADEX grid gives us a set of line intensities as a function of model parameters $p = (N_{12CO}/dv, n(H_2), T_{\text{kin}}, \Phi_A, X_{\text{mol}}/X_{12CO})$, which we then compare to our measured intensities $x$. To compare to column density per unit line width, we divide the measured intensities by 180 km s$^{-1}$, so that they are also per unit line width. The optical depth, and in turn the RADEX results, depend only on the ratio of column density to line width. Ward et al. (2003) found line widths of 180 km s$^{-1}$ for the NE component and 160 km s$^{-1}$ for the SW by resolving the structure in position–velocity diagrams for their study from CO $J = 1 \rightarrow 0$ to $J = 7 \rightarrow 6$, but we do not resolve the difference between the two and so we use the larger value. The Bayesian likelihood of the model parameters given the measurements is

$$P(\mathbf{p}|x) = \frac{P(\mathbf{p})P(x|\mathbf{p})}{P(x)},$$

where $P(\mathbf{p})$ is the prior probability of the model parameters (see Section 3.2), $P(x)$ is for normalization, and $P(x|\mathbf{p})$ is the probability of obtaining the observed data set given that the source follows the model described by $\mathbf{p}$. $P(x|\mathbf{p})$ is the product of Gaussian distributions in each observation,

$$P(x|\mathbf{p}) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left[ -\frac{(x_i - \tilde{I}_i(\mathbf{p}))^2}{2\sigma_i^2} \right].$$

where $\sigma_i$ is the standard deviation of the observational measurement for transition $i$ and $\tilde{I}_i(\mathbf{p})$ is the RADEX-predicted line intensity for that transition and model. For the total uncertainty, we take the statistical uncertainty in the total integrated intensity from the line fitting procedure and add 20%/10% calibration error for SSW/SLW in quadrature. To find the likelihood distribution of one parameter out of all of $\mathbf{p}$, we integrate over all other parameters to find, for example, $P(T_{\text{kin}})$.

3.1. Separate Components

We divide the lines fluxes into two components, one warmer and one cooler. Panuzzo et al. (2010) already showed that the high-J lines of 12CO trace a warmer component than those transitions available from the ground. However, some of the mid-J lines (especially CO $J = 4 \rightarrow 3$) may have significant contributions from both components. To separate the 12CO fluxes from each component, we follow an iterative procedure. We first model the lowest three 12CO lines from Ward et al. (2003); we take the sum of their measurements for the two observed lobes and scale the result by the ratio of their beam area to our 43" beam. The best-fit warm SLED is then subtracted from the low-J lines, producing the asterisks in the aforementioned figure. These triangles comprise the “warm component.” The best-fit warm SLED is then subtracted from the low-J lines, producing the asterisks in the aforementioned figure. These fluxes are refit to produce our results for the “cool component.”

We present a two-component model using just 12CO, as well as one including our high-J lines of 13CO along with the warm component. We did not model the low-J 13CO lines reported in Ward et al. (2003) due to the uncertainties presented in their Table 2 footnotes. We instead predicted the 13CO spectrum of the cool component, given its best-fit results and a 12CO/13CO ratio of 35 (in between previously found 30 to 40; Ward et al. 2003), and found a very small contribution to the higher-J lines we are modeling. These small contributions are subtracted from the warm component, as with 12CO in the previous paragraph.

We also sought to include [C I], which was assumed to be associated with the cool component, due to the low excitation temperature (~30 K) derived from the line ratios ($n_u/n_1 = g_u/g_1 \exp(-\Delta T/T_{\text{eq}})$, $n_1 \propto I_i (\text{W/cm}^2)/(A_i v_i)$). Here, we use

---

**Table 2: RADEX Model Parameters and Ranges**

| Parameter               | Range          | No. of Points |
|-------------------------|----------------|--------------|
| $T_{\text{kin}}$ (K)    | $10^6$–$10^7$ | 61           |
| $n(H_2)$ (cm$^{-3}$)   | $10^2$–$10^6$ | 41           |
| $N_{12CO}/\Delta V$ (cm$^{-2}$) | $10^4$–$10^{18}$ | 41 |
| $\Phi_A$               | $10^{-3}$–$10^0$ | 41           |
| $X_{13CO}/X_{12CO}$    | $10^{-3}$–$10^{-1}$ | 11 |
| $X_{\text{mol}}/X_{12CO}$ | $10^{-2}$–$10^2$ | 11 |
| $\Delta V$ (km s$^{-1}$) | 1.0 Fixed      | 11           |
| $T_{\text{background}}$ (K) | 2.73 Fixed     | 11           |

**Notes:** All parameters are sampled evenly in log space. Velocity is fixed because all modeling is column per unit line width.
Δ$T = 38.84$, $g_u = 5$, $g_l = 3$, $A_u = 2.65 \times 10^6$ s$^{-1}$, and $A_l = 7.88 \times 10^7$ s$^{-1}$. See Table 1 for the frequencies and unit conversion. However, this model produced some unphysical situations. We discuss our findings and their implications in Section 4.1.

This analysis necessarily assumes that all of the line emission for a given component is coming from one portion of gas described in bulk by the model parameters. In reality, there are likely a variety of physical conditions, existing in a continuum of parameters. However, the high-$J$ SPIRE data do not provide justification for modeling more than one warm component because the SLEDs are well described by one component. We did attempt a procedure to model multiple warm components of CO by first fitting the highest-$J$ lines and subtracting the predicted line fluxes for the mid-$J$ lines. Such a procedure has been used in Rangwala et al. (2011), for example. However, the predictions for the mid-$J$ lines either matched or were an overestimate of the observed fluxes, leaving no second component to be modeled. A range of conditions is definitely present in the molecular gas, yet these two (warm and cool) components are dominating the emission, within the observational and modeling uncertainties. We note that, with regards to the different molecules/atoms being modeled here, all three species have similar profiles, as shown from the HIFI (higher spectral resolution) spectra in Loenen et al. (2010).

### 3.2. Prior Probabilities

We use a binary prior probability, $P(p)$, to indicate either a physically allowed scenario ($P(p) = 1$) or an unphysical, and thus not allowed, scenario ($P(p) = 0$). In other words, all combinations of parameters that are deemed physical based on the following three conditions were given equal prior probability, and all others are given zero prior probability. The conditions are as follows.

1. The total length of the column ($L_{\text{col}}$) cannot exceed the length of the entire region. This assumes that the length in the plane of the sky is the same as that orthogonal to the plane of the sky; we chose an upper limit to the length of 900 pc because of the observed size (Ward et al. 2003). This is the most significant of all the priors, placing an upper limit on the column density and a lower limit on the density. This prior can be stated as

$$\frac{N_{12\text{CO}}}{n(H_2)\sqrt{\Phi_A X_{12\text{CO}}}} \leq 900 \text{ pc}. \quad (3)$$

For the relative abundance $X_{12\text{CO}}$ to molecular hydrogen, we assume $3.0 \times 10^{-4}$ (Ward et al. 2003).

2. The total mass in the emission region ($M_{\text{region}}$) cannot exceed the dynamical mass of the galaxy. We use the expression

$$M_{\text{region}} = \frac{A_{\text{region}} N_{12\text{CO}} \Phi_A 1.5m_H}{X_{12\text{CO}}} \quad (4)$$

to calculate the mass in the emission region, where 1.5 accounts for helium and other heavy elements, and $\Phi_A$ is the filling factor. We estimate the dynamical mass to be $2 \times 10^9 M_\odot$ (Ward et al. 2003; Naylor et al. 2010), calculated using rotational velocity and radius. The other assumed parameters in the above expression are listed in Table 3.

3. The optical depth of a line must be less than 100, as recommended by the RADEX documentation. The cloud excitation temperature can become too dependent on optical depth at high column densities, and so very high optical depths can lead to unreliable temperatures. We found that,
Probabilities. We also cannot account for 13CO cannot be reliably measured. We also cannot account for
used for scaling the line intensities. All other parameters used for prior analysis independently. (Note that due to the beam size being
have both an SLW and SSW spectrum and at least 5 12CO lines with S/N ≥ 10 (the convolution tends to increase the S/N of each pixel). Ninety pixels met this requirement (98 pixels would meet the requirement of 5 12CO lines with S/N ≥ 3, so little would be gained by going to lower S/N). Additionally, we do not find statistically significantly different results requiring only three lines, the minimum with which we could reasonably model the emission; to some extent, this requires at minimum the J = 6 → 5 transition, which means we will be tracing higher temperatures.

3.3. Likelihood Analysis of the Map

We run each map pixel through the aforementioned likelihood analysis independently. (Note that due to the beam size being larger than the pixels themselves, each pixel’s data are not independent of its neighbors.) We only model 12CO in the spectral maps because they are of lower integration time and 13CO cannot be reliably measured. We also cannot account for cool emission at different locations in our map.

To be run through the likelihood analysis, a pixel was required to have both an SLW and SSW spectrum and at least 5 12CO lines with S/N ≥ 10 (the convolution tends to increase the S/N in the presence of the other priors, this limit does not affect the likelihood results.

### Table 3

| Parameter                 | Value | Units   |
|---------------------------|-------|---------|
| Line width                 | 180   | (km s⁻¹) |
| Abundance (X₁₂CO/X₁₈H₂)   | 3.0 × 10⁻⁴ |        |
| Angular size scale         | 17    | (pc/°)  |
| Emission size              | 43.0  | (°)     |
| Length limit               | 900   | (pc)    |
| Dynamical mass limit       | 2 × 10⁹ | (M☉)   |

Notes.  

a Citations for parameters are in Section 3.2.  
b Used for scaling the line intensities. All other parameters used for prior calculations.  
c \( A_{\text{region}} = \pi \text{(angular size scale (pc/°) × source size (°)/2)}^2 \).

### Table 4

| Integrated Likelihood: Cool | Median | 1σ Range | 1D Max | 4D Max | Integrated Likelihood: Warm | Median | 1σ Range | 1D Max | 4D Max |
|-----------------------------|--------|----------|--------|--------|-----------------------------|--------|----------|--------|--------|
| \( T_{\text{kin}} \)        | 40     | 12–472   | 13     | 63     | 414                         | 335–518| 447      | 365    | 447    |
| \( n(H₂) \)                | 10⁻¹³  | 10⁸–10¹⁰ | 500    | 10⁴⁰   | 200                         | 10³⁵–10⁴⁰| 200      | 10³⁶–10⁴⁰|
| \( N_\text{co} \)          | 10¹⁰   | 10⁸–10¹² | 10¹⁰   | 10¹⁰   | 10⁸                         | 10⁸–10¹⁰ | 10¹⁰   | 10¹⁰   |
| \( P \)                    | 10⁻³⁴  | 10⁻⁵⁵–10⁻⁷⁶ | 10⁻²⁴ | 10⁻²⁴  | 10⁻⁶¹                       | 10⁻⁶⁶–10⁻⁸⁰ | 10⁻⁷⁵ | 10⁻⁷⁵ |
| \( N_{12\text{co}} \)      | 10¹⁵   | 10¹⁷–10²⁸ | 10¹⁷   | 10¹⁷   | 10¹⁶                         | 10¹⁶–10³⁷ | 10¹⁶   | 10¹⁶   |
| Mass                       | 10⁴⁷   | 10⁴⁷–10⁴⁷ | 10⁴⁷   | 10⁴⁷   | 10⁴⁷                         | 10⁴⁷–10⁴⁷ | 10⁴⁷   | 10⁴⁷   |

Notes. 1D Max refers to the maximum value that parameter after integrated over all the other parameters (the mode of the likelihood distributions). 4D Max refers to the single most probable grid point (mode of the entire multi-dimensional distribution). Median, 1σ lower and upper values refer to the integrated distribution, when the cumulative distribution function equals 0.5, ~0.16, and ~0.84, respectively. Note that the 1D Max may be outside the 1σ range because of asymmetry in the integrated likelihood. P and \( N_{12\text{co}} \) are calculated from the 2D distribution of \( T_{\text{kin}} \) and \( n(H₂) \) (or \( \Phi_\text{A} \) and \( N_{12\text{co}} \)), which is why we give the same value in both the 1D Max and 4D Max columns.

### Table 5

| Integrated Likelihood: CO | Median | 1σ Range | 1D Max | 4D Max | Integrated Likelihood: CO | Median | 1σ Range | 1D Max | 4D Max |
|---------------------------|--------|----------|--------|--------|---------------------------|--------|----------|--------|--------|
| \( T_{\text{kin}} \)        | 35     | 12–385   | 14     | 158    | 436                         | 344–548| 447      | 365    | 447    |
| \( n(H₂) \)                | 10⁻³⁴  | 10⁻⁴⁰–10⁻³⁵ | 10⁴⁰  | 10⁴⁰  | 10⁻⁸                         | 10⁻⁸–10⁻⁸ | 10⁻⁸ | 10⁻⁸ |
| \( N_\text{co} \)          | 10¹⁰   | 10⁸–10¹² | 10¹⁰   | 10¹⁰   | 10⁸                         | 10⁸–10¹⁰ | 10¹⁰   | 10¹⁰   |
| \( P \)                    | 10⁻³⁴  | 10⁻⁵⁵–10⁻⁷⁶ | 10⁻²⁴ | 10⁻²⁴  | 10⁻⁶¹                       | 10⁻⁶⁶–10⁻⁸⁰ | 10⁻⁷⁵ | 10⁻⁷⁵ |
| \( N_{12\text{co}} \)      | 10¹⁵   | 10¹⁷–10²⁸ | 10¹⁷   | 10¹⁷   | 10¹⁶                         | 10¹⁶–10³⁷ | 10¹⁶   | 10¹⁶   |
| Mass                       | 10⁴⁷   | 10⁴⁷–10⁴⁷ | 10⁴⁷   | 10⁴⁷   | 10⁴⁷                         | 10⁴⁷–10⁴⁷ | 10⁴⁷   | 10⁴⁷   |

### Table 6

| Integrated Likelihood: 12CO | Median | 1σ Range | 1D Max | 4D Max | Integrated Likelihood: 13CO | Median | 1σ Range | 1D Max | 4D Max |
|-----------------------------|--------|----------|--------|--------|-----------------------------|--------|----------|--------|--------|
| \( X_{12\text{co}}/X_\text{co} \) | −      | −        | −      | −      | \( X_{13\text{co}}/X_\text{co} \)  | −      | −        | −      | −      |
| \( N_{12\text{co}} \)      | −      | −        | −      | −      | \( N_{13\text{co}} \)               | −      | −        | −      | −      |

Notes. X₁₂co/Xco and N₁₂CO are not included in the likelihood analysis because 12CO cannot be reliably measured.
The Astrophysical Journal, 753:70 (17pp), 2012 July 1

Kamenetzky et al.

Figure 7. Bayesian likelihood analysis, primary parameter results, 12CO only. Each color represents a separate component: blue for cool and red for warm (see Section 3). Dashed/dotted vertical lines indicate the median/4D maximum of the distribution (see Table 4).

(A color version of this figure is available in the online journal.)

column density (⟨N_{12CO}⟩, the product of column density and filling factor). We note that in the results, the parameters of the most likely grid point (“4D Max”) are not necessarily the same as the median or the mode (“1D Max”) of the integrated likelihood distributions. The “4D Max” is describing one specific point, but the median, 1D Max, and associated error range are representative of the larger likelihood across all other parameters in the grid.

We first compare our two models, 12CO only versus multiple molecule. These are Tables 4 versus 5, Figures 7 versus 8, and Figures 9 versus 10.

The addition of 13CO to the warm component does not significantly change the temperature, but it does increase the likelihood of lower densities. It also decreases the likelihood of lower column densities. The consequences of these changes can be seen in the pressure and mass distributions in Figures 9 and 10. Adding 13CO increased the likelihoods of the “shoulders” of these distributions: the lower half of pressure and the upper half of mass. An examination of the contour plots in the bottom half of these figures illustrates why, statistically. In the 12CO only model, though column density and filling factor are not well-constrained independently, they are highly correlated; their contours run along an almost constant line of beam-averaged column density. Adding 13CO introduced the X_{13CO}/X_{12CO} parameter, which also impacts the absolute flux level of the models, like column density and filling factor. The results are likelihoods that are more constrained but not as highly correlated with one another. Therefore, the mass distribution is wider. In the 12CO-only model, the mass of the warm component is about 3.4%/6.3% (median/4D Max) the mass of the cool component. Rigopoulou et al. (2002) noted that “warm” gas is generally around 1% to 10% of the total gas mass for starburst galaxies, so this is about as expected. The addition of 13CO creates somewhat overlapping likelihood distributions for mass (the 1σ ranges are just touching), but the median and 4D Max now warm/cool ratios of 11% and 9%, respectively. One factor that may contribute to wider distributions is overestimated error; the error bars are dominated by our 20% calibration error, not statistical error.

After this point, we compare frequently to Panuzzo et al. (2010). The major factor responsible for the differences, just modeling 12CO alone, is the shape of the CO SLED at those lines with upper rotational number greater than J = 8. We also explicitly subtracted the cool component’s contribution from 12CO J = 4 → 3, whereas Panuzzo et al. (2010) simply underpredicted the total flux. We will also compare with Loenen et al. (2010), another high-J CO study of M82, in Section 4.4.

Our results are similar to Panuzzo et al. (2010), who found that these high-J CO lines trace a very warm gas component that is separate from the cold molecular gas traced by those lines below J = 4 → 3. Our best-fit temperature of the 12CO only model (at 447 K, Table 4) is close to their value at 545 K, but the overall likelihood for temperature, integrated over all parameters, yields a slightly lower 414 K. Given the size of the uncertainty (335–518 K) in the parameter, the two distributions are very similar, and therefore the result is not significantly different. Such warm gas has also been traced in the S(1) and S(2) transitions of H2, at 450 K with the Infrared Space Observatory (Rigopoulou et al. 2002) and 536 K with the Spitzer Infrared Spectrograph (Beirão et al. 2008).
We do find a slightly higher density than Panuzzo et al. (2010), with our best-fit value of $10^{4.1}$ compared to $10^{3.7}$ cm$^{-3}$, though the integrated likelihood distributions do overlap (see Figure 7). However, the temperature and density are degenerate; higher temperatures and lower densities may produce the same fluxes as lower temperatures and higher densities. Their product, the pressure, is better constrained. We seem to have collapsed/constrained the pressure distribution to the upper half of that presented in Panuzzo et al. (2010).

The column density is not as well constrained as presented in Panuzzo et al. (2010); we found that they had an error in calculating the expected fluxes of the higher-J lines for lower column density values. We have recalculated the fluxes for those column densities, and we find that, in fact, when properly calculated, these column densities have a non-zero likelihood. In the $^{12}$CO only model, the column density itself is not constrained. However, the column density and filling factor are degenerate, so it is their product (beam-averaged column density, $\langle N_{\text{co}} \rangle$) that is better constrained. Our best-fit value is $10^{16.6}$ cm$^{-2}$.

The total mass in the beam can be calculated using Equation (4) (and is presented as the top y-axis in Figures 9 and 10, upper right). As previously discussed, the $^{12}$CO only model produces the expected result of less gas mass in the warm component. In the cool component, we find a best-fit mass of $2.0 \times 10^7 M_\odot$ (median $4.7 \times 10^7 M_\odot$). This is smaller than the $2.0 \times 10^8 M_\odot$ traced by the LVG analysis of Ward et al. (2003) with lower-J CO lines. This difference is due to the fact that we subtract the contribution to the low-J flux from the warm component; in our initial modeling of the cold component, before this subtraction, our best-fit mass is $9.8 \times 10^7 M_\odot$ (with a range from 0.2 to $2.2 \times 10^8 M_\odot$). The warm component is a smaller fraction of the gas, with a best fit of $1.3 \times 10^6 M_\odot$, about 6.3% the mass of the cool component.

The $^{12}$CO/$^{13}$CO relative abundance is also a free parameter in our multi-species model; we find a best-fit value of about 32, similar to the 40 and 30 found previously for the NE and SW lobes, respectively (Ward et al. 2003).

As mentioned in Section 3.1, we also attempted to include our two [C i] lines with the cool component. We do not present the tables for this model because the mass distributions of the warm and cool components became overlapping, indicating the same amount of mass in both components, an unphysical situation. Additionally, the derived relative abundance of [C i] to $^{12}$CO was unusually high. We found a ratio of 0.48 to 3.3, which is higher than White et al. (1994; average value $\sim 0.5$), Schilke et al. (1993; 0.1–0.3), and Stutzki et al. (1997; 0.5), using other methods. Before subtracting the warm component’s contribution to the $^{12}$CO flux (when we just fit all of the low-J $^{12}$CO flux and [C i]), we find ratios more consistent with these values (best-fit 0.4, range 0.09–1.23). These two problems could be indicating that the assumption of CO and [C i] coming from the same component is flawed. The column density, temperature, and mass developed somewhat of a double-peaked structure; specifically, the addition of [C i] increases the likelihood of lower column densities and masses, but does not eliminate the previous likelihood peak.

It is unclear how much of the molecular CO and atomic C are truly cospatial and therefore how physical our results for modeling them all together as one bulk gas component may be.
Papadopoulos et al. (2004) presented results which argue that [C\textsc{i}] and CO are cospatial and trace the same hydrogen gas mass ([C\textsc{i}] doing so better than CO in many conditions). This conflicts with the theoretical picture of gas clouds (especially in PDRs) as a structured transition between molecular, atomic, and ionized gas, but new observational and theoretical evidence indicates that the types of gas are not so distinct (see references within Papadopoulos et al. 2004). For example, Howe et al. (2000) and Li et al. (2004) have found [C\textsc{i}] to trace 13CO well. If the ISM is clumpy, well mixed, and dynamic, then the [C\textsc{i}] and CO may be cospatial averaged over large scales. Strong stellar winds (and possibly the interaction with M81) could be contributing to the dynamic nature of the gas, so it is not unreasonable to believe that the gas has not achieved the simple layered pattern. Though the SPIRE FTS cannot resolve the two separate velocity components of M82, HIFI can, and observations of these two [C\textsc{i}] lines indicate a generally similar shape to the CO lines, namely two Gaussians with the SW component demonstrating higher flux (Loenen et al. 2010).

Wolfire et al. (2010) presented a model PDR which shows a cloud layer traced by atomic (not-yet-ionized) [C\textsc{i}] where the hydrogen is still molecular due to self-shielding effects. This “dark molecular gas” (called so because it is not traced by CO) would be less shielded and at a warmer temperature than the inner-most cloud layer of CO. It is possible that [C\textsc{i}] and CO are somewhat cospatial yet somewhat segregated, as in the “dark molecular gas model.” Our analysis is consistent with a picture in which the [C\textsc{i}] and 12CO do not completely overlap spatially.

The [C\textsc{i}] \(J = 1 \rightarrow 0\) emission can also be used to estimate the total hydrogen mass using Equation (12) of Papadopoulos et al. (2004). Using the median \(X_{\text{[C\textsc{i}]}/X_{\text{H}_2}\) of \(1.5 \times 10^{-4}\) from \(X_{\text{[C\textsc{i}]}/X_{\text{12CO}} = 0.5\) (White et al. 1994; Stutzki et al. 1997) and our assumed \(X_{\text{12CO}}/X_{\text{H}_2}\), we find a total gas mass of \(4.4 \times 10^7 \Omega_{10} M_\odot\). \(\Omega_{10}\) is the ratio of the column of the \(J = 1 \rightarrow 0\) emission to the total [C\textsc{i}] column (see Appendix A of Papadopoulos et al. 2004), which depends on the excitation conditions of the gas; for \(\Omega_{10} \sim 0.5\) (Papadopoulos & Greve 2004), the gas mass is \(8.8 \times 10^7 M_\odot\) but is uncertain by modeled uncertainty in \(X_{\text{[C\textsc{i}]alone}}\). This method of estimating the mass is higher than the total mass estimate of the cool component described earlier in this section. [C\textsc{i}] may be coming from a range of temperatures, but with only two lines, we cannot sort that out.

4.2. Physical Conditions: Map

Results for all of the same parameters for the deep spectrum were produced for each of the pixels in the map (that met the criteria in Section 3.3). We find that the beam size of SPIRE cannot resolve the structure in M82 as has been done with interferometric maps or high spectral resolution observations (which can resolve the velocity components of the NE and SW lobes). Because of the degeneracy between temperature/density and column density/filling factor, we present results of their products, pressure and beam-averaged column density in Figure 11. \(\langle N_{12\text{CO}} \rangle\) shows a radially decreasing trend, roughly corresponding to the decrease in the beam profile (plotted with a solid black line), implying an observational effect due to source–beam coupling. We would expect an off-centered beam to be able to probe the pressure of the
central region (because the relative ratios of the SLED would be preserved), so the lack of a radial trend also implies that we are not measuring different areas of M82 in each map pixel. This indicates that the SPIRE FTS map cannot resolve M82’s structure, and therefore the single “deep” spectrum is an adequate representation of the galaxy as a whole. Our results in Section 4.1 are descriptive of the bulk properties of the galaxy and we do not see trends on the scale of our map. This analysis is separate from the dust temperature gradients (which we also see in the continuum gradients of our map spectra) found in M82 by Roussele et al. (2010) and indicates that the dust and molecular gas are not coupled.

One difficulty with the map is that it has lower S/N; it must also be convolved up the largest beam size (43″, like the deep spectrum). However, as mentioned in Section 4.1, it is the highest-J fluxes in the SSW detector that largely constrain the results of the likelihood for the deep spectrum. Therefore, we also attempted to model the map with just those lines with upper-J level of nine or higher, without convolving. These lines all were measured with a beam FWHM of ~19″, offering higher resolution. However, these lines are also weaker, and with fewer, lower S/N lines, this method does not constrain any parameters as well.

Though the off-axis pixels may not provide new information about the physical trends of the galaxy, we can compare the deep spectrum to the center pixel of our map as a test of the source–beam coupling corrections (see Section 2.2). The results of the central pixel are very similar to those presented in Section 4.1, though the integrated likelihoods are generally wider (the parameters are less constrained). This is partially, but not entirely, due to larger error bars on the SSW lines.

### 4.3. Molecular Line Absorption

In addition to the Bayesian likelihood modeling, we can also briefly discuss the absorption lines presented in Table 1.

#### 4.3.1. Hydrogen Fluoride

Hydrogen fluoride (HF) is potentially a sensitive probe of the total molecular gas column because the HF/H$_2$ ratio is more reliably constant than $^{12}$CO/H$_2$ since the formation of HF is dominated by a reaction of F with H$_2$ (Monje et al. 2011). Furthermore, HF $J = 1 \rightarrow 0$ is generally seen in absorption because of its high $A$-coefficient, $2.42 \times 10^{-2}$ s$^{-1}$ (Monje et al. 2011). Assuming all HF molecules are in the ground state (generally true in the diffuse and dense ISM; Monje et al. 2011), the HF $J = 1 \rightarrow 0$ line yields the optical depth simply as $\tau = -\ln(F_l/F_c)$, where $F_l/F_c$ is the line-to-continuum ratio. In the case of HF, we mask out a nearby water emission line (1226 to 1229 GHz), though because the signal in each wavenumber bin is not independent due to the ringing, this can introduce added uncertainty. Therefore, the following discussion is meant to be approximate. We estimate the HF column density using

$$\int \tau dv = \frac{\Lambda_{ul} g_u \lambda^3}{8\pi g_l} N(\text{HF}),$$

where $g_u = 3$ and $g_l = 1$. This implies $\int \tau dv = 4.16 \times 10^{-13} N(\text{HF}) \text{ cm}^{-2} \text{ km s}^{-1}$. The HF line occurs in a part of our
Figure 11. Bayesian likelihood analysis for mapping observation. The x-axis is the map pixel’s radial distance (in pc) from the central detector’s position, with its clockwise angle (meaning degrees from north through west) represented by the color bar on the right. The y-axis is the median value (after integrating over all other parameters) and associated 68% error bars. The solid line represents the logarithm of the $^{12}$CO $J = 4 \rightarrow 3$ beam profile such that the peak corresponds with the result at the central pixel; it is only plotted to $50''$ from the center because of uncertainties in the profile beyond this region. The dashed line represents one half the beam FWHM.

(A color version of this figure is available in the online journal.)
4.4. Gas Excitation

At the high temperature of the warm component, the cooling will be dominated by hydrogen. Le Bourlot et al. (1999) modeled the cooling rates for H$_2$, and made their tabulated rates available with an interpolation routine for desired values of density, temperature, ortho- to para-H$_2$ ratio, and H to H$_2$ density ratio.\[12\] For our best-fit temperature and density, assuming n(H)/n(H$_2$) = 1 (recommended by Le Bourlot et al. 1999, for PDRs), this corresponds to a cooling rate of $10^{-22.54}$ erg s$^{-1}$ molecule$^{-1}$, or $3 L_\odot / M_\odot$ (using o/p H$_2$ = 1, though the number is only $\sim$3% lower for o/p H$_2$ = 3). Given the warm mass of $1.3 \times 10^6$ M$_\odot$ (using the $^{12}$CO model for the rest of this section), that implies a hydrogen luminosity of $3.9 \times 10^8 L_\odot$. The total observed hydrogen luminosity thus far has been higher than this number; adding the luminosities presented in Rigopoulou et al. (2002) and correcting for extinction as in Draine (1989), we find a total of $1.2 \times 10^7 L_\odot$ in the (0–0)S(0)–S(3), S(5), S(7), and (1–0)Q(3) lines. However, some of these hydrogen lines are tracing lower or higher temperature gas. We note that the mass range within one standard deviation of our likelihood results for the warm component is 0.93–5.2 $\times 10^6$ M$_\odot$, which corresponds to a predicted luminosity of 0.28–1.6 $10^7 L_\odot$, encompassing the measured hydrogen luminosity.

There are a few possibilities for the source of the excitation of the gas: X-ray photons, cosmic rays, UV excitation of PDRs, and shocks/collisional excitation. Hard X-rays from an AGN have already been ruled out by others in the literature due to the lack of evidence for a strong AGN and low X-ray luminosity (1.1 $\times 10^6$ L$_\odot$; Strickland & Heckman 2007).

The CO emission from M82 has previously been interpreted using PDR models. Beirão et al. (2008) noted with the Spitzer Infrared Spectrograph that the H$_2$ emission is correlated with polycyclic aromatic hydrocarbon (PAH) emission, indicating that it is mainly excited by UV radiation in PDRs. Loenen et al. (2010) combined HIFI data with ground-based detections in order to model $^{12}$CO $J = 1 \rightarrow 0$ to $J = 13 \rightarrow 12$ and $^{13}$CO $J = 1 \rightarrow 0$ to $J = 8 \rightarrow 7$. They reproduced the measured SLEDs with one low-density (log(n(H$_2$)) = 3) and two high-density (log(n(H$_2$)) = 5, 6) components with relative proportions of 70%, 29%, and 1%, respectively. The low-density component is largely responsible for the low-J emission, while the highest-density component is responsible for the highest-J emission. These high densities are not consistent with the results of our likelihood analysis detailed in Section 4.1; the likelihood of solutions for the warm component at log(n(H$_2$)) > 5 is essentially zero. We note that Loenen et al.’s (2010) Figure 3 shows the consistency between HIFI and SPIRE fluxes from Panuzzo et al. (2010). In other words, the difference is not due to discrepant line fluxes, but different models (PDR versus CO likelihood analysis).

There are two major differences in the approach of this work and Loenen et al. (2010). First, the order in which we approach the problem is different. We first analyze the CO excitation using likelihood analysis to determine the physical conditions. Once we have these conditions, we then look to the possible energy sources based on the conditions we have already derived, instead of first seeing under which conditions a certain energy source fits. Second, we look beyond the one best-fit solution: in addition to presenting the best-fit solution, our likelihoods analyze the relative probabilities for a larger parameter space.

We also attempted to reproduce our deep SLED with various PDR models. Meijerink et al. (2006) have added to their PDR and X-ray dominated region (XDR) models to include enhanced cosmic rays (at a rate of $5 \times 10^{-15}$ s$^{-1}$), near our assumed rate discussed later in this section. Such models are currently available for incident flux of log(G$_0$) of 2–4 (G$_0$ = $1.6 \times 10^{-3}$ erg cm$^{-2}$ s$^{-1}$) for log(n(H$_2$)) = 3 and log(G$_0$) of 3–5 for log(n(H$_2$)) of 4 and 5. By examining the ratios of $^{12}$CO $J = 9 \rightarrow 8$ with all higher-J lines (those largely driving the likelihood results, and also measured from similar beam sizes), none of the available nine PDR models are an ideal match, but the PDR scenario for log(n(H$_2$)) = 3, log(G$_0$) = 3 is the best match. Figure 12 compares the predicted and observed ratios. The ratios used are without source–beam coupling correction because the SSW already has similar beam sizes, but the ratios with beam correction are within 4%–8% of those presented in Figure 12.

---

\[12\] http://ccp7.dur.ac.uk/cooling_by_h2/
Additionally, we used a higher-resolution (in density and incident flux) grid of $^{12}$CO PDR models (Wolfire et al. 2010). The same line ratios previously discussed (those shown in Figure 12) can only be reproduced by higher densities. For $(J = 9 → 8)/(J = 10 → 9)$, the observed ratio is only found for $\log(n(H_2)) > 4.5$, $\log(G_0) > 2$, and for $(J = 9 → 8)/(J = 13 → 12)$, it is only found for $\log(n(H_2)) > 5$, $\log(G_0) > 2.5$.

To summarize, current PDR models can only explain the observed high-J $^{12}$CO emission with densities higher than those indicated by the likelihood analysis (even when all priors are excluded). The cosmic-ray-enhanced PDR models, though sparse, can come closer to reproducing the high-J line ratios at a lower density, though these are also below our likelihoods. There is also evidence that shocks enhance high-J lines far more than PDRs (Pon et al. 2012). In their models, almost all of the emission from the lowest-J lines came from unshocked gas, and most of the emission above $J = 7 → 6$ was from shocked gas. The combination of shocks and PDRs could be responsible for the high-J CO lines observed, while PDRs alone are adequate to explain the lower-J CO lines and PAH emission. In summary, these high-J lines are not consistent with current PDR models, but improved modeling of shocks and the effects of cosmic rays on PDRs may help explain their emission.

Cosmic rays (CRs) are another possibility for the excitation. The Very Energetic Radiation Imaging Telescope Array System (VERITAS) Collaboration recently reported that the cosmic-ray density in M82 is about 500 times the average Galactic (Milky Way) density (VERITAS Collaboration et al. 2009). By using the cosmic-ray ionization rate of the Galaxy (2–7 × 10$^{-13}$ s$^{-1}$; Goldsmith & Langer 1978; van Dishoeck & Black 1986), multiplied by 500, and then multiplied by the average energy per ionization (20 eV), one finds an energy deposition rate of 2–7 × 10$^{-13}$ eV s$^{-1}$ per H$_2$ molecule in M82. This implies a heating rate of 0.03–0.12 L$_\odot$/M$_\odot$, less than 5% of the required molecular hydrogen cooling rate. Thus, cosmic rays alone cannot excite the molecular gas.

Turbulent heating mechanisms may also play a role in M82. From Bradford et al. (2005), the turbulent heating per unit mass can be expressed in L$_\odot$/M$_\odot$ as

$$1.10 \left( \frac{v_{rms}}{25 \text{ km s}^{-1}} \right)^3 \left( \frac{1 \text{ pc}}{\Lambda_d} \right),$$

where $v_{rms}$ is the turbulent velocity and $\Lambda_d$ is the typical size scale of turbulent structures. We use the Jeans length for this size scale, calculated from the parameters of the likelihood results (Section 4.1), which is 0.9 pc. Given this size scale, the observed cooling rate could be replicated with a turbulent velocity of 33.7 km s$^{-1}$. This would imply a velocity gradient of approximately 37.5 km s$^{-1}$ pc$^{-1}$. When we calculate the velocity gradient using our model results (dv/dr = $\Delta n(H_2)/N(H_2)$), we find a 68% confidence lower limit of 16 km s$^{-1}$ pc$^{-1}$ if we use the model with $^{13}$CO), but the upper limit is unphysically high. The velocity required for turbulent heating seems reasonable in the context of our likelihood results. Panuzzo et al. (2010) used their calculated velocity gradient of 35 km s$^{-1}$ pc$^{-1}$ to determine that they could match the heating required with a sizescale of 0.3–1.6 pc. These velocity gradients seem large compared to Galactic star-forming sites (e.g., Imara & Blitz 2011), but M82 is known to have powerful stellar winds. According to Beirão et al. (2008), the starburst activity has decreased in the past few Myr, and this appears to be evidence of negative feedback (by stellar winds and supernovae), because M82 still has a large reservoir of gas available for SF. Therefore, there is evidence for turbulent heating mechanisms being in place. Additionally, Downes & Solomon (1998) found high turbulent velocities (30–140 km s$^{-1}$) in models of extreme star-forming galaxies.

None of the possibilities described seem to provide enough heating by themselves, with the exception of turbulent heating, which is based on a few approximations and assumptions. Likely, there is a combination of factors, namely PDRs and shocks/turbulent mechanisms. Such a situation has also been seen in other submillimeter-bright galaxies, discussed next. Interestingly, even a more quiescent galaxy like NGC 891 requires a combination of PDRs and shocks to explain mid-J CO transitions ($J = 6 → 5, J = 7 → 6$; Nikola et al. 2011).

4.5. Comparison to Other Starburst and Submillimeter Galaxies

Because only the first few lines in the CO ladder are easily visible from the ground for nearby galaxies, the high-J lines detected by Herschel represent new territory. Therefore, while adequate diagnostics of high-J CO lines are still being developed, it is useful to compare to other submillimeter-bright galaxies.

Mrk 231 contains a luminous (Seyfert 1) AGN. It also shows a strong high-J CO ladder, such that only the emission up to $J = 8 → 7$ is explained by UV irradiation from SF. Their high-J CO luminosity SLED, however, is flat (though ours for M82 are stronger than predicted, they are still decreasing with higher-J). Van der Werf et al. (2010) can explain this trend with either an XDR or a dense PDR. An additional difference between M82 and Mrk 231 is that OH$^+$ and H$_2$O$^+$ are both seen in strong emission in Mrk 231 (instead of absorption), indicative of X-ray-driven chemistry. Mrk 231 is also more face-on than M82.

The FTS spectrum of Arp 220 has many features not present in M82, such as strong HCN absorption, P-Cygni profiles of OH$^+$, H$_2$O$^+$, and H$_2$O, and evidence for an AGN (Rangwala et al. 2011). CO modeling, similar to the procedure done in this work, also indicates that the high-J lines trace a warmer component than low-J lines (~1350 K). Mechanical energy likely plays a large role in the heating of this merger galaxy as well. Though M82 has an outflow, it is not detected in P-Cygni profiles of the aforementioned lines.

The redshift of HLSW-01 allows the CO $J = 7 → 6$ to $J = 10 → 9$ lines to be observed from the ground, as has been done with Z-Spec (Scott et al. 2011). Unlike M82 (and others), the known CO SLED from $J = 1 → 0$ and up can be described by a single component at 227 K (1.2 × 10$^3$ cm$^{-3}$ density). If the velocity gradient is not constrained to be greater than or equal to that corresponding to virialized motion, then the best-fit solution becomes 566 K (0.3 × 10$^3$ cm$^{-3}$ density), closer to our temperature. We chose to exclude this prior due to uncertainties in the calculation of velocity gradient related to M82’s turbulent morphology. HLSW-01 appears to be unique in that a cold gas component is not required to fit the lower-J lines of the SLED, though two-component models can find a best-fit solution with a cold component.

In summary, M82 (like Arp 220 and HLSW-01) does not have the high CO excitation dominated by an AGN as seen in Mrk 231. Therefore, in addition to distinguishing between PDRs and shocks, high-J CO lines may also be used to indicate XDRs.
We have presented a multitude of molecular and atomic lines from M82 in the wavelength range (194–671 μm) accessible by the Herschel-SPIRE FTS (Table 1). After modeling $^{12}$CO, $^{13}$CO, and [C i], we find support for the high-temperature molecular gas component presented in the results of Panuzzo et al. (2010). The temperature traced by the warm component of $^{12}$CO is quite high (335–518 K), and the addition of $^{13}$CO slightly expands the likelihood ranges. The addition of [C i] produced results that indicate that the atom is not entirely tracing the same region as $^{12}$CO. Some of the emission from these molecules (especially [C i]) are likely tracing more diffuse gas less shielded from UV radiation.

The mapping observations did not resolve any significant gradients in physical parameters (except evidence for a slight drop-off in beam-averaged column density, consistent with the beam profile) indicating that the single “deep” spectrum is an adequate representation of the galaxy when limited by our beam size. However, the mapping observations were important in confirming the source–beam coupling factor utilized in Panuzzo et al. (2010) and here because, through convolution of the maps, we were able to confirm the central pixel’s results matched with the deep spectrum.

Molecular absorption traces lower column regions of the disk than those traced by CO emission, but contributes to the interpretation of the molecular gas of M82 being excited by a combination of sources. Despite the enhanced cosmic-ray density in M82, we do not find evidence that cosmic rays alone are sufficient to heat the gas enough to match the modeled hydrogen cooling rate. PDR models can only replicate the high-J CO line emission at high densities incompatible with those indicated by the likelihood analysis, though cosmic-ray-enhanced PDRs may be a closer match at lower densities. Turbulent heating from stellar winds and supernovae likely play a large role in the heating. More specifically, shocks are required to explain bright high-J line emission (Pon et al. 2012).

Like other submillimeter bright galaxies, Herschel has opened up new opportunities and questions about molecular and atomic lines that have never been observed before. Because of this, the diagnostic power of high-J CO lines is still in development, and newer models currently being developed may be able to explain the emission seen from M82 and other extreme environments.

SPIRE has been developed by a consortium of institutes led by Cardiff University (UK) and including: University of Lethbridge (Canada); NAOC (China); CEA, LAM (France); IFSI, University of Padua (Italy); IAC (Spain); Stockholm Observatory (Sweden); Imperial College London, RAL, UCL-MSSL, UKATC, University of Sussex (UK); and Caltech, JPL, NHSC, University of Colorado (USA). This development has been supported by national funding agencies: CSA (Canada); NAOC (China); CEA, CNES, CNRS (France); ASI (Italy); MCINN (Spain); SNSB (Sweden); STFC, UKSA (UK); and NASA (USA). J.K. also acknowledges the funding sources from the NSF GRFP. The research of C.D.W. is supported by grants from the Natural Sciences and Engineering Research Council of Canada. Thanks to the anonymous referee for comments which significantly improved this work.