The Astrophysical Journal, 693:216–235, 2009 March 1 doi:10.1088/0004-637X/693/1/216


# THE ATOMIC-TO-MOLECULAR TRANSITION IN GALAXIES. II: H I AND $H_2$ COLUMN DENSITIES

Mark R. Krumholz[1,2,5], Christopher F. McKee[3], and Jason Tumlinson[4]
[1] Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA; krumholz@ucolick.org
[2] Department of Astronomy & Astrophysics, University of California, Santa Cruz, Interdisciplinary Sciences Building, Santa Cruz, CA 95060, USA
[3] Departments of Physics and Astronomy, University of California, Berkeley, Campbell Hall, Berkeley, CA 94720-7304, USA; cmckee@astro.berkeley.edu
[4] Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA; tumlinson@stsci.edu
*Received 2008 August 26; accepted 2008 October 31; published 2009 March 2*

## ABSTRACT

Gas in galactic disks is collected by gravitational instabilities into giant atomic–molecular complexes, but only the inner, molecular parts of these structures are able to collapse to form stars. Determining what controls the ratio of atomic-to-molecular hydrogen in complexes is, therefore, a significant problem in star formation and galactic evolution. In this paper, we use the model of $H_2$ formation, dissociation, and shielding developed in the previous paper in this series to make theoretical predictions for atomic-to-molecular ratios as a function of galactic properties. We find that the molecular fraction in a galaxy is determined primarily by its column density and secondarily by its metallicity, and is to a good approximation independent of the strength of the interstellar radiation field. We show that the column of atomic hydrogen required to shield a molecular region against dissociation is $\sim 10\,M_\odot$ pc$^{-2}$ at solar metallicity. We compare our model to data from recent surveys of the Milky Way and of nearby galaxies, and show that the both the primary dependence of molecular fraction on column density and the secondary dependence on metallicity that we predict are in good agreement with observed galaxy properties.

*Key words:* galaxies: ISM – ISM: clouds – ISM: molecules – ISM: structure – molecular processes

## 1. INTRODUCTION

The formation of molecular hydrogen is a critical step in the transformation of interstellar gas into new stars. The neutral atomic interstellar medium (ISM) in galaxies is generally segregated into cold clouds embedded in a warm intercloud medium (McKee & Ostriker 1977; Wolfire et al. 2003), and the inner parts of some of these cold atomic clouds harbor regions where the gas is well shielded against dissociation by the interstellar radiation field (ISRF). In these regions molecules form, and once they do star formation follows.

A full theory of star formation requires as one of its components a method for expressing in terms of observables the fraction of a galaxy's ISM that is in the molecular phase (e.g. Krumholz & McKee 2005). No models published to date satisfy this requirement, but observations have yielded a number of empirical rules for galaxies' molecular content. Based on H I and CO mapping of nearby galaxies, Wong & Blitz (2002, hereafter WB02) and Blitz & Rosolowsky (2004, 2006, hereafter BR04 and BR06) infer that the molecular-to-atomic surface density ratio $R_{H_2} = \Sigma_{H_2}/\Sigma_{H\,I}$ in a galaxy varies with the interstellar pressure $P$ needed for hydrostatic balance in the ISM as $R_{H_2} \propto P^{0.92}$, and that the atomic surface density saturates at a maximum value of $\sim 10\,M_\odot$ pc$^{-2}$. The observed saturation and a similar dependence of molecular fraction on pressure are also seen in newer surveys such as the HERA CO-Line Extragalactic Survey (HERACLES) that cover a broader range of galaxy properties at higher spatial resolution (Walter et al. 2008; Bigiel et al. 2008; Leroy et al. 2008, hereafter L08). However, the physical origin of these patterns is unclear. The samples on which they are based are composed solely of nearby galaxies with a limited range of properties, and in the absence of a physical model it is uncertain how far they can safely be extrapolated to regimes of metallicity, surface density, or other properties not represented in samples of nearby galaxies.

Theoretical treatments of the problem to date do not yet make such an extrapolation possible. A number of authors have considered the microphysics of $H_2$ formation and the structure of photodissociation regions in varying levels of detail (e.g. van Dishoeck & Black 1986; Black & van Dishoeck 1987; Sternberg 1988; Elmegreen 1989; Draine & Bertoldi 1996; Neufeld & Spaans 1996; Spaans & Neufeld 1997; Hollenbach & Tielens 1999; Liszt & Lucas 2000; Liszt 2002; Browning et al. 2003; Allen et al. 2004), but none of these treatments addresses the problem of atomic-to-molecular ratios on galactic scales. Wyse (1986) and Wang (1990a, 1990b) present models for cloud formation in galactic disks, but these both rely on prescriptions for the rate of conversion of atomic-to-molecular gas that are based either on rates of cloud collisions or on Schmidt laws, not on physical models of $H_2$ formation and dissociation. Elmegreen (1993) gives a theory of the molecular fraction in galaxies that does include a treatment of the $H_2$ formation and self-shielding. However, his model neglects dust shielding, an order unity effect, and it also requires knowledge of a galaxy's ISRF strength, which cannot easily be determined observationally, and its interstellar pressure, which can only be inferred indirectly based on arguments about hydrostatic balance. This makes the model difficult to test or to apply as part of a larger theory of star formation. Schaye (2004) considers the conditions necessary to form a cold atomic phase of the ISM. The existence of such a phase is a necessary but insufficient condition for molecule formation, so while Schaye's model provides a minimum condition for star formation, it makes no statements about what fraction of the ISM goes into the molecular phase able to form stars. As a result, it is incapable of predicting, once star formation begins, at what rate it will occur.

Numerical models are in a similar situation. Hidaka & Sofue (2002) and Pelupessy et al. (2006) simulate galaxies using subgrid models for $H_2$ formation similar to those presented by Elmegreen (1993), and show that they can reproduce some qualitative features of the $H_2$ distribution in galaxies. Robertson & Kravtsov (2008) show that a simulation of a galaxy's ISM that

[5] Hubble Fellow.

includes radiative heating and cooling in the ionized and atomic phases, coupled with an approximate treatment of $H_2$ formation on grains and dissociation by the ISRF, can reproduce the observed molecular content of galaxies. This suggests that the simulations contain the necessary physical ingredients to explain the observations, but the simulations do not by themselves reveal how these ingredients fit together to produce the observed result. Moreover, like the observed empirical rules, the simulations are based on a very limited range of galaxy properties, and in the absence of a model we can use to understand the origin of the simulation results, it is unclear how to extrapolate. Extending the simulations to cover the full range of galaxy parameters in which we are interested would be prohibitively expensive in terms of both computational and human time.

Our goal in this paper is to remedy this lack of theoretical understanding by providing a first-principles theoretical calculation of the molecular content of a galactic disk in terms of direct observables. In Krumholz et al. (2008, hereafter Paper I), we lay the groundwork for this treatment by solving the idealized problem of determining the location of the atomic-to-molecular transition in a uniform spherical gas cloud bathed in a uniform, isotropic dissociating radiation field. In this paper, we apply our idealized model to atomic–molecular complexes in galaxies as a way of elucidating the underlying physical processes and parameters that determine the molecular content. We refer readers to Paper I for a full description of our solution to the idealized problem, but here we repeat a central point: for a spherical cloud exposed to an isotropic dissociating radiation field, if we approximate the transition from atomic to molecular as occurring in an infinitely thin shell separating gas that is fully molecular from gas of negligible molecular content, the fraction of a cloud's radius at which this transition occurs is solely a function of two dimensionless numbers:

$$\chi = \frac{f_{\rm diss}\sigma_{\rm d} c E_0^*}{n\mathcal{R}} \tag{1}$$

$$\tau_{\rm R} = n\sigma_{\rm d} R. \tag{2}$$

Here $f_{\rm diss} \approx 0.1$ is the fraction of absorptions of a Lyman–Werner (LW) band photons that produce $H_2$ dissociation rather than simply excitation and radiative decay to a bound state, $\sigma_{\rm d}$ is the dust absorption (not extinction) cross section per hydrogen nucleus in the LW band, $E_0^*$ is the free-space number density of LW photons (i.e. far outside our cloud), $n$ is the number density of hydrogen nuclei in the atomic shielding layer, $\mathcal{R}$ is the $H_2$ formation rate coefficient on dust grain surfaces, and $R$ is the cloud radius.

The quantity $\tau_{\rm R}$ is simply a measure of the size of the cloud. It is the dust optical depth that a cloud would have if its density throughout were equal to its density in the atomic region. We may think of $\chi$ as a dimensionless measure of the intensity of the dissociating radiation; formally, it is equal to the ratio of the rate at which LW photons are absorbed by dust grains to the rate at which they are absorbed by hydrogen molecules in a parcel of predominantly atomic gas in dissociation equilibrium in free space. This is a measure of the strength of the radiation field, because in strong radiation fields the gas contains very few molecules, so most LW photons are absorbed by dust and $\chi$ is large. In weak radiation fields, the gas contains more molecules, which due to their large resonant cross section dominate the absorption rate, making $\chi$ small.

Over the remainder of this paper, we apply the model of Paper I to atomic–molecular complexes in galaxies. In Section 2, we begin by considering giant clouds, which we may approximate as slabs, and in Section 3, we extend our treatment to clouds of finite size. In Section 4, we compare the model predictions of the previous two sections to observations of atomic and molecular gas. Finally, in Section 5, we summarize and discuss our conclusions.

## 2. THE ATOMIC ENVELOPES OF GIANT CLOUDS

In this section, we specialize to the case of giant clouds, which we define as those for which $e^{\tau_{\rm R}} \gg 1$. For these clouds, we show in Paper I that the dust optical depth from the cloud surface to the atomic–molecular transition, $\tau_{\rm H1}$, is a function of $\chi$ alone. We therefore begin our analysis with an estimate of $\chi$.

### 2.1. The Normalized Radiation Field

Of the quantities that enter into the normalized radiation field $\chi$, $f_{\rm diss}$ is the most certain, because it is only a very weak function of the spectrum of the dissociating radiation. We therefore take it to have a constant value $f_{\rm diss} \approx 0.1$ independent of environment (Draine & Bertoldi 1996; Browning et al. 2003; Paper I). Similarly, $\sigma_{\rm d}$ and $\mathcal{R}$ are functions of the properties of dust grains. These are both measures of the total surface area of dust grains mixed in the atomic shielding envelope around a molecular cloud; the former measures the area available for absorbing photons, while the latter measures the area available for catalyzing $H_2$ formation. There will, of course, be an additional dependence of these quantities on the optical and chemical properties of grains, but these effects likely provide only a small fraction of the total variation. To first order, therefore, we expect the ratio $\sigma_{\rm d}/\mathcal{R}$ to vary little with galactic environment, and we can simply adopt the value from the solar neighborhood. This is

$$\frac{\sigma_{\rm d}}{\mathcal{R}} = 3.2\times10^{-5}\frac{\sigma_{\rm d,-21}}{\mathcal{R}_{-16.5}}\ {\rm s\ cm^{-1}}, \tag{3}$$

where $\sigma_{\rm d,-21} = \sigma_{\rm d}/10^{-21}$ cm$^2$, $\mathcal{R}_{-16.5} = \mathcal{R}/10^{-16.5}$ cm$^3$ s$^{-1}$, and our best estimates for the solar neighborhood give $\sigma_{\rm d,-21} \approx \mathcal{R}_{-16.5} \approx 1$ (Draine & Bertoldi 1996; Wolfire et al. 2008).

Unfortunately, $n$ and $E_0^*$ are considerably harder to determine, since we cannot easily make direct measurements of the atomic density around a molecular cloud or the dissociating radiation field to which it is subjected, particularly for clouds in extragalactic space. (It is possible to determine these quantities for photodissociation regions (PDRs) being produced by individual star clusters—see Smith et al. (2000) and Heiner et al. (2008a, 2008b)—but these methods are generally not able to determine mean radiation fields around giant clouds.) However, we can still gain considerable insight into the ratio $E_0^*/n$ that enters into $\chi$ if we realize that $n$ is not free to assume any value for a given $E_0^*$. The atomic gas in a galaxy generally comprises regions of both cold and warm gas (cold neutral medium (CNM) and warm neutral medium (WNM), respectively) in approximate pressure balance (e.g. McKee & Ostriker 1977; Wolfire et al. 2003). Molecular clouds form in regions where the gas is primarily cold. This is because the effective opacity to LW photons provided by the small population of molecules found in a given element of predominantly atomic gas varies as $n^2$, so the cold phase, due to its higher density, is far more effective at shielding from LW photons than the warm phase. Thus, the $n$ we are concerned with is not the mean density of a galaxy's atomic ISM, it is the density in the cold phase only. In the presumably

dense gas in the vicinity of a molecular cloud, most of the mass is likely in the cold rather than the warm phase in any event.

We can estimate the CNM density by using the condition of pressure balance between the cold and warm phases. Wolfire et al. (2003) show that, for a given ambient far-ultraviolet (FUV) radiation intensity $G_0$ (given in units of the Habing (1968) field, corresponding to a number density $E_0^* \approx 4.4 \times 10^{-4}$ LW photons cm$^{-3}$), ionization rate from extreme ultraviolet (EUV) radiation and X-rays $\zeta_t$, abundance of dust and polycyclic aromatic hydrocarbons $Z_d$, and gas phase metal abundance $Z_g$, the minimum number density $n_{\rm min}$ at which CNM can exist in pressure balance with WNM is well approximated by

$$n_{\rm min} \approx 31 G_0' \frac{Z_d'/Z_g'}{1 + 3.1(G_0' Z_d'/\zeta_t')^{0.365}} \ {\rm cm}^{-3}, \quad (4)$$

where the primes denote quantities normalized to their values in the solar neighborhood. Wolfire et al. (2003) obtain this expression by constructing a temperature–density relation, determined by balancing the rate of grain photoelectric heating against cooling by the fine structure lines of C II and O I. Once they have constructed the $T-n$ curve, they identify the temperature at which the pressure is minimized. This is the warmest temperature at which the CNM can be in pressure balance, and thus the corresponding density is the lowest possible CNM density. The primary uncertainty in this expression arises from the abundance, size distribution, and reaction properties of polycyclic aromatic hydrocarbons (PAHs), but changes in PAH properties generally change $n_{\rm min}$ only at the factor of $\sim 2$ level (see Figure 8 of Wolfire et al. 2003).

In a galaxy where young stars provide the dominant source of radiation and the IMF is constant, the FUV heating rate and the EUV/X-ray ionization rate are likely to be proportional to the star-formation rate, and therefore to each other. We therefore assume that $\zeta_t' = G_0'$. Furthermore, if the physics of dust formation does not vary strongly from galaxy to galaxy, then the dust and gas phase metal abundances are likely proportional to the total metallicity $Z$, so we adopt $Z_d' = Z_g' = Z'$. With these approximations, the minimum CNM density becomes

$$n_{\rm min} \approx 31 \frac{G_0'}{1 + 3.1 Z'^{0.365}} \ {\rm cm}^{-3}. \quad (5)$$

We caution at this point that both the assumptions that $\zeta_t' = G_0'$ and $Z_d' = Z_g' = Z'$ are unlikely to hold in elliptical galaxies, where young stars are not the dominant sources of EUV or X-ray radiation, and where the amount of dust per unit metallicity is known to be different than in spirals. Thus, Equation (5) is unlikely to hold in ellipticals.

The CNM can exist in pressure balance at densities higher than $n_{\rm min}$, so we take the typical CNM density to be

$$n_{\rm CNM} = \phi_{\rm CNM} n_{\rm min}. \quad (6)$$

We adopt $\phi_{\rm CNM} \approx 3$ as our fiducial value, which gives a CNM density of $n_{\rm CNM} = 22$ cm$^{-3}$ and (using Wolfire et al.'s $T-n$ relation, Equation (18) of this paper) a temperature of $T_{\rm CNM} = 105$ K, consistent with observations of typical CNM properties in the solar neighborhood. (Near a GMC, we expect $G_0' \sim 10$ rather than $G_0' = 1$, due to the proximity of sites of star formation (M.G Wolfire et al. 2009, in preparation) but this does not affect our results, since we only care about the ratio $G_0'/n$.) In practice, $\phi_{\rm CNM}$ cannot be much larger than this, because pressure balance between the CNM and WNM is possible only over a limited range of CNM densities. If the CNM densities exceeds $n_{\rm min}$ by more than a factor of $\sim 10$, the CNM and WNM again cannot be in pressure balance because it is impossible for the warm phase to have a high enough pressure.

Using Equation (3) for $\sigma_{\rm d}/\mathcal{R}$ and Equation (6) for $n$, and noting that the LW photon number density in the solar neighborhood is roughly $7.5 \times 10^{-4}$ cm$^{-3}$ (Draine 1978; Paper I), we find a total estimate for the dimensionless radiation field strength

$$\chi = 2.3 \left( \frac{\sigma_{\rm d,-21}}{\mathcal{R}_{-16.5}} \right) \frac{1 + 3.1 Z'^{0.365}}{\phi_{\rm CNM}}. \quad (7)$$

Note that all explicit dependence on the dust properties, the radiation field, and the atomic gas density have cancelled out of this expression.

Dependence on the dust properties has dropped out for the simple reason explained above: $\sigma_{\rm d}$ and $\mathcal{R}$ are both measures of the dust surface area, so their ratio is nearly constant. We can understand the somewhat more subtle reason that dependence on radiation field and the atomic gas density cancel by examining the physics behind Equation (4). As noted already, the minimum possible density in the cold atomic phase of the ISM corresponds to the density and temperature at which the pressure reaches a local minimum. Because the dependence of the cooling rate on gas temperature is determined almost entirely by the quantum mechanical constants and element abundances that determine the shapes of the C II and O I cooling curves, and the photoelectric heating rate is essentially independent of temperature, the temperature at which this minimum pressure occurs is nearly fixed at $\approx 240$ K, and does not depend on the background radiation field (c.f. Equation (34) and Appendix C of Wolfire et al. 2003). Thus, the density minimum will simply be the density at which the temperature reaches $\approx 240$ K. Since the heating rate varies as $nE_0^*$ and the cooling rate as $n^2$, it immediately follows that the density at which a fixed temperature is reached varies as $n \propto E_0^*$. This explains why $n/E_0^*$ is nearly constant in the CNM. There is only a weak dependence on metallicity, which arises because the heating rate depends on the charge state of PAHs, and this in turn depends weakly on metallicity.

Before moving on, we should note that we have neglected the role of internal radiation in determining where a cloud changes from atomic to molecular. This is justified because most stars that contribute significant amounts of dissociating radiation are born in molecular clouds, but they do not stay internal to those clouds for very long. Most dissociating photons come from massive stars born in clusters that burrow their way out of their parent molecular clouds via their winds and H II regions in only a few Myr. Thus, most of the dissociating radiation to which a molecular cloud is subjected is delivered externally rather than internally, even if it comes from stars born in that cloud.

This does, however, raise another cautionary point. We have computed the CNM density based on an implicit assumption of pressure balance, and we must consider under what circumstances pressure balance might not hold. One situation in which gas might not reach pressure balance is if it is subjected to hydrodynamic perturbations such as supernova shocks that create rapid and substantial changes in pressure, pushing gas into the unstable regime of density and pressure. Such gas is subject to an instability in which pockets of stable CNM condense within it, leaving behind a lower density ambient medium that expands

to become stable WNM (e.g. Audit & Hennebelle 2005). Thus, the typical gas density will be significantly different than the value $n_{\rm CNM}$ that we have estimated only if the time between successive shocks that drive gas into instability is small compared to the time required for this instability to operate. This is of order the cooling timescale, which Wolfire et al. (2003) estimate to be

$$t_{\rm cool} \approx 7.7 \left(\frac{T}{10^4 \ {\rm K}}\right)^{1.2} \left(\frac{nT}{3000 \ {\rm K \ cm^{-3}}}\right)^{-0.8} {\rm Myr}, \quad (8)$$

where $n$ and $T$ are the gas number density and temperature. For typical CNM conditions near the solar circle this is $\sim$ 40 kyr, while for typical WNM conditions it is slightly under 10 Myr. Higher values of $G_0'$, as are expected near GMCs, reduce these to a few kyr and a few Myr, respectively. Thus, perturbations that produce velocities $\lesssim$ 5 km s$^{-1}$, i.e., not fast enough to induce shocks in the WNM, are essentially ineffective at driving the gas out of equilibrium since any CNM gas they disturb will re-equilibrate very quickly. Stronger perturbations such as supernova blast waves that drive WNM unstable can keep substantial amounts of gas out of pressure balance only if they recur on timescales of a few Myr or less. For both supernova blast waves and the shocks induced by the ubiquitous turbulence in the atomic ISM, the recurrence time is $\sim$ 10 Myr, larger (although not hugely so) than the equilibration time (Wolfire et al. 2003). We can therefore conclude that the typical atomic envelope around a molecular cloud is likely to be close to pressure balance between CNM and WNM. Some envelopes will have been subjected to a strong shock in the last few Myr, and these may have CNM densities substantially different than $n_{\rm CNM}$, but they will be in the minority.

Alternatively, one could consider a galaxy in which the average galactic environment is so extreme that no two-phase equilibrium is possible, i.e. the pressure is so low that only WNM is stable, or the pressure is so high that only CNM is stable. Wolfire et al. and Schaye (2004) both find that a pressure so low that only WNM exists is consistent with vertical hydrostatic balance only in the very diffuse outer parts of galactic disks. For the Milky Way, Wolfire et al. find that a CNM phase can exist everywhere the mean ISM density is $\gtrsim 0.2$ cm$^{-3}$, which is everywhere in the Milky Way inside $\sim$ 15 kpc in Galactocentric radius; Schaye estimates that hydrostatic balance requires the existence of a CNM phase any time the local gas surface density exceeds 3–10 $M_\odot$ pc$^{-2}$, which is true over a similar region. Thus we can conclude that a pure WNM is unlikely to exist anywhere except in the far outer regions of galactic disks; in such regions, Equation (7) underestimates the dimensionless radiation field strength, and thus we will overestimate the molecular fraction. The converse possibility is a galaxy in which the pressure is so high that no WNM is present, only CNM; however, Wolfire et al. find that WNM can be present any time the mean ISM density $\lesssim$ 70 cm$^{-3}$, unless the CNM is entirely confined by a surrounding hot ionized medium. The case where no WNM exists because $n \gtrsim 70$ cm$^{-3}$ corresponds to an entire galaxy whose mean density matches that of a typical molecular cloud in the Milky Way, and such conditions are only found in starburst systems where the molecular fraction is essentially unity. In such cases, our models will provide only an upper limit to the amount of H I present, but even then our model may apply if the atomic shielding layer occurs far from the galactic midplane where the density is lower. The latter possibility of an ISM consisting solely of cold atomic gas and hot ionized gas appears not to be realized in nature. Thus, in summary, we expect our two-phase model to apply everywhere in galaxies except in their far outer parts, where the density is so low that no cold phase exists, and at the midplanes of starburst galaxies, where the density is so high that no warm phase can exist.

### 2.2. The Shielding Column

The normalized radiation field $\chi$ is the primary factor controlling the size of the atomic gas column that is needed to shield a molecular cloud against dissociation, and thus in determining what fraction of the gas in a galaxy is molecular. This quantity in turn determines what fraction of a galaxy's ISM is molecular, and, therefore, available for star formation, and what fraction is atomic, and thus without star formation. The invariance of $\chi$ across galactic environments has important consequences for atomic–molecular complexes. First, $\chi$ measures the relative importance of dust shielding and $H_2$ self-shielding, so our result that $\chi \sim 1$ across galaxies implies that dust and self-shielding contribute nearly equally in essentially all galactic environments. We do not expect to find clouds where either dust or molecular shielding completely dominate except near strong sources of dissociating radiation where the atomic ISM is not in pressure balance or in regions of such low or high density that the atomic ISM does not have two phases.

Second, the dust optical depth through the H I shielding layer, $\tau_{\rm H\,I}$, is primarily a function of $\chi$; the dependence on $\tau_{\rm R}$ arises from geometric effects, and is greater than order unity only if the molecular region inside a cloud is a small fraction of its size. (Strictly speaking this is the optical depth of the CNM, not of all the atomic gas. As noted above, most of the atomic gas around a molecular cloud is probably CNM, but there could be significant amounts of WNM along the line of sight that is not associated with that cloud, and which does not contribute to shielding it.) This implies that the dust optical depth through the atomic envelopes of molecular clouds should be roughly constant across galactic environments, at least as long as this optical depth is not close to that of the entire atomic–molecular complex. The only significant variation will be a weak increase in $\tau_{\rm H\,I}$ with metallicity. Correspondingly, the total H I gas column will decrease with metallicity to a power less than unity, since $\tau_{\rm H\,I}$ increases slightly with metallicity, but the column of H I required to achieve this dust optical depth decreases with metallicity. To illustrate these effects, we solve for $\tau_{\rm H\,I}$ as a function of $Z$ in the limit $\tau_{\rm R} \rightarrow \infty$, using the formalism of Paper I, and plot the result in Figure 1.

As the plot shows, for our fiducial model $\phi_{\rm CNM} = 3$, we predict that the H I layer around a molecular cloud in the Milky Way, $Z' = 1$, should have an absorption optical depth of $\tau_{\rm H\,I} = 0.40$ to LW photons, corresponding to an H I column $N_{\rm H\,I} = 4.0 \times 10^{20}$ cm$^{-2}$ (mass column density $\Sigma = 4.5 \ M_\odot$ pc$^{-2}$), assuming a dust absorption cross section per H nucleus of $\sigma_{\rm d} = 10^{-21}$ cm$^{-2}$. It is important to note that all of these values represent the absorption column on *one side* of a giant cloud. A 21 cm observation would detect the shielding column on both sides for a cloud exposed to the ISRF on both sides, so the detected H I column would be *double* the values given in Figure 1. As one can see in Section 3, the column is somewhat larger for a cloud of finite size.

It is important to note that the shielding column we have calculated is somewhat different than the atomic-to-molecular transition column density reported for the Milky Way by Savage et al. (1977, $\log N(H) = 20.7$) and for the Large Magellanic Cloud (LMC) and Small Magellanic Cloud (SMC) by Tumlinson et al. (2002, $\log N(H) \geqslant 21.3$ and $\geqslant 22$,

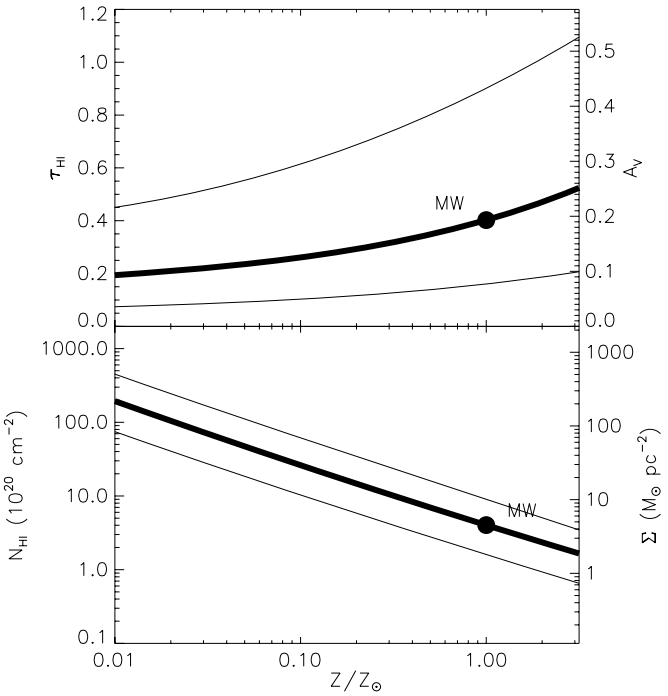


**Figure 1.** The dust optical depth of the H I shielding layer $\tau_{\rm H\,I}$ (upper panel) and the corresponding H I column $N_{\rm H\,I}$ (lower panel) for very large clouds, $\tau_{\rm R} \rightarrow \infty$. We show these results computed for $\phi_{\rm CNM} = 3$ (thick lines) and for $\phi_{\rm CNM} = 1$ and 10 (upper and lower thin lines, respectively). The circles indicate our values for the Milky Way, $Z' = 1$, for our fiducial $\phi_{\rm CNM} = 3$: $\tau_{\rm H\,I} = 0.40$, $N_{\rm H\,I} = 4.0 \times 10^{20}$. To compute $N_{\rm H\,I}$ from $\tau_{\rm H\,I}$, we assume a dust absorption cross section per H nucleus in the LW band of $\sigma_{\rm d} = 10^{-21} Z/Z_\odot$. We also show the visual extinction $A_V$ corresponding to our $\tau_{\rm H\,I}$ and the mass column density $\Sigma$ corresponding to $N_{\rm H\,I}$. For the former we have assumed $A_V/\tau_{\rm H\,I} = 0.48$, following the models of Draine (2003a, 2003b, 2003c) as explained in the text. To compute the latter we assume a mean particle mass per H nucleus of $2.34 \times 10^{-24}$ g cm$^{-2}$, corresponding to a standard cosmic mixture of H and He.

respectively). These values are the total column densities along pencil-beam lines of sight at which the fraction of the gas column in the form of $H_2$ reaches about 10% of the total. In contrast, in the two-zone approximation we adopt in Paper I, we assume that the atomic-to-molecular transition is infinitely sharp, and under this approximation the shielding column we report is the column at which the gas goes from fully atomic to fully molecular. Were the transition truly infinitely sharp as we have approximated it to be, the ratio of $H_2$ to total column density would be zero at our computed shielding column. Comparing our theoretical shielding columns to the detailed numerical radiative transfer models we present in Paper I shows that in reality, for conditions typical of the Milky Way, the ratio of $H_2$ column to total column at our calculated transition column $N_{\rm H\,I}$ is roughly 20%. Since this is a factor of 2 larger than the 10% ratio used in the observationally defined transition column, and in our simple model the $H_2$ fraction increases linearly with total column density once we pass our predicted transition point, we expect our shielding column to be a factor of $\sim 2$ larger than the values reported by Savage et al. and Tumlinson et al. We compare our model predictions to these data sets in more detail in Section 4.2.

Using the extinction and absorption curves of Draine (2003a, 2003b, 2003c), the ratio of visual extinction to 1000 Å absorption is $A_V/\tau_{\rm H\,I} = 0.48$ for Draine's $R_V = 4.0$ model, so the visual extinction corresponding to $\tau_{\rm H\,I} = 0.40$ is $A_V = 0.19$. Adopting the $R_V = 5.5$ curve instead, appropriate for denser clouds, gives $A_V = 0.28$, while $R_V = 3.1$, for diffuse regions, gives $A_V = 0.13$. Our estimates for the LW dust optical depth and visual extinction vary little with metallicity, changing by only a factor of 2.7 for a metallicity ranging from $10^{-2}\ Z_\odot$ to $10^{0.5}\ Z_\odot$.

The variation between the curves with $\phi_{\rm CNM} = 1$, 3, and 10 shows the full plausible range of variation in shielding column arising from our uncertainty about the true density in the atomic envelopes of molecular clouds. The $\phi_{\rm CNM} = 1$ and 10 curves are both within a factor of 2.6 of the fiducial model, so this is an upper bound on our uncertainty. The actual error is likely to be smaller than this, since $\phi_{\rm CNM} = 1$ and 10 correspond to the extreme assumptions that the CNM assumes its minimum or maximum possible equilibrium densities.

We can obtain a quick approximation to the results shown in Figure 1 simply by noting that at solar metallicity our fiducial normalized radiation field is $\chi = 3.1$, and we show in Paper I that for a giant cloud with $\chi < 4.1$ (corresponding to $Z' < 2.5$ for our fiducial parameters), the LW dust optical depth through the atomic shielding layer is

$$\tau_{\rm H\,I} = \frac{\psi}{4}, \tag{9}$$

where

$$\psi = \chi \frac{2.5 + \chi}{2.5 + \chi e}. \tag{10}$$

The dust-adjusted radiation field $\psi$ is a function only of metallicity; for our fiducial parameters $\phi_{\rm CNM} = 3$ and $\sigma_{\rm d,-21}/\mathcal{R}_{-16.5} = 1$, and Milky Way metallicity $Z' = 1$, we obtain $\psi = 1.6$. Moreover, the dependence on metallicity is weak: at $Z' = 1/10$, $\psi = 1.0$, while at $Z' = 1/100$, $\psi = 0.77$. Because $\psi$ depends only on metallicity, we can also express the characteristic H I shielding column on one side of a giant cloud solely as a function of $Z'$:

$$\Sigma_{\rm H\,I} = \frac{\mu_{\rm H}}{\sigma_{\rm d}} \tau_{\rm H\,I}(Z', \phi_\chi) \tag{11}$$

$$= 4.5\ M_\odot\ {\rm pc}^{-2} \frac{f(Z', \phi_\chi)}{\sigma_{0,-21} Z'}, \tag{12}$$

where $\mu_{\rm H} = 2.34 \times 10^{-24}$ g is the mean mass per hydrogen nucleus, and $\sigma_{0,-21}$ is the dust absorption cross section at Milky Way metallicity ($\log Z' = 0$) in units of $10^{-21}$ cm$^2$. The function $f(Z', \phi_\chi)$ is given by

$$f(Z', \phi_\chi) = 0.54(0.32 + Z'^{0.385})\phi_\chi^{-1} \times \left(\frac{1.05\phi_\chi + 0.42 + Z'^{0.385}}{0.39\phi_\chi + 0.42 + Z'^{0.385}}\right), \tag{13}$$

where

$$\phi_\chi \equiv \left(\frac{\phi_{\rm CNM}}{3}\right)\left(\frac{\mathcal{R}_{-16.5}}{\sigma_{\rm d,-21}}\right), \tag{14}$$

$f(1, 1) = 1$, and for our fiducial parameters $\phi_\chi = 1$. The numerical factors that appear in $f(Z', \phi_\chi)$ are derived simply by substituting Equation (7) for $\chi$ into Equation (10) and thence into Equation (11). These equations, and therefore the numerical values in the function $f$, depend solely on microphysical constants that describe the properties of molecular hydrogen and the chemistry of its formation on grain surfaces (which set $f_{\rm diss}$ and $\sigma_{\rm d}/\mathcal{R}$) and the shapes of the C II and O I cooling curves (which set the ratio $E_0^*/n$). We have therefore calculated the shielding column to a good approximation solely in terms of microphysical constants.

## 3. THE ATOMIC ENVELOPES OF FINITE CLOUDS

### 3.1. Formulation of the Problem

To account for the fact that clouds have finite sizes and column densities, and that these can be quite small in dwarf galaxies or other low-pressure environments, we must examine the second dimensionless number that characterizes $H_2$ formation and shielding: $\tau_R = n\sigma_d R$. Consider a cloud of known, fixed column density $\Sigma_{comp}$. If atomic–molecular complexes were of uniform density, then we could find $\tau_R$ simply by multiplying $\Sigma_{comp}$ by the dust cross section per unit mass $\sigma_d/\mu_H$, where $\mu_H \approx 2.34 \times 10^{-24}$ g is the mean mass per H nucleus. However, the atomic region is warmer and has a lower mean mass per particle than the molecular one, and thus has a correspondingly lower density. This reduces the dust optical depth through it. Since it is the density and dust optical depth through the atomic shielding layer that matters, we must estimate $\tau_R$ using the value of $n$ appropriate for the atomic gas rather than the mean density in the complex. In other words, the quantity we want is

$$\tau_R = n_{CNM}\sigma_d R, \tag{15}$$

and we define

$$\phi_{mol} \equiv \frac{n_{mol}}{n_{CNM}} \tag{16}$$

as the ratio of densities. Here $n_{mol}$ and $n_{CNM}$ are the number densities of hydrogen nuclei in the molecular and CNM phases of the ISM, respectively.

In the Milky Way, typical molecular cloud densities are $n_{mol} \approx 100\,cm^{-3}$ (McKee & Ostriker 2007), while observations of the giant H I clouds around these molecular regions find typical densities $n_{CNM} \approx 10\ cm^{-3}$ (Elmegreen & Elmegreen 1987), suggesting $\phi_{mol} \approx 10$. We do not expect this ratio to vary strongly between galaxies, so we should generally find $\phi_{mol} \approx 10$.

We can can make an independent argument for $\phi_{mol} \approx 10$ by considering thermal pressure balance across the atomic–molecular interface. This argument only applies to gas near an atomic–molecular transition surface, which may or may not include the bulk of the gas in a cloud, but it does provide an estimate for the density ratio near the interface. Pressure balance requires that

$$\phi_{mol} = 1.8\frac{T_{CNM}}{T_{mol}}, \tag{17}$$

where $T_{mol}$ and $T_{CNM}$ are the temperatures in the molecular and cold neutral atomic media, respectively, and the factor of 1.8 accounts for the difference in the mean number of particles per H nucleus in the two phases. Across a very wide range of galactic environments, the temperature in the molecular phase of the ISM is $T_{mol} \approx 10 - 20\,K$, as a result of the balance between grain photoelectric heating and CO cooling. Since we are interested in gas at the edge of the molecular region, we adopt $T_{mol} = 20\,K$ as typical. (Temperatures are somewhat higher in starburst systems, but in these galaxies the molecular fraction is essentially unity in any event.) Using the model of Wolfire et al. (2003) for the atomic medium and making the approximations $Z'_d = Z'_g = Z'$ and $\zeta'_t = G'_0$ as in Section 2.1, the density–temperature relation in the atomic gas is

$$n_{CNM} \approx \frac{20\,G'_0 T_{CNM,2}^{-0.2} e^{1.5/T_{CNM,2}}}{1 + 2.6\left(T_{CNM,2}^{1/2} Z'\right)^{0.365}}, \tag{18}$$

where $T_{CNM,2} = T_{CNM}/(100\,K)$. Combining this with Equations (5) and (6) enables us to write an implicit equation for the CNM temperature in terms of $\phi_{CNM}$ and $Z'$:

$$\frac{20\,T_{CNM,2}^{-0.2} e^{1.5/T_{CNM,2}}}{1 + 2.6\left(T_{CNM,2}^{1/2} Z'\right)^{0.365}} = \phi_{CNM}\frac{31}{1 + 3.1 Z'^{0.365}}. \tag{19}$$

Substituting the solution to this equation in Equation (17) immediately gives us $\phi_{mol}$, the ratio of the number densities of H nuclei in the molecular and CNM gas. For our fiducial $\phi_{CNM} = 3$, we find $\phi_{mol} = 9.6$ at $Z' = 1$, varying by only a few percent for metallicities in the range $Z' = 10^{-2} - 10^{1}$. Given the encouraging agreement between this and the value $\phi_{mol} \approx 10$ we find observationally, we adopt the value of $\phi_{mol}$ given by Equations (17) and (19) as our standard value for the remainder of this work.

A spherical cloud that consists of a molecular core of number density $n_{mol}$ and an outer atomic envelope of number density $n_{CNM}$ has a mean column density

$$\Sigma_{comp} = \frac{4}{3}\mu_H n_{CNM} R\left[1 + (\phi_{mol} - 1)\,x_{H_2}^3\right], \tag{20}$$

where $x_{H_2}$ is the fraction of the cloud radius at which it transitions from molecular to atomic, i.e. $x_{H_2} = 1$ corresponds to a cloud that is molecular throughout and $x_{H_2} = 0$ to one that is atomic throughout (see Paper I). It is convenient to rewrite this in terms of an optical depth

$$\tau_c \equiv \frac{3}{4}\left(\frac{\Sigma_{comp}\sigma_d}{\mu_H}\right) \tag{21}$$

$$\rightarrow 0.067\,Z'\Sigma_{comp,0}, \tag{22}$$

where $\Sigma_{comp,0} = \Sigma_{comp}/(1\,M_\odot\,pc^{-2})$, and the arrow in the second step indicates that we have used our fiducial $\sigma_d = 10^{-21}\,Z'\,cm^2$. Equations (20) and (21) imply

$$\tau_c = \tau_R\left[1 + (\phi_{mol} - 1)\,x_{H_2}^3\right]. \tag{23}$$

Note that neither $\tau_R$ nor $\tau_c$ is the true center-to-edge dust optical depth of the complex; $\tau_R$ is the optical depth the complex would have if its density were $n_{CNM}$ throughout, and $\tau_c$ is the optical depth it would have if its atomic and molecular gas were mixed uniformly rather than spatially segregated.

We are now in a position to compute the shielding column and the atomic and molecular fractions in finite clouds. If we consider a complex of a given column density $\Sigma_{comp}$ and metallicity $Z'$, and we take the dust opacity to be given by its Milky Way value adjusted for metallicity, $\sigma_d = \sigma_{d,MW}Z'$ with $\sigma_{d,MW} = 10^{-21}\ cm^{-2}$, then Equations (21) and (23) give one constraint on the unknowns $\tau_R$ and $x_{H_2}$ from pressure balance between the atomic and molecular phases. Dissociation–formation equilibrium, as computed in Paper I, gives a second constraint. We show in Paper I how to compute the value of $x_{H_2}$ for a given $\tau_R$ and $\chi$: its value is given implicitly by the solution to Equations (33) and (37), or (43) and (44), of that paper. Since we have already computed $\chi$ in terms of the metallicity (Equation (7)), a choice of $\Sigma_{comp}$ and $Z'$ fully determine the two unknowns $\tau_R$ and $x_{H_2}$.

### 3.2. Numerical Solution

We can either solve this system of nonlinear algebraic equations numerically, or approximate the solution analytically.

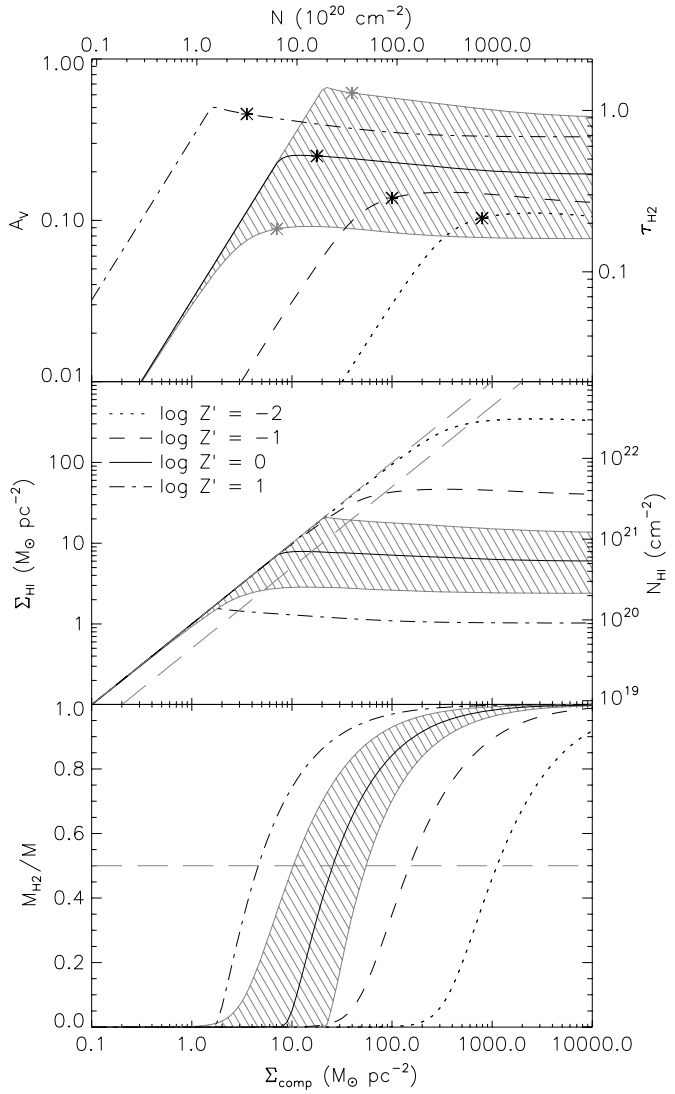


**Figure 2.** Visual extinction $A_V$ (top panel), H I column density $\Sigma_{\rm H\,I}$ (middle panel), and $H_2$ fraction $M_{H_2}/M$ (bottom panel) in finite cloud complexes, as a function of complex mass column density $\Sigma_{\rm comp}$ (or number column density $N$). In each panel the curves shown are for metallicities $Z'$ running from $10^{-2}-10^{1}$, as indicated. The hatched regions centered on the $Z' = 0$ curve indicate the range of models with $\phi_{\rm CNM} = 1-10$, while all other curves are for our fiducial value $\phi_{\rm CNM} = 3$. In the top panel, the asterisks on the curves indicate the value of $\Sigma_{\rm comp}$ for which the molecular fraction falls to 1/2. The curves to the left of this point should be regarded as lower limits on the visual extinction to the molecular region. In the middle panel the two parallel dashed lines indicate ratios of $M_{\rm H\,I}/M = 1/2$ and 1. Since values of $M_{\rm H\,I}$ below 1/2 are lower limits, once our models curves cross the lower of these lines, their true solution could be anywhere between them. Similarly, the dashed line in the bottom panel corresponds to $M_{H_2}/M = 1/2$, and when the model curves fall below this they should be treated as upper limits.

We first show the results of a numerical calculation for a variety of values of $\Sigma_{\rm comp}$ and $Z'$ in Figure 2. Rather than giving $\tau_{\rm R}$ and $x_{H_2}$ directly, which are not particularly interesting because we cannot measure them directly, we plot three derived quantities of interest which in at least some circumstances we can observe: the LW optical depth and visual extinction from the cloud surface to the atomic–molecular transition surface along a radial path, the H I column averaged over the entire cloud, and the total $H_2$ mass fraction over the entire cloud.

It is important to point out that the value of $A_V$ we report is measured differently than the H I column density $\Sigma_{\rm H\,I}$, or than the column densities we will use in Section 4. The column density is measured by averaging the mass per unit area over the entire complex, while $A_V$ is measured along a single pencil beam from the surface of the cloud to the atomic–molecular transition along a radial trajectory. The former quantity is more analogous to what is measured in an observation using a telescope beam that only marginally resolves or does not resolve a complex, while the latter is more closely analogous to a measurement of the extinction of a background point source through a cloud. We also caution that, for reasons we discuss in Section 3.3, our predictions are only accurate for molecular mass fractions $> 1/2$. (This is in the worst case of very low metallicity and intermediate $\phi_{\rm CNM}$; our accuracy range expands as metallicity increases toward solar and as $\phi_{\rm CNM}$ gets smaller or larger than 3.) Below this limit, our calculations yield only upper limits on the molecular fraction, not firm predictions. This confidence limit is shown in Figure 2.

The plots immediately yield a number of interesting results. First, our prediction of nearly constant $A_V$ through the atomic shielding layers around molecular clouds continues to hold whenever there is a significant molecular fraction, even for finite clouds. Our prediction of a characteristic $A_V \approx 0.2$ through atomic shielding envelopes of molecular clouds therefore continues to apply.

We also find that there is a saturation in the H I column density at roughly 6 $M_\odot$ pc$^{-2}$ for solar metallicity, which rises by a factor of a somewhat less then 10 for every decade by which the metallicity declines. The H I column saturates simply because once $\Sigma_{\rm comp}$ is large enough, the cloud column densities become so large that they are effectively in the infinite cloud limit. At this point, the shielding column is geometry independent, and is determined solely by the normalized strength of the radiation field, $\chi$, a value that does not vary much from galaxy to galaxy. Once $\Sigma_{\rm comp}$ is sufficiently large to put a complex in the large cloud limit, adding additional mass simply increases the size of the shielded molecular layer, so the $H_2$ fraction just rises smoothly. The saturation value of 6 $M_\odot$ pc$^{-2}$ at solar metallicity is set by a combination of the fundamental constants describing $H_2$ formation and dissociation, the shape of the C II and O I cooling curves (which determine the CNM density and temperature), and the properties of interstellar dust grains, which set the ratio $\sigma_{\rm d}/\mathcal{R}$.

### *3.3. Geometric Uncertainties for Finite Clouds*

In Section 4.7 of Paper I we show that our method for determining the molecular abundance in finite clouds suffers from a systematic uncertainty arising from our imperfect knowledge of the opacity along rays that pass through the atomic envelope of a cloud. For our fiducial model, we take the opacity along these rays due to molecules mixed into the atomic gas to be set by the value of the dissociation radiation field at the surface of the zone where molecules dominate the opacity. Our results depend on this approximation very little except at low molecular volume fraction, $x_{H_2}^3 \lesssim 0.2$; in that case, the uncertainty about this approximation means that our model enables us to predict only an upper limit on the molecular fraction, not an exact value.

In this paper, we are concerned with molecular mass rather than volume fractions, so we must quantify that uncertainty. To do so, we proceed as in Section 4.7 of Paper I: we adopt the opposite assumption that opacity along rays passing through the region of the cloud where molecules dominate the opacity but still constitute a small fraction of all H nuclei is infinite. We then repeat the calculations of Section 3 following this assumption: for a given $\Sigma_{\rm comp}$ and $Z'$, rather than solve the

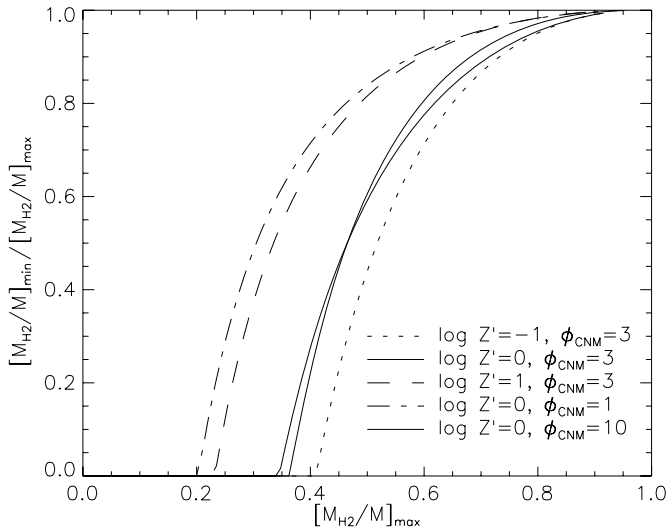


**Figure 3.** Ratio of minimum to maximum predicted $H_2$ mass fraction vs. maximum predicted $H_2$ mass fraction for a variety of values of $Z'$ and $\phi_{\rm CNM}$, as indicated. Curves for $\log Z' < -1$ are not shown because they are indistinguishable from those for $\log Z' = -1$.

system of equations formed by Equation (23) of this paper and Equations (33) and (37) or (43) and (44) of Paper I, we instead solve Equation (23) together with Equations (69) and (70) or (71) of Paper I. Doing so gives a lower bound on the molecular content for a given $\Sigma_{\rm comp}$ and $Z'$. By comparing the results in this case to our fiducial calculations as presented in Section 3, we obtain an estimate of the uncertainty of our results.

In Figure 3, we show the results of this exercise. On the $y$-axis we show the ratio of the $H_2$ mass fraction predicted using our maximum opacity assumption, which represents the minimum possible molecular content, divided by the value produced by our fiducial model, which represents the maximum. This gives an estimate of our uncertainty. The $x$-axis indicates the $H_2$ mass fraction predicted using the fiducial assumption we make elsewhere in the paper. As the plot shows, the two calculations differ most at low $Z'$ and intermediate $\phi_{\rm CNM}$. In this case, the calculations differ by a factor of a few for $H_2$ mass fractions below $\sim 0.5$. If we adopt a factor of 3 as an accuracy goal, this means that for cases where we predict an $H_2$ mass fraction below 0.5, and at low metallicity, our predictions should be taken only as upper limits. For solar metallicity or higher our confidence range extends down to molecular mass fractions around 0.4, and we attain upper limits below this.

### 3.4. Analytic Approximation

We can gain additional insight into the behavior of the solution by constructing an analytic approximation. The ratio of the molecular mass $M_{\rm H_2}$ to the total complex mass $M$ is

$$f_{\rm H_2} \equiv \frac{M_{\rm H_2}}{M} = \frac{\phi_{\rm mol} x_{\rm H_2}^3}{1 + (\phi_{\rm mol} - 1)\, x_{\rm H_2}^3}. \tag{24}$$

We wish to obtain an approximation for this in terms of the known quantities $\psi$ (given in terms of metallicity by Equations (7) and (10)) and $\tau_{\rm c}$ (given in terms of complex column density by Equation (21)). We show in Paper I that for $\psi \lesssim 3$ and molecular volume fractions $x_{\rm H_2}^3 \gtrsim 0.15$, a range in parameter space that includes most of our models for realistic cloud parameters, the molecular volume is well approximated by

$$x_{\rm H_2}^3 \approx 1 - \frac{3\psi}{4\tau_{\rm R\Delta}}, \tag{25}$$

where for convenience we have defined

$$\tau_{\rm R\Delta} \equiv \tau_{\rm R} + a\psi, \tag{26}$$

and $a = 0.2$ is a numerical parameter that is optimized for agreement between the approximate and numerical solutions. Substituting this approximation into the condition for pressure balance, Equation (23), gives

$$\tau_{\rm c} = \tau_{\rm R}\left[\phi_{\rm mol} + \frac{3\psi}{4\tau_{\rm R\Delta}}\,(1 - \phi_{\rm mol})\right]. \tag{27}$$

As in Paper I, our approach to obtaining an analytic solution is to perform a series expansion in $a$. We therefore define

$$\tau_{\rm c\Delta} \equiv \tau_{\rm c}\left(1 + \frac{a\psi}{\tau_{\rm R}}\right), \tag{28}$$

so that $\tau_{\rm c\Delta}/\tau_{\rm c} = \tau_{\rm R\Delta}/\tau_{\rm R}$. Using this definition of $\tau_{\rm c\Delta}$ together with Equation (27) for $\tau_{\rm c}$ implies that

$$\tau_{\rm c\Delta} = \phi_{\rm mol}\left(\tau_{\rm R\Delta} - \frac{3}{4}\psi\right) + \frac{3}{4}\psi. \tag{29}$$

If we now use our approximation (25) and rewrite the result in terms of $\tau_{\rm c\Delta}$ using Equation (29), we obtain

$$f_{\rm H_2} = 1 - \frac{3\psi}{4\tau_{\rm c\Delta}}. \tag{30}$$

We must now express $\tau_{\rm c\Delta}$ in terms of $\psi$, $\tau_{\rm c}$, and $a$ alone. Thus,

$$\tau_{\rm c\Delta} = \tau_{\rm c} + a\psi\left(\frac{\tau_{\rm c}}{\tau_{\rm R}}\right) \tag{31}$$

$$= \tau_{\rm c} + a\psi\left(\frac{\tau_{\rm c\Delta}}{\tau_{\rm R\Delta}}\right). \tag{32}$$

The second term on the right-hand side still involves the unknowns $\tau_{\rm c\Delta}$ and $\tau_{\rm R\Delta}$, but because they are already multiplied by $a$, we now need only determine them to zeroth order in $a$. Solving Equation (29) for $\tau_{\rm R\Delta}$ gives

$$\frac{\tau_{\rm R\Delta}}{\tau_{\rm c\Delta}} = \frac{1}{\phi_{\rm mol}} + \left(1 - \frac{1}{\phi_{\rm mol}}\right)\frac{3\psi}{4\tau_{\rm c\Delta}} \tag{33}$$

$$\approx \frac{1}{\phi_{\rm mol}} + \left(1 - \frac{1}{\phi_{\rm mol}}\right)\frac{3\psi}{4\tau_{\rm c}}, \tag{34}$$

where, in the second step, we have dropped a term of order $a$ to obtain an expression that is accurate to zeroth order in $a$. Substituting this into Equation (32), and thence into Equation (30) gives our final expression for the molecular mass fraction, accurate to first order in $a$:

$$f_{\rm H_2} = 1 - \frac{3\psi}{4\tau_{\rm c}}\left[1 + \frac{4a\psi\phi_{\rm mol}}{4\tau_{\rm c} + 3(\phi_{\rm mol} - 1)\psi}\right]^{-1}. \tag{35}$$

Comparison of this approximate expression with the numerical solution illustrated in Figure 2 shows that for our fiducial

$\phi_{\rm CNM} = 3$ and metallicities from $Z' = 10^{-2} - 10$, it is accurate to better than 30% whenever the approximation analytic solution gives $f_{\rm H_2} > 0.25$, but that it goes to zero too sharply at low molecular fraction. We can improve the approximation by forcing the $H_2$ fraction to approach zero smoothly rather than sharply at low column density. Experimentation shows that the expression

$$f_{\rm H\,I}^{-3} = 1 + \left\{ \left( \frac{4\tau_{\rm c}}{3\psi} \right) \left[ 1 + \frac{4a\psi\phi_{\rm mol}}{4\tau_{\rm c} + 3(\phi_{\rm mol} - 1)\psi} \right] \right\}^3 \quad (36)$$

matches the numerical result for $f_{\rm H\,I} \equiv M_{\rm H\,I}/M$ for $\phi_{\rm CNM} = 3$ to better than 20% for all $Z' < 10$ regardless of the value of $f_{\rm H\,I}$. (However, note that, as we show in Section 3.3, for $f_{\rm H_2} \lesssim 1/2$ our estimate of the molecular content is only an upper limit, and this is true of Equation (36) as well.) Using Equation (21) to replace $\tau_{\rm c}$ with $\Sigma_{\rm comp}$, and substituting in our fiducial values $\phi_{\rm CNM} = 5$, $a = 0.2$, and $\sigma_{\rm d} = 10^{-21} Z'$ cm$^{-2}$, Equation (36) becomes

$$f_{\rm H\,I} \rightarrow \left[ 1 + \left( \frac{s}{11} \right)^3 \left( \frac{125 + s}{96 + s} \right)^3 \right]^{-1/3}, \quad (37)$$

where

$$s \equiv \frac{\Sigma_{\rm comp,0} Z'}{\psi} \quad (38)$$

and $\Sigma_{\rm comp,0} = \Sigma_{\rm comp}/(1\ M_\odot\ {\rm pc}^{-2})$. Note that our result indicates that, to a good approximation, the molecular content of an atomic–molecular complex depends only on the combination of input parameters $Z'\Sigma_{\rm comp,0}/\psi$; the numerator $Z'\Sigma_{\rm comp,0}$ is simply the dust column density of the complex up to a scaling factor, while the denominator $\psi$ is the dimensionless radiation field which Equations (7) and (10) give solely as a function of metallicity.

From Equation (37), it also immediately follows that the $H_2$ to H I ratio $R_{\rm H_2} \equiv f_{\rm H_2}/f_{\rm H\,I}$ is

$$R_{\rm H_2} \approx \left[ 1 + \left( \frac{s}{11} \right)^3 \left( \frac{125 + s}{96 + s} \right)^3 \right]^{1/3} - 1. \quad (39)$$

For $R_{\rm H_2} > 1$, which is the regime for which our models apply with high confidence, we can use an even simpler expression

$$R_{\rm H_2} \approx 0.08 s = 0.08 \frac{\Sigma_{\rm comp,0} Z'}{\psi}, \quad (40)$$

which is accurate to $\sim 30\%$.

Similarly, we show in Paper I that the dust absorption optical depth through the atomic layer for a finite cloud is well approximated by

$$\tau_{\rm H\,I} \approx \frac{\psi}{4} \left[ \frac{1}{1 - (a'/4)(\psi/\tau_{\rm R})} \right], \quad (41)$$

where $a' = \frac{3}{2} - 4a = 0.7$. If we treat $a'/4$ as a small parameter and perform a series expansion around it, then we need only approximate $\tau_{\rm R}$ to zeroth order in $a$. We can do this simply by using Equation (34) with $a = 0$, which allows us to set $\tau_{\rm R} = \tau_{\rm R\Delta}$ and $\tau_{\rm c} = \tau_{\rm c\Delta}$. Thus, to zeroth order in $a$, we have

$$\tau_{\rm R} = \tau_{\rm R\Delta} \left[ \frac{1}{\phi_{\rm mol}} + \left( 1 - \frac{1}{\phi_{\rm mol}} \right) \frac{3\psi}{4\tau_{\rm c}} \right], \quad (42)$$

and to first order in $a$ or $a'$, we have

$$\tau_{\rm H\,I} = \frac{\psi}{4} \left[ 1 - \frac{a'\psi\phi_{\rm mol}}{4\tau_{\rm c} + 3(\phi_{\rm mol} - 1)\psi} \right]^{-1}. \quad (43)$$

As with approximation (35) for the molecular mass fraction, this expression works well whenever the molecular fraction is not too low, and may be improved by forcing the optical depth to approach the total cloud optical depth smoothly when the column density becomes low. The expression

$$\tau_{\rm H\,I}^{-2} = \tau_{\rm c}^{-2} + \frac{16}{\psi^2} \left[ 1 - \frac{a'\psi\phi_{\rm mol}}{4\tau_{\rm c} + 3(\phi_{\rm mol} - 1)\psi} \right]^2 \quad (44)$$

is accurate to better than 35% for all $Z' < 10$, and to better than 25% for $Z' < 1$.

It is also convenient to invert our analytic expressions to determine column density as a function of molecular content and metallicity. The term $(125 + s)/(96 + s)$ in Equation (37) is generally close to unity except at extremely high column densities, so if we neglect second-order corrections to the difference between this term and unity, we can solve Equation (37) for $f_{\rm H\,I}$ to obtain

$$\frac{\Sigma_{\rm comp,0} Z'}{\psi} \rightarrow 11 \left( f_{\rm H\,I}^{-3} - 1 \right)^{1/3} \frac{8.7 + \left( f_{\rm H\,I}^{-3} - 1 \right)^{1/3}}{11 + \left( f_{\rm H\,I}^{-3} - 1 \right)^{1/3}}. \quad (45)$$

This expression matches the numerical solution at the $\sim 20\%$ level for $f_{\rm H\,I} < 0.75$. Note that this result implies that the column density $\Sigma_{\rm comp}$ at which a given H I fraction is reached depends on metallicity both explicitly through the $Z'$ term in the numerator, representing the effect of metallicity on dust content, and implicitly through $\psi$ (Equations (7) and (10)), representing the effect of metallicity on the ratio of radiation intensity to CNM density. Since $\psi$ is an increasing function of metallicity, for a given molecular fraction $\Sigma_{\rm comp}$ has a weaker than linear dependence on metallicity. For example, evaluating Equation (45) with $f_{\rm H\,I} = 0.5$ at solar metallicity $Z' = 1$ indicates that we expect the gas to be half molecular for complexes with $\Sigma_{\rm comp} = 27\ M_\odot\ {\rm pc}^{-2}$. (The exact numerical solution is $\Sigma_{\rm comp} = 25.5\ M_\odot\ {\rm pc}^{-2}$.) At one-third solar metallicity, $Z' = 1/3$, half molecular content is reached at $\Sigma_{\rm comp} = 67\ M_\odot\ {\rm pc}^{-2}$ (using Equation (45); numerically $\Sigma_{\rm comp} = 55.3\ M_\odot\ {\rm pc}^{-2}$), somewhat less than a factor of 3 higher.

## 4. COMPARISON WITH OBSERVATIONS

Our model makes strong predictions for the relative fractions of H I and $H_2$ as a function of total surface density and metallicity, and in this section we compare these results to a variety of galactic and extragalactic observations.

### 4.1. Extragalactic Observations

#### 4.1.1. Data Sets

We use three extragalactic data sets for comparison to our models. Two are recent surveys that have mapped nearby galaxies in 21 cm H I and 2.6 mm CO(1 → 0) emission at overlapping positions, and therefore provide an ideal laboratory in which to test our model. The first of these is the work of WB02, BR04, and BR06, who report H I and $H_2$ surface densities on a pixel-by-pixel basis in 14 nearby galaxies (including the Milky Way). The $H_2$ surface densities are inferred from CO

observations taken as part of the BIMA SONG survey (Regan et al. 2001; Helfer et al. 2003), while the H I observations are from the VLA. WB02 and BR06 supplement these data with stellar surface density measurements from 2MASS (Jarrett et al. 2003), which together with the equations given in BR06 can be used to derive a pressure in each pixel, if one assumes that the gas is in hydrostatic balance, that the stellar-scale height greatly exceeds the gas-scale height, and that the gas velocity dispersion has a known value. The galaxies in the sample are all molecule-rich spirals with metallicities within 0.5 dex of solar.

The second extragalactic data set we use is compiled by L08, who combine H I measurements from the THINGS survey with CO data partly taken from BIMA SONG and partly from the ongoing HERACLES survey. The authors also include 2MASS stellar surface densities in their compilation, and give mean pressures. Unlike the BR06 sample, the data reported are averages over galactocentric rings rather than individual pixels, although point-by-point maps at sub-kpc resolution are in preparation (F. Walter, 2008, private communication). The sample includes 23 galaxies, of which roughly half are large spirals and roughly half are low-mass, H I-dominated dwarfs. The galaxies in the data set partly overlap with those in the sample of BR06, but extend over a wider range of metallicities and molecule fractions.

We summarize the galaxies in the samples in Table 1. We also report metallicities for each galaxy where these are available in the literature. We do not include NGC 4414 in the analysis because no gas phase metallicity is available for it. In the comparison that follows, we neglect the presence of metallicity gradients within these galaxies because gradients are only available for some of them. On top of this, we note that the metallicities themselves are probably uncertain at levels from 100ths to 10ths of a dex, depending on the galaxy and the analysis technique used, which adds additional scatter on top of that already introduced by our neglect of metallicity gradients.

In both data sets, the uncertainty in the H I column densities is generally $\sim$ 10%. In the CO data, the formal uncertainties are generally $\sim$ 20%–30%, but the dominant uncertainty is probably systematic: the *X* factor used to convert observed CO luminosities into $H_2$ column densities. This is uncertain at the factor of $\sim$ 2 level (Blitz et al. 2007), and almost certainly varies with metallicity (e.g. Bolatto et al. 2008). Thus, although for clarity we will suppress error bars in the plots that follow, recall that the molecular data are uncertain at the factor of $\sim$ 2 level.

The third data set we use in Section 4.1.4 is the *Spitzer* survey of the Small Magellanic Cloud ($S^3$MC) (Bolatto et al. 2007; Leroy et al. 2007). This survey differs from the SONG and THINGS data sets in that those surveys infer the presence of $H_2$ via CO emission, whereas $S^3$MC measures molecular hydrogen using measurements of dust from *Spitzer* combined with H I measurements by Stanimirović et al. (1999) and Stanimirović et al. (2004). The basic idea behind the technique is that one determines the dust-to-gas ratio in a low column density region where molecules are thought to contribute negligibly to the total column. Then by comparing the H I and dust-column density maps, one can infer the presence of $H_2$ in pixels where the total dust column exceeds what one would expect for the observed H I column and a fixed dust-to-gas ratio. The reason for using this technique is that, at the low metallicity of the SMC ($\log[O/H] + 12 = 8.0$ (Dufour 1984), i.e. 0.76 dex below solar), CO may cease to be a reliable tracer of molecular gas. The SMC represents the lowest metallicity galaxy for which we have $H_2$ detections rather than upper limits; the BR06 sample does not contain any galaxies with metallicities as low as the SMC, and no CO was detected in any of the L08 galaxies with metallicities comparable to or lower than the SMC. Thus, the SMC represents a unique opportunity to test our models at very low metallicity.

Before comparing to these data sets, it is worth commenting briefly on one additional extragalactic data set to which we will not compare our models: observations of $H_2$ column densities along sightlines in the LMC and SMC using *FUSE* (Tumlinson et al. 2002). We do not use this data set for comparison because it includes only sightlines with low column densities that are strongly dominated by atomic gas; the highest reported $H_2$ fraction is below 10%. Sightlines with significantly higher molecular content than this absorb too much background starlight to allow *FUSE* to make a reliable measurement of the $H_2$ column. The low column densities of the clouds that *FUSE* can observe in these galaxies place almost all of them into the regime where our theory yields only upper limits. Those limits

**Table 1**
Galaxy Metallicities

| Galaxy | log(O/H) + 12[a] | Sample[b] | Reference |
|---|---|---|---|
| Solar (Milky Way) | 8.76 | B | 7 |
| DDO154 | 7.67 | L | 3 |
| HOI | 7.54 | L | 6 |
| HOII | 7.68 | L | 6 |
| IC10 | 8.26 | B | 1 |
| IC 2574 | 7.94 | L | 6 |
| NGC 0598 | 8.49 | B | 5 |
| NGC 0628 | 8.51 | L | 5 |
| NGC 0925 | 8.32 | L | 5 |
| NGC 2403 | 8.39 | L | 5 |
| NGC 2841 | 8.81 | L | 5 |
| NGC 2976 | 8.30 | L | 8 |
| NGC 3077 | 8.64 | L | 8 |
| NGC 3184 | 8.72 | L | 5 |
| NGC 3198 | 8.42 | L | 5 |
| NGC 3351 | 8.80 | L | 5 |
| NGC 3521 | 8.49 | BL | 5 |
| NGC 3627 | 9.25 | BL | 4 |
| NGC 4214 | 8.22 | L | 2 |
| NGC 4321 | 8.71 | B | 5 |
| NGC 4414 | ... | B | ... |
| NGC 4449 | 8.31 | L | 2 |
| NGC 4501 | 8.78 | B | 5 |
| NGC 4736 | 8.50 | BL | 5 |
| NGC 5033 | 8.68 | B | 5 |
| NGC 5055 | 8.68 | BL | 5 |
| NGC 5194 | 8.75 | BL | 5 |
| NGC 5457 | 8.44 | B | 5 |
| NGC 6946 | 8.53 | L | 5 |
| NGC 7331 | 8.48 | BL | 5 |
| NGC 7793 | 8.34 | L | 5 |

**Note.**
[a] We take the metallicity relative to solar to be proportional to the O/H ratio, i.e. $\log Z' = [\log(O/H) + 12] - 8.76$.
[b] B = galaxy is in BR06 sample; L = galaxy is in L08 sample.
**References.** 1: Garnett (1990); 2: Martin (1997); 3: van Zee et al. (1997); 4: Ferrarese et al. (2000); 5: Pilyugin et al. (2004); 6: Walter et al. (2007); 7: Caffau et al. (2008); 8: Walter et al. (2008). An entry ... indicates that there is no gas-phase metallicity reported in the literature.

are generally consistent with the data, but the comparison is not particularly illuminating.

#### 4.1.2. Column Density and Metallicity Dependence

In this section, we compare our models predictions of the H I and $H_2$ mass fractions as a function of total column density and metallicity to our two extragalactic data sets. To perform this comparison, for convenience we first break the sample into four metallicity bins: $-1.25 < \log Z' <$-0.5, $-0.5 < \log Z' <$-0.25, $-0.25 < \log Z' < 0$, and $0 < \log Z' < 0.5$, where $Z' = Z/Z_\odot$ and we adopt $\log(\mathrm{O/H}) + 12 = 8.76$ as the value corresponding to solar metallicity (Caffau et al. 2008). These bins roughly evenly divide the data; we do not include the galaxies for which metallicities are not available in the literature.

Next we must consider how finite spatial resolution will affect our comparison. Given the beam sizes in the observed data, which range from $\sim 0.3$ kpc for the nearest galaxies to a few kpc for the most distant, each pixel (for BR06) or ring (for L08) in the observed data set is likely to contain multiple atomic–molecular complexes. The atomic and molecular observations are convolved to the same resolution, so this does not bias the measurement of the atomic-to-molecular ratio, although it does mean that this ratio is measured over an averaging scale set by the beam size. Since individual atomic–molecular complexes presumably represent peaks of the galactic column density, however, the total (H I + $H_2$) observed gas surface density $\Sigma_{\rm obs}$ reported for each pixel or ring represents only a lower limit on $\Sigma_{\rm comp}$, the total surface density of individual complexes. In our model, the fraction of the cloud in a molecular form is a strictly increasing function of $\Sigma_{\rm comp}$ (and of metallicity $Z$), while the atomic fraction is a strictly decreasing function. Because we have only the observed column density $\Sigma_{\rm obs}$ available, not the column density of an individual complex $\Sigma_{\rm comp}$, we have no choice but to take $\Sigma_{\rm comp} = \Sigma_{\rm obs}$ as our best estimate. The error in this approximation probably ranges from tens of percent for nearby galaxies where the beam size is not much larger than the size of a complex up to an order of magnitude or more for ring averages in distant galaxies. As a result, though, we expect to overestimate the atomic fraction and underestimate the molecular fraction. Physically, we may think of this as a clumping effect: the clumpier the gas is, the better it is able to shield itself against dissociating radiation. Since our observations smooth over scales larger than the characteristic gas clumping scale, we will miss this effect.

There are also two effects which go in the other direction, however. First, as noted above, shielding comes primarily from cold atomic gas, not warm gas. Although most of the gas in the immediate vicinity of a single molecular region is probably cold, observations that average over many molecular cloud complexes are likely to include a fair amount of WNM gas as well. Since we have only the total H I column densities including both phases, we are prone to overestimate $\Sigma_{\rm comp}$ and, therefore, the molecular fraction, because the WNM raises the H I column but does not provide much shielding. This effect is not important at moderate to high column densities, where the atomic gas does not dominate the total mass budget, but it could be significant at lower densities.

Second, there may be gas along our line of sight through a galaxy that is not associated with an atomic–molecular complex along that line of sight. This effect will increase in severity as the galaxy comes close to edge-on, since this will increase the path length of our line of sight through the galaxy. As with WNM gas, this extra material contributes to $\Sigma_{\rm obs}$ but not to the complex surface density $\Sigma_{\rm comp}$, and it therefore leads us to overestimate the molecular fraction. We can perform a very simple calculation to estimate the size of this effect. Consider a simple self-gravitating gas disk characterized by the standard vertical density profile $n \propto \mathrm{sech}^2(0.88z/h)$, where $h$ is the half-height of the gas. At some point in this disk is an atomic–molecular complex, centered at the midplane. Since complexes are found preferentially at the midplane unless the galaxy is very close to edge-on, then we need not consider the possibility of our line of sight intersecting multiple independent complexes. Since the complex is formed from a large-scale gravitational instability in the disk, its characteristic size is $\sim h$, and we therefore consider gas to be "associated" with the complex if it is within a distance $h$ of it in the plane of the disk. Suppose this galaxy has an inclination $i$. Making the worst-case assumption, that there is no density enhancement due to the presence of the self-gravitating complex, we can then compute what fraction of the gas we see along our line of sight is not associated with it. This is simply

$$\frac{\int_{h\cot i}^{\infty} \mathrm{sech}^2(0.88z/h)\,dz}{\int_0^{\infty} \mathrm{sech}^2(0.88z/h)\,dz} = 1 - \tanh(1.13\cot i). \quad (46)$$

This is less than 0.5 for all inclinations less than 64°. Only a handful of the galaxies in the SONG and THINGS surveys have inclinations larger than this, so even in the worst-case scenario, where complexes do not represent any enhancement of the gas density, we expect nonassociated gas to produce an error smaller than a factor of $\sim 2$ for the great majority of the galaxies to which we are comparing.

Given the limitations imposed by finite resolution, we proceed as follows. Using the method described in Section 3, we compute the molecular mass fraction

$$f_{\rm H_2}(\Sigma_{\rm comp}, Z) \equiv \frac{M_{\rm H_2}}{M} \quad (47)$$

as a function of complex column density $\Sigma_{\rm comp}$ and metallicity $Z$. (For these and all subsequent predictions we use our fiducial value of $\phi_{\rm CNM} = 3$.) We then generate predicted H I and $H_2$ column densities for each metallicity bin and each observed column density $\Sigma$ via

$$\Sigma_{\rm H_2,predicted} = f_{\rm H_2}(\Sigma_{\rm obs}, Z_{\rm max})\Sigma_{\rm obs}, \quad (48)$$

$$\Sigma_{\rm H\,I,predicted} = \left[1 - f_{\rm H_2}(\Sigma_{\rm obs}, Z_{\rm max})\right]\Sigma_{\rm obs}, \quad (49)$$

where $Z_{\rm max}$ is the maximum metallicity for that bin. Since $f_{\rm H_2}(\Sigma_{\rm comp}, Z)$ is a strictly increasing function of $\Sigma_{\rm comp}$ and $Z$, and we know that $\Sigma_{\rm obs} < \Sigma_{\rm comp}$ and $Z < Z_{\rm max}$, we expect $f_{\rm H_2}(\Sigma_{\rm obs}, Z_{\rm max})$ to be a lower limit on the true molecular fraction. We therefore expect $\Sigma_{\rm H_2,predicted}$ to be a lower limit on the observed data, and $\Sigma_{\rm HI,predicted}$ to be an upper limit. The possible exception to this statement is at low column densities, where a significant fraction of the H I column may be in the form of WNM.

We plot the data against our theoretical prediction for the upper envelope of $\Sigma_{\rm H\,I}$ in Figure 4, and we show the corresponding predicted lower envelopes for $\Sigma_{\rm H_2}$ and $R_{\rm H_2} \equiv \Sigma_{\rm H_2}/\Sigma_{\rm H\,I}$ in Figures 5 and 6. For the BR06 data set, rather than plotting the tens of thousands of individual pixels it contains, for each galaxy we show the data averaged over 20 logarithmically spaced column density bins, running from the minimum to the maximum

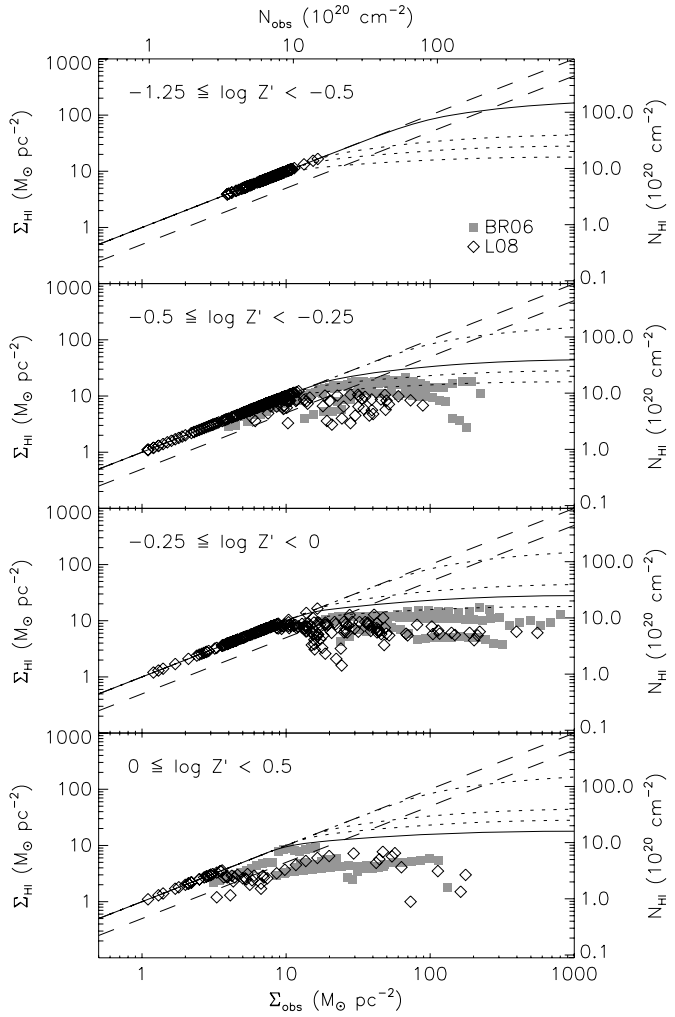


**Figure 4.** H I column density $\Sigma_{\rm H\,I}$ vs. total column density $\Sigma$ for galaxies in metallicity bins $-1.25 < \log Z' < -0.5$, $-0.5 < \log Z' < -0.25$, $-0.25 < \log Z' < 0$, and $0 < \log Z' < 0.5$, as indicated. In each panel we plot the values of $\Sigma_{\rm H\,I}$ and $\Sigma$ from the samples of BR06 (squares) and L08 (diamonds) and show our model predictions of $\Sigma_{\rm H\,I}$ as a function of $\Sigma$ for $\log Z' = -0.5$, $-0.25$, 0, and 0.5 (lines, highest to lowest). The curve for the value of $\log Z'$ equal to the maximum $\log Z'$ for each bin, which should represent the upper envelope of the data, is shown as a solid line. The rest are shown as dotted lines. The four lines are the same in each panel. To maximize readability we omit the error bars on the data points. The parallel slanted dashed lines, $\Sigma_{\rm H\,I} = \Sigma_{\rm obs}/2$ and $\Sigma_{\rm H\,I} = \Sigma_{\rm obs}$, show the range of values of $\Sigma_{\rm H\,I}$ for which our predictions should be treated as upper limits for the reasons discussed in Section 3.3. Parts of our model curves above the $\Sigma_{\rm H\,I} = \Sigma_{\rm obs}/2$ line could be as high as the $\Sigma_{\rm H\,I} = \Sigma_{\rm obs}$ line.

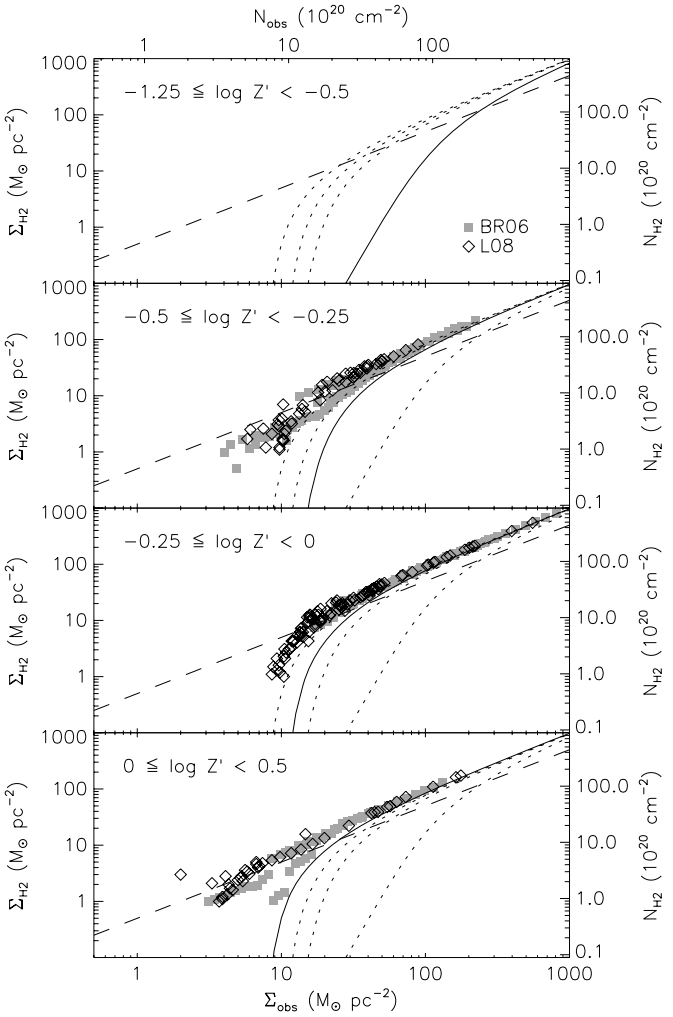


**Figure 5.** Same as Figure 4, except that we show $H_2$ rather than H I column densities. For clarity we do not show nondetections. Note that $N_{\rm H_2}$ is the column density of H nuclei in molecular form, which is twice the column density of $H_2$ molecules. The slanted dashed line is $\Sigma_{\rm H_2} = \Sigma_{\rm obs}/2$; as discussed in Section 3.3, above this line our model curves may be taken as predictions, while below it they should be taken as upper limits.

value of $\Sigma_{\rm obs}$ reported for that galaxy. As the plots show, our model predictions for the upper envelope of the H I surface density and the corresponding lower envelope of the $H_2$ surface density as a function of total surface density and metallicity agree very well with the data. The data fill the space up to our predicted envelopes, but for the most part do not cross them, even when the predicted H I envelope becomes flat for $\Sigma_{\rm obs} \gtrsim 10\ M_\odot\ {\rm pc}^{-2}$. Moreover, our model recovers not only the primary dependence of $\Sigma_{\rm H\,I}$ on the total observed column density $\Sigma_{\rm obs}$, but also the secondary dependence on metallicity. For example, rings in the lowest metallicity bin in the L08 data set reach total mean column densities of almost 20 $M_\odot\ {\rm pc}^{-2}$, but still show no detectable molecular component. On the other hand, in the highest metallicity bin, the molecular fraction is close to 70% in rings with $\Sigma_{\rm obs} \approx 20 M_\odot\ {\rm pc}^{-2}$. Our models reproduce this effect: at a metallicity of $\log Z' = -0.5$, we predict that the gas will be 94% atomic even at a surface density of 20 $M_\odot\ {\rm pc}^{-2}$, whereas for $\log Z' = 0.5$ we predict an atomic fraction of only 30% at that column density, in agreement with the data.

We caution that we can only predict upper limits on the molecular content in regions of parameter space when our predicted molecular fraction falls below $\sim 1/2$, for the reasons discussed in Section 3.3. We have indicated the regions where our model predictions convert to upper limits in Figures 4–6. Alternatively, one can express this uncertainty as giving a minimum column density at which we can predict a value rather than an upper limit for molecular content. For reference, at the metallicities of $\log Z' = -1.25$, $-0.5$, $-0.25$, 0, and 0.5, which define the edges of our metallicity bins, the minimum column densities for which we can predict numerical values to better than a factor of few confidence are 250, 58, 38, 25, and 11 $M_\odot\ {\rm pc}^{-2}$, respectively.

Finally, we note that the future HERACLES/THINGS data set represents an opportunity to perform an even stronger test of our model. In Figure 4, a significant fraction of the data points falls below our predicted upper limits, and correspondingly these points are above our lower limits in Figures 5 and 6.

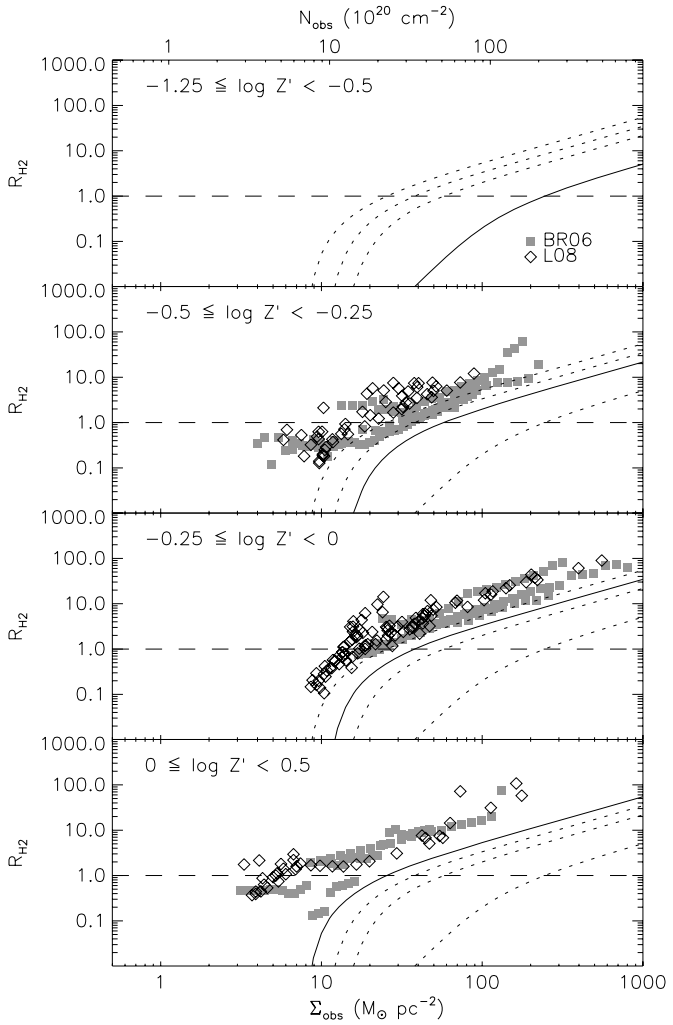


**Figure 6.** Same as Figures 4 and 5, except that we show $R_{H_2} \equiv \Sigma_{H_2}/\Sigma_{H\,I}$ rather than column densities. For clarity we do not show nondetections. The horizontal dashed line corresponds to $R_{H_2} = 1$; as discussed in Section 3.3, above this line our model curves may be taken as predictions, while below it they should be taken as upper limits.

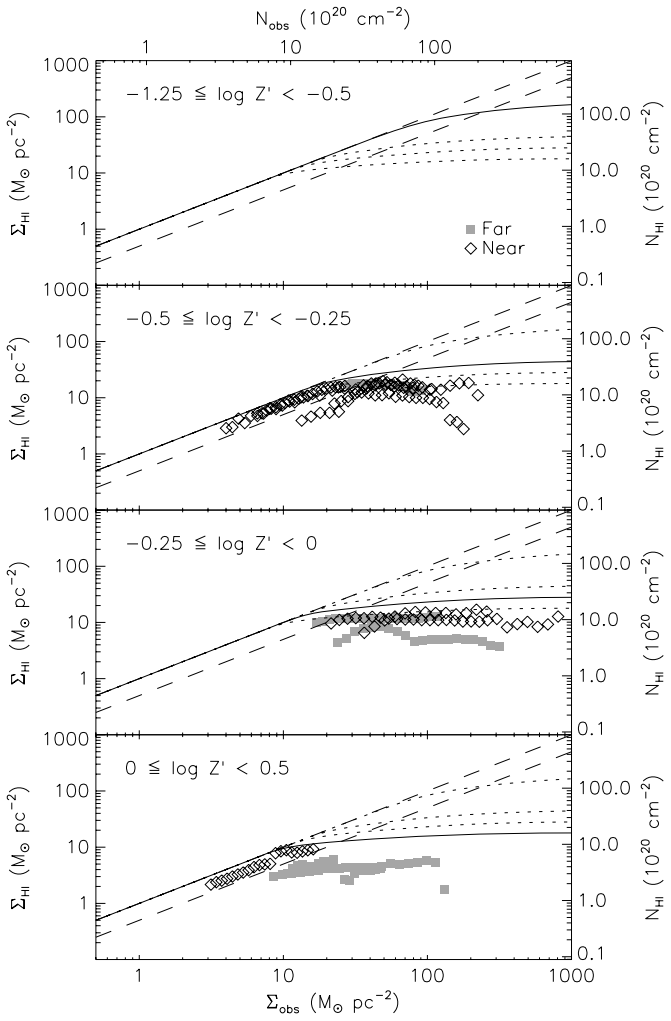


**Figure 7.** Same as Figure 4, except that we show only data from BR06, and we have divided this data set into "near" galaxies, those for which the survey resolution is $< 1$ kpc, and "far" galaxies, those for which the resolution is $> 1$ kpc.

We hypothesize that these data points represent rings or pixels within which the gas is significantly clumped, so that the averaged column density $\Sigma_{obs}$ seen in the observation is significantly lower than the column density at which most of the molecular gas in that beam or ring is found, and our calculation for a complex with $\Sigma_{comp} = \Sigma_{obs}$ overestimates the H I column. In reality, these regions probably consist of patches of high column density where most of the molecules reside, embedded in a lower density ambient medium that has a lower molecular fraction than we determine by averaging over large scales. If we could observe these regions at higher resolution, in Figure 4 the high column, high molecular content points will lie on the right of and slightly above the low-resolution points, since both the total and H I column densities will be higher than for the lower resolution observation, but the increase in the total column will be larger than in the H I column. Conversely, the low column, low molecule patches will lie on the left and slightly downward from the low-resolution points, since both the total and H I columns will decline, but the H I by less, since the atomic fraction rises. These changes will bring the data points closer to our model curves. Indeed, the existing data already hint that such an effect is present: the single beam-averaged observations of the BR06 data set scatter away from our limit lines noticeably less than the ring-averaged observations from L08. Similarly, we can divide the point-by-point data from BR06 into a "near" sample, consisting of galaxies for which the resolution is smaller than 1 kpc, and a "far" sample, consisting of galaxies with larger resolutions. We plot a version of Figure 4 using only this divided data set from BR06 in Figure 7. The comparison is quite noisy, since only the $-0.25 < \log Z' < 0$ metallicity bin contains more than one galaxy in both the near and far sets, but the data in that bin do at least seem consistent with the hypothesis that the near data fall closer to the model lines than the far data.

The full HERACLES survey, currently underway, will report measurements of the molecular surface density for individual patches $\sim 0.5$ kpc in size, generally smaller than the beam patches of the BIMA / SONG survey. While this is still considerably larger than the $\lesssim 0.1$ kpc size of a typical atomic–molecular complex in the Milky Way, at the higher spatial resolution of the full HERACLES data set beam-smearing effects should be reduced and the column densities reported for each patch should be closer to the true column densities $\Sigma_{comp}$ of the individual atomic–molecular complexes. We therefore predict that in the full HERACLES data set, the $\Sigma_{H\,I} - -\Sigma_{obs}$ relation should be closer to our theoretical upper and lower limit curves than the lower resolution or azimuthally averaged data shown in

Figures 4–6.

#### 4.1.3. Pressure Dependence

BR04 and BR06 find that the molecular fraction in a galaxy correlates with the midplane gas pressure as

$$R_{\rm H_2} \equiv \frac{\Sigma_{\rm H_2}}{\Sigma_{\rm H\,I}} = \left( \frac{P/k_{\rm B}}{3.5 \pm 0.6 \times 10^4 \ {\rm K \ cm^{-3}}} \right)^{0.92 \pm 0.07}, \quad (50)$$

while L08 find

$$R_{\rm H_2} = \left( \frac{P/k_{\rm B}}{2.0 \times 10^4 \ {\rm K \ cm^{-3}}} \right)^{0.8} \quad (51)$$

for their sample. Although our model does not directly give a prediction for the dependence of the molecular fraction on pressure, and we argue that surface density and metallicity are the physical variables that directly control the molecular fraction, we wish to check whether our model is consistent with the observed correlation.

To avoid introducing any bias in performing this check, we must determine pressures from observable quantities in the same way that BR06 and L08 do, following an approximation introduced by BR04.[6] The BR04 approximation treats the galaxy as an infinite thin disk of uniform gas and stars in vertical hydrostatic balance. For such a disk, the pressure is related to surface density by

$$P = \frac{\pi}{2} G \Sigma_{\rm g} \left( \Sigma_{\rm g} + \Sigma_* \frac{v_{\rm g}}{v_*} \right), \quad (52)$$

where $\Sigma_{\rm g}$ is the total (H I plus $H_2$) gas surface density, $\Sigma_*$ is the stellar surface density, and $v_{\rm g}$ and $v_*$ are the vertical velocity dispersions of the gas and stellar components, respectively. The gas velocity dispersion is roughly constant, $v_{\rm g} \approx 8$ km s$^{-1}$, for the galaxies in the BR06 and L08 samples. The term $\Sigma_*/v_*$ presents more difficulty, however. It varies by orders of magnitude from galaxy centers to edges, and between galaxies, so we cannot pick a single value for it. Since in our model the molecular fraction is a function only of metallicity and $\Sigma_{\rm g}$, this means that we do not predict a single-valued relationship between $R_{\rm H_2}$ and $P$ that we can directly compare to the empirical fits.

However, we can still compare our model to the data in two ways. First, we can pick a range of values of $\Sigma_*/v_*$ consistent with the range in the observed sample, and demonstrate that the resulting range of predictions for $R_{\rm H_2}$ covered by our model is consistent with the observational data. For this purpose, we follow BR06 in adopting a constant value $v_{\rm g} = 8$ km s$^{-1}$ and using a simplified version of Equation (52), which follows from assuming that both the surface density and scale height of the stars are much greater than those of the gas:

$$\frac{P}{k_{\rm B}} = 272 \ {\rm cm^{-3} \ K} \left( \frac{\Sigma_{\rm obs}}{M_\odot \ {\rm pc^{-2}}} \right) \left( \frac{\Sigma_*}{M_\odot \ {\rm pc^{-2}}} \right)^{0.5} \times \left( \frac{v_{\rm g}}{\rm km \ s^{-1}} \right) \left( \frac{h_*}{\rm pc} \right)^{-0.5}, \quad (53)$$

[6] It is important to note that the midplane pressure which the observers attempt to estimate is not the same as the pressure in the molecular gas or the CNM which we balanced in Section 2.1; the total midplane pressure includes contributions from all phases of the ISM rather than just the CNM or molecular clouds, as well as contributions from magnetic fields, cosmic rays, and bulk motions. The pressure we use in Section 2.1 includes thermal pressure in the CNM or molecular gas only.

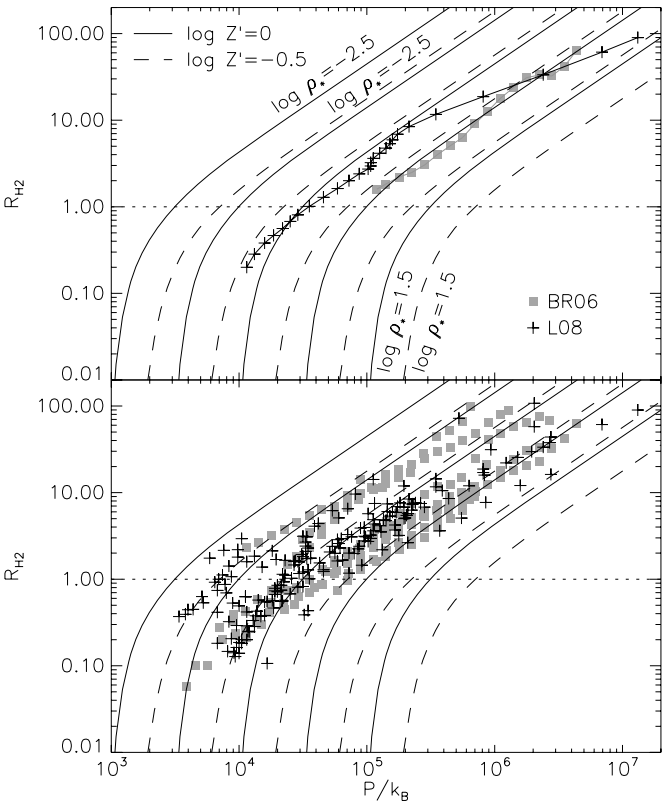


**Figure 8.** Molecular-to-atomic ratio $R_{\rm H_2}$ vs. pressure $P$. We show the data sets of BR06 (squares) and L08 (diamonds), and predictions for models with $\log \rho_* = -2.5, -1.5, -0.5, 0.5$, and 1.5 in units of $M_\odot$ pc$^{-3}$ (highest to lowest lines, as indicated), and with metallicities $\log Z' = -0.5$ and 0 (dashed and solid lines, respectively). In the upper panel for clarity we show only a single galaxy from each data set (NGC 4736 from BR06 and HO II from L08), while in the lower panel we show the data for all galaxies; the model curves in each panel are identical. For BR06 the data we show are averaged over 0.1 dex wide bins in pressure, as in BR06's Figure 3. For the L08 data, we only plot rings for which both molecular and atomic gas are detected. The dotted horizontal lines show $R_{\rm H_2} = 1$; below this line our predictions should be taken as upper limits only.

where $\Sigma_{\rm obs}$ is the observed value of the total gas surface density $\Sigma_{\rm g}$, $h_*$ is the stellar scale height, and for convenience, we define $\rho_* = \Sigma_*/h_*$. We consider values of $\log \rho_*$ from $-2.5$ to 1.5 in units of $M_\odot$ pc$^{-3}$, a range that covers almost all of the samples of BR06 and L08. For a given value of $\rho_*$, Equation (53) gives a single-valued relationship between $P$ and $\Sigma_{\rm g}$, so for each $P$ and a choice of metallicity, we can use our model to make a prediction for $\Sigma_{\rm H\,I}$ and thus for $R_{\rm H_2}$. As with our calculations in Section 4.1.2, we expect to systematically underpredict $R_{\rm H_2}$ because the observed column density $\Sigma_{\rm obs}$ is an underestimate of the true column density of atomic–molecular complexes that are not resolved by the telescope beam.

We show the results of this computation in Figure 8. As expected, the model covers an area consistent with the observed data. First note that there is a systematic offset between the results from BR06 and L08; this is likely because some of the structural parameters that are used in Equation (53) are quite uncertain, particularly $h_*$, and the values adopted by the two surveys are not necessarily the same. We should keep this uncertainty in mind as we proceed, since it suggests an upper limit on the level of agreement we can expect.

Nonetheless, the correlation between $R_{\rm H_2}$ and $P$ shown in both observational surveys, $R_{\rm H_2} \propto P^{0.8-0.9}$, is somewhat flatter than any of our model curves for a particular choice of $\rho_*$,

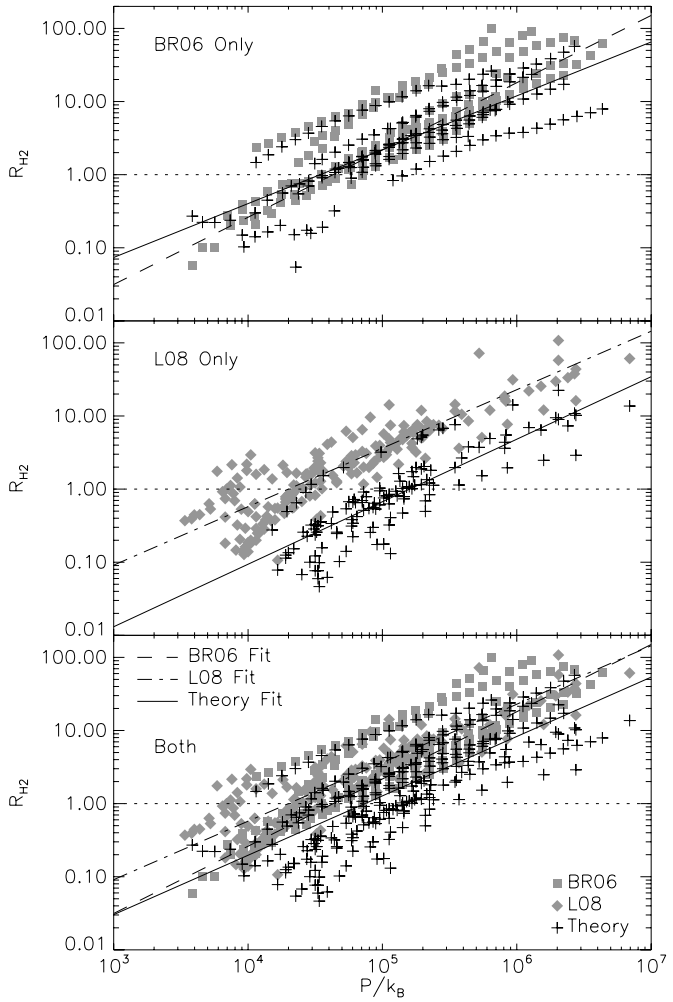


**Figure 9.** Molecular-to-atomic ratio $R_{H_2}$ vs. pressure $P$. We show the data sets of BR06 (squares) and L08 (diamonds), and predictions from theory for the value of $\Sigma_g$ and $Z'$ for each data point (plus signs). In the top panel we show only the BR06 data and the theoretical predictions corresponding to it. In the middle we show only the L08 data and predictions, and in the bottom panel we show both data sets together. In all panels we only plot rings or points for which both molecular and atomic gas are detected for the observations, and the theoretical points are corrected for finite sensitivity as described in the text. We do not plot theoretical predictions for galaxies without measured metallicities. We also show the best power-law fits of the BR06 data (Equation (50); dashed line), the L08 data (Equation (51); dot-dashed line), and to the model predictions (Equation (54); solid line). In the upper panel the theory line uses only the BR06 data, in the middle it uses only the L08 data, and in the bottom panel we show a line fit to our predictions for both data sets. The dotted horizontal lines show $R_{H_2} = 1$; below this line our predictions should be taken as upper limits only.

which all approach $R_{H_2} \propto P^1$ at large $P$. However, this is to be expected: low pressures and gas column densities are found preferentially in the outer parts of galaxies, where $\rho_*$ is small, so at low $P$ we expect to be closer to the higher model curves, which have low $\rho_*$. Conversely, high pressures and values of $\Sigma_g$ are found preferentially in regions with higher stellar densities $\rho_*$, so high $P$ values should be closer to the lower model curves, which have larger $\rho_*$. This covariance between $P$ and $\rho_*$ results in a flattening of the $R_{H_2} \propto P$ relation that we would predict if all galaxies had fixed $\rho_*$.

We can eliminate this covariance effect and produce a stronger test than that shown in Figure 8 by using our model to generate predictions for molecular content directly, and plotting those against the inferred pressure. To do this, for each galactocentric ring in the L08 data set, we take the observed total gas surface density $\Sigma_{obs}$ and the metallicity given in Table 1 (which we treat as constant in each galaxy) and use our model to predict $R_{H_2}$. Similarly, for the BR06 data set, we take the observed gas surface density $\Sigma_{obs}$ in each pixel and use our model to generate a prediction for $R_{H_2}$. Since the pressure in each ring or pixel is known, by this procedure we generate a synthetic set of data points ($P$, $R_{H_2}$) which we can compare to the observations, and from which we can generate a fit for $R_{H_2}$ as a function of $P$.

Before making the comparison, however, we must also model the effects of finite telescope sensitivity, which make it impossible to detect molecules below a certain minimum column density. Since, as Figure 8 shows, our model predicts that the slope of $R_{H_2}$ versus $P$ varies with $P$, this can affect the fit we generate from our synthetic data. We include this effect using a procedure nearly identical to that of BR06. These authors first estimate the minimum value of $R_{H_2}$ as a function of pressure that they could detect in each galaxy based on their telescope noise limits. For our synthetic data, we adopt a minimum $R_{H_2,min} = 0.05$ for all galaxies, slightly below the lowest average $R_{H_2}$ that BR06 report for any of the galaxies they analyze. For each galaxy, we then identify the lowest pressure $P_{min}$ for which all the pixels/rings have $R_{H_2} > R_{H_2,min}$. At pressures above $P_{min}$, the sample should be nearly complete, in the sense that no pixels or rings will have molecular nondetections that could bias a fit of $R_{H_2}$ versus $P$. At pressures below $P_{min}$, the data are incomplete, and we therefore drop all pixels or rings whose pressures are below $P_{min}$. This produces a sensitivity-corrected set of synthetic data. Finally, following BR06, we then bin the pixels by pressure, using bins 0.1 dex wide, and for each pressure bin in each galaxy we compute an average value of $R_{H_2}$. We do not bin the rings in the L08 sample.

We overplot the synthetic and real data in Figure 9. Fitting our synthetic data to a power-law function for $R_{H_2}$ versus $P$ gives

$$R_{H_2} = \left(\frac{P/k_B}{9.0 \times 10^3 \text{ K cm}^{-3}}\right)^{0.81}, \quad (54)$$

if we include our models for both the BR06 and L08 data sets. Using only one or the other gives

$$R_{H_2} = \begin{cases} \left(\dfrac{P/k_B}{2.2 \times 10^2 \text{ K cm}^{-3}}\right)^{0.74}, & \text{(BR06)} \\ \left(\dfrac{P/k_B}{2.8 \times 10^5 \text{ K cm}^{-3}}\right)^{0.85}, & \text{(L08)}. \end{cases} \quad (55)$$

As expected, we obtain a slope shallower than unity as a result of the systematic increase of $\rho_*$ with $P$. The scatters in the real and synthetic data are also comparable. As Figure 9 shows, our model predictions overlap with the observed data reasonably well, particularly for the BR06 data; the best fit for the BR06 data set is reasonably close to the observational best fit, while for L08 the theoretical predictions and the best fit to them give a slope similar to the observed value, but are systematically shifted to lower $R_{H_2}$. The displacement between our model and the L08 data probably exists for the same reason that our predictions for $\Sigma_{H_2}$ and $R_{H_2}$ in Section 4.1.2 are lower limits: spatial averaging leads us to an underestimate of the true column densities of the atomic–molecular complexes in a galaxy, which in turn leads us to slightly underpredict the molecular content. The averaging is significantly worse if done over rings rather than over individual beam pointings, so the offset is noticeably larger for the ring-averaged data set.

It is also worth cautioning that in performing this fit, we have included model points where we predict $R_{H_2} < 1$, a region of

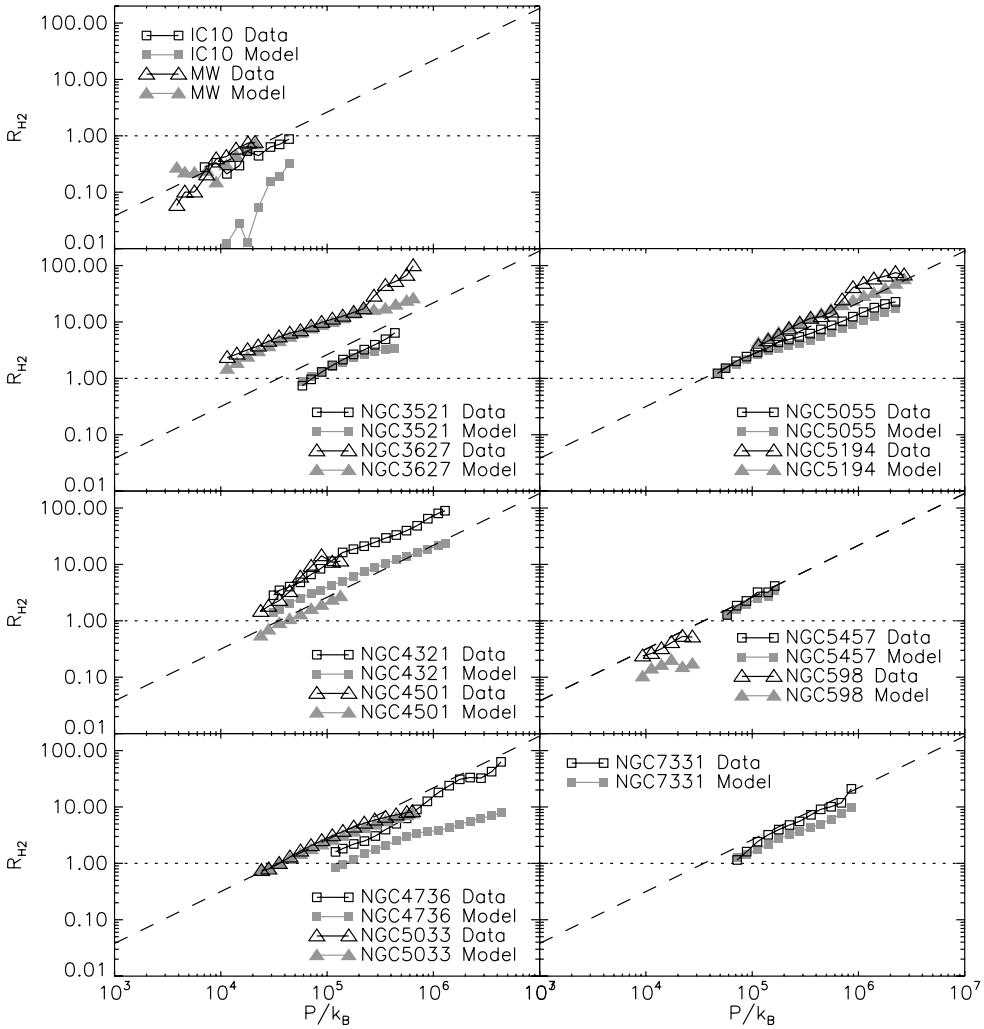


**Figure 10.** Same as Figure 9, except that we show only the BR06 data, and we plot only two galaxies per panel, as indicated, so that we can compare on a galaxy-by-galaxy basis. In each panel the black open points are the observed data, while the gray filled points are our model predictions. The dashed line shows the BR06 empirical fit.

parameter space where the predictions of our theory should be taken as upper limits. If we exclude these points from both our model predictions and from the observed data, all the slopes become significantly more shallow, but they remain consistent with one another. In this case, our model predictions give a best fit of $R_{H_2} = [(P/k_B)/(83\ \mathrm{K\ cm^{-3}})]^{0.47}$ for the combined BR06 and L08 samples, while the BR06 and L08 data sets give $[(P/k_B)/(120\ \mathrm{K\ cm^{-3}})]^{0.51}$ and $[(P/k_B)/(300\ \mathrm{K\ cm^{-3}})]^{0.53}$, respectively. These fits are performed without weighting by the errors or properly including the effects of upper limits, so they should only be taken as general indications, but the results do show that even if we limit our fit to the part of parameter space where we can apply our theory with high confidence, we obtain good agreement between model predictions and observations.

We can further check the level of agreement between the model and the data, and see how well our theory compares to the purely empirical fit, by using the BR06 sample where there is no (or less) systematic offset due to unresolved clumping. In Figure 10, we plot the same data as in Figure 9, but show the data and model predictions for the galaxies one by one. As the plots show, our model not only fits the general trend between $R_{H_2}$ and $P$, for most galaxies we obtain a good match on a bin-by-bin basis. Even in those galaxies where the agreement is poor, our model still gives a correct lower limit for $R_{H_2}$. The level of agreement between our predicted curves and the data is as good as the purely empirical fit between $R_{H_2}$ and pressure in BR06, without the need for any free parameters. It is not clear why one model or the other does better for particular galaxies. Our model appears to provide a substantially better fit for NGC 3627, NGC 5033, and NGC 5055, while doing noticeably worse for IC 10, NGC 0598, and NGC 4736; for the rest of the sample, the BR06 empirical fit and our model calculations are nearly equally good fits. It is not clear what galaxy properties favor one model or the other.

#### *4.1.4. The Small Magellanic Cloud*

As noted above, the SMC's metallicity of 12+log(O/H) = 8.0 (Dufour 1984) makes it the lowest metallicity galaxy in our sample for which we have detections rather than upper limits on the $H_2$ content. To analyze this data set, we take the maps of H I and $H_2$ column density determined by Leroy et al. (2007) and break the data into 20 bins in total gas column density (including He). In each bin, we compute the mean H I and $H_2$ column density. It is worth noting that the technique used to infer $H_2$ column densities in this data set has $\sim 50\%$ systematic uncertainties, so we must proceed with this caution in mind.

To test our theoretical model, we generate a prediction for the $H_2$ mass fraction as a function of total gas column density

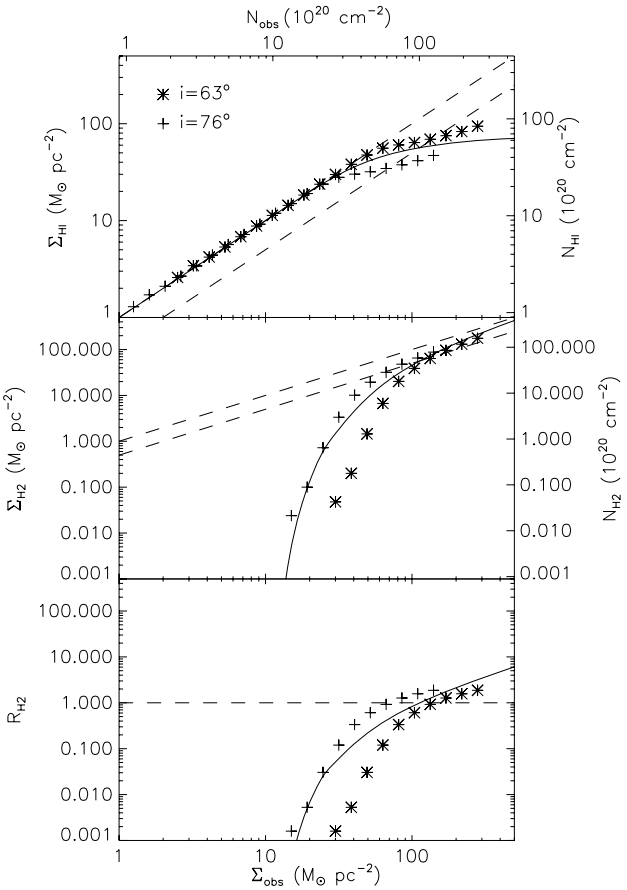


**Figure 11.** Column densities of H I and $H_2$ (upper two panels) and $H_2$ to H I ratio (bottom panel) as a function of total column density $\Sigma_{obs}$ in the SMC. The solid curve shows the model prediction, while the dashed lines show our confidence regions, as in Figures 4–6. The asterisks and plus signs represent the observed column densities using inclination corrections of 63° and 76°, respectively.

$\Sigma_{comp}$ as in Equation (48), using a metallicity relative to solar of $Z' \approx 0.2$. Before comparing the model curve to the data, we face the difficult problem of inclination correction. The SMC is a triaxial structure with an aspect ratio of roughly 1:2:4, with 4 representing the direction along the line of sight (Crowl et al. 2001). The large extent along the line of sight means that the inclination correction is very significant, but the triaxiality of the galaxy means that no single number describes the inclination, as is the case for a disk. We therefore compare the model and the data using two different inclination corrections, one assuming an inclination $i = 76°$ (corresponding to a 1:4 aspect ratio) and one assuming an inclination $i = 63°$ (corresponding to a 1:2 aspect ratio). These should bracket the true inclination correction. Another caution we should make here is that, due to these large inclinations, there may be significant amounts of gas along a given line of sight that are not associated with whatever atomic–molecular complex it intersects, so the H I column densities may be overestimates (see Section 4.1.2).

Figure 11 shows a comparison of the data and the model curve. As expected, the data for the two different inclination corrections bracket the model curve quite well. We should be cautious about reading too much into the agreement at low $\Sigma_{obs}$ ($\lesssim 30\ M_\odot\ pc^{-2}$), where both the data and our model are quite uncertain. Indeed, $R_{H_2}$ is such a sharp function of $\Sigma_{obs}$ at low column density that the factor of 2 variation in the total column density introduced by the uncertainty in the inclination translates to a factor of $\sim 50$ difference in the value of $R_{H_2}$ at a fixed $\Sigma_{obs}$. Thus, our model need only be correct within a factor of 50 to lie in between the two curves for the different inclination corrections. At higher $\Sigma_{obs}$ ($\gtrsim 80\ M_\odot\ pc^{-2}$), on the other hand, the inclination correction only produces a factor of $\sim 2$ uncertainty in $R_{H_2}$, and the ability of our model to match the data here is significant. It shows that we are capable of predicting the total column density at which $R_{H_2} \sim 1$ to better than a factor of 2 accuracy even in a galaxy with metallicity $\sim 1/10$ solar.

### 4.2. Galactic Observations

#### 4.2.1. Data Sets

We next compare our model to observations of the molecular content of clouds in the Milky Way, as measured by the *Copernicus* (Spitzer & Jenkins 1975) and the *Far Ultraviolet Spectroscopic Explorer (FUSE)* (Moos et al. 2000; Sahnow et al. 2000) missions. These satellites measured absorption of ultraviolet light from background stars, and in some cases AGNs, in the Lyman–Werner bands, enabling them to estimate the population of molecular hydrogen along a given line of sight.

Savage et al. (1977) and Bohlin et al. (1978) (for *Copernicus*) and Rachford et al. (2002, 2008) (for *FUSE*) report measurements of the molecular hydrogen columns in the Milky Way disk for lines of sight at low galactic latitude ($b \lesssim 10°$), using stars as background sources. They combine these with H I observations along the same lines of sight to determine molecular hydrogen fractions for clouds along these sightlines. Gillmon et al. (2006) report *FUSE* observations of sightlines at high galactic latitude ($b \gtrsim 20°$) using AGNs as background sources. These observations probe clouds lying above or below the galactic plane, which are presumably illuminated only on one side by stars in the disk. As with the low-latitude samples, Gillmon et al. combine these observations with measurements of the H I column density along the same sightlines to determine molecular fractions.

These observations are different than the extragalactic ones in that they are pencil-beam measurements rather than averages over atomic–molecular complexes. We possess little information about the geometry of the clouds these beams probe, and this ignorance complicates comparison of the data to our model. A given observed total (atomic plus molecular) column density $\Sigma_{obs}$ might be the result of a line of sight passing through the center of a small cloud, or might be the result of a beam that passes only tangentially through a much larger cloud. Our models do not predict the same $H_2$ content in these two cases, so in the absence of additional information there is no way to map a given $\Sigma_{obs}$ to a unique prediction for the molecular fraction. Similarly, there is good evidence in the data sets from the rotational excitation of the $H_2$ that at least some lines of sight probe multiple clouds that are separated in space (e.g., Browning et al. 2003). Our model predicts a lower molecular fraction for this case than for the case of a single cloud with the same $\Sigma_{obs}$. Attempts to remove these effects statistically are complicated by the fact that the *FUSE* and *Copernicus* lines of sight are not an unbiased sample. Each of the samples we use was chosen specifically to probe a certain range in column density or other properties, and all the selections are biased against high column densities and high molecular fractions, both of which produce high extinctions of LW photons that make determining a molecular column very costly or altogether impossible. Below we discuss how we deal with the problems of geometric uncertainty and selection bias for the galactic plane and high-latitude samples.

A second complication for this data set is that the assumption we make in our model that clouds are subjected to a relatively uniform background dissociating radiation field (see Krumholz et al. 2008) may not be appropriate in this context. Treating the ISRF as uniform is reasonable for the giant cloud complexes with masses $\gtrsim 10^4\ M_\odot$ and sizes $\gtrsim 100$ pc probed by extragalactic measurements; these sample the radiative output of many stars and star clusters. In contrast, the entire sample of high-latitude clouds observed by Gillmon et al. (2006) has a total mass $\sim 3000\ M_\odot$ spread out over a $\sim 100$ pc$^2$ area (Gillmon & Shull 2006), making them tiny in comparison to giant complexes. The galactic clouds probed by the *Copernicus* and *FUSE* observations are of unknown size, but are likely to be similarly small. This is because any line of sight passing through a true giant cloud complex would be completely opaque in the Lyman–Werner bands, thus preventing *Copernicus* or *FUSE* from making any measurement of the $H_2$ column. The dissociating flux in the vicinity of such small clouds may be dominated by a single nearby star or star cluster.

Due to the effects of uncertain geometry and nonuniform radiation fields, we expect this data set to show significantly more scatter than the extragalactic one.

#### 4.2.2. The Galactic Plane Data

Given the limitations of pencil-beam measurements, we must compare our model to the data in a way that accounts for our geometric uncertainty. The most straightforward comparison we can make is by considering a limiting case. Consider an observation that measures a total column density $\Sigma_{\rm obs}$. According to our model, the fraction of gas in molecular form will depend on the geometry of the complex; if the line of sight probes the outer parts of a very large cloud, always staying within the atomic region, then in principle there could be no molecules present no matter how large $\Sigma_{\rm obs}$ might be. The reverse is not true, however. Even if the line of sight passes exactly through the center of an atomic–molecular complex, it must still pass through the atomic shielding layer, and there will therefore be some H I along that line of sight. In our model, therefore, there is a minimum amount of H I that we predict must be present for a given column of $H_2$; this minimum corresponds to the case of a line of sight that passes directly through the center of a complex. This implies that we can make a prediction as follows: we consider a complex of mean column density $\Sigma_{\rm comp}$, and use our model to predict the atomic, molecular, and total column densities, $\Sigma_{\rm H\,I,cen}$, $\Sigma_{\rm H_2,cen}$, and $\Sigma_{\rm cen}$, that one would see along a line of sight passing through the center of the complex. The curves $\Sigma_{\rm H\,I,cen}$ versus $\Sigma_{\rm cen}$ and $\Sigma_{\rm H_2,cen}$ versus $\Sigma_{\rm cen}$ that we generate through this procedure should then be lower and upper limits, respectively, on the observed distributions of $\Sigma_{\rm H\,I}$ versus $\Sigma_{\rm obs}$ and $\Sigma_{\rm H_2}$ versus $\Sigma_{\rm obs}$.

Instead of computing a limit, we can also make an estimate for the "typical" atomic and molecular fractions we should see at a given $\Sigma_{\rm obs}$. If we knew the true distribution of column densities for atomic–molecular complexes in the Milky Way, we could do this by integrating all lines of sight through that distribution and computing the mean atomic and molecular columns at a given total column. However, we do not know the true distribution, and even if we did this procedure would not address effects of observational bias in selecting sightlines. Instead, we make a much rougher calculation. Using our model we can compute the radius $x_{\rm H_2}(\Sigma_{\rm comp})$ at which the cloud whose true mean column density is $\Sigma_{\rm comp}$ transitions from atomic to molecular for solar metallicity ($Z = 1$). The atomic and molecular columns along a sightline that strikes the cloud of radius $R$ at a distance $\beta R$ from its center are

$$\Sigma_{\rm H\,I} = \frac{2\tau_{\rm R}}{\sigma_{\rm d}} \begin{cases} \sqrt{1-\beta^2} - \sqrt{x_{\rm H_2}^2 - \beta^2}, & \beta < x_{\rm H_2} \\ \sqrt{1-\beta^2}, & \beta > x_{\rm H_2} \end{cases} \quad (56)$$

$$\Sigma_{\rm H_2} = \frac{2\tau_{\rm R}\phi_{\rm mol}}{\sigma_{\rm d}} \begin{cases} \sqrt{x_{\rm H_2}^2 - \beta^2}, & \beta < x_{\rm H_2} \\ 0, & \beta > x_{\rm H_2} \end{cases} \quad (57)$$

and the $H_2$ fraction averaged over all pencil beams passing through the cloud is

$$f_{\rm H_2,beams}(\Sigma_{\rm comp}) = 2\int_0^1 \beta \frac{\Sigma_{\rm H_2}}{\Sigma_{\rm H_2} + \Sigma_{\rm H\,I}} d\beta. \quad (58)$$

To estimate a "typical" atomic or molecular content for a given observed total gas column $\Sigma_{\rm obs}$, we can simply take $\Sigma_{\rm H\,I,obs} \approx [1 - f_{\rm H_2,beams}(\Sigma_{\rm obs})]\Sigma_{\rm obs}$ and $\Sigma_{\rm H_2,obs} \approx f_{\rm H_2,beams}(\Sigma_{\rm obs})\Sigma_{\rm obs}$. This is equivalent to saying that our sightlines do indeed probe random impact parameters on cloud complexes, and that when we observe a sightline of column density $\Sigma_{\rm obs}$, most of the time we are observing a complex whose mean column density is also about $\Sigma_{\rm obs}$. This assumption could fail if low column density clouds were rare compared to high column density ones, so that most of the sightlines that produce a given $\Sigma_{\rm obs}$ were really tangential paths through high column clouds rather than beams passing close to the centers of low column clouds. However, there is no evidence that high column density atomic–molecular complexes outnumber low column ones, and in general in the ISM more mass tends to be in diffuse than dense structures, so we proceed with our assumption that $\Sigma_{\rm comp} \approx \Sigma_{\rm obs}$.

We plot both our limits on the H I and $H_2$ columns and our estimates for their typical values against the *Copernicus* and *FUSE* data sets in Figure 12. As the plot shows, both our limits and our typical columns match the observed data reasonably well. As with the extragalactic case, we recover the overall trend that the gas is mostly H I until a total column of a few $M_\odot$ pc$^{-2}$, and then mostly molecular thereafter.

#### 4.2.3. The High-Latitude Data

The comparison to the high-latitude data set is somewhat less straightforward, because the clouds at high latitude are only illuminated from one side. Indeed, Gillmon et al. (2006) find that these clouds become molecular at column densities a factor of $\sim 2$ lower than do clouds in the galactic plane, probably as a result of this one-sided illumination. We have not considered the case of clouds of finite size illuminated over only half their surfaces. However, we can obtain a reasonable approximation to this configuration using the case of *semi-infinite* clouds subjected to an isotropic, uniform dissociation radiation field, which we considered in Section 2. A semi-infinite cloud blocks dissociating radiation over $2\pi$ sr, and thus the depth of the atomic shielding layer at its surface is determined only by photons that arrive from the "front side." This is, therefore, a close analogy to the case of a high-latitude cloud, although it differs in the cloud geometry.

In Paper I we show that using a semi-infinite slab in place of a finite cloud tends to produce errors in estimating the exact location of a transition from atomic to molecular in a finite cloud, but that the semi-infinite calculation does give a good estimate of the column at which clouds transition from mostly atomic

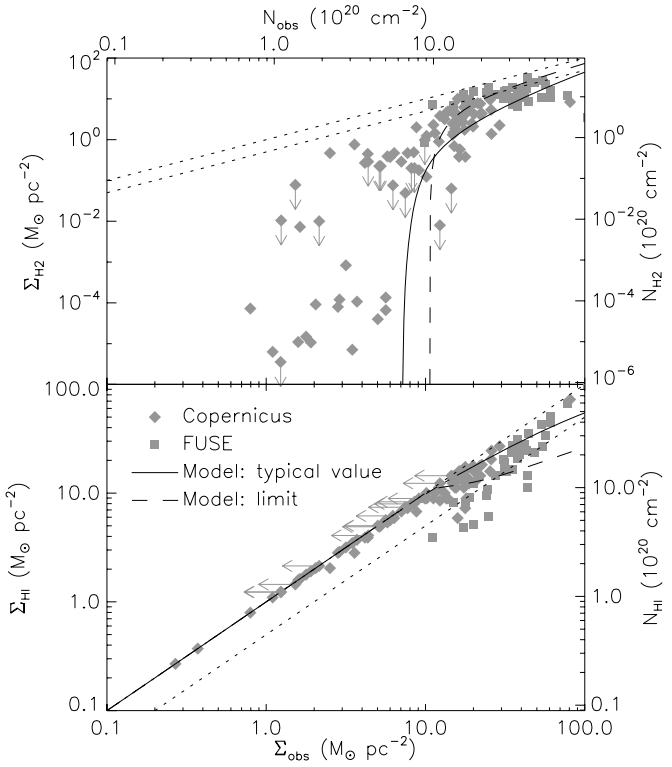


**Figure 12.** Column density of $H_2$ (upper panel) and H I (lower panel) vs. total column density for the *Copernicus* (diamonds) and *FUSE* (squares) data, and for the limits (dashed line) and typical values (solid line) we compute from our model. The dashed line is an upper limit in the upper panel, and a lower limit in the lower one. For the data, we only plot points with measured H I columns. Arrows indicate lines of sight for which the H I column is measured but only an upper limit is available for the $H_2$ column. In the upper panel the dotted lines show $\Sigma_{H_2} = \Sigma$ and $\Sigma_{H_2} = \Sigma/3$ (the edge of our region of confident prediction), while in the lower panel they show $\Sigma_{H\,I} = \Sigma$ and $\Sigma_{H\,I} = 2\Sigma/3$. Note that $N_{H_2}$ represents the column of H nuclei in molecular form; the number of $H_2$ molecules is half this.

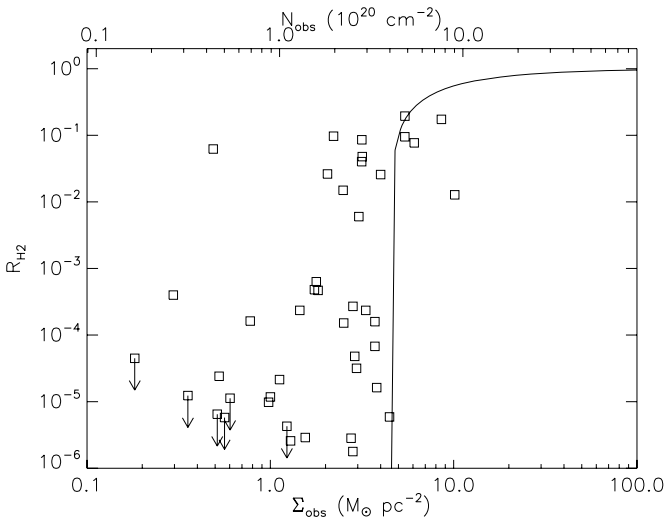


**Figure 13.** $R_{H_2}$ vs. column density for the *FUSE* high-latitude sample (squares; Gillmon et al. 2006) and for our model (line). Arrows indicate lines of sight for which only an upper limit is available for the $H_2$ column. Note that $N_{obs}$ represents the total column of H nuclei in either atomic or molecular form.

to mostly molecular. We therefore expect to obtain roughly the right transition column and thus the right column of H I for clouds that are mostly molecular, but not particularly accurate predictions for the exact molecular column. For a semi-infinite cloud of solar metallicity, as we show in Section 2 we predict a molecular to atomic ratio of

$$R_{H_2} = \begin{cases} 0, & \Sigma < 4.5\ M_\odot\ \mathrm{pc}^{-2} \\ (\Sigma - 4.5\ M_\odot\ \mathrm{pc}^{-2})/\Sigma, & \Sigma > 4.5\ M_\odot\ \mathrm{pc}^{-2} \end{cases} \quad (59)$$

where $\Sigma$ is the total column density. In other words, for a semi-infinite cloud of solar metallicity, the first 4.5 $M_\odot$ pc$^{-2}$ are atomic, and the rest are molecular. We show this prediction overplotted with the *FUSE* high-latititude data in Figure 13. As the plot shows, we indeed do not get very accurate predictions for the exact ratio of H I to $H_2$, but our calculation agrees quite well with the general value of $\Sigma$ for which the transition from molecular to atomic occurs. In particular, our one-sided model recovers the observational result that a cloud illuminated from one side shows a lower atomic-to-molecular transition column than an isotropically illuminated cloud.

## 5. SUMMARY AND CONCLUSIONS

We present a first-principles calculation of the molecular gas content of galactic disks in terms of the observable properties of those galaxies. Our calculations build on the simple model for photodissociation fronts in finite clouds presented in Paper I, in which we show that the amount of atomic material required to shield a molecular cloud against dissociation by the interstellar radiation field (ISRF) can be characterized by two parameters: $\chi$, a radiation field strength normalized by the gas density and the properties of dust grains, and $\tau_R$, a measure of the dust optical depth of a cloud. We show that, due to the way the density in the cold phase of the atomic ISM varies with ISRF, the normalized radiation field strength takes a characteristic value $\chi \approx 1$ in all galaxies where a two-phase atomic ISM is present, with only a weak dependence on metallicity.

The existence of a characteristic normalized radiation field strength, and its weak dependence on metallicity, has a number of important consequences. First, it enables us to give a simple analytic approximation (Equations (36) or (37)) for the fraction of mass in an atomic–molecular complex that will be in the atomic or molecular phases solely in terms of the column density of the complex and the metallicity of the gas. This makes it easy to test our calculations against observations. Second, we show that as a consequence of $\chi$ assuming a nearly fixed value, the atomic envelopes of molecular clouds have a characteristic visual extinction $A_V \approx 0.2$ at solar metallicity. Similarly, the transition from atomic to molecular gas occurs at a characteristic shielding column of $\Sigma_{H\,I} \approx 10\ M_\odot$ pc$^{-2}$. These quantities both vary sub linearly with metallicity, with $A_V$ declining and $\Sigma_{H\,I}$ increasing as metallicity does. We calculate these values and their metallicity dependence solely in terms of the microphysical constants that describe the properties of $H_2$ formation and dissociation, the cooling curves of C II and O I, and the properties of interstellar dust grains. Our model does not depend on unobservable parameters such as the gas volume density or the ISRF strength in a galaxy.

Our model compares favorably with observations of the atomic and molecular content of clouds both in the Milky Way and in nearby galaxies. We are able to reproduce both the characteristic column density at which clouds transition from being primarily atomic to primarily molecular, and the way that this characteristic column depends on metallicity and on whether clouds are illuminated on one side or on both sides. We are also able to reproduce the observed correlation between molecular content and interstellar pressure.

The development of a predictive model for the molecular content of galaxies that does not rely on unknown and generally unobservable quantities such as the radiation fields or gas volume densities in those galaxies opens up new possibilities to advance our understanding of galactic evolution. In low-density dwarfs or high-redshift galaxies containing few metals, the formation of molecular clouds may be the rate limiting step in star formation. On the other hand, previous work has shown that, once molecular gas forms, it converts itself into stars at a rate of a few percent of the mass per free-fall time independent of its density or environment (Krumholz & McKee 2005; Krumholz & Tan 2007; McKee & Ostriker 2007). Thus, a theory of molecule formation creates the possibility of developing a theory of the star-formation rate capable of making predictions that can be applied not only in the relatively molecule-rich nearby galaxies for which empirical star-formation laws have been determined (e.g. Kennicutt 1998), but also in the more distant and lower metallicity universe where these laws are known to break down (e.g. Wolfe & Chen 2006). We plan to explore such a theory in future work.

We thank B. Rachford and A. Leroy for providing copies of their data, and F. Bigiel, A. Leroy, S. Faber, A. Sternberg, and F. Walter for helpful discussions and comments on the manuscript. We thank the referee, E. Rosolowsky, for providing comments that improved the quality of the paper. Support for this work was provided by: NASA through Hubble Fellowship grant #HSF-HF-01186 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555 (M.R.K); NASA, as part of the *Spitzer Space Telescope* Theoretical Research Program, through a contract issued by the Jet Propulsion Laboratory, California Institute of Technology (M.R.K); and by the National Science Foundation through grants AST-0807739 (to M.R.K), AST-0606831 (to C.F.M), and PHY05-51164 (to the Kavli Institute for Theoretical Physics, where M.R.K, C.F.M, and J.T. collaborated on this work).

## REFERENCES

Allen, R. J., Heaton, H. I., & Kaufman, M. J. 2004, ApJ, 608, 314
Audit, E., & Hennebelle, P. 2005, A&A, 433, 1
Bigiel, F., Leroy, A., Walter, F., Brinks, E., de Blok, W. J. G., Madore, B., & Thornley, M. D. 2008, AJ, 136, 2846
Black, J. H., & van Dishoeck, E. F. 1987, ApJ, 322, 412
Blitz, L., Fukui, Y., Kawamura, A., Leroy, A., Mizuno, N., & Rosolowsky, E. 2007, in Protostars and Planets 5, ed. B. Reipurth, D. Jewitt, & K. Keil (Tucson, AZ: Univ. of Arizona Press), 81
Blitz, L., & Rosolowsky, E. 2004, ApJ, 612, L29
Blitz, L., & Rosolowsky, E. 2006, ApJ, 650, 933
Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132
Bolatto, A. D., Leroy, A. K., Rosolowsky, E., Walter, F., & Blitz, L. 2008, ApJ, 686, 948
Bolatto, A. D., et al. 2007, ApJ, 655, 212
Browning, M. K., Tumlinson, J., & Shull, J. M. 2003, ApJ, 582, 810
Caffau, E., Ludwig, H.-G., Steffen, M., Ayres, T. R., Bonifacio, P., Cayrel, R., Freytag, B., & Plez, B. 2008, A&A, 488, 1031
Crowl, H. H., Sarajedini, A., Piatti, A. E., Geisler, D., Bica, E., Clariá, J. J., & Santos, J. F. C., Jr. 2001, AJ, 122, 220
Draine, B. T. 1978, ApJS, 36, 595
Draine, B. T. 2003a, ARA&A, 41, 241
Draine, B. T. 2003b, ApJ, 598, 1017
Draine, B. T. 2003c, ApJ, 598, 1026
Draine, B. T., & Bertoldi, F. 1996, ApJ, 468, 269
Dufour, R. J. 1984, in IAU Symposium 108, Structure and Evolution of the Magellanic Clouds, ed. S. van den Bergh & K. S. D. Boer (Dordrecht: Reidel), 353
Elmegreen, B. G. 1989, ApJ, 338, 178
Elmegreen, B. G. 1993, ApJ, 411, 170
Elmegreen, B. G., & Elmegreen, D. M. 1987, ApJ, 320, 182
Ferrarese, L., et al. 2000, ApJS, 128, 431
Garnett, D. R. 1990, ApJ, 363, 142
Gillmon, K., & Shull, J. M. 2006, ApJ, 636, 908
Gillmon, K., Shull, J. M., Tumlinson, J., & Danforth, C. 2006, ApJ, 636, 891
Habing, H. J. 1968, Bull. Astron. Inst. Netherlands, 19, 421
Heiner, J. S., Allen, R. J., Emonts, B. H. C., & van der Kruit, P. C. 2008a, ApJ, 673, 798
Heiner, J. S., Allen, R. J., Wong, O. I., & van der Kruit, P. C. 2008b, A&A, 489, 533
Helfer, T. T., Thornley, M. D., Regan, M. W., Wong, T., Sheth, K., Vogel, S. N., Blitz, L., & Bock, D. C.-J. 2003, ApJS, 145, 259
Hidaka, M., & Sofue, Y. 2002, PASJ, 54, 223
Hollenbach, D. J., & Tielens, A. G. G. M. 1999, Rev. Mod. Phys., 71, 173
Jarrett, T. H., Chester, T., Cutri, R., Schneider, S. E., & Huchra, J. P. 2003, AJ, 125, 525
Kennicutt, R. C. 1998, ApJ, 498, 541
Krumholz, M. R., & McKee, C. F. 2005, ApJ, 630, 250
Krumholz, M. R., Mc Kee, C. F., & Tumlinson, J. 2008, ApJ, 689, 865
Krumholz, M. R., & Tan, J. C. 2007, ApJ, 654, 304
Leroy, A., Bolatto, A., Stanimirovic, S., Mizuno, N., Israel, F., & Bot, C. 2007, ApJ, 658, 1027
Leroy, A., Walter, F., Bigiel, F., Brinks, E., de Blok, W. J. G., & Madore, B. 2008, AJ, 136, 2782
Liszt, H. 2002, A&A, 389, 393
Liszt, H., & Lucas, R. 2000, A&A, 355, 333
Martin, C. L. 1997, ApJ, 491, 561
McKee, C. F., & Ostriker, E. C. 2007, ARA&A, 45, 565
McKee, C. F., & Ostriker, J. P. 1977, ApJ, 218, 148
Moos, H. W., et al. 2000, ApJ, 538, L1
Neufeld, D. A., & Spaans, M. 1996, ApJ, 473, 894
Pelupessy, F. I., Papadopoulos, P. P., & van der Werf, P. 2006, ApJ, 645, 1024
Pilyugin, L. S., Vílchez, J. M., & Contini, T. 2004, A&A, 425, 849
Rachford, B. L., et al. 2008, ApJ, submitted
Rachford, B. L., et al. 2002, ApJ, 577, 221
Regan, M. W., Thornley, M. D., Helfer, T. T., Sheth, K., Wong, T., Vogel, S. N., Blitz, L., & Bock, D. C.-J. 2001, ApJ, 561, 218
Robertson, B. E., & Kravtsov, A. V. 2008, ApJ, 680, 1083
Sahnow, D. J., et al. 2000, ApJ, 538, L7
Savage, B. D., Bohlin, R. C., Drake, J. F., & Budich, W. 1977, ApJ, 216, 291
Schaye, J. 2004, ApJ, 609, 667
Smith, D. A., Allen, R. J., Bohlin, R. C., Nicholson, N., & Stecher, T. P. 2000, ApJ, 538, 608
Spaans, M., & Neufeld, D. A. 1997, ApJ, 484, 785
Spitzer, Jr, L., & Jenkins, E. B. 1975, ARA&A, 13, 133
Stanimirović, S., Staveley-Smith, L., Dickey, J. M., Sault, R. J., & Snowden, S. L. 1999, MNRAS, 302, 417
Stanimirović, S., Staveley-Smith, L., & Jones, P. A. 2004, ApJ, 604, 176
Sternberg, A. 1988, ApJ, 332, 400
Tumlinson, J., et al. 2002, ApJ, 566, 857
van Dishoeck, E. F., & Black, J. H. 1986, ApJS, 62, 109
van Zee, L., Haynes, M. P., & Salzer, J. J. 1997, AJ, 114, 2479
Walter, F., Brinks, E., de Blok, W. J. G., Bigiel, F., Kennicutt, Jr., R. C., Thornley, M. D., & Leroy, A. K. 2008, AJ, 136, 2563
Walter, F., et al. 2007, ApJ, 661, 102
Wang, Z. 1990a, ApJ, 360, 529
Wang, Z. 1990b, ApJ, 360, 543
Wolfe, A. M., & Chen, H.-W. 2006, ApJ, 652, 981
Wolfire, M. G., McKee, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003, ApJ, 587, 278
Wolfire, M. G., Tielens, A. G. G. M., Hollenbach, D., & Kaufman, M. J. 2008, ApJ, 680, 384
Wong, T., & Blitz, L. 2002, ApJ, 569, 157
Wyse, R. F. G. 1986, ApJ, 311, L41