Excitation of Molecular Clouds and the Emission from Molecular Hydrogen*

Michael G. Burton

Anglo Australian Observatory, Epping Laboratory, P.O. Box 296, Epping, N.S.W. 2121, Australia.

Abstract

The observation and theory of UV excitation of molecular clouds (i.e. photodissociation regions) and the shock excitation of molecular clouds is reviewed, with particular emphasis placed on the emission from molecular hydrogen. Recent advances in near-IR instrumentation have engendered new observations which are challenging our theoretical understanding of these fundamental physical processes.

1. Introduction

Molecular hydrogen (H$_2$) in the disk of our galaxy is the major active component of the interstellar medium (ISM). It is found in molecular clouds, gravitationally bound bodies quite distinct from the hotter, lower pressure components of the three phase ISM model (e.g. McKee 1990). Star formation occurs within molecular clouds, in obscured regions with few direct signatures. We must observe the byproducts of the star formation to learn how the process works and what the necessary conditions are. This leads to the study of the excitation processes for the emitted radiation, to derive diagnostic measures of the protostellar environment and thus determine the conditions under which a star is born.

As reviewed by Hollenbach (1988), there are a number of mechanisms which heat molecular clouds above the 3K blackbody temperature of the universe. Cosmic rays can maintain the interior of dark clouds at 10K in the absence of any other sources. Ambipolar diffusion provides frictional heating in collapsing clouds as the contracting neutrals slip past the magnetically-supported ions. Gas–grain interactions provide heat if the grain temperature is higher than the gas (e.g. via radiant heating from nearby stars). X-rays can penetrate into molecular clouds and photoionise the gas, imparting kinetic energy to the electrons to heat the gas through collisions. Far-UV photons ($6 \leq h\nu \leq 13\cdot6$ eV) penetrate clouds to optical depths $\lambda_\nu \sim$ a few. They heat the gas through the photoelectric ejection of hot electrons from grain surfaces, which then collide with the gas particles. Far-UV photons can also directly excite H$_2$ electronically. The molecule then decays to an excited vibrational level of the ground electronic state, followed by

* Paper presented at the Workshop on Star Formation in Different Environments, held at the University of Sydney, 9–11 October 1991.
Fig. 1. Energy level diagram for the ground electronic state of the hydrogen molecule. Vibrational number (from \(v = 0\) to 14) increases along the x-axis, and rotational number increases vertically from \(J = 0\) within each vibrational level. The energy level is given in degrees Kelvin (1 K \(\equiv 1.4 \text{ cm}^{-1}\)). Several lines which are discussed in the text are shown; the \(v = 1-0\) S(1) line at 2.122 \(\mu\)m, 2-1 S(1) at 2.248 \(\mu\)m, 3-2 S(3) at 2.201 \(\mu\)m, 4-3 S(3) at 2.345 \(\mu\)m, 1-0 O(7) at 3.808 \(\mu\)m, 0-0 S(13) at 3.846 \(\mu\)m, 0-0 S(0) at 28.22 \(\mu\)m and the 8-4 S(5) at 8269 \(\AA\). The notation is that the change in vibrational number is written first (e.g. 1-0) and the rotational change second. The number in parentheses shows the final rotational level, and S denotes a change of \(\Delta J = -2\), Q of 0 and O of +2 (e.g. S(1) means \(J\) goes from 3 to 1).

fluorescent emission as it cascades back to the ground vibrational state. In dense gas \(n \gtrsim 10^5 \text{ cm}^{-3}\) this provides a new heating source, \(\text{H}_2\) vibrational heating, as collisions depopulate these excited levels before they radiate. Finally, shock heating may heat the gas impulsively.
The hot gas is then seen via radiative emission as it cools. In opaque cloud cores ($T \sim 10$–$30$ K), low rotational lines of CO are the dominant coolant. In the photodissociated surfaces of molecular clouds, fine structure emission from [OI] (63 $\mu$m) and [CII] (158 $\mu$m) usually dominates the cooling. If the gas is heated to $\sim$1000 K or greater, H$_2$ line emission itself is a significant coolant.

This review will concentrate on the two dominant excitation processes in molecular clouds, UV excitation or photodissociation regions, where both photoelectric and H$_2$ vibrational heating contribute, and shock heating. Particular attention will be paid to emission from the hydrogen molecule itself, and the observational diagnostics it provides. X-ray excitation of molecular hydrogen is not covered, as it has yet to be observed, although it may prove to be important in certain environments (e.g. SNRs). Extensive models for this mechanism do, however, exist; see Draine and Woods (1989).

Molecular hydrogen line emission traces the location of energetic events in clouds. The molecule’s low moment of inertia means the levels are well separated, with no excited levels populated in cold clouds ($T \lesssim 100$ K). Thus, it does not radiate in quiescent clouds, unlike other molecular species such as CO. It also has no permanent dipole moment, so radiates through quadrupole transitions. This results in relatively weak line strengths, but has the great advantage of producing optically thin radiation (column densities of order $10^{24}$ cm$^{-2}$ are needed for $n_{H_2} \sim 1$). Thus, observation of a line leads straightforwardly to a determination of the level column density, which can then be directly compared with model predictions for the excitation process at work. An energy level diagram showing the rotational and vibrational levels of the ground electronic state of molecular hydrogen is shown in Fig. 1. Some observed emission lines are shown for illustrative purposes, and will be discussed further in the text.

2. Photodissociation Regions

(2a) General

Photodissociation regions (PDRs) are interstellar regions of predominantly neutral gas occurring where the far-UV radiation field dominates the heating and chemistry of the gas. Since the FUV field significantly affects the chemistry to optical depths $A_v \sim 4$, which is also the typical half-thickness of a molecular cloud through the Galaxy, PDRs represent the typical environment of most molecular clouds, outside of opaque cores. They are most commonly observed in massive star forming regions where the emission is brightest, but the theory of PDRs applies even to clouds illuminated by just the local interstellar field. Recent reviews of this topic include Hollenbach (1990), concentrating more on the theoretical aspects, and Genzel et al. (1989), on the observational side. The theory of PDRs is detailed in Tielens and Hollenbach (1985); recent major contributions include Black and van Dishoeck (1987) (especially for H$_2$ in low density clouds), Sternberg and Dalgarno (1989) and Burton et al. (1990) (extending the models to high density, and treating H$_2$ in detail), and Hollenbach et al. (1991) (extending the models to low FUV fields).

Fig. 2 (reproduced from Tielens and Hollenbach 1985) illustrates a PDR schematically. Photons more energetic than 13.6 eV are absorbed in an HII region, and the PDR begins at a sharp H$^+$/H interface. The FUV flux can
photodissociate molecules and ionise atoms (e.g. C, S, Fe and Si) with ionisation potentials less than 13·6 eV to $A_v \gtrsim 1$–2. Going into the cloud from the neutral surface, hydrogen becomes molecular typically at $A_v \sim 1$–2, $C^+$ becomes C, then CO at $A_v \sim 2$–4, and $Si^+$ becomes Si by $A_v \sim 3$–5. Finally, any O not in CO is converted to O$_2$, O and H$_2$O deep into the molecular cloud, by $A_v \sim 5$–10. Given this chemical structure, the surface layers, typically at $T \sim 500$ K, cool mainly through the [OI] 63 μm and [CII] 158 μm lines, while deeper in the cloud the molecular gas ($T \lesssim 50$ K) cools by low-$J$ lines of CO. UV-pumped H$_2$ emission in the near-IR arises in the warm, mainly atomic zone; i.e. while H$_2$ is still a trace component of the gas. Other significant emission lines near the surface include [SiIII] 35 μm, [OI] 146 μm, [CI] 9850 and 8727 Å, and somewhat deeper into the cloud [CI] 370 and 609 μm. Nevertheless, the bulk of the radiation from PDRs is far-IR grain continuum, re-radiated from the far-UV absorbed. A clear observational diagnostic of a PDR is that the total line emission (mostly [OI] + [CII]) equals about 0·1–1% of the far-IR continuum.

An important subclass of PDRs exists when both the gas density is sufficiently high compared to the radiation field and the radiation field is fairly high (i.e. $n/G_0 \gtrsim 40$ cm$^{-3}$, $G_0 \gtrsim 10^3$, see later). When this occurs self-shielding of the molecules can move the C$^+/CO$ and H/H$_2$ transitions close to the surface of
Fig. 3. Contour plots of intensity ratios of bright emission lines from PDRs, as a function of hydrogen nucleus number density \( n \) and far-UV field \( G_0 \) (relative to the local interstellar value). These can be used as diagnostic plots to derive physical parameters for clouds from line observations. The \([\text{OI} + \text{CII} + \text{SiIII}] / G_0\) plot (top right) is the 'line efficiency', equivalent to the total line emission divided by the far-IR continuum. A typical value is \(0.1\%\).

Despite this seemingly complex structure, most PDRs may be described by varying just two parameters, the gas hydrogen nucleus number density, \( n \), and the far-UV field, \( G_0 \) (measured in units of the local interstellar radiation field, an equivalent 1D flux of \(1.6 \times 10^{-3}\) erg s\(^{-1}\) cm\(^{-2}\), Habing 1968). Other parameters, such as the chemical abundances, only have a secondary effect on the structure. We may therefore use observations of the emission lines as diagnostics of the physical state of the gas. Fig. 3 (from Burton et al. 1990) shows a series of line ratio plots for this purpose. For instance, the \([\text{OI}] / [\text{CII}]\) and CO line ratios are principally density indicators, with the presence of high-\( J \) CO indicating
very dense gas. The H$_2$ 1–0/2–1 S(1) ratio is commonly used to distinguish between fluorescently- and shock-excited gas. Over much of the parameter space shown in Fig. 3 this ratio is indeed constant (1·7), but for $n \gtrsim 10^5$ cm$^{-3}$ it is density-dependent and can approach the 'shock' value of 10. In the absence of other indicators that might favour fluorescence over shocks, this ratio cannot simply be used to discriminate between these two excitation mechanisms.

(2b) Some Applications

In this section we will outline some of the applications of the PDR model to interpreting observations of a wide range of sources, and their role in regulating star formation. The details will be found in the associated references (and see especially Hollenbach 1990).

Galactic Sources

The neutral, molecular and far-IR continuum of a number of Galactic sources are well described by the PDR model. Ranging from HII regions such as M17 and Orion, reflection nebulae such as NGC 2023, planetary nebulae such as NGC 7027, and the molecular ring around the Galactic centre (see Black and van Dishoeck 1987; Sternberg and Dalgarno 1989; Burton et al. 1990 and references therein), a PDR with far-UV field $G_0$ of $10^4$–$10^5$, mostly illuminating gas of moderate density $10^4$–$10^5$ cm$^{-3}$, but with a small fraction, typically 5–10%, at much higher density, $10^6$–$10^7$ cm$^{-3}$, provides a good fit to the data. These 'clumpy' PDRs are a ubiquitous feature of the ISM. The clumps may form an evolutionary sequence between dense condensations observed in molecular cloud cores and those seen in ionised gas, and be self-gravitating entities. Could they be the seeds for future star formation?

The $C^+$–CO Correlation

A tight linear correlation has been observed between the intensity of the [CII] (158 $\mu$m) and $^{12}$CO J = 1–0 lines in many galactic and extra-galactic sources (Crawford et al. 1985). This can be explained by a PDR model (see Wolfire et al. 1989), despite the C$^+$ arising from the warm surface layer, $A_V \lesssim 1$–2, and the CO originating deeper into the cloud. For low density both the [CII] and CO line intensities scale with density, and for high density the surface intensities reach constant values. The beam filling factors for the emitting gas are the same as they arise from the same regions. Hence a linear correlation exists between the observed intensities.

The ISM of IR-Bright Galaxies

IR and submillimetre observations of the central kpc of many nearby IR-bright galaxies can be explained using the PDR model, and thereby determine the average physical conditions in the neutral ISM of the nuclei (see Wolfire et al. 1990). The number, size, and density of molecular clouds, and the radiation field can be derived, as well as the mass of molecular gas, the mass of warm atomic gas, and the filling factors for these clouds in the galaxies. For instance, in M82, modelling the observations yields $3 \times 10^5$ small (0·4 pc), dense ($5 \times 10^4$ cm$^{-3}$) clouds, irradiated by a FUV field $G_0 = 8 \times 10^3$. The atomic gas (C$^+$) is at
Excitation of Molecular Clouds

\[ \sim 400 \text{ K} \text{ and accounts for } 10\% \text{ of the molecular gas mass of } 6 \times 10^7 M_\odot. \text{ The area filling factor of the clouds is } 0.3 \text{ and their volume filling factor in the nucleus is } 4 \times 10^{-4}. \]

The Regulation of Low and High Mass Star Formation

The observed constancy of optical depth of \( A_v \sim 7.5 \) through molecular clouds (Solomon et al. 1987) can be explained using a PDR model (see McKee 1989). The assumption is made that the rate of low mass star formation is controlled by ambipolar diffusion, and that newly formed stars inject energy into the clouds which support the clouds against gravitational collapse. For a cloud with low optical depth exposed to the local radiation field, a high ionisation rate is maintained through the cloud, which leads to a low ambipolar diffusion rate and hence a low star formation rate. The support decreases and the cloud collapses, raising the optical depth. On the other hand, for large initial optical depths, the interior is shielded from the FUV, the ionisation fraction is low, and ambipolar diffusion proceeds leading to increased star formation. The subsequent injection of energy causes the cloud to expand, thus lowering the optical depth. McKee showed that equilibrium is achieved for \( A_v \sim 7.5 \), and that the external FUV flux regulates the column density and the low mass star formation rate.

A feedback mechanism, based on the PDR model, has been proposed by Parravano (1988) to regulate high mass star formation. There are two stable states in the pressure–temperature phase diagram for the neutral gas of the ISM, a warm phase (\( \sim 10^4 \text{ K} \)) and a cold phase (\( \sim 100 \text{ K} \)), with a minimum pressure \( P_{\text{min}} \) needed for the cold phase to exist. The assumption is made that star formation takes place in the cold phase (e.g. through the coalescence of molecular clouds), and that grain photoelectric heating dominates so that \( P_{\text{min}} \) increases monotonically with the UV field \( G_0 \) (i.e. the PDR model). Suppose first that \( P > P_{\text{min}} \) so the cold phase exists, molecular clouds grow and OB stars form, increasing \( G_0 \) and raising \( P_{\text{min}} \). If then \( P_{\text{min}} > P \), the cold phase no longer exists, star formation drops, \( G_0 \) drops and \( P_{\text{min}} \) falls. OB star formation is therefore regulated so that \( P \sim P_{\text{min}} \) in galaxies.

(2c) Molecular Hydrogen Line Emission in PDRs

Fluorescent molecular hydrogen was first identified in the reflection nebula NGC 2023 (Gatley et al. 1987) on the basis of its rich vibrational–rotational near-IR spectrum, and in particular by the strength of the the 2–1 S(1) line compared with the 1–0 S(1) line (25% of its intensity compared with a typical 10% observed in shocks). The emission comes from a thin shell illuminated by the B1.5 star HD 37903 [see Fig. 8 (top), from Burton et al. 1989b]. Since then numerous other fluorescent sources have also been identified, including some which were first believed to be shock-excited (e.g. The Orion Bar, Hayashi et al. 1985). The interpretation of fluorescent \( \text{H}_2 \) emission is complicated by the need to consider two processes, radiative excitation of the \( \text{H}_2 \) molecule by far-UV, and its collisional de-excitation by H atoms. \( \text{H}_2 \) absorbs photons in the 11–13.6 eV range and is electronically excited. The decay back to the ground electronic state leads to the vibrational continuum (\( \text{H}_2 \) photodissociation) 10% of the time and to a bound, vibrationally excited state 90% of the time (typically at \( \sim 2 \text{ eV} \) above ground). In pure radiative fluorescence, as first investigated by Black and Dalgarno (1976) and recently in considerable detail by Black and van Dishoeck (1987), the
molecule cascades down the vibrational–rotational ladder, with a timescale of 1 year per decay, to the ground state, fluorescently emitting near-IR photons. The 1–0/2–1 $S(1)$ line ratio is 1.7. If, however, $n > n_{\text{crit}}$, collisional de-excitation by H atoms can proceed faster than radiative decay, with much of the vib–rotational energy transformed into heat. The spectrum is modified as the levels are pushed toward local thermal equilibrium (LTE), and can resemble a thermal spectrum as seen in shocks. Two recent models to explore this process in detail are those of Sternberg and Dalgarno (1989) and Burton et al. (1990). They are illustrated in Fig. 4, which shows a column density versus energy level diagram for a high density PDR from the latter model ($n = 10^6 \text{ cm}^{-3}$, $G_0 = 10^4$), and a jump shock ($n = 10^6 \text{ cm}^{-3}$, $V_s = 10 \text{ km s}^{-1}$) for comparison. A thermal distribution of energy levels is seen in the shock and the vibrational series cannot be distinguished in the diagram. In the PDR however, just the $v = 1$ series has merged with

---

**Fig. 4.** H$_2$ column densities, divided by the level degeneracy, as a function of energy level (in Kelvin), for a PDR (density $n = 10^6 \text{ cm}^{-3}$, UV-field $G_0 = 10^4$, top) and a jump shock ($n = 10^6 \text{ cm}^{-3}$, $V_s = 10 \text{ km s}^{-1}$, bottom). For the PDR each vibrational series is labelled, and within each series the rotational levels increase from $J = 0$ (except for $v = 0$ where they start from $J = 2$). In the shock the levels are thermalised, as is the case for the lower levels in the PDR; however the higher levels in the PDR ($v \geq 2$) retain a characteristic fluorescent distribution.
the \( v = 0 \) series, where collisions have thermalised their populations. Higher vibrational levels (\( \geq 2 \)) still retain their fluorescent distribution, with populations considerably greater than the thermal values. Excitation temperatures are often derived from the ratios of lines from different vibrational series, but as Fig. 4 demonstrates, their interpretation should be treated with care. For instance, between the 1–0 and 2–1 S(1) lines, a temperature of 2500 K would be derived in this case, between 3–2 and 2–1 S(1) it would be 8800 K, and between 4–3 and 3–2 S(1) 9500 K. A naive interpretation of data with such characteristics might be a combination of shocks and fluorescence, or even that there is a small fraction of very hot gas producing the highest excitation lines, when in fact a PDR, with collisional de-excitation of UV-pumped molecules in dense gas, provides a natural explanation.

The calculation of H\(_2\) line intensities in PDR models includes the contribution from UV pumping and collisional excitation, with the uncertain (small?) contribution of H\(_2\) formation normally neglected. The \( n \) and \( G_0 \) dependences of the FUV pumping contribution to an H\(_2\) line intensity, \( I_{\text{pump}} \), can be derived in various limits depending on whether self-shielding dominates dust attenuation of the FUV (\( n/G_0 \lesssim 40 \text{ cm}^{-3} \)) and whether collisional de-excitation of the FUV-pumped lines are important (\( n \gg n_{\text{crit}} \)) (see Burton et al. 1990):

\[
I_{\text{pump}} \propto n, \quad n \ll n_{\text{crit}}, \quad n/G_0 \lesssim 40 \text{ cm}^{-3},
\]

\[
I_{\text{pump}} \propto n_{\text{crit}}, \quad n \gg n_{\text{crit}}, \quad n/G_0 \lesssim 40 \text{ cm}^{-3},
\]

\[
I_{\text{pump}} \propto G_0, \quad n \ll n_{\text{crit}}, \quad n/G_0 \gtrsim 40 \text{ cm}^{-3},
\]

\[
I_{\text{pump}} \propto G_0 n_{\text{crit}}/n, \quad n \gg n_{\text{crit}}, \quad n/G_0 \gtrsim 40 \text{ cm}^{-3}.
\]

The thermal contribution, \( I_{\text{thermal}} \), is given by

\[
I_{\text{thermal}} \propto n N_2 \gamma_{\text{ex}}, \quad n \ll n_{\text{crit}},
\]

\[
I_{\text{thermal}} \propto N_2 e^{-E_u/kT}/Z(T), \quad n \gg n_{\text{crit}}.
\]

with \( E_u \) the energy of the upper state of the transition, \( N_2 \) the column density of warm H\(_2\), \( \gamma_{\text{ex}} \) the collisional excitation rate and \( Z(T) \) the partition function. Since the H\(_2\) rotational levels are \( \geq 500 \text{ K} \) above ground, and the vibrational levels \( \geq 6000 \text{ K} \) from ground, \( T \gtrsim 100 \text{ K} \) gas is needed for collisions to significantly excite the pure rotational spectrum and \( T \gtrsim 1000 \text{ K} \) to populate the vibrational levels. Since \( T \) increases for increasing \( n \) and \( G_0 \), the thermal contribution becomes more important, relative to the pump contribution, as \( n \) and \( G_0 \) increase. However, the column density \( N_2 \) of warm molecules appropriate for a particular line is a complicated function of \( n \) and \( G_0 \), and therefore the precise dependence of the thermal contribution on \( n \) and \( G_0 \) requires detailed numerical computations.

It is only recently that observational capabilities have progressed to the point where these models can be tested in detail. For instance, Geballe et al. (1992) have obtained spectra over the J, H and K atmospheric windows (1.1 to 2.5 \( \mu \text{m} \)) at the peak of the H\(_2\) emission from NGC 2023, measuring over 50 lines originating from all the vibrational levels from \( v = 1 \) to 7. The spectra are shown in Fig. 5, together
Fig. 5. Observations of the fluorescent $\text{H}_2$ emission spectrum in the reflection nebula NGC 2023, at J, H and K (solid lines), together with the best-fitting PDR model to the data (dotted lines), which uses a far-UV field $G_0 = 10^4$ and a clumpy density structure, with surface filling factors for the gas of $3 \times 10^4$ cm$^{-3}$ and $0.7 \times 10^6$ cm$^{-3}$ in the beam. Although a reasonable fit is obtained at K, it deteriorates at shorter wavelengths.

with the best-fitting PDR model to the data. The model uses a FUV field $G_0 = 10^4$, appropriate to the source, and a clumpy density structure. The area filling factors for the beam derived from the fit can then be used to constrain geometrical models for the source. As can be seen from the data in the K window (1.9–2.5 $\mu$m), which is dominated by the lower vibrational levels (i.e. $v = 1$ and 2), a good fit is obtained. At shorter wavelengths (and higher excitation) however, the fit is only moderate. A number of lines, particularly from the $v = 2$, 3 and 4 levels, are much stronger than the model would predict. Clearly the model does not calculate the full physics correctly, which is not entirely surprising since a number of simplifications and extrapolations are made, based on the behaviour at low excitation. A possibility is that $\text{H}_2$ formation is occurring, a process that has been neglected from the models. $\text{H}_2$ molecules are formed on the surfaces
of grains with about 2 eV of energy, placing them in a state with \( v = 4 \), and then are ejected from the surface with kinetic energy \( \sim 0.2 \text{ eV} \) (e.g. Duley and Williams 1986). The Duley and Williams model predicts we would observe newly formed H\(^2\) after collisional relaxation, via an overpopulation of the \( v = 0 \) and 1 levels with \( J \leq 6 \), but the details are uncertain. The excess emission observed from the first few vibrational levels in NGC 2023 may possibly come directly from the formation of new H\(^2\) molecules.

Despite the advances in IR technology that now make it possible to obtain spectra like those shown in Fig. 5, the interpretation is hindered by the relatively low spectral resolution of the data. Many lines are blended and cannot be unambiguously identified. Although the near-IR observations will continue to improve, there is another possibility for further progress, namely to observe at shorter wavelengths (\( \lambda < 1 \mu \text{m} \), and even higher excitation) where the great sensitivity and uniformity of optical CCDs can be exploited. On the other hand, the lines are very much weaker and the extinction greater, so the observations are not easy. However, guided by high spatial resolution images now attainable in the near-IR, it is possible to observe H\(^2\) at ‘near-optical’ wavelengths. Fig. 6 presents such a spectrum (from Burton et al. 1992a), from 7600 to 8700 Å, at the same location in NGC 2023 as Fig. 5. Over 30 resolved lines are seen from the \( v = 3 \), 4, 7 and 8 levels (up to 45,000 K above ground). Such observations await detailed

![Fig. 6. The fluorescent H\(^2\) emission in NGC 2023 near 8000 Å. Lines from the \( v = 3 \), 4, 7 and 8 series are labelled, together with a strong [CI] line, which is also of PDR origin. (The crosses denote artifacts left over after subtraction of strong OH night sky lines.) One line remains unidentified at 8449 Å.](image-url)
modelling, but they allow us to define the characteristic emission spectrum of fluorescently excited molecular hydrogen in great detail, and thus will tightly constrain models of the physical processes at work.

As a final example of H$_2$ line emission from PDRs, we will consider the emission from the lowest rotational levels of H$_2$, the 0–0 S(0) and S(1) lines at 28·2 and 17·0 μm respectively, arising from levels only 510 and 1015 K above ground. Their behaviour is somewhat different from the near-IR vibrational lines. They can be excited at cooler temperatures, have lower critical densities, and have the potential of being powerful probes of the extensive regions of warm (>100 K) molecular gas that exist in the Galaxy. On the other hand, these mid-IR lines are intrinsically weak ($\tau_{\text{decay}} \gtrsim 100$ years), and are difficult to observe due to the large thermal backgrounds and poor atmospheric transmission at their emission wavelengths. The 0–0 S(1) line was recently detected for the first time, in both the Orion shock and PDR (Parmar et al. 1991), but the fundamental 0–0 S(0) line has not yet been seen. With the imminent launch of the Infrared Satellite Observatory (ISO), and its spectroscopic capability from 2·5 to 200 μm (covering virtually all the PDR lines of interest), we may soon expect the 0–0 S(0) line to be measured. In Fig. 7 are shown some predictions of the surface brightnesses expected from the lowest two H$_2$ lines (from Burton et al. 1992b), and their contrast with the continuum (the IR continuum from PDRs peaks near the wavelength of the 28 μm line, so high spectral resolution is also necessary to detect it).

---

**Fig. 7.** Predictions of H$_2$ intensities for the 0–0 S(0) (squares) and 0–0 S(1) (diamonds) lines, at 28·2 and 17·0 μm respectively, in a PDR of density $10^6$ cm$^{-3}$, as a function of the far-UV field $G_0$. The left axis (solid lines) shows the intensity, and the right axis (dotted lines) the line-to-continuum ratio, for a spectral resolution 20,000, appropriate for the ISO. The fundamental 0–0 S(0) line has not yet been observed.
3. Shock Excitation in Molecular Clouds

(3a) General Considerations

Interstellar shock waves are generated by the supersonic injection of mass into the ISM, usually from young stellar objects, stellar winds or supernova remnants. Since the sound speed of cold interstellar gas is low, around 1 km s$^{-1}$, shock waves are common phenomena, and play a crucial role in determining the structure and dynamics of the ISM. The formation and destruction of molecular clouds, the heating of the hot phase of a three-phase medium, and the possible triggering of star formation are all influenced by shock waves. Since the bulk of the kinetic energy of the impact goes directly into heating the gas and is then radiated as the gas cools, shock waves are strong line emitters accessible to spectroscopic observation. Thermal emission from grains is usually relatively weak, in contrast to that from PDRs. The total intensity, from all cooling mechanisms, in the direction normal to the shock is given by

$$I_{\text{shock}} \sim 1.5 \times 10^{-1} \frac{\mu}{2.3} \frac{n_0}{10^6 \text{ cm}^{-3}} \left(\frac{V_s}{10 \text{ km s}^{-1}}\right)^3 \text{ erg cm}^{-2} \text{s}^{-1} \text{sr}^{-1},$$

where $n_0$ is the pre-shock hydrogen nucleus number density, $V_s$ is the shock speed, and $\mu$ the number of gas particles per hydrogen nucleus. Optical and UV-lines can dominate this emission, especially for fast shocks ($V_s \gtrsim 80 \text{ km s}^{-1}$), but many IR transitions have been identified from shocks, including H$_2$ vibrational–rotational lines, high-$J$ CO, $v = 1$–0 CO, low-$J$ OH, atomic and ionic fine-structure and forbidden transitions (e.g. [FeII] 1.64 $\mu\text{m}$, [OI] 63 $\mu\text{m}$, [SII] 35 $\mu\text{m}$), and atomic hydrogen (e.g. Br $\alpha$ and $\gamma$). The total H$_2$ luminosity may be large, often in the range of a 100–2000 $L_\odot$ for Galactic sources, and can be the dominant coolant in particular cases. We can also estimate a typical H$_2$ line intensity. For instance, the 1–0 S(1) line emission accounts for $\sim$7% of the total H$_2$ line emission from a shock, and if we assume that H$_2$ line radiation dominates the cooling, then $I_{1-0S(1)} \sim 10^{-2}$ erg cm$^{-2}$ s$^{-1}$ sr$^{-1}$ when $n_0 \sim 10^6 \text{ cm}^{-3}$, $V_s \sim 10 \text{ km s}^{-1}$. More realistically, H$_2$ cooling could only dominate when the gas temperature is in the range 1000–4000 K (ignoring possible contributions from CO, [O] and H$_2$O), and if the post-shock temperature reaches say 10,000 K then the 1–0 S(1) line intensity will be $\sim 30\%$ of this value, or $\sim 3 \times 10^{-3}$ erg cm$^{-2}$ s$^{-1}$ sr$^{-1}$.

Shock waves in molecular clouds have mostly been observed when outflowing gas from a young stellar object impacts ambient molecular material. The H$_2$ is heated to 1000 K or higher, and since it cools within a year observations of H$_2$ line emission directly trace the locations where the supersonic material interacts with the ambient gas. For instance, in the massive outflows of DR 21 and NGC 2071 (Garden et al. 1990; see also Fig. 8b), the H$_2$ delineates two well-collimated lobes along the outflow direction. In detail, however, the morphology is complex. Numerous irregularly spaced knots lie within distinct regions of emission in each lobe, with apparent changes of outflow direction between the regions. The H$_2$ may be showing us the inhomogeneous, clumpy nature of the molecular cloud as a precessing jet impinging on its inner surface. Instabilities within a jet are almost
Fig. 8. Contour maps of emission in the $v = 1-0$ S(1) line of H$_2$ (2.122 µm) in (a) the reflection nebula NGC 2023, (b) the bipolar outflow DR 21 and (c) the supernova remnant IC 443. NGC 2023 is a fluorescent source, excited by UV photons from a B1-5 star some 80'' N of the emission peak. The H$_2$ contours are shaded and overlaid by those of the 3.3 µm 'unidentified' emission feature. DR 21 is shocked, powered by the outflow from a massive young stellar object at the centre of the lobes. The crosses denote stars and the shading continuum emission. Finally, IC 443 is another shocked source, where the expanding blast wave from a SN is running into a molecular cloud, illuminating a molecular ring over 25 pc long in H$_2$ line emission.
Excitation of Molecular Clouds

---

**Fig. 9.** Illustrative sketches showing the characteristic structure of a J-shock (top) and a C-shock (bottom). Distance increases along the x-axis, from pre-shock to post-shock, with the y-axis showing the variation in three physical parameters: temperature $T$ (thick line), density $n$ (dotted line) and flow velocity $V$ (dashed line) in the frame of the ambient gas. For the C-shock, both the ion velocity ($V_i$) and the neutral velocity ($V_n$) are shown, illustrating the characteristic drift which occurs. This causes drag heating of the gas over a distance much greater than the mean free path of the gas, the scale length of the shock front in a jump shock. Also shown is the form of the line profile which would be produced by a single J- or C-shock.

---

certainly responsible for chains of evenly spaced knots seen in some HH objects, such as HH 7–11, with the leading knot having a bow shock morphology as it ploughs into the ambient cloud. Not all molecular shocks are associated with star formation, however. In the SNR IC 443 (Burton et al. 1988 and Fig. 8c) the expanding blast wave from a SN is overrunning the remnant of the molecular cloud from which the pre-SN star formed, producing a ring of H$_2$ emission over 25 pc long.

**3b) Jump and Continuous Shocks**

As will become apparent in the next section, there is considerable uncertainty surrounding our understanding of the shock mechanism, and in particular the structure of the shock front. Two basic forms of shock have been proposed, jump (J-type) and continuous (C-type), and in this section we will briefly review their properties. More extensive discussions can be found in the reviews of Hollenbach et al. (1989) and Draine (1991), and the references therein.

**J-Shocks**

These occur following a pressure disturbance in a medium where the sound speed is less than the wave velocity and where no information can be transported upstream of the disturbance (e.g. when the magnetic field is small). An irreversible transition occurs in which the energy of bulk flow is converted into random thermal energy of particles. In a distance less than the mean free path of the
molecules, a discontinuous jump or J-shock occurs in which the temperature, density and velocity of the gas change suddenly (see Fig. 9, top). Emission from the shock front itself is negligible, and the gas cools downstream with most of the emission coming out close to the velocity of the shock. It can be modelled by applying the Rankine–Hugoniot conditions to the jump, and then simply following the cooling and compression of the post-shock gas. High speed shocks \((V_s \gtrsim 40-50 \text{ km s}^{-1})\) into molecular gas are J-type, dissociating and possibly even ionising the gas. Lower velocity shocks with relatively high ionisation \((\gtrsim 10^{-6}\) relative to the number of nuclei), or with weak magnetic fields present, are also J-type.

**C-shocks**

A different type of shock occurs at low velocities and moderate to low ionisations (depending on the magnetic field), namely a magnetically-mediated continuous or C-type shock. In a partially ionised medium the gas should not be regarded as a single fluid, but instead the neutral material (containing most of the inertia) and the ionised material (bearing the magnetic field) must be treated separately. For a quiescent cloud these components are essentially comoving, but if a disturbance is applied then ‘slip’ velocities can arise between the neutral and ionised fluids. Essentially, if the Alfvén velocity in the ionised material is faster than the shock velocity of a disturbance, information can propagate ahead of the shock front warning the gas of its impending approach. The ions move under the influence of the magnetic field, and as they begin to compress, a drift velocity is set up between them and the neutrals. Drag heating (i.e. ambipolar diffusion) then heats the neutrals, and effectively softens the shock front. The flow in the ionised material must vary continuously, but the flow in the neutral gas can do so too if radiation is significant from the drag heating region, suppressing the increase in gas temperature. A C-shock is formed (see Fig. 9, bottom), with no discontinuous changes in the physical state of the gas, and the flow remaining supersonic. Emission arises from the entire drag heating region, and the line profiles reflect the whole velocity range from rest to the shock velocity. To calculate the emission and the profiles it is, however, necessary to compute the heating and cooling rates throughout the flow.

**(3c) A History of Shocked Molecular Hydrogen Research**

This section will discuss our current understanding (or lack thereof) of (slow) molecular shocks by way of a brief historical overview of observations and theory of the brightest \(\text{H}_2\) emission line source in the sky, the Orion Molecular Cloud (OMC-1). This is associated with the massive star formation of the BN-KL nebula and the deeply embedded source IRc2, lying behind the Orion Nebula. Fast, dissociative shocks, which can produce \(\text{H}_2\) line emission from reformed molecules downstream of the shock front, are not covered by this review (but see Hollenbach and McKee 1989; Neufeld and Dalgarno 1989). We do note, however, that in certain circumstances the spectrum might mimic that of fluorescently excited \(\text{H}_2\)!

Nor has the excitation of \(\text{H}_2\) in a magnetic precursor of a fast shock in very low density gas \((\sim 5 \text{ cm}^{-3})\) been considered. This has been suggested by Graham et al. (1991) as the explanation for the \(\text{H}_2\) line emission in the Cygnus Loop SNR, and may present yet further problems to our formulation of shock theory.
**Discovery and J-Shock Models**

Molecular hydrogen was first observed in emission through the 2.122 $\mu$m 1–0 S(1) line by Gautier et al. (1976), in OMC-1. Beckwith et al. (1978) mapped the distribution of the emission, finding two irregular lobes whose distribution roughly parallels the bipolar structure seen in CO, and suggesting a relation to the outflow. The 2–1 S(1) line (2.248 $\mu$m) was measured to be 10% the strength of the 1–0 S(1) line, ruling out fluorescence (see earlier) and prompting a number of hydrodynamic J-shock models to explain the emission (Hollenbach and Shull 1977; Kwan 1977; London et al. 1977; Shull and Hollenbach 1978). These models reproduced the 1–0/2–1 S(1) line ratio for shocks moving at $\sim$10 km s$^{-1}$ into gas of densities $\geq 10^5$ cm$^{-3}$, but also predicted that all the molecular gas would be dissociated for shock speeds $\geq$24 km s$^{-1}$.

**Further Observations and C-Shock Models**

As more observations of the OMC-1 shock were made, it became clear that that the simple J-shock picture above was not adequate. The 1–0 S(1) line was found to extend over 150 km s$^{-1}$ (Nadeau and Geballe 1979), much greater than the dissociation velocity. Estimates of the extinction to the emission region came out between 2 and 4 magnitudes, and to obtain these absolute intensities high pre-shock densities are required ($n > 10^6$ cm$^{-3}$). High-J CO had been observed (Storey et al. 1981) and such densities would produce excessive CO emission in very high levels ($J \gtrsim 35$), in conflict with the observations. C-shock models were then proposed to explain the observations (Draine and Roberge 1982; Chernoff et al. 1982; Draine et al. 1983), and fits were obtained for the CO and H$_2$ intensities with shock speeds of 36–38 km s$^{-1}$, pre-shock densities of $(2–7) \times 10^5$ cm$^{-3}$, and transverse magnetic fields of 0.5–1.5 mG. The fits were, however, extremely sensitive to the initial conditions and changing these by more than 20% leads to quite different predicted intensities and line ratios. In addition, although C-shocks alleviated the problem of the large line widths, by increasing the dissociation velocity to $\sim$45 km s$^{-1}$, they did not remove it.

**Infrared Spectroscopy and J-Shock Models Once More**

The sensitivity of near-IR spectrometers improved considerably in the mid-1980s, and four observations were made in OMC-1 which conflicted with the C-shock picture. The first was a map at high resolution of the ratio of the 0–0 S(13) and 1–0 S(7) lines of H$_2$ (Brand et al. 1989a), both being emitted close to 3.8 $\mu$m (and thus being simultaneously observable, since their separation, 0.04 $\mu$m, is within the bandpass of high resolution spectrometers), yet coming from quite different upper energy levels (17,000 and 8000 K above ground, respectively). Their ratio was expected to be sensitive to the local conditions, and thus to vary across OMC-1 as these changed. In fact it remained constant, despite the intensities varying by more than an order of magnitude from one position to another. Secondly, a spectrum was obtained across the K and L windows (2–4 $\mu$m) which contained over 25 H$_2$ emission lines, originating from vibrational levels from $v = 0$ to 4, and up to $J = 19$ (Brand et al. 1988; Burton et al. 1989a). From the intensities it was clear that the C-shock models predicted insufficient
emission for both the very lowest excitation [0–0 S(2) at 12.3 μm] and highest excitation [0–0 S(17) at 3.49 μm and 4–3 S(3) at 2.34 μm] transitions of H₂. The temperature of the gas in C-shocks was simply not high enough to excite the columns of high-excitation gas observed. Thirdly, line profiles from the ν = 1, 2 and 3 levels (Moorhouse et al. 1990) proved to be identical, suggesting that the high-excitation lines are excited in exactly the same shocks as those of lower excitation, and not in separate shocks. Finally, lines from the CO ν = 1–0 band (4.7 μm) were detected from OMC-1 (Geballe and Garden 1987), with similar linewidths to the H₂ lines. This could be reproduced in a C-shock only by increasing the density by an order of magnitude from the model fit.

Brand et al. (1988) proposed a very simple model to account for these observations, a non-magnetic jump shock into dense gas (n > 10⁶ cm⁻³) where H₂ line emission and molecule dissociation controlled the cooling, and not emission from any other species (such as O, CO or H₂O) (see Fig. 10). Essentially, molecule dissociation rapidly brings the post-shock gas temperature down to ~4000 K, and from there the cooling through H₂ lines is the same for all shocks. Only one free parameter is required in the model [after normalising to the 1–0 S(1) line intensity], the shock driving pressure, which controls the degree of dissociation and thus the intensity of the highest excitation lines, which arise from the hottest

Fig. 10. Molecular hydrogen line ratios in OMC-1, plotted as column density versus energy level, and labelled by transition. The column density (normalised so the 1–0 S(1) line is unity) is divided by the column expected for gas in LTE at 2000 K. The excess of hot gas (i.e. above 2000 K) is seen by the rise at higher energy level. C-shock models, by contrast, predict a deficit. The continuous line is the best-fitting J-shock model, and depends on only one free parameter, the driving pressure behind the shock.
gas. Line ratios are generally quite insensitive for $V_s$ in the range $10^{-2} - 10^{-3}$ km s$^{-1}$, and thus line profiles are also the same. Improved estimates of the extinction at 2 $\mu$m had reduced its value to 1 mag so the absolute intensities of the lines could be reproduced. The high-\emph{J} CO lines still presented a problem (too much emission would be produced in the dense gas), but the combination of difficulty in comparing the large-beam CO data with the small-beam H$_2$ data, and the uncertain contribution of the Orion photodissociation to the high-\emph{J} lines, placed less emphasis on the problem. Accounting for the wide line profiles still posed a serious problem.

\textit{Line Profiles and Bow Shock Models for C-Shocks}

Despite the impressive success by Brand \emph{et al.} (1988) in accounting for the H$_2$ line ratios and their constancy with both position and as a function of velocity, understanding how such a J-shock could exist poses severe theoretical problems. Firstly, the theory predicts that J-shocks should only exist under these conditions if the magnetic field is unreasonably low ($B \lesssim 10 \mu$G), whereas our expectations are for fields of order 1 mG, or if the ionisation fraction is unreasonably high (perhaps $> 10^{-5}$ for $n > 10^6$ cm$^{-3}$, compared with a typical scaling assumed for the ionisation fraction of $10^{-5} n^{-0.5}$). A field of this strength must have an important effect on the dynamics, necessitating the need for a multifluid model. Secondly, the J-shock model assumes that H$_2$ dominates the cooling, whereas all theoretical chemical studies show that a large fraction of the O is converted to OH and H$_2$O, which then would dominate the cooling. A quite different H$_2$ emission spectrum is produced for such a J-shock.

We might speculate that external UV or X-rays may keep the ionisation fraction relatively high in the pre-shock gas (perhaps with all C as C$^+$, which would maintain it at $\sim 10^{-4}$), and at the same time suppress the formation of H$_2$O post-shock, thereby permitting J-shocks with H$_2$ cooling. However, the J-shock model of Brand \emph{et al.} simply cannot explain the H$_2$ line profiles without some \emph{ad hoc} assumptions. For instance, the H$_2$ 1-0 S(1) line profile has been measured with 1" spatial and 12 km s$^{-1}$ spectral resolution in OMC-1 (Burton \emph{et al.} 1991). There is little variation from pixel to pixel across the source. With few exceptions it has extreme width ($\sim 150$ km s$^{-1}$), is smooth with an enhanced blue wing, and has a peak velocity at rest with respect to the ambient cloud. This can be reproduced in a J-shock model by hypothesising that an ensemble of emitters, each with its own velocity profile, contributes to the observed profile along every line of sight where it has been measured (see Brand \emph{et al.} 1989b). However, this requires the individual cloudlets to achieve velocities up to $\pm 75$ km s$^{-1}$, within a cylinder of cross section $7 \times 10^{15}$ cm everywhere within the source. Such an efficient redistribution of momentum seems unlikely.

In a series of papers (Smith and Brand 1990a, 1990b, 1990c; Smith \emph{et al.} 1991a, 1991b; Smith 1991) the merits of C-shocks are reconsidered, but now in fast bow shocks rather than the planar shocks previously studied. A dense knot of gas is envisaged as being driven at high velocity into ambient molecular gas, creating a bow shock structure around the leading edge of the clump. The emission at each point on the bow shock is computed as in a planar shock, using the appropriate components of the shock velocity and magnetic field, and then the emission is summed over the entire bow structure. At the head of the
Fig. 11. Sketch of a bow shock, as used by Smith et al. in their C-shock model of OMC-1. A dense knot or 'bullet' is driven at \( V_w \sim 250 \text{ km s}^{-1} \) into quiescent molecular gas (or, in reverse, a fast wind might overrun a dense clump of gas), resulting in the formation of a paraboloidal bow shock around its leading edge. All the gas in the cap of the bow is dissociated. The shock velocity at each point along the bow is the normal component of the wind velocity at that point. There is a region behind the cap where \( \text{H}_2 \) is excited; far enough along the bow the normal velocity is too small to shock the gas. The model requires a strong magnetic field (\( \sim 30 \text{ mG} \)) and dense gas (\( n \sim 10^6 \text{ cm}^{-3} \)), and the field and observer are orientated at some arbitrary angle to the bullet. The calculated line intensities are the sum of all emission from the bullet.

bow, the shocks are dissociative, and the bulk of the \( \text{H}_2 \) emission comes from the outer part of the bow, where the velocity normal to the surface is below the breakdown velocity (see Fig. 11). The gas gets cooler moving from the head to the tail of the bow, with atomic optical line emission arising from hot gas at the head, then fine structure \([\text{OI}]\) 63\( \mu \text{m} \) and progressively lower excitation line emission further along, to a cool region in the tail emitting low-\( J \) \( \text{CO} \). With a shock velocity of 250 km s\(^{-1}\) for the bow, a high magnetic field strength of 30 mG, and a clump density of \( 10^6 \text{ cm}^{-3} \), Smith et al. have modelled (i) the observed 150 km s\(^{-1}\) wide line profile, (ii) the similarity of profiles from different energy levels, (iii) smooth, symmetric profiles centred on the rest velocity, provided the orientation is within 30° of the plane of the sky, (iv) the line ratios, (v) the constancy of the line ratios across the source, and (vi) the observed [OI], OH and CO (rotational + vibrational) line strengths and widths.

This is an impressive range of achievements for the model. There are some cautionary matters, however, which must be addressed before the model can be accepted as explaining the shock in OMC-1. Firstly, the strong magnetic field, equivalent to an Alfvén speed \( \sim 45 \text{ km s}^{-1} \), is about 30 times larger than we would expect for these densities; our limited knowledge of the magnetic field in interstellar clouds appears to be roughly consistent with \( B_0 \sim b_0 n^{0.5} \mu \text{G} \), with \( b_0 \sim 1 \) (Troland and Heiles 1986). Secondly, a paraboloidal shape is assumed for
the bow (indeed, a spherical shock is shown not to work). Can it be expected to remain paraboloidal (e.g. instabilities?) and, if so, what effect will this have, especially on the line ratios? Thirdly, in order to measure the integrated emission, the bow structure must not be resolved by the observations, which constrains the size of the 'bullets' to be <1–5" (i.e. less than the beam sizes of the line profile and line ratio observations). Conversely, the radius of curvature of the 'bullet' must be large compared with the shock thickness, which is of order $10^{15}$ cm, or 0.1" at the distance to Orion. Already the size of the bullets is tightly constrained, and future observations, at higher spatial resolution, will be able to test this bow shock model by searching for changes in line ratios on small spatial scales.

To end this discussion on a final salutary note, a recent observation of shocked [C+] 158 μm line emission from the SNR IC 443 (Haas et al. 1990) presents another challenge to all shock models. The line is clearly not of PDR origin, and fast shocks, which could produce the emission, are ruled out due to the lack of IR continuum. None of the shock models described above can produce the intensity seen in IC 443. C+ also is seen in OMC-1, but determination of the contribution, if any, from the shocks is hindered by the strong emission from the Orion PDR.

4. Conclusions

Far-UV photons and shock waves are the major excitation sources in molecular clouds, resulting in a rich spectrum of atomic fine-structure and molecular lines at infrared and submillimetre wavelengths. The morphology, chemical composition and physical state of the molecular clouds is determined in these processes, with consequent effect on the conditions that star formation occurs in. Emission from molecular hydrogen, the major constituent of the clouds, has only been observed from these two processes. Since the H2 line emission is optically thin, observations of the emission from a range of energy levels yields the population of excited molecules, which can be directly compared with the theoretical expectations from shocks and PDRs. It has become clear that the standard method of distinguishing between shocked and fluorescent H2 line emission, i.e. measuring the ratio of the 1–0 and 2–1 S(1) lines, is no longer appropriate without additional evidence for either mechanism. Such evidence may include (i) larger line widths in shocks, (ii) larger line to continuum ratios in shocks or (iii) stronger high-excitation lines ($v \geq 3$) in fluorescence. With the recent major advances in the capabilities of IR instrumentation, we are now able to obtain H2 spectra of unprecedented quality, and thereby thoroughly test the models. The first such detailed observations in a fluorescent source, NGC 2023, have found excess emission in several highly-excited lines, which may be a signature of H2 formation. Similar observations in the shocked source OMC-1 have led to major revisions of the theory of molecular shocks, which still is the subject of considerable change.

Acknowledgments

This paper has resulted from research conducted over a number of years and I am grateful to many colleagues and collaborators from the University of Edinburgh, the Royal Observatory, the Joint Astronomy Centre Hawaii, NASA Ames Research Center and the Anglo Australian Observatory for their support and advice over this time. I particularly wish to thank David Allen and David
Hollenbach and the referee for their comments on the manuscript, which have greatly improved its presentation.

References


