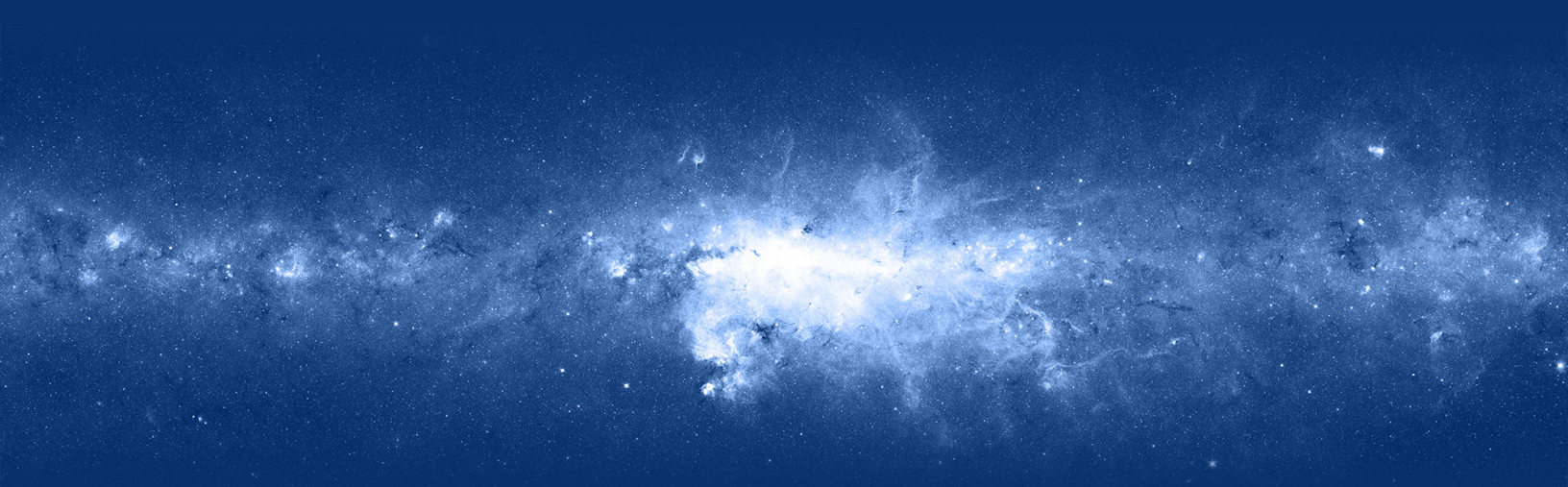
ASTR 670 • The interstellar medium

by Benedikt Diemer (University of Maryland)

Updated May 6, 2025



Contents

0 About these notes

These notes are a guide to the interstellar medium (ISM) portion of ASTR 670 at the University of Maryland. They are less self-contained than the hydrodynamics notes for the first half of the course (§A.1 gives an ultra-brief introduction to the necessary hydrodynamics). In particular, these notes frequently refer to textbooks (see below) and they do not reproduce much of the information conveyed via lecture slides.

0.1 Acknowledgments and commonly used references

This course is based on a previous version taught by my colleague Alberto Bolatto, whose overall course design and notes I relied on. We will frequently refer to the seminal textbook by Draine (2011, hereafter Draine), as well as the textbooks of Lequeux (2005, hereafter Lequeux) and Osterbrock & Ferland (2006, hereafter O&F). I also took inspiration from the excellent ISM slides by Karin Sandstrom (UCSD).

The image on the title slide is artistically adapted from the Spitzer GLIMPSE survey of the ISM in the Milky Way.

0.2 Notation

We try to adapt a consistent notation follows the hydrodynamics notes wherever possible, with some noteworthy exceptions. Tables 1 and 2 list the meaning of all frequently used symbols and constants.

Sym.	Units	Meaning	§
A_{21}	s^{-1}	Einstein A coefficient for spontaneous emission	3.4
α_{AB}	cm^3/s	Recombination coefficient (often Case A or B)	5.2
B_{12}	$cm^3/erg\ s$	Einstein B coefficient for absorption	3.4
c_s	cm/s	Sound speed	A.1
E	erg/cm^3	Total energy per unit volume, or other energy	A.1
ε	$erg/g = cm^2/s^2$	Internal energy per unit mass (often thermal, $\varepsilon = \varepsilon_{th}$)	A.1
$f(u)$	s/cm	Velocity distribution function, often Maxwell-Boltzmann	3.3
g_X	—	Statistical weight of state X	3.1
G_{pi}	s^{-1}	Ionization rate	5.1
γ	—	Ideal gas equation of state parameter	A.1
Γ	$erg/cm^3/s$	Heating rate per unit volume	A.1
j_ν	$erg/cm^3/s/Hz$	Emissivity of radiation	4.2
J_ν	$erg/cm^2/s/Hz$	Mean intensity of radiation	5.1
κ_ν	cm^{-1}	Attenuation coefficient	4.2
L_ν	erg/s/Hz	Luminosity at frequency ν	6.1
λ	cm	Wavelength	A.1
λ_{mfp}	cm	Mean free path	A.1
Λ	$erg/cm^3/s$	Cooling rate per unit volume	2.2
m	g	Mass in a fluid element	A.1
m_{ptl}	g	Mass of individual particles	A.1
$\dot{M}$	$M_\odot/yr$	Mass accretion or outflow rate	2.5
μ	—	Mass of particles in units of proton mass m_p	A.1
n	cm^{-3}	Number density	1.2
n_e	cm^{-3}	Number density of electrons	1.2
n_p	cm^{-3}	Number density of protons (ionized hydrogen atoms)	1.2
n_H	cm^{-3}	Number density of hydrogen nuclei or atoms	1.2
$\bar{n}_\gamma$	—	Photon occupation number	3.3
N_X	cm^{-2}	Column density of species X	4.2
ν	$Hz = s^{-1}$	Frequency of EM radiation	5.1
P	$dyne/cm^2 = erg/cm^3$	Pressure in a fluid element	1.3
q_{12}	cm^3/s	Collisional excitation/deexcitation rate	3.5
r	cm	Radial coordinate in Eulerian space	—
R_e	—	Reynolds number (relative importance of viscosity)	9.3
ρ	g/cm^3	Mass density	A.1
σ_{12}	cm^2	Cross-section for collisional excitation or deexcitation	3.5
σ_{pi}	cm^2	Cross-section to photoionization for frequency ν	5.1
σ_u	cm/s	Velocity dispersion	4.2

Table 1: Continued on next page.

Sym.	Units	Meaning	§
t	s	Time coordinate	—
t_X	s	Lifetime of state X or time for X to happen	—
T	K	Temperature	A.1
τ	—	Optical depth	—
Υ_{12}	—	Velocity-averaged collision strength	3.5
u	cm/s	Particle velocity (differs from the hydro notes)	5.2
$\boldsymbol{u}$	cm/s	Fluid velocity	A.1
U_ν	erg/cm^3/Hz	Energy density of radiation	3.3
$\boldsymbol{x}$	cm	Coordinate position vector in Eulerian space	—

Table 1: Definition of symbols used throughout the text. The section given in the right column typically refers to the section where a quantity is first defined. We are using the CGS unit system throughout.

Sym.	Value	Units	Meaning
yr	3.16×10^{7}	s	Year in seconds
eV	1.60×10^{-12}	erg	Electron volt unit
Å	10^{-8}	cm	Angstrom
AU	1.50×10^{13}	cm	Astronomical unit
pc	3.09×10^{18}	cm	Parsec
kpc	3.09×10^{21}	cm	Kiloparsec
Mpc	3.09×10^{24}	cm	Megaparsec
a_0	5.29×10^{-9}	cm	Bohr radius
α	7.30×10^{-3}	—	Fine-structure constant (about 1/137)
k_B	1.38×10^{-16}	erg/K	Boltzmann constant
h	6.63×10^{-27}	erg s	Planck constant; $\hbar \equiv h/2\pi$
c	3.00×10^{10}	cm/s	Speed of light
G	6.67×10^{-8}	cm^3/g/s^2	Gravitational constant
q_e	4.80×10^{-10}	statC	Electron charge
m_e	9.11×10^{-28}	g	Electron mass
m_p	1.67×10^{-24}	g	Proton mass
Ryd	2.18×10^{-11}	erg	Rydberg, hydrogen ionization energy, 13.6 eV
σ_SB	5.67×10^{5}	erg/s/cm^2/K^4	Stefan-Boltzmann constant
$M_\odot$	1.99×10^{33}	g	Solar mass
$R_\odot$	6.96×10^{10}	cm	Solar radius

Table 2: Physical and astronomical constants and unit conversions in Gaussian CGS units as used throughout the text.

1 Introduction: What is the ISM?

The ISM is not one thing: it consists of different gas "phases" that exhibit rather diverse properties. We begin with an overview of these phases (§1.1 and §1.2). Arguably, their most important interaction is that they hydrodynamically push against each other to establish an approximate pressure equilibrium (§1.3). Similarly, different forms of energy in the ISM also store similar amounts of energy (§1.4). Finally, we consider the electromagnetic (EM) spectrum and the kinds of observations on which our knowledge of the ISM rests (§1.5).

1.1 Gas phases

Figure 1 shows a list of the main gas phases that constitute the ISM. Their temperatures and densities should be taken to be rough indications because the phases exhibit a range of properties (some with larger scatter than others) and because some of the values are poorly constrained. Since the majority of cosmic gas is made of hydrogen, its state is a key feature of the phases. We interchangeably use the symbols H II and H^+ to denote ionized hydrogen atoms. Since they are, simply put, protons, their properties are sometimes abbreviated with the subscript p. Atomic hydrogen is written as either H I, H^0, or simply H, depending on the context. We consider m_p to be the mass of a hydrogen atom regardless of its ionization state (ignoring the negligible contribution from electrons). Diatomic molecular hydrogen is always denoted as H_2.

The ISM is a highly active system, where gas continuously cycles between different phases. Figure 2 shows an overview of some of the most important conversion processes that we will study in these notes. When gas first enters the galaxy, it has likely already been shock-heated to hot temperatures (§2), although it can also directly enter the WNM via "cold streams." Stars back-react on the gas via "feedback," a summary term that includes radiation, stellar winds, planetary nebulae, and supernovae.

Phase		T (K)	n_H (cm^{-3})	f_V -	P/k_B (K/cm^3)	Comments
H II 23%	Hot ionized medium (HIM)	$10^{5.7}$	0.004	0.5	4400	Collisionally ionized, shock-heated by supernovae and stellar winds
	H II regions	10000	0.1-10^4	0.01	varies	Photo-ionized nebulae around stars; density and pressure vary across these bubbles
	Warm ionized medium (WIM)	8000	0.2	0.1	4400	Diffuse photo-ionized gas, large scatter in temperature and density
H I 60%	Warm neutral medium (WNM)	8000	0.5	0.4	4400	About 60% of HI by mass; in pressure equilibrium witn CNM
	Cool neutral medium (CNM)	100	40	0.01	4400	Significant fraction of the mass despite small volume filling fraction
H_2 17%	Diffuse molecular gas	50	150	0.001	4400	Self-shielded against dissociation, but not dense enough to form stars
	Molecular clouds	10-50	10^3-10^6	0.0001	>10000	The site of star formation; more or less gravitationally bound

Figure 1: Overview of the phases of the ISM. The left column distinguishes the three fundamental states of hydrogen: ionized (H II), atomic (H I), and molecular (H_2); the percentages indicate their approximate mass fractions. All densities are given as n_H (§1.2), and f_V denotes the volume filling fraction. Most of the numbers are based on Table 1.3 in Draine. Many of them are uncertain, or typical values for quantities with large scatter, or both. The exact values were picked to highlight the approximate pressure equilibrium between the HIM, WIM, WNM, and CNM (where $P/k_B = nT$ rather than $n_H T$, a conversion that depends on whether the gas is ionized, atomic, or molecular).

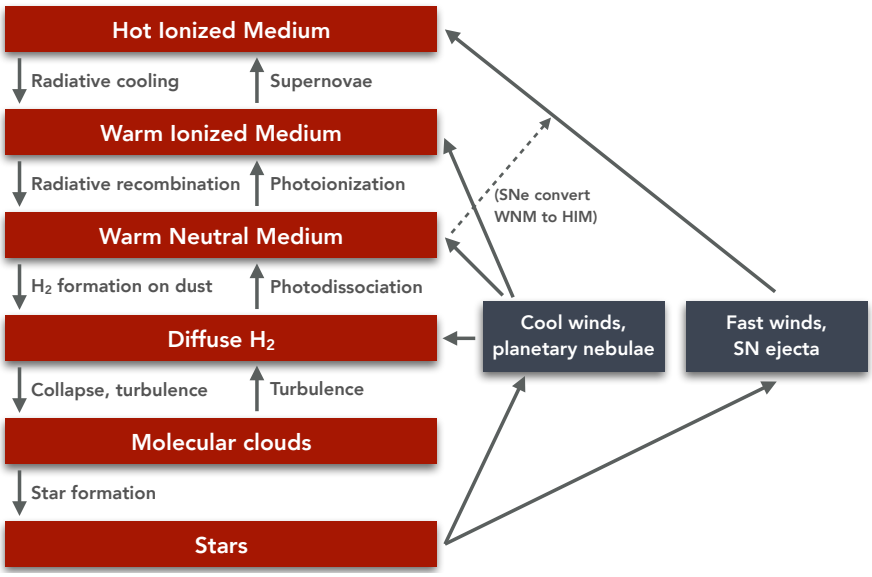


Figure 2: The most important processes that convert gas from one phase to another, although more than one process is responsible in many cases.

1.2 Conversion between densities

We need to be careful with the densities of different states of hydrogen, or when converting density to number density. The conversion depends on the Helium fraction, metallicity, and ionization state. After Big Bang nucleosynthesis, the Universe has a hydrogen fraction of $f_\mathrm{H} = 0.76$ by mass, meaning that 24% of the mass is in Helium. Given that Helium nuclei are four times heavier, this would correspond to a hydrogen fraction by number of

$$f_\mathrm{H,n} = 1 - \frac{1 - f_\mathrm{H}}{4} = 0.94\,, \tag{1.1}$$

meaning that 6% of particles are Helium nuclei. Throughout this course, however, we assume the Solar abundance of 10% Helium by number. This estimate takes into account that ISM gas is likely enriched above cosmic values, although we ignore heavier metals for simplicity. We then have $f_\mathrm{H,n} = 0.9$ and $f_\mathrm{H} = 0.6$, or about 40% of the mass in Helium.

In order to convert between ρ and n, we need the mean particle weight. This number now depends on whether the gas is ionized, neutral atomic, or molecular because freeing electrons or forming molecules changes the number of particles without changing their mass. For atomic gas, we have

$$\mu_\mathrm{atomic} = f_\mathrm{H,n} \times 1 + (1 - f_\mathrm{H,n}) \times 4 = 0.9 + 0.1 \times 4 = 1.3\,. \tag{1.2}$$

In molecular gas, we assume that all H atoms are in H_2. The Hydrogen fraction by number becomes $f_\mathrm{H,n} = 1 - 0.1/(0.9/2 + 0.1) \approx 0.82$, which gives a mean particle weight

$$\mu_\mathrm{molecular} = f_\mathrm{H,n} \times 2 + (1 - f_\mathrm{H,n}) \times 4 = 0.82 \times 2 + 0.18 \times 4 \approx 2.36\,. \tag{1.3}$$

For ionized gas, Hydrogen has a mass of half a proton per particle and Helium has mass 4/3 per particle (since there is one nucleus and two electrons), so we get

$$\mu_\mathrm{ionized} = 0.9 \times 0.5 + 0.1 \times 4/3 = 0.58\,. \tag{1.4}$$

These changing conversions are annoying, which is why we will generally use the number of protons in hydrogen independent of its state, $n_{\rm H}$, rather than n. The approximate conversion is

$$n_{\rm H} \equiv n({\rm H}^0) + n({\rm H}^+) + 2n({\rm H}_2) = \frac{\rho}{\mu/0.9 \times m_{\rm p}}\,, \tag{1.5}$$

where the factor of 0.9 again represents $f_{\rm H,n}$ for atomic gas, since $n_{\rm H}$ is define to count Hydrogen atoms (protons). For atomic gas, we thus have

$$n_{\rm H} = \frac{\rho}{\mu_{\rm atomic}/0.9 \times m_{\rm p}} \approx \frac{\rho}{1.4\ m_{\rm p}}\,. \tag{1.6}$$

We also need to be careful when converting between $n_{\rm H}$ and n. We get

$$n_{\rm atomic} = n_{\rm H} + 0.1\ n_{\rm H} = 1.1\ n_{\rm H}\,, \tag{1.7}$$

whereas totally molecular gas would obey the relation

$$n_{\rm molecular} = 0.5\ n_{\rm H} + 0.1\ n_{\rm H} = 0.6\ n_{\rm H}\,. \tag{1.8}$$

In ionized gas, we need to count one electron per hydrogen and two per Helium, so we get

$$n_{\rm e} = n_{\rm H} + 2 \times 0.1\ n_{\rm H} = 1.2\ n_{\rm H} \tag{1.9}$$

and

$$n_{\rm ionized} = n_{\rm e} + n_{\rm i} = 1.2\ n_{\rm H} + 1.1\ n_{\rm H} = 2.3\ n_{\rm H}\,. \tag{1.10}$$

In general, we will neglect the contribution of higher elements to these densities because they are small (Table 1.4 in Draine) and because they would introduce a dependence on the metallicity. We will write the number density of ionized hydrogen interchangeably as $n({\rm H}^+)$ or as the number density of protons, $n_{\rm p}$, depending on the context. In the case of pure hydrogen, $n_{\rm e} = n_{\rm p}$, but that is not the case once we include helium and metals as demonstrated above.

1.3 Pressure equilibrium

In hydrodynamics, we thought of pressure as an energy density in units of erg/cm^3, equivalent to thermal, kinetic, or magnetic energy. When studying the ISM, we are often more interested in pressure as a proxy for the density and temperature of a medium. We convert

$$P = nk_{\rm B}T \qquad \Longrightarrow \qquad P/k_{\rm B} = nT \tag{1.11}$$

and write the pressure as $P/k_{\rm B}$, with units of K/cm^3. While these units are not terribly meaningful, the interpretation of the value of $P/k_{\rm B}$ is straightforward: it is the product of number density per cm^3 and temperature in K.

One of the most noteworthy features of the densities and temperatures listed in Figure 1 is that they result in very similar pressures for many phases, around 4400 K/cm^3 for the HIM, WNM, and CNM (the estimates for WIM and diffuse molecular gas vary quite a bit). On some level, this result is expected because phases with higher pressure would expand and displace those with lower ones until they reach an approximate equilibrium (e.g., Spitzer 1956). Molecular clouds tend to have higher pressure because they experience self-gravity.

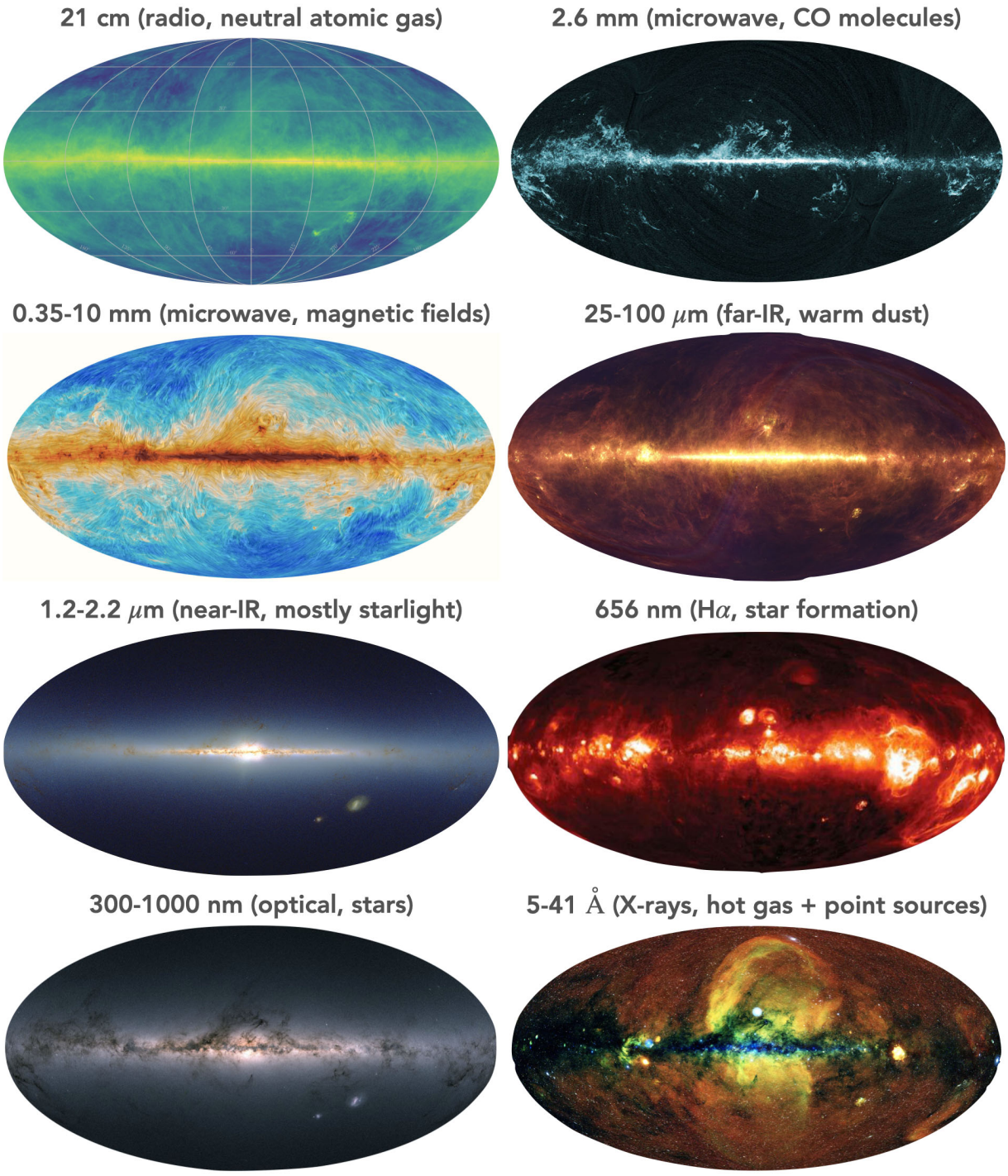


Figure 3: All-sky maps of the Milky Way in various wave bands, from radio (top left) to X-rays (bottom right). The projections are such that the center of the images shows the center of the Galaxy. The images are from (1) the Effelsberg-Parkes H I survey, (2) the Planck satellite measurements of carbon-monoxide (CO) line emission, (3) the Planck polarization measurements, which show magnetic field strength and direction, (4) the IRAS satellite, (5) the 2MASS survey, (6) Finkbeiner (2003), (7) the GAIA satellite, and (8) the eROSITA all-sky survey. If a wavelength range is given, the images are composites of various wavelengths.

Component	E (eV/cm^3)	Comments
Cosmic microwave background (CMB)	0.27	
Far-IR radiation from dust	0.31	
Starlight	0.54	For energies $h\nu < 13.6$ eV
Thermal energy	0.49	
Kinetic energy	0.22	Largely from turbulence
Magnetic energy	0.89	For a median $B \approx 0.6\mu G$
Cosmic ray energy	1.39	

Table 3: Estimates of the energy per unit volume in the local ISM. The different forms of energy are remarkably close to equipartition. Adapted from Table 1.5 in Draine.

1.4 Energetics

Besides the densities and pressures considered in the previous sections, another fundamental property of the ISM is the energy density in various forms (such as thermal energy, light, gas motions, cosmic rays, and magnetic fields). Some of the corresponding energy densities are hard to measure, but Table 3 shows our best guesses for the local ISM. Many of the components are expected to vary strongly across the galaxy. For example, the intensity of starlight depends on the distance to the nearest massive star.

All of the energy densities listed are surprisingly close to equipartition, with no more than a factor of a few between the smallest and largest. In some cases, this must be a coincidence. For example, the intensity of the CMB is set by cosmology and has nothing to do with the ISM. If we happened to live at $z = 1$, the CMB energy density would be $(1 + z)^4 = 16$ times higher.

In other cases, however, there are good reasons to expect approximate equipartition. For example, galaxies act as magnetic dynamos that will, after a sufficient number of rotations, bring the magnetic field into approximate equipartition with the kinetic energy (see hydro notes). If the kinetic energy is largely caused by turbulent motions, we would also expect an approximate equipartition with thermal energy. As mentioned, the intensity of starlight could be much smaller or greater depending on our exact location, but if it were orders of magnitude larger everywhere, it would push ISM gas out of the galaxy, which would in turn slow down star formation and reduce the stellar radiation field. In other words, some of the similarities between energy densities are probably explained by equilibria between different processes such as star formation and gravity. We will return to such overarching questions in §10.

1.5 Observations across the EM spectrum

Our knowledge of the galactic ISM is based on observations across almost the entire EM spectrum, perhaps with the exception of γ-rays. Figure 3 shows an overview of all-sky projections across this vast range of wavelengths. Figure 4 shows a schematic of the EM spectrum and the names typically associated to certain wavelength ranges. These denominations are not always unique; for example, the ranges of "far-infrared," "sub-millimeter," and "microwave" often overlap somewhat.

Generally speaking, the energy of the radiation is connected to the energy (and thus temperature) of the emitting processes. On the low-energy end of the spectrum, the spin-flip transition of atomic hydrogen emits 21 cm radiation (§4), and rotational transitions in molecules emit photons in the mm range (§8, first and second panels of Figure 3). The polarization of the CMB can tell us about magnetic fields because of polarized emission from aligned dust grains (third panel). Dust

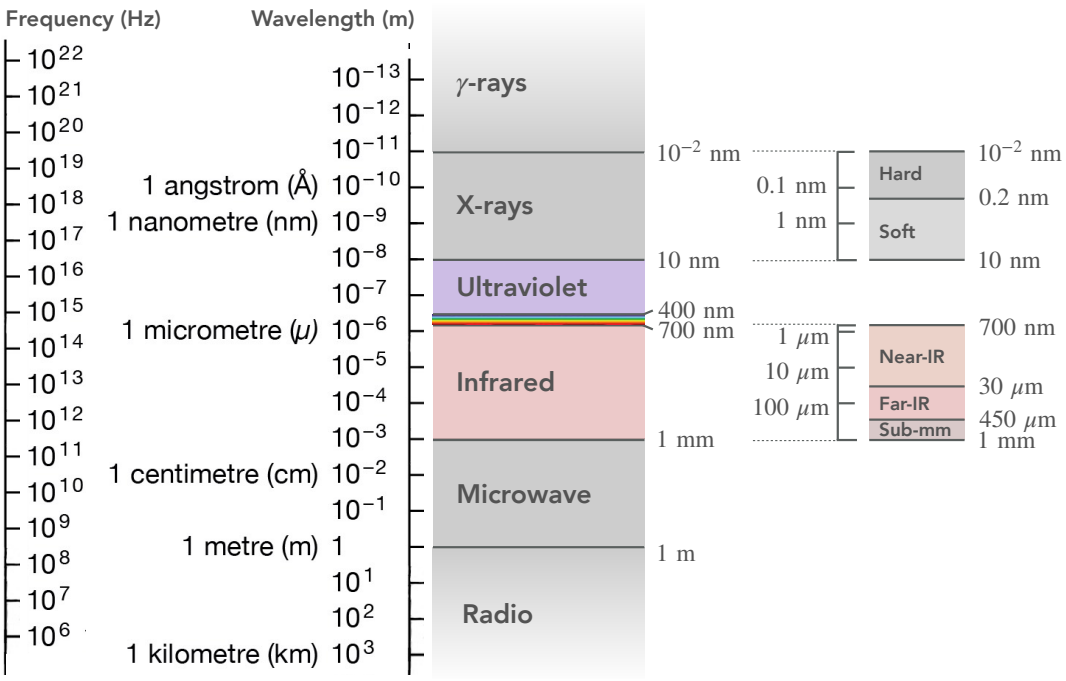


Figure 4: Overview of the electromagnetic spectrum. The bands are named according to astronomical convention, although some divisions are not unique. For example, the ranges that are called "far-IR" and "sub-millimeter" often overlap.

generally emits in the infrared (§8.3, fourth panel), whereas starlight dominates the near-IR and optical bands (fifth and seventh panels). The hydrogen $n = 3 \rightarrow 2$ transition is known as Hα, and is a good indicator of star formation (§6.5, sixth panel). Although UV all-sky maps now exist, this waveband is hard to observe because it is very efficiently absorbed by dust. Finally, hot gas emits X-rays (§2, eighth panel). In these notes, we will generally tackle the different phases progressing from hot to cold, though with some exceptions.

2 The hot ionized medium

The term hot ionized medium (HIM) denotes gas with $10^5 < T < 10^7$ K. While the HIM has a very low density (of order $10^{-3}/\mathrm{cm}^3$), it fills about half the ISM volume in the Milky Way (Figure 1). The main physical distinction of "hot" gas is that it is **collisionally ionized** (§2.1), a state that leads to radiative cooling (§2.2 and §2.3). The necessary heat is mostly generated and maintained by a combination of shock heating during infall into the galaxy, supernovae, and stellar winds (§2.4 and §2.5). The HIM is critical in setting the equilibrium pressure of the ISM (§2.6).

2.1 Collisional ionization equilibrium (CIE)

Collisional ionization means that the gas is so hot that particles (most notably, free electrons) have sufficient kinetic energy to ionize hydrogen and possibly higher elements. The average energy per particle reaches this threshold when

$$\langle E_{\mathrm{kin}} \rangle = \frac{3}{2} k_{\mathrm{B}} T = 13.6 \text{ eV} \quad \Longrightarrow \quad T_{\mathrm{ion}} \approx 10^5 \text{ K}\,, \tag{2.1}$$

but in reality even a relatively small number of high-energy particles is sufficient to ionize the gas. Thus, the characteristic temperature where hydrogen begins to be ionized is about 10^4 K. Elements with more protons will naturally be ionized at higher temperatures, but partial ionization states are reached at lower temperatures. For example, it takes about 24.6 eV to remove one electron from Helium ("singly ionized") and 54.4 eV to remove the second ("fully ionized").

On the other hand, excited atoms will quickly return to their lower energy state (de-excite) and free electrons will recombine with ions (§3.1). Assuming that photo-ionization is not important and that enough time has passed, the gas will settle into **collisional ionization equilibrium** (CIE), a steady state with fixed abundances of each ionization state of each element. The number of both ionizations and recombinations scales with $n_{\mathrm{e}} n_{\mathrm{i}}$, where n_{i} is the number density of ions. De-excitations occur regardless of the number density, and thus follow the frequency of excitations, which also scale with $n_{\mathrm{e}} n_{\mathrm{i}}$. We conclude that, in CIE, the ionization fractions depend only on temperature!

Assuming that we know the elemental composition of a given gas, the independence from density gives us a powerful observational tool to measure the temperature. Each (partially ionized) ion has a temperature where it is most abundant, e.g., 10^5 K for C IV. If we observe a higher ionization state, we know that the temperature must be higher than the maximum-abundance temperature of the next-lower ion. For example, if we see C IV, $T > T_{\mathrm{max}}(\mathrm{C\ III})$. Moreover, emission and absorption in a collisionally ionized plasma depend differently on density (as n^2 and n along the line of sight, respectively). If we have both emission and absorption lines, we can thus measure n and get T based on the highest ionization states present.

2.2 Cooling rates

While the continuous excitations and ionizations of ions are fueled by the kinetic energy of particles (mostly electrons), the reverse processes of de-excitation and recombination both emit photons. Thus, the net effect of CIE is **cooling**, that is, the removal of heat by radiating it out of the galaxy (or at least out of the gaseous system we are considering). Since all processes involved scale as $n_{\mathrm{e}} n_{\mathrm{i}}$, we typically express the cooling rate as

$$\Lambda = \left(\frac{\Lambda}{n_{\mathrm{e}} n_{\mathrm{i}}} \right) \times n_{\mathrm{e}} n_{\mathrm{i}}\,. \tag{2.2}$$

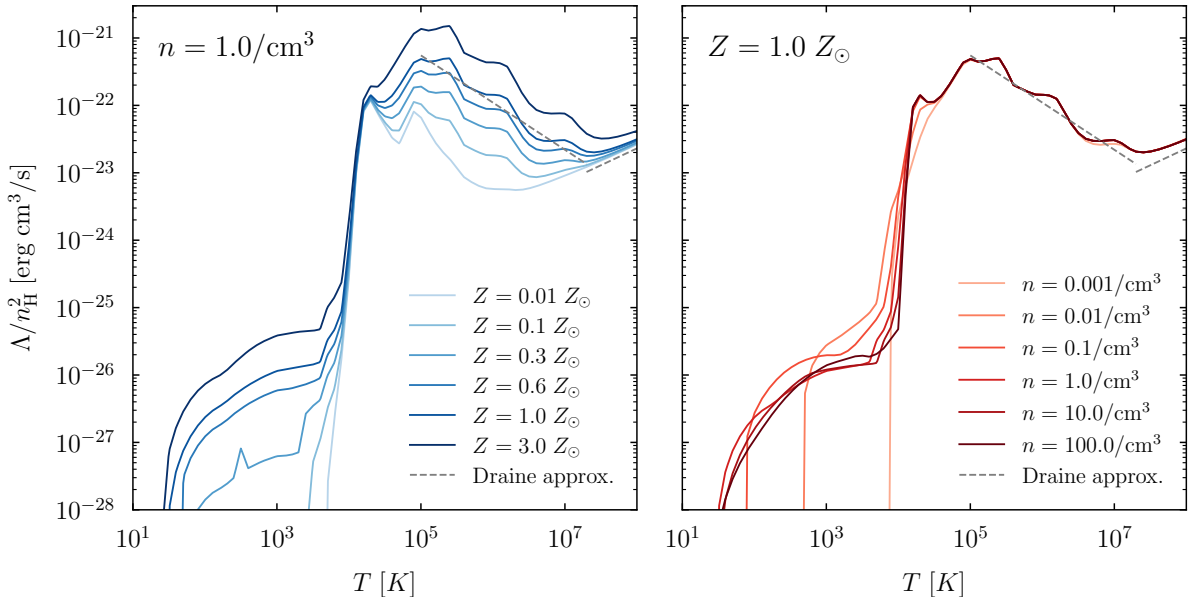


Figure 5: Numerically calculated cooling curves as a function of metallicity (left) and density (right) based on the cooling rate tables of Ploeckinger & Schaye (2020). These tables approximately take into account radiation backgrounds, cosmic rays, self-shielding, and many other effects. The cooling rates have already been divided by n_H^2, meaning that the right panel shows only the residual dependence of this approximately density-independent quantity. The dashed lines show the Draine approximation for $Z = Z_\odot$.

The quantity $\Lambda/n_\mathrm{e}n_\mathrm{i}$ is now a function of metallicity and temperature only (at least approximately). Draine §34.1 gives an overview of cooling processes in hot gas. There are fundamentally two regimes: line cooling at temperatures below about 3×10^7 K, and free-free emission (Bremsstrahlung) at higher temperature. The latter scales as $\Lambda_\mathrm{ff}/n_\mathrm{e}n_\mathrm{i} \propto \sqrt{T}$, which can be derived from first principles (though including some quantum mechanics, see Draine §10.1 for details).

For the line emission that dominates below 3×10^7 K, we assume that collisions dominate over photon interactions such as absorption and stimulated emission (§3, see §3.7 for details on how line cooling rates are computed). Figure 34.3 in Draine shows the contributions of different ions to the total cooling assuming CIE. Figure 5 shows more realistic cooling curves down to lower temperature, which were computed based on the tables of Ploeckinger & Schaye (2020). These tables do contain dependencies on density, but the right panel confirms that this dependence is very weak at high temperatures. The most striking feature in the cooling curves is the drop in Λ/n_H^2 by many orders of magnitude at $T \approx 10^4$ K, where hydrogen begins to a) get collisionally excited and thus contribute to cooling and b) be ionized and thus provide free electrons that can, in turn, excite and ionize higher elements. As a result, gas can cool to 10^4 K relatively quickly but may then get "stuck" at this temperature, depending on the metallicity.

2.3 Cooling times

The cooling rate is not a particularly intuitive quantity, which is why it is commonly converted to a **cooling time** (Draine §34.2), defined as the time it would take gas to cool to zero given its current cooling rate,

$$t_\mathrm{cool} \equiv \frac{T}{|\mathrm{D}T/\mathrm{D}t|}\,. \tag{2.3}$$

This time is an idealization because the cooling rate would change as soon as the temperature changes, and also because we are neglecting conduction and heating. Nevertheless, the cooling time gives us an order-of-magnitude of the kind of timescales on which a gas can cool significantly. To derive a formula for the cooling time from hydrodynamics, we consider the Euler equation for the energy of a Lagrangian fluid element (see also Draine Equation 35.24),

$$\frac{\mathrm{D}\varepsilon}{\mathrm{D}t} = -\frac{P}{\rho}\boldsymbol{\nabla}\cdot\boldsymbol{u} + \frac{\Gamma}{\rho} - \frac{\Lambda}{\rho}\,. \tag{2.4}$$

We neglect heating for now ($\Gamma = 0$). At first sight, things look bad because the equation depends on the velocity field $\boldsymbol{u}$, which is unknown for a general parcel of ISM gas. However, we recall that the $P\boldsymbol{\nabla}\cdot\boldsymbol{u}$ term physically represents $P\mathrm{d}V$ work. When we derived the Euler equations, we used the divergence theorem to convert fluxes through the surface of the fluid element into divergences. We now revert this logic to convert the divergence of velocity into a change in volume (using that $\boldsymbol{\nabla}\cdot\boldsymbol{u}$ is constant over a microscopic fluid element),

$$\boldsymbol{\nabla}\cdot\boldsymbol{u} = \frac{1}{V}\int_V \boldsymbol{\nabla}\cdot\boldsymbol{u}\,dV = \frac{1}{V}\int_S \boldsymbol{u}\cdot d\boldsymbol{S} = \frac{1}{V}\frac{\mathrm{D}V}{\mathrm{D}t}\,. \tag{2.5}$$

In the last step, we have made the physical connection that a net positive velocity perpendicular to the surface of the fluid element means that the fluid element is expanding (similar to the mass flux argument used when we derived the continuity equation). We now convert all quantities to P, T, V, and the constant mass of the fluid element, m, total number of particles, N, and the degrees of freedom, N_{dof}. Given equipartition, we have a total internal energy of $m\varepsilon = N \times N_{\mathrm{dof}} \times k_{\mathrm{B}}T/2$. We multiply Equation 2.4 by m,

$$\frac{\mathrm{D}(m\varepsilon)}{\mathrm{D}t} = -PV\boldsymbol{\nabla}\cdot\boldsymbol{u} - V\Lambda \quad\Longrightarrow\quad \frac{N_{\mathrm{dof}}}{2}\frac{\mathrm{D}(Nk_{\mathrm{B}}T)}{\mathrm{D}t} = -P\frac{\mathrm{D}V}{\mathrm{D}t} - V\Lambda\,. \tag{2.6}$$

The challenge is now to convert the derivative of V (which is meaningless, since the fluid element is an imaginary quantity) into other derivatives. Thus, instead of pulling Nk_{B} out of the derivative on the LHS, we recognize that $N = nV$ and $Nk_{\mathrm{B}}T = PV$. We add $\mathrm{D}(PV)/\mathrm{D}t$ to both sides and expand it on the RHS,

$$\left(\frac{N_{\mathrm{dof}}}{2}+1\right)\frac{\mathrm{D}(Nk_{\mathrm{B}}T)}{\mathrm{D}t} = -P\frac{\mathrm{D}V}{\mathrm{D}t} - V\Lambda + P\frac{\mathrm{D}V}{\mathrm{D}t} + V\frac{\mathrm{D}P}{\mathrm{D}t} = V\frac{\mathrm{D}P}{\mathrm{D}t} - V\Lambda \tag{2.7}$$

and, using $n = N/V$ to eliminate the remaining V terms,

$$\frac{\mathrm{D}T}{\mathrm{D}t} = \frac{2}{(N_{\mathrm{dof}}+2)nk_{\mathrm{B}}}\left(\frac{\mathrm{D}P}{\mathrm{D}t} - \Lambda\right) = \frac{(\gamma-1)}{\gamma nk_{\mathrm{B}}}\left(\frac{\mathrm{D}P}{\mathrm{D}t} - \Lambda\right)\,. \tag{2.8}$$

As the temperature decreases, we can imagine that the pressure is constant and the fluid element contracts, or that the density is constant and the pressure decreases accordingly. In the "isobaric" (constant pressure) case, $\mathrm{D}P/\mathrm{D}t = 0$ and

$$\left(\frac{\mathrm{D}T}{\mathrm{D}t}\right)_{\mathrm{isobaric}} = -\frac{(\gamma-1)}{\gamma nk_{\mathrm{B}}}\Lambda\,. \tag{2.9}$$

In the "isochoric" (constant density) case, we have $\mathrm{D}n/\mathrm{D}t = 0 \implies \mathrm{D}P/\mathrm{D}t = nk_{\mathrm{B}}\mathrm{D}T/\mathrm{D}t$ and thus

$$\frac{\mathrm{D}T}{\mathrm{D}t}\left(1 - \frac{\gamma-1}{\gamma}\right) = \frac{\mathrm{D}T}{\mathrm{D}t}\left(\frac{1}{\gamma}\right) = -\frac{(\gamma-1)}{\gamma nk_{\mathrm{B}}}\Lambda\,, \tag{2.10}$$

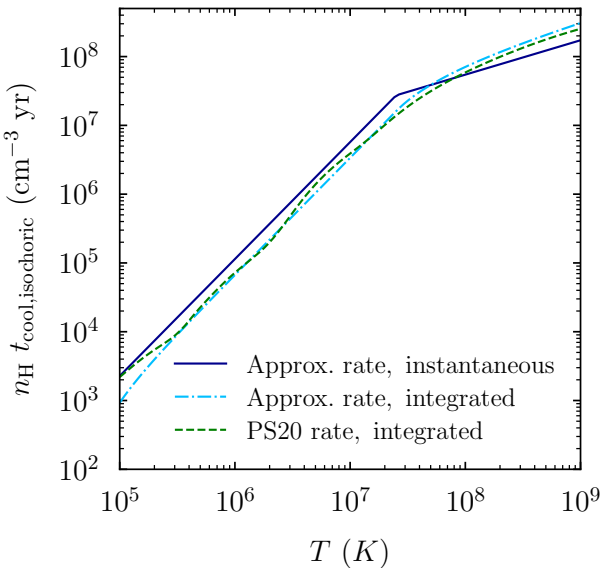


Figure 6: Different estimates of the isochoric cooling time from a temperature T to 10^5 K at $Z = Z_\odot$. The dark blue line is estimated from the instantaneous cooling rate at T and relies on Draine's approximation for the cooling rate (Equation 34.2). The light-blue line was calculated by integrating over the evolution of the cooling rate with temperature. For the green line, the approximate cooling rates were replaced by the Ploeckinger & Schaye (2020) functions (Figure 5). The approximate cooling time agrees quite well with the more sophisticated calculations.

and thus

$$\left(\frac{\mathrm{D}T}{\mathrm{D}t}\right)_{\text{isochoric}} = \gamma \left(\frac{\mathrm{D}T}{\mathrm{D}t}\right)_{\text{isobaric}} = -\frac{(\gamma-1)}{nk_\mathrm{B}}\Lambda\,. \tag{2.11}$$

Given that we are dealing with gas that is too hot to contain molecules, we can safely assume $\gamma = 5/3$ to get the cooling time,

$$\boxed{t_{\text{cool,isobaric}} = \frac{5}{2}\frac{nk_\mathrm{B}T}{\Lambda}} \quad \text{and} \quad \boxed{t_{\text{cool,isochoric}} = \frac{3}{2}\frac{nk_\mathrm{B}T}{\Lambda}} \tag{2.12}$$

It makes sense that the isochoric cooling time is shorter, since the density remains higher than in the isobaric case, meaning more cooling. We recall that Λ/n^2 is a function of temperature only, meaning that $t_\text{cool} \times n$ is also independent of density. Thus, the cooling time actually scales as $f(T)/n$. At low densities, the cooling time can get long!

We can also derive exact expressions for the cooling time by integrating Equation 2.3 over T instead of assuming a constant Λ. In the isochoric case, we can keep n outside of the integral to get

$$t_{\text{cool,isochoric}} = \frac{3}{2}nk_\mathrm{B}\int_{T_\text{low}}^{T_\text{high}} \frac{1}{\Lambda(T')}\mathrm{d}T'\,, \tag{2.13}$$

where T_high is the starting temperature and T_low the lower final temperature. In the isobaric case, we need to convert density to pressure, $n = P/k_\mathrm{B}T$ since the latter is constant, which gives

$$t_{\text{cool,isobaric}} = \frac{5}{2}P\int_{T_\text{low}}^{T_\text{high}} \frac{1}{T' \times \Lambda(T')}\mathrm{d}T'\,, \tag{2.14}$$

It does not make sense to integrate to $T_\text{low} = 0$, since the cooling rates become small at low T (Figure 5), meaning that gas can never actually cool to $T = 0$. Figure 6 shows the isochoric

cooling time calculated both based on the instantaneous cooling rate (Equation 2.12) and based on integration (Equation 2.13). The instantaneous estimate compares surprisingly well to the more accurate integral.

2.4 Radiative corrections to the Sedov-Taylor phase of supernovae

We now turn to the processes that heat gas and keep it hot. In the hydrodynamics part of the course, we treated the self-similar expansion phase of a supernova remnant, known as the Sedov-Taylor (ST) phase. From self-similarity and the given quantities of ejecta energy and surrounding density, we found that the shock radius must scale as

$$R_\mathrm{s} \propto \left(\frac{E_\mathrm{SN}}{\rho_1}\right)^{1/5} t^{2/5} \,. \tag{2.15}$$

The requirement that the ST phase starts when the ejecta has swept up its own mass in material sets the self-similar unit system. However, we neglected the effects of cooling. Based on our recently acquired knowledge of the cooling function, we can compute at what point the ST solution becomes invalid due to significant radiative energy losses. The following derivations follow Draine §39. We use Draine's way of quantifying the numbers in terms of characteristic input quantities,

$$E_{51} \equiv \frac{E_\mathrm{SN}}{10^{51}\ \mathrm{erg}} \qquad n_0 \equiv \frac{n_1}{1/\mathrm{cm}^3} \qquad t_3 \equiv \frac{t}{1000\ \mathrm{yr}} \tag{2.16}$$

where $n_1 = \rho_1/1.4m_\mathrm{p}$ (see §1.2). In this convenient notation,

$$\begin{aligned} R_\mathrm{s} &= 5.0\ \mathrm{pc} \times E_{51}^{1/5} n_0^{-1/5} t_3^{2/5} \\ u_\mathrm{s} &= 1950\ \mathrm{km/s} \times E_{51}^{1/5} n_0^{-1/5} t_3^{-3/5} \\ T_\mathrm{s} &= 5.25 \times 10^7\ \mathrm{K} \times E_{51}^{2/5} n_0^{-2/5} t_3^{-6/5} \,, \end{aligned} \tag{2.17}$$

where T_s is the temperature right behind the shock front. To get a sense of when radiative losses become important, we compute the total energy radiated away,

$$\Delta E = -\int_0^t \mathrm{d}t \int_0^{R_\mathrm{s}} \mathrm{d}r\ 4\pi r^2 \Lambda(\rho, T) \,. \tag{2.18}$$

The cooling function can be approximated as

$$\Lambda \approx 1.1 \times 10^{-22}\ \frac{\mathrm{erg\ cm}^3}{\mathrm{s}} \left(\frac{T}{10^6\ \mathrm{K}}\right)^{-0.7} n_\mathrm{e} n_\mathrm{H} \tag{2.19}$$

in the relevant temperature regime (Draine Equation 34.2). We convert using the expressions from §1.2,

$$n_\mathrm{H} = n_1 \left(\frac{\rho}{\rho_1}\right) \qquad n_\mathrm{e} = 1.2 \times n_1 \left(\frac{\rho}{\rho_1}\right) \tag{2.20}$$

and solve the radial integral,

$$\Delta E = -1.2 \times 1.1 \times 10^{-22}\ n_0^2 \left(\frac{T_\mathrm{s}}{10^6\ \mathrm{K}}\right)^{-0.7} \frac{\mathrm{erg}}{\mathrm{cm}^3\ \mathrm{s}} \int_0^t \mathrm{d}t\ \frac{4\pi R_\mathrm{s}^3}{3} \left\langle \left(\frac{\rho}{\rho_1}\right)^2 \left(\frac{T}{T_\mathrm{s}}\right)^{-0.7} \right\rangle \,. \tag{2.21}$$

The average term is now purely a function of the self-similar ST solution and comes out to 1.817. We plug in the forms of Equation 2.17 to get the fraction of the total supernova energy lost to radiation,

$$\frac{\Delta E}{E_\mathrm{SN}} \approx -2.38 \times 10^{-6} \times n_0^{-0.55} E_{51}^{-0.68} t_3^{3.04} \,. \tag{2.22}$$

We can decide at what fraction we declare the ST phase for no longer valid. We follow Draine in setting this fraction to 1/3, which is reached after a time

$$t_{\rm rad} = 4.9 \times 10^4 \text{ yr} \times n_0^{-0.55} E_{51}^{0.22} \,. \tag{2.23}$$

At that time, the shock has reached a radius

$$R_{\rm s,rad} = 24 \text{ pc} \times n_0^{-0.42} E_{51}^{0.29} \,, \tag{2.24}$$

which is a factor of five or so smaller than the ≈ 100 pc radius at which the shock can no longer be supported without cooling (see hydro notes). Clearly, the radiative corrections make a big difference in the ST solution!

2.5 Stellar winds

Intuitively, it might seem like winds could never be as energetic as supernovae, but a simple calculation shows that that's not necessarily true:

$$E_{\rm tot} = \frac{1}{2} M_{\rm w} u_{\rm w}^2 = \frac{1}{2} \dot{M} \, \Delta t \, u_{\rm w}^2 \approx 10^{50} \text{ erg} \times \dot{M}_6 \, u_3^2 \left(\frac{\Delta t}{10^7 \text{ yr}} \right) \,, \tag{2.25}$$

where we have defined additional quantities of convenience for the mass loss rate, wind velocity, and surrounding density,

$$\dot{M}_6 \equiv \frac{\dot{M}}{10^{-6} \, M_\odot/\text{yr}} \qquad u_3 \equiv \frac{u_{\rm w}}{1000 \text{ km/s}} \qquad n_3 \equiv \frac{n_{\rm H}}{10^3/\text{cm}^3} \,. \tag{2.26}$$

The total energy is not categorically different from the energy of a supernova. The terminal wind velocity turns out to be about the escape velocity of the star. O stars have a mass of 60 $M_\odot$ and $1.5 < u_3 < 2.5$; for B stars, we have 18 $M_\odot$ and $0.3 < u_3 < 1.5$.

The next questions we might ask are how large the volume influenced by the stellar wind is, and whether it grows forever or reaches a particular size. For simplicity, we assume that the stellar wind turns on suddenly at $t = 0$ and expands into an already existing H II region (hence the higher environmental density n_3 compared to the supernova calculation). As for a supernova blast wave, the wind first expands freely before it sweeps up a mass comparable to its own mass. For the supernova, this mass was constant, but for a wind it grows with the outflow rate. Thus, the time when the free expansion phase ends is approximately described by

$$\dot{M} t \approx \frac{4\pi}{3} \rho_0 u_{\rm w}^3 t^3 \tag{2.27}$$

which we solve to find

$$t_0 = \left(\frac{3\dot{M}}{4\pi \rho_0 u_{\rm w}^3} \right)^{1/2} \,. \tag{2.28}$$

We convert ρ to n as described in §1.2 and plug in our default numbers,

$$t_0 \approx 2.5 \text{ yr} \times n_3^{-1/2} \dot{M}_6^{1/2} u_3^{-3/2} \,. \tag{2.29}$$

Clearly, the free expansion phase is so short that it will not be observable (and probably not physical given that the star does not actually turn on instantaneously). As for the Sedov-Taylor phase of

supernovae, we could derive the scaling of the shock radius in the self-similar phase purely from dimensions. If we start at t_0 and $l_0 = t_0 u_\mathrm{w}$, we get

$$R_\mathrm{s} = l_0 \left(\frac{t}{t_0}\right)^{3/5} = \left(\frac{3}{4}\frac{\dot{M} u_\mathrm{w}^2}{\pi \rho_0}\right)^{1/5} t^{3/5} \approx 0.09 \text{ pc} \times t_3^{3/5} n_3^{-1/5} \dot{M}_6^{1/5} u_3^{2/5} \,. \tag{2.30}$$

This scaling is slightly different from the 2/5 for the ST phase because the stellar wind energy is injected continuously, leading to the extra factor of t in Equation 2.27. The shock moves outward with a velocity

$$u_\mathrm{s} = \frac{\partial R_\mathrm{s}}{\partial t} = \frac{3 R_\mathrm{s}}{5t} \approx 55 \text{ km/s} \times t_3^{-2/5} n_3^{-1/5} \dot{M}_6^{1/5} u_3^{2/5} \,. \tag{2.31}$$

In Appendix B.1 we present a more sophisticated derivation of the normalization of the shock radius and speed (Draine §38.1), but the final result differs by only 20%.

As the shock decelerates and the Mach number drops, the gas is shock-heated less and its cooling time decreases. Thus, the shock becomes "radiative," meaning that the temperature quickly decreases to the pre-shock temperature. This happens when the cooling time becomes comparable to t. From Figure 36.2 in Draine, we learn that the cooling times for slow shock speeds are fairly short. Indeed, an approximate calculation (Draine Equation 38.6) shows that the shock becomes radiative after only about 7 years for the parameters assumed above! At this transition, the shock velocity briefly drops, but the velocity and radius quickly recover to the same scalings as above.

As with the ST phase of supernovae, the energy-conserving phase ends when the shock speed decreases to the sound speed of the surrounding ISM (about 15 km/s in H II regions). From Equation 2.31, that happens at $t \approx 4 \times 10^4$ yr, at which point the shock radius is $R_\mathrm{s} \approx 1$ pc. In summary, stellar wind bubbles are much smaller than late-stage supernova remnants, but they still inject a significant amount of energy into the ISM.

2.6 The supernova rate and the pressure of the hot ISM

To understand the importance of supernovae for the ISM as a whole, we need to know the supernova rate. Observationally, we believe the rate in the MW to be about one per 40 years, or one per 60 years in the disk (within a radius of 15 kpc and thickness of 200 pc). The rate per volume is then

$$S = \Big(60 \text{ yr} \times \pi (15 \text{ kpc})^2 \times 200 \text{ pc}\Big)^{-1} = 1.2 \times 10^{-13} \frac{1}{\text{pc}^3 \text{ yr}} \equiv 1.2 \; S_{13} \,. \tag{2.32}$$

Using the final lifetime and size of SN remnants (t_fade and R_fade, Draine §39.1), we can now determine the average number of SNe affecting a random position in the ISM. We write

$$\begin{aligned} N_\mathrm{SN} &= S \times \frac{4\pi}{3} R_\mathrm{fade}^3 t_\mathrm{fade} \\ &\approx 0.24 \; S_{13} \; n_0^{-1.47} E_{51}^{1.26} \left(\frac{c_\mathrm{s}}{10 \text{ km/s}}\right)^{-2.6} \,. \end{aligned} \tag{2.33}$$

The numerical pre-factor is remarkably close to unity! This fact led McKee & Ostriker (1977) to postulate that N_SN is not coincidentally close to one, but that supernovae maintain the ISM pressure at a value that balances that of the SN remnants. If the pressure is lower, supernova bubbles expand farther, raising the pressure. Once bubbles begin to overlap, on average, they encounter a pressure similar to their interior and are stalled. This is a very simplistic assumption, but it makes a reasonable prediction for the pressure in the ISM. We convert the sound speed in Equation 2.33 to pressure,

$$P = \frac{c_\mathrm{s}^2 \rho}{\gamma} = \frac{c_\mathrm{s}^2 \times 1.4 \; n_\mathrm{H} \; m_\mathrm{p}}{\gamma} \,, \tag{2.34}$$

and use our main assumption ($N_{\rm SN} = 1$) to solve for the pressure,

$$P/k_{\rm B} \approx 5700 \text{ K/cm}^3 \times S_{13}^{0.77} n_0^{-0.13} E_{51}^{0.97} \,. \tag{2.35}$$

This estimate is remarkably close to the observed ISM pressure of about $P/k_{\rm B} \approx 4000$ K/cm^3! Moreover, the estimate is almost independent of density, meaning that the only inputs are the observed supernova rate and energy, as well as known physics such as hydrodynamics and the cooling function. We conclude that supernovae are likely to contribute a large part to the pressurization of the ISM. We explore this line of argument further in §10.2.

3 Atomic physics I: Energy levels and transitions

In the ISM, the state of atoms, ions, electrons, molecules, and radiation plays a critical role. One of the key processes is the excitation of atoms and ions into higher energy levels by either photons or collisions with other particles. In preparation for the following topics, we dedicate this section to studying the physics of atoms, and hydrogen in particular.

We begin by recapitulating the energy levels of atoms and transitions between them (§3.1 and §3.2, Draine §4). We lay the necessary foundations in statistical mechanics in §3.3 (Draine §3). We introduce the basic equations for radiative absorption and emission in §3.4 (Draine §6) and for collisional excitation in §3.5 (O&F §3.5). We combine those mechanisms to describe the level populations of a two-level system in §3.6 (Draine §17) and apply our insights to line cooling in §3.7.

3.1 Energy levels of atoms

To understand the ionization structure of gas in the ISM, we cannot only consider the ground and fully ionized states because atoms (including hydrogen) have many intermediate energy levels. In quantum mechanics, the orbitals of electrons are characterized by a principle quantum number, $n \geq 1$, and an orbital angular momentum quantum number l, where $0 \leq l < n$. This number determines the quantized angular momentum of an electron, $L_\mathrm{e} = l\hbar$. This angular momentum can point in different directions, which we quantify via the component of the angular momentum that points in the z-direction, $m\hbar$, with $-l < m < l$. In addition to their orbital angular momentum, electrons have a spin of $\hbar s$, where $s = \pm 1/2$.

The Pauli exclusion principle states that only one electron can occupy a given state. Each combination of n, l, and m can host two electrons whose spins point in opposite directions (spin-up and spin-down). As a result, a given combination of n and l, which we call a **subshell**, can host $2(2l+1)$ electrons with distinct quantum numbers. For historic reasons, the values of $l = 0, 1, 2, 3$ are labeled s, p, d, f (which stand for sharp, principal, diffuse, and fundamental). These subshells can host at most 2, 6, 10, and 14 electrons, respectively. We can write down the occupation of subshells with superscripts, e.g., $1s^2 2s^2 2p^6 3s$ for the $Z = 11$ electrons of Sodium in its ground state (Figure 7).

However, this notation does not tell us which states (values of m and s) are filled in an incomplete subshell. It is thus not particularly useful if we want to understand the transitions between different energy levels, which depend on the total change in energy and angular momentum between the previous and new state of the electrons. Instead, we use **spectroscopic notation** to describe the configuration of electrons in a subshell of a particular atom or ion. Depending on the ionization state, a number of subshells will typically be filled entirely or partially. We are generally interested in the "outermost" electrons because those are easiest to excite to a higher level (see, e.g., Table 4.1 in Draine). We sum the angular momenta of the electrons in a subshell in vector space so that $L\hbar$ is the total orbital angular momentum, $S\hbar$ is the total spin, and $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}$ is the total angular momentum. A possible state of the subshell is called a **term**, which we write in spectroscopic notation as

$$\boxed{n^{2S+1}L^p_J} \tag{3.1}$$

where L is written as a letter (S, P, D, F), e.g., 3^2P_1. The p stands for "parity" and can be blank (even parity) or "o" (odd parity). This flag indicates whether the wave function changes sign if all electrons are reflected through the origin. In principle, the number of possible combinations in a term (called the "multiplicity") is $g = (2S+1)(2L+1)$, but not all of those possibilities are necessarily distinct or allowed. The selection rules get complicated, and it is generally best to just

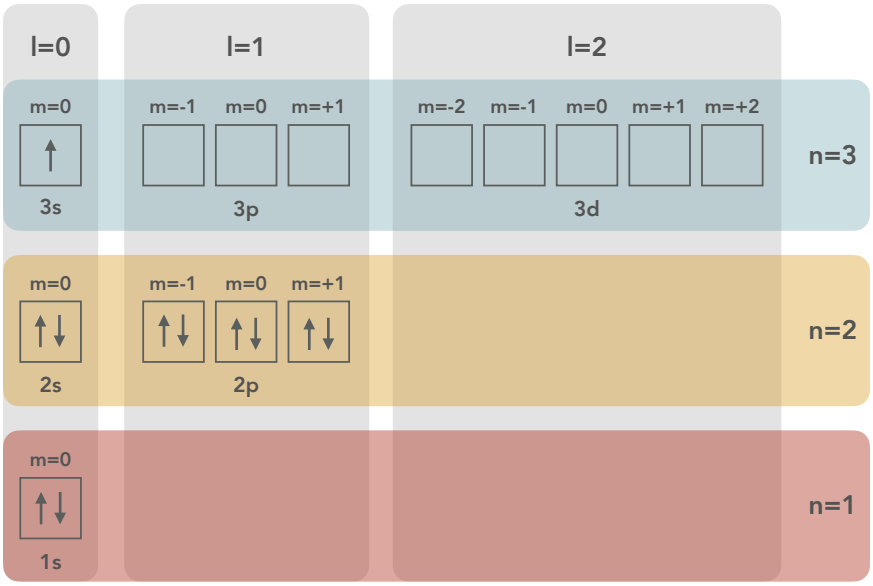


Figure 7: Shells ($n = 1, 2, 3$) and subshells ($l = 0, 1, 2$) as filled in the ground state of Sodium ($Z = 11$). Each arrow represents an electron with spin up or down.

look up which terms are present in different atoms and ions (see Draine §4.5 for details).

The J-number describes the different ways in which orbital and spin angular momenta can combine in vector space, with the total angular momentum ranging from $J = |L-S|$ to $J = |L+S|$. From the perspective of the electrons, the protons in the nucleus are orbiting and thus creating a magnetic field, which interacts with the spin of the electrons. This so-called **spin-orbit coupling** leads to **fine structure splitting** between orbitals that would otherwise have the same energy (same n and l). For example, four electrons in a ${}^3P$ state ($L = 1$) can have $J = 0, 1, 2$, leading to a "triplet" (as opposed to a "singlet" or "doublet" for one or two possible values of J). The number of distinct combinations that leads to the same state is called the "degeneracy," $g = 2J + 1$. Even more subtly, the magnetic moments from the spin of the electrons and the protons in the nucleus interact, leading to **hyperfine splitting** of the energy levels based on the relative spin orientations (parallel or anti-parallel).

We can roughly estimate the differences between the various types of split in the energy levels. The primary levels are determined by the Coulomb interactions between electrons and nuclei. The relevant energy unit is the Rydberg, where[1]

$$\mathrm{Ryd} \equiv \frac{2\pi^2 m_\mathrm{e} q_\mathrm{e}^4}{h^2} = 13.6 \text{ eV} \,. \tag{3.2}$$

The primary energy levels of electrons are given by the quantum number n and have $E_\mathrm{primary} \approx 13.6 \text{ eV}(Z^2/n^2)$. For hydrogen in the ground state ($Z = 1$, $n = 1$), the ionization energy of 13.6 eV corresponds to a UV wavelength of 912Å. The hyperfine differences are depressed by the "fine structure constant,"

$$\alpha \equiv \frac{q_\mathrm{e}^2}{\hbar c} \approx \frac{1}{137} \,, \tag{3.3}$$

with an energy of $E_\mathrm{fine} \approx 13.6 \text{ eV}(\alpha^2 Z^4/n^5)$. Finally, the hyperfine splitting is suppressed by an

[1] We use Gaussian cgs units, where the often quoted factor of $\epsilon_0 \to 1/4\pi$. The difference is balanced by a different value of q_e. See the hydro notes for details.

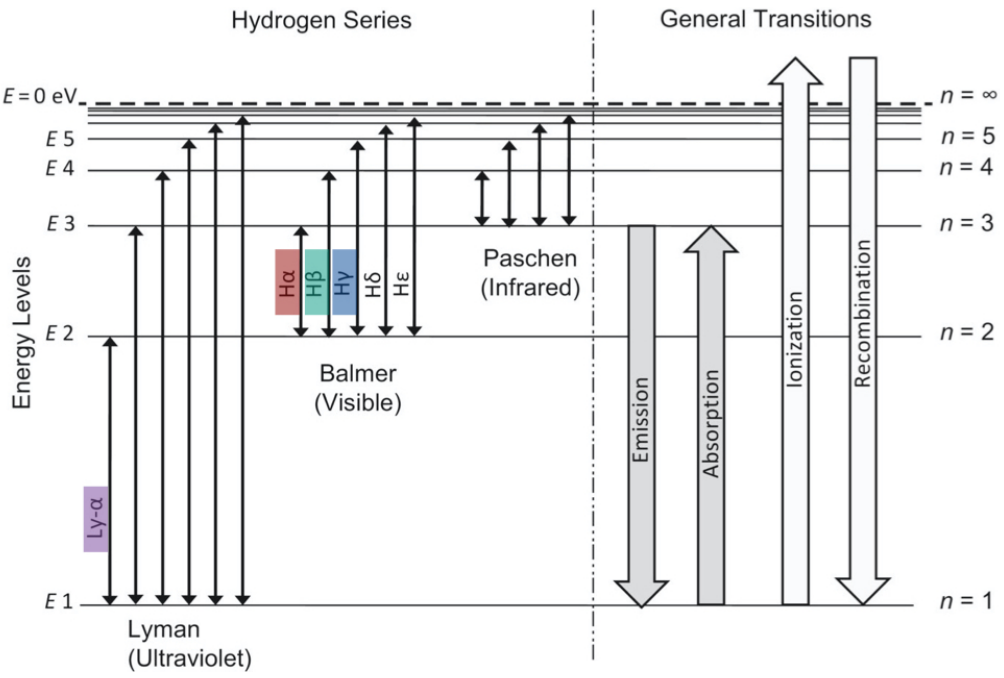


Figure 8: Level transitions in the hydrogen atom. The ultraviolet $n = 2 \to 1$ transition (called Lyman-α) connects the $1s$ and $2p$ states because there is no allowed one-photon transition from $2s$ to $1s$. The Hα and Hβ lines are important optical tracers. Adapted from Trypsteen & Walker (2017).

additional factor of $m_e/m_p \approx 1/2000$. For hydrogen, the hyperfine split is of order 10^{-6} eV, which corresponds to the famous 21 cm radiation (§4.1).

3.2 Transitions between levels

An atom can transition between energy levels by absorbing or emitting a photon of the right frequency. It does not matter which electron in a subshell transitions — what matters are the states of the subshell before and after the transition. For example, the strongest (most likely) transitions are **electric dipole transitions**, but those can only occur between two terms that satisfy five selection rules:

- Parity must change
- $\Delta L = 0, \pm 1$, but $L = 0 \to 0$ is forbidden
- $\Delta J = 0, \pm 1$, but $J = 0 \to 0$ is forbidden
- $\Delta S = 0$
- If a single electron is involved, $\Delta l = \pm 1$

Transitions that break only the $\Delta S = 0$ rule are called **semi-forbidden transitions** and are written with one square bracket, e.g., "NII] 2143.4Å $^5S^o_2 \to {}^3P_2$." Transitions that violate at least one of the other rules are called **forbidden transitions** and are written with two square brackets, e.g., "[NII] 6549.9Å $^1D_2 \to {}^3P_1$." There are a few more levels of "forbidden-ness," but the name is misleading in that forbidden transitions can occur — they are just much less likely than "allowed" transitions. Each transition occurs with a probability per time of $A_{21}(nL, n'L')$, where n and L are the original quantum numbers and n' and L' the new ones (more on that in §3.4). The decay rate of a given energy level is then given by the sum over all possible transitions to a lower level. We

can invert this decay rate to get the lifetime of level nL,

$$t_{\rm nL} = \frac{1}{\sum_{n'<n} \sum_{L'=L,L\pm1} A_{21}(nL, n'L')} \,. \tag{3.4}$$

For hydrogen, the transition rates are roughly between 10^4/s and 10^8/s, which translates into lifetimes of 10^{-8} to 10^{-4} s for the excited levels.

Figure 8 shows an overview of the energy levels and transitions of hydrogen. One significant restriction imposed by the transition rules is that there is no allowed transition from $2s$ to $1s$, meaning that a hydrogen atom in the 2^2S state cannot decay to the 1^2S ground state using an allowed transition! However, even the lifetime of the 2^2S state is only about 0.12 s. In §6.1, we will see that the time it takes for the average atom to be ionized in typical astrophysical scenarios is much longer. Thus, the vast majority of neutral atoms are in the ground state at any given time.

3.3 Thermodynamic equilibrium and blackbody radiation

In many situations, we will be interested in the state of atoms in equilibrium, that is, after enough time that short-lived fluctuations have settled down. In this section, we give a very brief treatment of the necessary background; more details can be found in Draine §3. We start from one of the fundamental principals of statistical mechanics, the **partition function** of a system,

$$Z_{\rm int}(T) = \sum_i g_{\rm i}\ e^{-E_{\rm i}/k_{\rm B}T} \,, \tag{3.5}$$

where i runs over all possible internal states, $E_{\rm i}$ is the energy associated with state i, and $g_{\rm i}$ is the **statistical weight** or "degeneracy" of the state (i.e., how many distinct quantum states share the energy level i). The system could, for example, be an atom embedded in a gas with temperature T. In addition to the internal quantum states, an atom (or other system) can also move in space. Integrating over the possible velocities leads to an extra factor in the **total partition function per unit volume**,[2]

$$p(X,T) = \left(\frac{2\pi m_{\rm X} k_{\rm B} T}{h^2}\right)^{3/2} Z_{\rm int}(T) \,, \tag{3.6}$$

where $m_{\rm X}$ is the mass of the "system" X (Draine §3.1). We can think of the partition function as the volume in phase space occupied by a given state i, which essentially translates into a probability of the system being in this state given a temperature T. This principle applies to a reaction between different states, say $A + B \leftrightarrow C$. If we assume **local thermodynamic equilibrium (LTE)**, i.e., that the system has had enough time to settle into a balance of constant number densities $n_{\rm A}$, $n_{\rm B}$, and $n_{\rm C}$, then those densities are related by the **law of mass action**,

$$\frac{n_{\rm C}}{n_{\rm A} n_{\rm B}} = \frac{p(C)}{p(A)p(B)} \,. \tag{3.7}$$

This logic can be extended to more reactant species, with densities and partition functions that keep multiplying (Draine §3.2). A particularly elegant expression arises for a two-state system where state "2" has a higher energy than state "1" by a difference E_{21}. The kinetic term in the total partition function cancels, leaving only the exponential dependence on temperature and the statistical weights,

$$\boxed{\frac{n_2}{n_1} = \frac{g_2}{g_1} e^{-E_{21}/k_{\rm B}T}} \tag{3.8}$$

[2]The partition function is commonly written as $f(X,T)$, but we use the notation $p(X,T)$ to avoid confusion with the velocity distribution, $f(u)$.

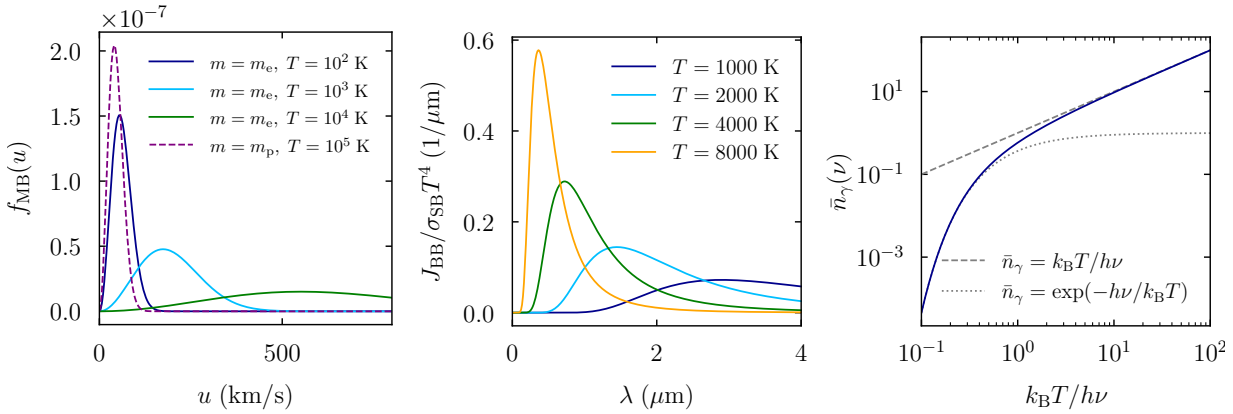


Figure 9: Equilibrium distributions of particle and photon energies. *Left:* The Maxwell-Boltzmann distribution for electrons (solid) and protons (dashed) at different temperatures. *Center:* The blackbody distribution for different temperatures as a function of wavelength, normalized such that all curves enclose the same total energy. Note that J gives an intensity per frequency or per wavelength, which means that the normalized distributions has units of inverse length. *Right:* The photon occupation number for a blackbody distribution.

This equation will turn out to be extremely useful because it applies to any system with quantized energy levels, including the atomic energy levels as discussed in §3.1. In thermodynamical equilibrium, it can be shown that the partition function implies that the distribution of velocities follows the **Maxwell-Boltzmann distribution**,[3]

$$\boxed{f_{\rm MB}(u) = \sqrt{\frac{2}{\pi}} \left(\frac{m_{\rm e}}{k_{\rm B} T} \right)^{3/2} u^2 e^{-m_{\rm e} u^2 / 2 k_{\rm B} T}} \tag{3.9}$$

The left panel of Figure 9 shows Maxwell-Boltzmann distributions for electrons and protons at different temperatures (on a linear scale). If radiation is also in thermal equilibrium then its intensity follows the **blackbody distribution**,

$$\boxed{J_{\nu,\rm BB} = \frac{2 h \nu^3}{c^2} \frac{1}{e^{h\nu / k_{\rm B} T} - 1}} \tag{3.10}$$

The mean intensity J_ν has units of erg/cm^2/s/Hz/sr = erg/cm^2 (see Appendix 1 of O&F for a summary of the different quantities describing intensity, flux, and so on). Since J_ν gives the intensity per steradian, $4\pi J_\nu$ is the intensity from all directions (assuming an isotropic flux). The center panel of Figure 9 shows the blackbody distribution at different temperatures, normalized by the total energy emitted, $\sigma_{\rm SB} T^4$ (which is the definition of the Stefan-Boltzmann constant). In some cases, it will be convenient to replace the intensity with the **photon occupation number**,

$$\bar{n}_\gamma \equiv \frac{c^3}{8\pi h \nu^3} U_\nu = \frac{c^2}{2 h \nu^3} J_\nu \,, \tag{3.11}$$

[3]We use u to denote the speed of particles in these notes, even though we used w in the hydro notes. The expression given here refers the isotropic MB distribution for one-dimensional velocities, which is responsible for the u^2 factor. We omit the factor of n that is included in the hydro notes. This difference corresponds to whether the distribution is normalized such that its 6D phase-space integral gives the total particle number N in a unit volume or unity. For these notes, the latter is more convenient. We note that technically, the MB distribution is valid only if there are enough collisions between particles to establish equilibrium but not so many that they lead to collisional exchange of momentum and energy (Appendix A.1). The missing factor of 2 in O&F Equation 2.6 appears to be a typo.

where $U_\nu = 4\pi J_\nu/c$ is the energy density of radiation at a frequency ν (Draine §17.1). The photon occupation number is the number of photons per mode per polarization (Draine §7.1), which provides a dimensionless way to describe a radiation spectrum. The bar symbol highlights that the radiation is assumed to be isotropic, and it also distinguishes $\bar{n}_\gamma$ from the symbol for number densities (which it is not). For a blackbody, the pre-factors of $J_{\nu,\mathrm{BB}}$ and $\bar{n}_\gamma$ cancel, giving

$$\boxed{\bar{n}_{\gamma,\mathrm{BB}} = \frac{1}{e^{h\nu/k_\mathrm{B}T} - 1}} \tag{3.12}$$

To understand the meaning of $\bar{n}_\gamma$, we can consider the limits of the blackbody distribution. If the temperature is cold compared to the photon energy, $k_\mathrm{B}T \ll h\nu$, we approach the usual expression for the partition function, $\bar{n}_\gamma \to e^{-h\nu/k_\mathrm{B}T}$. If $k_\mathrm{B}T \approx h\nu$ then $\bar{n}_\gamma \approx 1$. For hot temperatures, $k_\mathrm{B}T \gg h\nu$, we get $\bar{n}_\gamma \to k_\mathrm{B}T/h\nu$, meaning that the occupation number increases linearly with temperature. These limits are highlighted in the right panel of Figure 9.

3.4 Absorption and emission

In this and the following sections, we consider systems with different, quantized energy levels and the transitions between them. As in Equation 3.8, we assume a two-state system with an energy difference E_{21}. Some of the conclusions apply for multi-level systems, but we restrict our analysis to two for simplicity. In this section, we consider the emission and absorption of photons and the corresponding excitation and deexcitation of energy levels. The system in question could be an atom, an ion, a molecule, or even a dust grain. The derivation largely follows Draine §6.1.

There are three fundamental ways by which the interaction with photons can change the state of a system: absorption, spontaneous emission, and stimulated emission. **Absorption** means that a photon kicks the system from the ground state to the excited state. The rate of this process must be proportional to the product of the density of ground-state particles and photons of the right frequency,

$$\left.\frac{\mathrm{d}n_2}{\mathrm{d}t}\right|_{1\to2} = -\left.\frac{\mathrm{d}n_1}{\mathrm{d}t}\right|_{1\to2} = n_1 B_{12} U_\nu(\nu_{21})\,, \tag{3.13}$$

where $\nu_{21} = E_{21}/h$ is the photon frequency that is needed for the transition and B_{12} is the **Einstein B coefficient** for the transition. This coefficient has units of $\mathrm{cm}^3/\mathrm{erg\ s}$ because it multiplies the energy density U_ν. One could also define it with respect to the number density of photons, for example.

The second important process is **spontaneous emission**, the opposite of absorption. This process occurs at a rate that is independent of the radiation field at a rate, $n_2 A_{21}$, where the latter is called the **Einstein A coefficient** (which has units of 1/s, see §3.2). The situation is symmetric thus far, but there is a third process that only occurs for deexcitation: **stimulated emission**, where a photon of frequency ν_{21} triggers a deexcitation and emission of another photon. Combining the two rates, we get a total rate for $2 \to 1$ transition of

$$\left.\frac{\mathrm{d}n_1}{\mathrm{d}t}\right|_{2\to1} = -\left.\frac{\mathrm{d}n_2}{\mathrm{d}t}\right|_{2\to1} = n_2\left[A_{21} + B_{21}U_\nu(\nu_{21})\right]\,. \tag{3.14}$$

In principle, A_{21}, B_{12}, and B_{21} need to be calculated from the quantum mechanics of the system in question (see Appendix 4 of O&F for details). However, they are related and reduce to a single coefficient. We can show this by assuming that many particles form a system in thermal equilibrium where

$$U_\nu(\nu_{21}) = \frac{4\pi}{c} J_{\nu,\mathrm{BB}}(\nu_{21}) = \frac{8\pi h\nu_{21}^3}{c^3}\frac{1}{e^{E_{21}/k_\mathrm{B}T} - 1}\,. \tag{3.15}$$

Moreover, we know that the level populations must follow

$$\frac{n_2}{n_1} = \frac{g_2}{g_1} e^{-E_{21}/k_\mathrm{B}T} \tag{3.16}$$

in thermal equilibrium (Equation 3.8). We balance the three transition mechanisms,

$$n_1 B_{12} U_\nu(\nu_{21}) = n_2 \left[A_{21} + B_{21} U_\nu(\nu_{21})\right] , \tag{3.17}$$

and insert the blackbody energy density and level populations,

$$B_{12} \frac{8\pi h\nu_{21}^3}{c^3} \frac{1}{e^{E_{21}/k_\mathrm{B}T} - 1} = \frac{g_2}{g_1} e^{-E_{21}/k_\mathrm{B}T} \left[A_{21} + B_{21} \frac{8\pi h\nu_{21}^3}{c^3} \frac{1}{e^{E_{21}/k_\mathrm{B}T} - 1}\right] . \tag{3.18}$$

This equation looks complicated but simplifies in the limits of zero and infinite temperature, in which it must still be valid. In the limit of $T \to 0$, the -1 terms in the denumerators become irrelevant, the exponentials on the LHS and RHS cancel, and the B_{21} term on the RHS vanishes,

$$B_{12} \frac{8\pi h\nu_{21}^3}{c^3} = \frac{g_2}{g_1} A_{21} . \tag{3.19}$$

Physically, this means that stimulated emission plays no role any longer when the radiation field approaches zero, which allows us to relate B_{12} and A_{21}. We insert Equation 3.19 into 3.18,

$$A_{21} \frac{1}{e^{E_{21}/k_\mathrm{B}T} - 1} = e^{-E_{21}/k_\mathrm{B}T} \left[A_{21} + B_{21} \frac{8\pi h\nu_{21}^3}{c^3} \frac{1}{e^{E_{21}/k_\mathrm{B}T} - 1}\right] , \tag{3.20}$$

and multiply by $e^{E_{21}/k_\mathrm{B}T} - 1$ to find

$$A_{21} = (1 - e^{-E_{21}/k_\mathrm{B}T}) A_{21} + B_{21} \frac{8\pi h\nu_{21}^3}{c^3} . \tag{3.21}$$

In the $T \to \infty$ limit, the exponential on the RHS approaches unity and we obtain (together with Equation 3.19)

$$\boxed{B_{21} = \frac{c^3}{8\pi h\nu_{21}^3} A_{21} \quad \text{and} \quad B_{12} = \frac{g_2}{g_1} B_{21} = \frac{g_2}{g_1} \frac{c^3}{8\pi h\nu_{21}^3} A_{21}} \tag{3.22}$$

We can write the $1 \to 2$ and $2 \to 1$ transition rates in a particularly elegant form using the photon occupation number (Equation 3.12),

$$\begin{aligned} \left.\frac{\mathrm{d}n_1}{\mathrm{d}t}\right|_{2\to1} &= n_2 A_{21} (1 + \bar{n}_\gamma) \\ \left.\frac{\mathrm{d}n_2}{\mathrm{d}t}\right|_{1\to2} &= n_1 A_{21} \frac{g_2}{g_1} \bar{n}_\gamma . \end{aligned} \tag{3.23}$$

Balancing these rates would return the expression for the level population (Equation 3.8). In summary, we can now calculate the rates of absorption and emission in a radiation field given only one system-dependent parameter, A_{21}. This parameter can be looked up in tables, e.g., for the energy levels of atoms (O&F §3.5).

3.5 Collisional excitation

For atoms, another important source of excitations is provided by collisions with electrons (or other particles, but we assume electrons for concreteness). This mechanism can contribute to radiative cooling, where kinetic energy from electrons is converted into excited states that decay and emit photons, which escape (§3.7). The kinds of atomic lines that are most relevant depends on the composition and temperature of a gas.

Once again, we wish to calculate the rate at which such collisions happen and the populations in the ground state and excited levels. As in the previous section, we consider a two-level system. The cross-section for an excitation, $\sigma_{12}(u)$, depends on the electron velocity u and the properties of a given atom. Near the lower velocity threshold set by $m_e u^2/2 \geq E_{21}$, $\sigma \propto 1/u^2$ because the Coulomb force "focuses" slower electrons more effectively (O&F §3.5). This insight allows us to write the cross-section as

$$\sigma_{12}(u) = \frac{\pi\hbar^2}{m_e u^2}\frac{\Omega_{12}(u)}{g_1}\,, \tag{3.24}$$

where $\Omega_{12}(u)$ is called the collision strength for the $1 \to 2$ transition (which is only weakly dependent on the velocity). There is a similar cross-section, $\sigma_{21}(u)$, for the opposite process, collisional deexcitation. To relate this "reverse" cross-section to σ_{12}, we take the concept of equilibrium even further that previously. LTE means not only that the level populations are in equilibrium, but also that the distribution of electron energies is constant. Thus, excitations and deexcitations must **balance at each energy**, which is called the principle of **detailed balance**. The rate of collisions is proportional to the cross-section times the electron velocity (i.e., the volume swept out per time by a moving electron). By balancing the $1 \to 2$ and $2 \to 1$ rates we get

$$n_e n_1 u_1 \sigma_{12}(u_1) f(u_1) \mathrm{d}u_1 = n_e n_2 u_2 \sigma_{21}(u_2) f(u_2) \mathrm{d}u_2\,, \tag{3.25}$$

where the difference in velocities makes up for the energy difference between the levels,

$$\frac{m_e u_1^2}{2} = \frac{m_e u_2^2}{2} + E_{21}\,. \tag{3.26}$$

From §3.3, we know that the level populations in equilibrium are related by

$$\frac{n_2}{n_1} = \frac{g_2}{g_1} e^{-E_{21}/k_B T} \tag{3.27}$$

and that the velocity distribution $f(u)$ follows the Maxwell-Boltzmann distribution (Equation 3.9). The exponentials from Equation 3.26 and the ratios of $f(u)$ cancel, and we are left with the particularly simple relation

$$g_1 u_1^2 \sigma_{12} = g_2 u_2^2 \sigma_{21}\,. \tag{3.28}$$

To get the actual transition rate q_{21}, we need to integrate over the electron velocities. This step is easier for the deexcitation rate because there is no minimum velocity threshold. The rate per unit volume per unit time is $n_e n_2 q_{21}$, where

$$q_{21} = \int_0^\infty u \sigma_{21} f_{\mathrm{MB}}(u) \mathrm{d}u = \sqrt{\frac{2\pi}{k_B T}}\frac{\hbar^2}{{m_e}^{3/2}}\frac{\Upsilon_{12}(T)}{g_2}\,, \tag{3.29}$$

where $\Upsilon_{12}(T)$ is the **velocity-averaged collision strength**,

$$\Upsilon_{12}(T) \equiv \frac{m_e}{k_B T}\int_0^\infty \Omega_{12}(u) e^{-m_e u^2/2k_B T} u \,\mathrm{d}u\,. \tag{3.30}$$

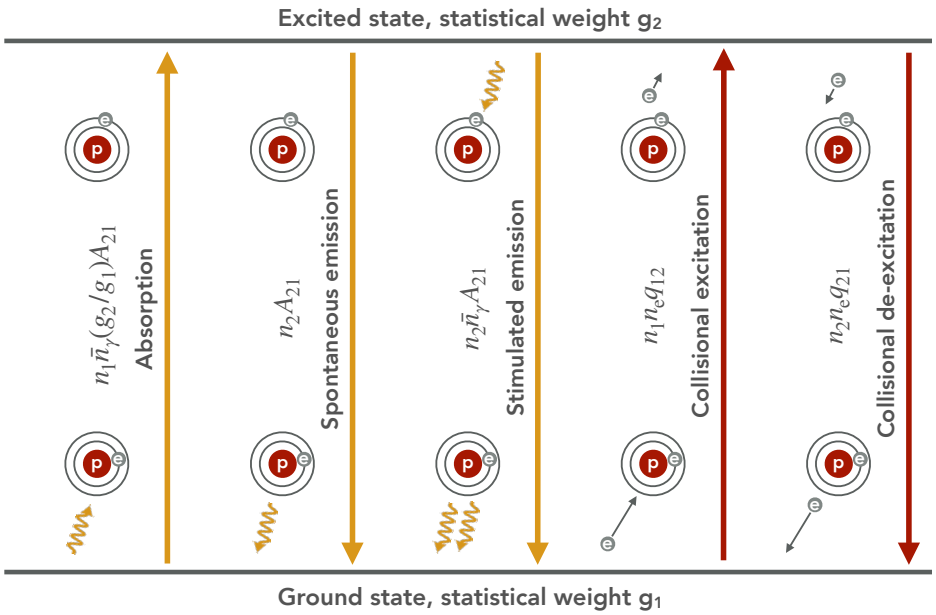


Figure 10: Schematic depiction of the five channels of excitation and deexcitation. Radiative processes are represented with yellow arrows and collisional processes with red arrows. A hydrogen atom is used to depict a system with two energy levels.

These collision strengths are dimensionless and can be found in lookup tables (e.g., O&F §3.5). We use Equation 3.25 to work out the equivalent excitation rate,

$$q_{12} = \frac{g_2}{g_1} q_{21} e^{-E_{21}/k_\mathrm{B}T} = \sqrt{\frac{2\pi}{k_\mathrm{B}T}} \frac{\hbar^2}{{m_\mathrm{e}}^{3/2}} \frac{\Upsilon_{12}(T)}{g_1} e^{-E_{21}/k_\mathrm{B}T} \,. \tag{3.31}$$

As with the radiative transitions in §3.4, we have reduced the number of unknown factors in the collisional excitation and deexcitation rates to a single quantum-mechanical constant per transition.

3.6 Two-level atom with radiative and collisional excitation

In the previous two sections, we have treated the cases of photon and collisional excitation, which give a total of five possible mechanisms for excitation and deexcitation (Figure 10). With all of this machinery in place, we can now write down the balance of states in terms of both rates,

$$\underbrace{n_\mathrm{e} n_1 q_{12}}_{\text{coll. exc.}} + \underbrace{n_1 \bar{n}_\gamma \frac{g_2}{g_1} A_{21}}_{\text{absorption}} = \underbrace{n_\mathrm{e} n_2 q_{21}}_{\text{coll. deexc.}} + \underbrace{n_2 (1 + \bar{n}_\gamma) A_{21}}_{\text{spont.+stim. em.}} \tag{3.32}$$

where we have written the radiative transition rates in terms of the photon occupation number as in Equation 3.23. The ratio of level populations is then

$$\frac{n_2}{n_1} = \frac{n_\mathrm{e} q_{12} + \bar{n}_\gamma (g_2/g_1) A_{21}}{n_\mathrm{e} q_{21} + (1 + \bar{n}_\gamma) A_{21}} \tag{3.33}$$

This ratio does not necessarily equal the thermal equilibrium expression (Equation 3.26) because we did not assume equilibrium: the radiation field does not need to be a blackbody, and the velocity

distribution of electrons could differ from Maxwell-Boltzmann. To gain insight into the meaning of this equation, it helps to define the **critical density** of electrons where the collisional deexcitation rate matches the radiative one,

$$\boxed{n_{\rm crit} \equiv \frac{(1+\bar{n}_\gamma)A_{21}}{q_{21}}} \tag{3.34}$$

This expression also includes stimulated emission. We can develop further intuition by taking physically meaningful limits. When there is **no radiation field**, $\bar{n}_\gamma = 0$, we get

$$\frac{n_2}{n_1} = \frac{n_{\rm e} q_{12}}{n_{\rm e} q_{21} + A_{21}} = \frac{g_2}{g_1} e^{-E_{21}/k_{\rm B}T} \frac{1}{1 + \frac{A_{21}}{n_{\rm e} q_{21}}} = \frac{g_2}{g_1} e^{-E_{21}/k_{\rm B}T} \left(1 + \frac{n_{\rm crit}}{n_{\rm e}}\right)^{-1} . \tag{3.35}$$

The Einstein A_{21} coefficient is still present because spontaneous emission happens even in the absence of any radiation field. The expression above highlights that the states are in thermodynamic equilibrium except for the spontaneous emission term. We can think of this term as a comparison between the electron density and the critical density. If $n_{\rm e} \gg n_{\rm crit}$, we recover the thermodynamical equilibrium expression because the radiative term becomes insignificant. If $n_{\rm e} \ll n_{\rm crit}$, the upper level is suppressed by $n_{\rm e}/n_{\rm crit}$ compared to the equilibrium expression because there are not enough collisions to keep up with the radiative deexcitations.

Another interesting limit of Equation 3.33 is the **low-density limit**, $n_{\rm e} \to 0$, in which case the level populations are purely determined by the radiation field,

$$\frac{n_2}{n_1} = \frac{g_2}{g_1} \frac{\bar{n}_\gamma}{(1+\bar{n}_\gamma)} . \tag{3.36}$$

For a blackbody, we recover the thermodynamic equilibrium expression of Equation 3.26 once again. Interestingly, the latter case applies more generally: as $\bar{n}_\gamma$ approaches the blackbody distribution, the occupation ratio approaches

$$\begin{aligned} \frac{n_2}{n_1} &= \frac{n_{\rm e} q_{21} (g_2/g_1) e^{-E_{21}/k_{\rm B}T} + (g_2/g_1) \frac{1}{e^{E_{21}/k_{\rm B}T}} A_{21}}{n_{\rm e} q_{21} + \left(1 + \frac{1}{e^{E_{21}/k_{\rm B}T} - 1}\right) A_{21}} \\ &= \frac{g_2}{g_1} e^{-E_{21}/k_{\rm B}T} \frac{n_{\rm e} q_{21} + \frac{1}{1 - e^{-E_{21}/k_{\rm B}T}} A_{21}}{n_{\rm e} q_{21} + \left(\frac{e^{E_{21}/k_{\rm B}T}}{e^{E_{21}/k_{\rm B}T} - 1}\right) A_{21}} = \frac{g_2}{g_1} e^{-E_{21}/k_{\rm B}T} . \end{aligned} \tag{3.37}$$

We recover the thermodynamic equilibrium level occupation regardless of $n_{\rm e}$! When photons and particles are in equilibrium, the photons alone bring the system into equilibrium and the additional collisions with electrons do not change the level populations.

3.7 Line cooling

In §2.2, we simply stated that cooling rates can be computed from first principles. To do so, we neglect absorption and stimulated emission, meaning that collisional excitation dominates. Collisional deexcitation can play a significant role though, and the residual radiation fields cause departures from exact equilibrium. The cooling rate for a given line can be worked out from the equations in §3.6. If the electron density is very low, we can assume that each collisional excitation is directly followed by the emission of a photon (O&F §3.5), and

$$\Lambda_{\rm ce,12} = n_{\rm e} n_1 q_{12} h\nu_{21} . \tag{3.38}$$

If, however, the density is high enough for collisional deexcitation to be important, we need to use Equation 3.35, which gives a cooling rate of

$$\Lambda_{\mathrm{ce},12} = n_2 A_{21} h\nu_{21} = n_\mathrm{e} n_1 q_{12} h\nu_{21} \frac{1}{1 + \frac{n_\mathrm{e} q_{21}}{A_{21}}} \,. \tag{3.39}$$

In the limit of high electron density, we recover the thermodynamic equilibrium cooling rate,

$$\Lambda_{\mathrm{ce},12} = n_1 A_{21} h\nu_{21} \frac{g_2}{g_1} e^{-E_{21}/k_\mathrm{B}T} \,. \tag{3.40}$$

Either way, we sum over all possible lines to get the total line cooling rates shown in Figure 5.

4 Neutral atomic gas

The majority of gas in the galactic ISM, about 60%, is in the neutral atomic phase, or H I. This gas is cool enough to not be collisionally ionized, far enough from O stars to not be part of an H II region or otherwise photoionized (§6), but not cold enough to form molecules. However, the ionization fraction is not identically zero because photons leaking out from O stars, cosmic rays, and X-rays ionize a small fraction of the gas (Draine §16 and §29).

The H I is further sub-divided into two distinct phases, the **Cold Neutral Medium (CNM)** and **Warm Neutral Medium (WNM)**. The WNM slightly dominates, with about 60% of the H I. As the names suggest, these phases are distinguished by their very different temperatures, roughly 100 K and $5000 - 8000$ K (we will discuss the reasons in §10.1). The real temperatures vary and are somewhat uncertain for the WNM, but it is clear that the two phases are in rather different thermodynamic states. Given that all ISM phases are in approximate pressure equilibrium, the roughly equal mass fractions and different temperatures translate into different number densities of about $30 - 40/\mathrm{cm}^3$ and $0.5/\mathrm{cm}^3$, respectively. As a result, the WNM occupies a significant fraction of the ISM volume, about 50% near the midplane of the Milky Way (Draine §29).

Given the relatively low temperature of the CNM, we might expect it to be hard to detect H I directly, or that its signature might depend strongly on its temperature. Thankfully, nature has provided an extremely useful observational diagnostic for H I: the 21 cm line. In the following two sections, and using insights from §3, we derive the properties of this line and see how a given density and temperature translate into observable 21 cm radiation.

4.1 Hyperfine splitting of hydrogen and the 21 cm line

The spin flip transition arises because the ground state of hydrogen splits into two different states, with the proton and electron spins either anti-aligned (the lower level) or aligned (the upper level). The latter state has $g_2 = 2S + 1 = 3$ while the true ground state has $g_1 = 1$. The energy difference between spin states is much smaller than the lines we have encountered so far, $E_{21} = 5.87 \times 10^{-6}$ eV. This corresponds to a wavelength of 21.11 cm or a frequency of 1420 MHz. The spontaneous decay rate is also much smaller than those of normal atomic levels, $A_{21} = 2.88 \times 10^{-15}/\mathrm{s}$, with an equivalent lifetime $t_{21} \approx 10^7$ yr (Draine §17.3).

What is the level population for this system? The energy difference corresponds to a temperature $T_{21} = E_{21}/k_\mathrm{B} = 0.0682$ K, meaning that the 21 cm line is always collisionally excited as long as the density is sufficient. However, since H I gas is mostly neutral by definition, electrons tend to play a minor role; the majority of collisions is with other hydrogen atoms. The rate for this process has been worked out and is approximately

$$q_{21}(T_\mathrm{gas}) = 10^{-10}\ \frac{\mathrm{cm}^3}{\mathrm{s}} \times \begin{cases} 1.19\ T_2^{0.74-0.20\ln T_2} & \forall\ 20\ \mathrm{K} < T_\mathrm{gas} < 300\ \mathrm{K} \\ 2.24\ T_2^{0.207} e^{-0.876/T_2} & \forall\ 300\ \mathrm{K} < T_\mathrm{gas} < 1000\ \mathrm{K}\,, \end{cases} \tag{4.1}$$

where $T_2 = T_\mathrm{gas}/100$ K (Draine §17.3). We can get the collisional excitation rate q_{12} from Equation 3.31,

$$q_{12} = q_{21} \frac{g_2}{g_1} e^{-E_{21}/k_\mathrm{B}T_\mathrm{gas}} = q_{21} \times 3 e^{-0.0682\ \mathrm{K}/T_\mathrm{gas}}\,. \tag{4.2}$$

For the average patch of H I gas in the Galaxy, we would not assume that any strong radiation sources are present. However, the 21 cm line has such low energy that even the Cosmic Microwave Background (CMB) can excite it! Besides the blackbody field with $T_\mathrm{CMB} = 2.725$ K, another important source of radiation is galactic radio background due to synchrotron emission (basically

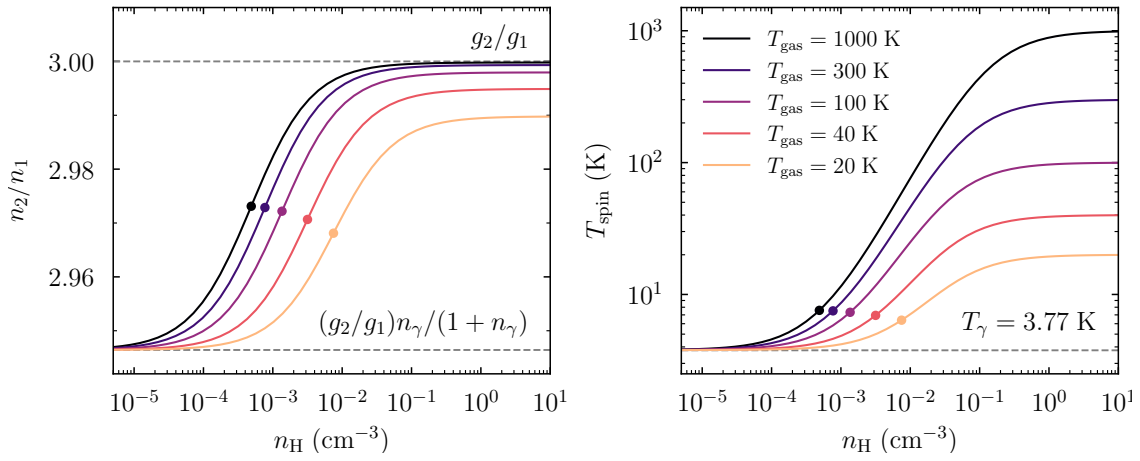


Figure 11: Level population (left) and the equivalent spin temperature (right) for the 21 cm hyperfine transition of hydrogen. Each line shows a different gas temperature. At very low densities, the line is governed by the background radiation field, which is modeled as a blackbody with $T_{\rm A} = 3.77$ K and $\bar{n}_\gamma = 55$.

magnetic Bremsstrahlung), which has an equivalent temperature of $T_{\rm syn} \approx 1.04$ K. Summing those contributions,[4] we get an "antenna temperature" of $T_\gamma \approx 3.77$ K (Draine §17.3). At the frequency of the 21 cm transition, the photon occupation number is then

$$\bar{n}_\gamma = \frac{1}{e^{E_{21}/k_{\rm B}T_\gamma} - 1} \approx \frac{k_{\rm B}T_\gamma}{E_{21}} \approx \frac{3.77\ {\rm K}}{0.0682\ {\rm K}} \approx 55\,. \tag{4.3}$$

The fact that $\bar{n}_\gamma \gg 1$ reminds us that $k_{\rm B}T_\gamma \gg E_{21}$. Given the large $\bar{n}_\gamma$, radiative excitation and deexcitation will be important despite the small A_{21}. We solve for the level populations using Equation 3.33, but replacing $n_{\rm e}$ with $n_{\rm H}$,

$$\frac{n_2}{n_1} = \frac{n_{\rm H}q_{12} + \bar{n}_\gamma(g_2/g_1)A_{21}}{n_{\rm H}q_{21} + (1+\bar{n}_\gamma)A_{21}}\,. \tag{4.4}$$

Figure 11 shows the solution for a range of gas temperatures and densities. The right panel shows the equivalent **spin temperature**, which is defined as the temperature we would infer if the system was in thermodynamic equilibrium (which it is most likely not),

$$\frac{n_2}{n_1} = \frac{g_2}{g_1}e^{-E_{21}/k_{\rm B}T_{\rm spin}} \quad \Longrightarrow \quad T_{\rm spin} = \frac{E_{21}}{k_{\rm B}\ln(g_2 n_1/g_1 n_2)} = \frac{0.0682\ {\rm K}}{\ln(3n_1/n_2)}\,. \tag{4.5}$$

To understand the solution, we recall that the critical density (Equation 3.34) represents the density at which radiation and collisions switch as the dominant factor setting the level populations,

$$n_{\rm crit} = \frac{(1+\bar{n}_\gamma)A_{21}}{q_{21}}\,. \tag{4.6}$$

The critical density is shown with colored points in Figure 11. We note that $n_{\rm crit}$ refers to the hydrogen density $n_{\rm H}$ rather than the electron density in this case because the majority of collisions are

[4] In general, blackbody spectra of different temperatures do not simply add. However, in the so-called Rayleigh-Jeans limit when $h\nu \ll k_{\rm B}T$, we have $e^{h\nu/k_{\rm B}T} \approx 1 + h\nu/k_{\rm B}T$ and $J_{\nu,{\rm BB}} \approx (2\nu^2/c^2)k_{\rm B}T$, so that two blackbody spectra do add. This case applies here because we are interested in the blackbody intensity at the frequency of the 21 cm line, for which $h\nu$ is tiny.

with atoms. At low densities, $n_\mathrm{H} \ll n_\mathrm{crit}$, radiative excitations dominate and the level populations reflect the CMB+synchrotron temperature of the radiation. The solution tends to

$$\frac{n_2}{n_1} \to \frac{g_2}{g_1}\frac{\bar{n}_\gamma}{(1+\bar{n}_\gamma)} = \frac{g_2}{g_1} e^{-E_{21}/k_\mathrm{B}T_\gamma} = 3e^{-0.0682\ \mathrm{K}/3.77\ \mathrm{K}} \approx 2.946\,. \tag{4.7}$$

The spin temperature approaches the radiation temperature, $T_\mathrm{spin} \to T_\gamma$. At $n_\mathrm{H} \gg n_\mathrm{crit}$, the level populations reflect the thermodynamic temperature of the H I gas,

$$\frac{n_2}{n_1} \to \frac{q_{12}}{q_{21}} = \frac{g_2}{g_1} e^{-E_{21}/k_\mathrm{B}T_\mathrm{gas}} = 3e^{-0.0682\ \mathrm{K}/T_\mathrm{gas}}\,. \tag{4.8}$$

Equivalently, the spin temperature approaches the gas temperature, $T_\mathrm{spin} \to T_\mathrm{gas}$. Whenever $T_\mathrm{gas} \gg 0.0682$ K, the exponential tends to unity and $n_2/n_1 \to 3$. Overall, we find that $T_{21} \approx 0.0682$ K is so small that even very cold gas or a weak radiation field such as the CMB is sufficient to push the level population towards $g_2/g_1 = 3$, meaning that about 3/4 of the hydrogen atoms are in the excited state regardless of density and temperature.

4.2 Emission and absorption in H I gas

Observationally, the finding that the level population is more or less constant has an important consequence: the **21 cm emissivity of H I gas does not depend on density and temperature**, even in the regime where collisions dominate over radiation. Specifically, the emissivity[5] is

$$j_\nu = \frac{1}{4\pi} n_2 A_{21} h\nu_{21} \phi_\nu \approx \frac{3}{16\pi} n_\mathrm{H} A_{21} h\nu_{21} \phi_\nu\,, \tag{4.9}$$

where ϕ_ν is the normalized line profile that accounts for the velocity distribution of the H I (Draine §8.1). This is a key result because it means that the total emission is proportional to the total mass of H I observed. However, a direct translation to a column density is possibly only if the observed H I region is optically thin. To figure that out, we compute the **attenuation coefficient**,

$$\kappa_\nu = n_1\sigma_{12} - n_2\sigma_{21}\,, \tag{4.10}$$

which has units of cm^{-1} and quantifies the chance of a photon to be absorbed along its path. Here, σ_{12} and σ_{21} refer to the cross-sections for radiative excitation and deexcitation. We did not define those because we used the Einstein B coefficients in §3.4, but it can be shown that B_{12} is proportional to the integral of σ_{12} over frequency (Draine §6.2). Since B_{12} is directly related to A_{21} (§3.4), we can write

$$\sigma_{12} = \frac{g_2}{g_1}\frac{c^2}{8\pi\nu_{21}^2}\phi_\nu A_{21} \tag{4.11}$$

and $\sigma_{21} = (g_1/g_2)\sigma_{12}$, in analogy with the B coefficients (Equation 3.22). Putting these expressions into Equation 4.10, we get

$$\kappa_\nu = n_1\frac{g_2}{g_1}\frac{c^2\phi_\nu A_{21}}{8\pi\nu_{21}^2}\left(1 - \frac{g_1}{g_2}\frac{n_2}{n_1}\right) = n_1\frac{g_2}{g_1}\frac{c^2\phi_\nu A_{21}}{8\pi\nu_{21}^2}\left(1 - e^{E_{21}/k_\mathrm{B}T_\mathrm{spin}}\right)\,. \tag{4.12}$$

The exponential form of the level populations is always valid by definition of the spin temperature (Equation 4.5). Once again, we use that $k_\mathrm{B}T_\mathrm{spin} \gg E_{21}$, as well as $n_1 = n_\mathrm{H}/4$, to write

$$\kappa_\nu \approx \frac{3}{32\pi}\frac{c^2 h}{\nu_{21}}\frac{n_\mathrm{H} A_{21}\phi_\nu}{k_\mathrm{B}T_\mathrm{spin}}\,. \tag{4.13}$$

[5] The emissivity has units of energy density per time per Hz per steradian, so that integrating it along a path gives an intensity J_ν, which has units of energy per area per time per Hz per steradian. In this formula, ϕ_ν has units of $1/\mathrm{Hz} = \mathrm{s}$ and is normalized such that $\int \phi_\nu \mathrm{d}\nu = 1$.

The absorption coefficient depends on the spin temperature! This may seem surprising because the level populations barely deviate from g_2/g_1 as we saw in Figure 11. The key is that the absorption depends on the difference between the rates of emission and absorption (Equation 4.10). As $T_{\rm spin} \to \infty$, stimulated emission ($n_2\sigma_{21}$) becomes exactly as likely as absorption ($n_1\sigma_{12}$), and the net result is nil.

To compute an optical depth for H I, we assume that the 21 cm line is broadened by particle velocities u with a normalized Gaussian line profile,

$$\phi_\nu = \frac{c}{\sqrt{2\pi}\nu_{21}\sigma_{\rm u}} e^{u^2/2\sigma_{\rm u}^2}\,, \tag{4.14}$$

where $\sigma_{\rm u}$ is the velocity dispersion; we insert this expression into Equation 4.13. We multiply κ_ν by a distance to get an optical depth, $\tau_\nu = \kappa_\nu L$, which we express in terms of the column density, $N_{\rm H} = n_{\rm H} L$. At the center of the line (i.e., ignoring the u-dependence of the velocity profile), we get

$$\tau(\nu_{21}) = 0.22 \left(\frac{N_{\rm H}}{10^{20}/{\rm cm}^2}\right)\left(\frac{T_{\rm spin}}{100\ {\rm K}}\right)^{-1}\left(\frac{\sigma_{\rm u}}{1\ {\rm km/s}}\right)^{-1}\,. \tag{4.15}$$

We conclude that column densities up to $N_{\rm H} \approx 10^{20}/{\rm cm}^2$ are optically thin, meaning that the total intensity of 21 cm radiation is an accurate proxy for the total column density of H I.

5 Atomic physics II: Photoionization and recombination

In §3, we studied how photons can excite and deexcite energy levels in quantized systems such as atoms. In this chapter, we complete the picture by also considering the ionization of atoms by photons. We work out the cross-section for this process in §5.1 (Draine §13.1) and consider its opposite, recombination, in §5.2 (Draine §14.1). We balance the two processes in equilibrium to find the ionization state of a gas in §5.3 (Draine §3). A short overview of some of the topics can also be found in O&F §2.2.

5.1 Photoionization

When studying H II regions in §6, we will need to calculate the cross-sections for the ionization and recombination of hydrogen atoms. We assume that all ionizations occur from the ground state, which turns out to be justified by a comparison of the typical lifetime of excited states (small fractions of a second, §3.2) and ionization times (about a year, §6.1).

The actual calculation of the ionization cross-section is complicated and quantum-mechanical in nature, but it turns out that an analytical expression can be derived for systems with only one electron and a nuclear charge Z (which is 1 for hydrogen),

$$\sigma_{\rm pi}(\nu) = \frac{2^9\pi^2\alpha a_0^2}{3e^4}\frac{1}{Z^2}\left(\frac{Z^2\nu_0}{\nu}\right)^4 \frac{e^{4-4\tan^{-1}(x)/x}}{1-e^{-2\pi/x}} \qquad x \equiv \sqrt{\frac{\nu}{Z^2\nu_0}-1}\,, \tag{5.1}$$

where ν is the frequency of the (potentially) ionizing photons, ν_0 is the frequency above which photons can ionize hydrogen, α is the fine-structure constant, and a_0 is the Bohr radius,

$$h\nu_0 \approx 13.6\text{ eV} \qquad \alpha \equiv \frac{q_{\rm e}^2}{\hbar c} \approx \frac{1}{137} \qquad a_0 \equiv \frac{\hbar^2}{m_{\rm e}q_{\rm e}^2} \approx 5.3\times10^{-9}\text{ cm}\,. \tag{5.2}$$

We can interpret the expression for $\sigma_{\rm pi}$ as follows. The ionization energy is $Z^2\nu_0$. As the photon energy approaches the exact ionization energy, we have $x \to 0$ and thus

$$\sigma_{\rm pi}(\nu \to Z^2\nu_0) \to \frac{2^9\pi^2\alpha a_0^2}{3e^4}\frac{1}{Z^2} \approx \frac{0.07}{Z^2}\pi a_0^2 \approx 6.3\times10^{-18}\text{ cm}^2\left(\frac{1}{Z^2}\right)\,. \tag{5.3}$$

We can think of the cross-section as a fraction of the "Bohr area" spanned by the pseudo-classical radius of a hydrogen atom. Towards higher photon energies, the cross-section decreases steeply due to the ν^{-4} term (Figure 2.2 in O&F or Figure 13.1 in Draine). In other words, the photon frequency needs to be fairly finely tuned to the ionization energy. For systems with more than one electron, the calculation gets much more complicated because the electron that is being removed from the ion could have originated from different subshells. The resulting cross-sections may no longer monotonously decrease with frequency because higher energies can open up lower-energy shells for ionization (Figure 13.2 in Draine).

In astrophysical scenarios, we are less interested in the cross-section but rather in the number of ionizations per unit volume per unit time given a certain density. To compute this number, we need to integrate over all possible photon energies. We define the **ionization rate per atom** per unit time for hydrogen,

$$\boxed{G_{\rm pi} \equiv \int_{\nu_0}^{\infty}\frac{4\pi J_\nu}{h\nu}\sigma_{\rm pi}(\nu){\rm d}\nu} \tag{5.4}$$

where J_ν is the mean intensity and the factor of 4π integrates over solid angles. Since $G_{\rm pi}$ has units of s^{-1}, $nG_{\rm pi}$ has the desired units of a rate per unit volume per unit time.

5.2 Recombination

We now consider the reverse process to ionization, recombination. As for ionization, we are interested in a rate of recombinations per unit volume per unit time. Given that a proton and an electron must meet for a recombination to happen, it is clear that we can write this rate as

$$\frac{\text{recombinations}}{\text{unit time} \times \text{unit volume}} = n_\mathrm{e} n_\mathrm{p}\ \alpha(T) \tag{5.5}$$

where $\alpha(T)$ is called the **recombination coefficient**. We have abbreviated the density of ionized hydrogen as $n(H^+) = n_\mathrm{p}$. This coefficient has units of cm^3/s so that its product with n^2 gives a rate per unit volume per unit time, as for the ionization rate. There are a number of ways in which an electron can recombine with a an ion (see Draine §14), but we will focus on radiative recombination, where a photon carries away the energy difference between the (moving) electron and the state it settles into. That state, however, is not always the ground state: it is common for an electron to settle into higher energy shells, which then quickly decay to the ground state (§3.1). The recombination rate to a given level nl is the velocity-dependent cross-section multiplied by the electron velocity u, i.e., the volume swept out by a moving electron per time. The total recombination coefficient to that level is the integral over electron velocities,

$$\alpha_\mathrm{nl}(T) = \int_0^\infty u \sigma_\mathrm{nl}(u) f(u) \mathrm{d}u\,, \tag{5.6}$$

where $\sigma_\mathrm{nl}(u)$ is the cross-section for recombination into level nl and $f(u)$ is the velocity distribution. A priori, this distribution might not be straight-forward because it contains electrons that were freed from atoms by ionization and are thus not in thermal equilibrium. However, the cross-section for elastic scattering between electrons is much larger than the ionization cross-section, about 10^{-13} cm^2. Thus, the photoionization electrons are quickly thermalized, and we can assume the Maxwell-Boltzmann distribution (Equation 3.9). Alternatively, we can define the electron kinetic energy, $E = m_\mathrm{e} u^2/2$, and rewrite the coefficient as

$$\alpha_\mathrm{nl}(T) = \sqrt{\frac{8 k_\mathrm{B} T}{\pi m_\mathrm{e}}} \int_0^\infty \sigma_\mathrm{nl}(E) \frac{E}{k_\mathrm{B} T} e^{-E/k_\mathrm{B}T} \frac{dE}{k_\mathrm{B} T}\,. \tag{5.7}$$

We defer the calculation of the cross-section σ_nl to §5.3, but it is roughly proportional to E^{-1}. The extra $1/T$ factor means that approximately $\alpha_\mathrm{nl}(T) \propto T^{-1/2}$. These rates are tabulated for a few temperatures in Table 14.1 in Draine, and we can sum over all n and l to get the total coefficient,

$$\boxed{\alpha_\mathrm{A}(T) = \sum_{n=1}^{\infty} \sum_{l=0}^{n-1} \alpha_\mathrm{nl}(T)} \tag{5.8}$$

This calculation, however, is not always valid. If the recombination is directly into the ground state, the resulting photon has $\nu > \nu_0$ and can ionize another hydrogen atom. If the surrounding medium is optically thin to to ionizing radiation, this photon will escape and Equation 5.8 is correct. If, on the other hand, the surrounding medium is optically thick to ionizing photons, the photon will immediately ionize another atom, and the preceding recombination has no net effect. Contrary to the optically thin **Case A** of Equation 5.8, the optically thick scenario is called **Case B** and results in a coefficient

$$\boxed{\alpha_\mathrm{B}(T) = \sum_{n=2}^{\infty} \sum_{l=0}^{n-1} \alpha_\mathrm{nl}(T) = \alpha_\mathrm{A}(T) - \alpha_\mathrm{1s}(T)} \tag{5.9}$$

Approximations for $\alpha_\mathrm{A}(T)$ and $\alpha_\mathrm{B}(T)$ are listed in Draine Table 14.1. They are not well described by a simple power law (Draine §14.2).

5.3 Ionization balance and the Milne relation

Our final task is to calculate the cross-section for recombination, $\sigma_{\rm nl}$. It turns out that we can achieve this goal by invoking an equilibrium balance between ionizations and recombinations. The derivation is fairly lengthy; additional details can be found in Draine §3 and O&F Appendix 2. Without loss of generality, we can imagine a system in **photoionization equilibrium**, meaning that the rates of ionization and recombination are equal,

$$\boxed{n_0\ G_{\rm pi} = n_{\rm e} n_{\rm p}\ \alpha(T)} \tag{5.10}$$

As described in §5.1 and §5.2, we have multiplied the ionization rate by the number density of neutral hydrogen atoms, $n_0 \equiv n(H^0)$, and the recombination coefficient by $n_{\rm e} n_{\rm p} = n_{\rm e}^2$ to obtain units of $\rm cm^{-3} s^{-1}$ on both sides. This equation will be the foundation of our modeling of the ionized regions around stars (§6), where we use a known $G_{\rm pi}$ and $\alpha(T)$ to infer the relative numbers of neutral and ionized atoms. For now, our purpose is the opposite: we wish to relate the rates of ionization and recombination. To achieve this, we need to know the densities. We use the idea of thermodynamic equilibrium, specifically the law of mass action (Equation 3.7), which we apply to the ionization/recombination reaction, $e^- + H^+ \leftrightarrow H^0$,

$$\frac{n_0}{n_{\rm e} n_{\rm p}} = \frac{p(H^0)}{p(e^-)p(H^+)} = \left(\frac{h^2}{2\pi m_{\rm e} k_{\rm B} T}\right)^{3/2} \frac{\sum_i g_{0,i}\ e^{-E_{0,i}/k_{\rm B}T}}{g_{e,i} \sum_i g_{p,i}\ e^{-E_{+,i}/k_{\rm B}T}}\,. \tag{5.11}$$

The symbols $E_{0,i}$ and $E_{+,i}$ denote the energies of the different states of neutral and ionized hydrogen. Note that the internal partition function of a single electron or proton is just $g_e = g_{\rm p} = 2$ (spin up and spin down). The degeneracy of the neutral state depends on the quantum number n, $g_0 = g_{\rm n}$; for the $1s$ state, $g_0 = 4$ (both proton and electron can be either spin-up or spin-down).

We now make the key assumption that only the lowest-energy state (the first term in each sum) contributes significantly, which corresponds to a relatively low temperature and thus suppressed partition functions for states with $E \gg k_{\rm B}T$. With this simplification, we obtain the **Saha equation** for ionization,

$$\boxed{\frac{n_0}{n_{\rm e} n_{\rm p}} = \frac{g_{\rm n}}{g_{\rm e} g_{\rm p}} \left(\frac{h^2}{2\pi m_{\rm e} k_{\rm B} T}\right)^{3/2} e^{(E_+ - E_{\rm n})/k_{\rm B}T}} \tag{5.12}$$

where $E_{\rm n}$ is the energy in shell n. Similar equations can be found for general scenarios; for hydrogen we have $E_+ - E_0 = 13.6$ eV. We will now use the Saha equation to infer the recombination cross-section based on an imaginary system in perfect equilibrium. In practice, however, we should not assume that the equation applies to gases in the ISM because they are often out of equilibrium (for example, due to non-blackbody radiation fields; Draine §3.3).

Combined with the Saha equation, the ionization balance establishes a relation between $G_{\rm pi}$ and $\alpha(T)$, but those rates are averaged over photon energies and electron velocities. Thus, we do not yet have an expression for the recombination cross-section at a particular velocity. To solve this issue, we once again take advantage of the concept of detailed balance (§3.5): there must be a hypothetical system where the rate at which photons in an energy interval $h{\rm d}\nu$ are removed from a fluid by ionization is balanced by the rate at which they are replenished by recombination. Thus, we equate $n_0 G_{\rm pi}$ and $n_{\rm e} n_{\rm p} \alpha(T)$ as in Equation 5.10 but at each energy, i.e., without the integrals in Equations 5.4 and 5.6,

$$n_0 \frac{4\pi J_\nu}{h\nu} \sigma_{\rm pi}(h\nu) {\rm d}\nu = n_{\rm e} n_{\rm p} u f_{\rm MB}(u) \sigma_{\rm nl}(u) [1 + \bar{n}_\gamma] {\rm d}u\,. \tag{5.13}$$

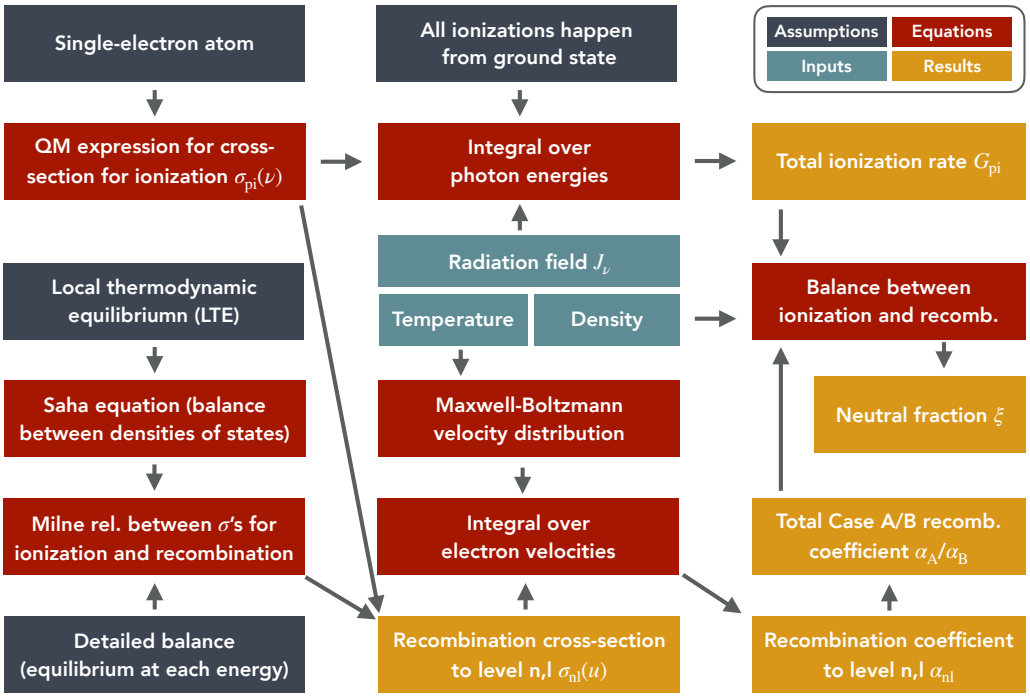


Figure 12: Flowchart for the derivation of the ionization and recombination rates, and the resulting ionization fraction. Dark blue fields represent assumptions that go into the respective derivations, cyan fields represent input quantities, red fields represent expressions or equations, and yellow fields represent results.

The $1 + \bar{n}_\gamma$ factor accounts for the possibility of stimulated emission, where a photon of the right frequency triggers a recombination that produces a second photon of the same frequency (analogous to stimulated emission, §3.4). To satisfy the principle of detailed balance, we need to consider the electron velocity that corresponds to the difference between the photon energy $h\nu$ and the ionization energy,

$$\frac{m_e u^2}{2} = h(\nu - \nu_0) \quad \Longrightarrow \quad u = \sqrt{\frac{2h(\nu-\nu_0)}{m_e}} \quad \Longrightarrow \quad \mathrm{d}u = \frac{h}{m_e u}\mathrm{d}\nu\,. \tag{5.14}$$

Moreover, we know that the radiation field must be a blackbody in thermodynamic equilibrium, $J_\nu = J_{\nu,\mathrm{BB}}$ (Equation 3.10). We rearrange Equation 5.13 to solve for the ratio of the recombination and ionization cross-sections and insert the Saha equation (5.12) and the Maxwell-Boltzmann distribution (3.9),

$$\begin{aligned}
\frac{\sigma_{nl}(u)}{\sigma_{pi}(h\nu)} &= \frac{n_0}{n_e n_p}\frac{4\pi J_\nu}{h\nu}\frac{1}{(1+\bar{n}_\gamma)}\frac{1}{u f_{MB}(u)}\frac{m_e u}{h} \\
&= \frac{g_n}{g_e g_p}\left(\frac{h^2}{2\pi m_e k_B T}\right)^{3/2} e^{h\nu_0/k_B T}\frac{4\pi}{h\nu}\frac{2h\nu^3}{c^2}\frac{1}{(e^{h\nu/k_B T}-1)}\frac{1}{1+\frac{1}{e^{h\nu/k_B T}-1}} \\
&\quad \times \frac{\sqrt{\pi}}{4}\left(\frac{2k_B T}{m_e}\right)^{3/2}\frac{1}{u^2}e^{m_e u^2/2k_B T}\frac{m_e}{h} \\
&= \frac{g_n}{g_e g_p}\frac{2\pi^{3/2}\nu^2 m_e}{hc^2}\left(\frac{h^2}{\pi m_e{}^2}\right)^{3/2}\frac{1}{u^2}\exp\left[\frac{h\nu_0}{k_B T} - \frac{h\nu}{k_B T} + \frac{h(\nu-\nu_0)}{k_B T}\right]\,.
\end{aligned} \tag{5.15}$$

The exponentials all cancel and we are left with the **Milne relation**,

$$\boxed{\frac{\sigma_{\rm nl}(u)}{\sigma_{\rm pi}(h\nu)} = 2\frac{g_{\rm n}}{g_{\rm e}g_{\rm p}}\left(\frac{\nu h}{cm_{\rm e}u}\right)^2} \tag{5.16}$$

This expression is impressively simple! In particular, it does not depend on temperature. For hydrogen, we recall that $g_{\rm n} = 4$ for the ground state and $g_{\rm e} = g_{\rm p} = 2$, so that the g factors cancel. We are left with a dimensionless combination of ν, u, and physical constants.

We are now in a position to solve the radiative ionization and recombination balance for hydrogen. Equation 5.1 gives us an analytical expression for the ionization cross-section at a given frequency, $\sigma_{\rm pi}(\nu)$. Using Equation 5.16, we can infer the corresponding recombination cross-section $\sigma_{\rm nl}(u)$ at the corresponding electron velocity. Given that virtually all neutral atoms are in the ground state, the ionization rate $G_{\rm pi}$ can be found by integrating the radiation field and $\sigma_{\rm pi}$ over energy (Equation 5.4). For recombinations, we similarly integrate over velocity or energy (Equations 5.5 and 5.6), but we need to repeat this procedure for all levels nl and sum the coefficients (Equations 5.8 and 5.9). Given a number density, an ionization rate, and a recombination coefficient, we can use ionization equilibrium (Equation 5.10) to figure out the fraction of atoms that is in a neutral state. Figure 12 shows a flow chart for this process.

6 Photoionized gas and H II regions

When we considered hot gas in §2, the ionization state was simple: we assumed that all hydrogen was ionized at the relevant temperatures. Inside an SN remnant, even higher elements are largely ionized. In contrast, the densest parts of the ISM, molecular clouds, are basically neutral. Unfortunately, not all gas phases present us with such simple ionization states. For example, H I is partially photoionized by stars and cosmic rays, with an ionization fraction that depends on density and other parameters.

A particularly interesting situation arises in so-called H II regions, bubbles around massive, blue stars that ionize the gas that surrounds them. In this section, we consider the ionization state (§6.1–6.3) and thermal balance (§6.4) in H II regions, as well as general observational diagnostics of ionized gas (§6.5) and diffuse ionized gas in the ISM (§6.6).

6.1 Ionization equilibrium around a star

We begin with an example to get a sense of the order of magnitude of some of the quantities introduced in §5. In particular, we consider a pure-hydrogen bubble surrounding a single, hot star (adapted from O&F §2.1). After a very short time, each place in the "nebula" will settle into ionization equilibrium, meaning that the number of ionizations and recombinations will balance. Given the discussion in §5, we need three inputs: the radiation field (for the ionization rate), temperature (for the recombination coefficient), and density (for the ionization balance). In our simple example, we assume that the nebula is optically thin to ionizing radiation. This will hold close to the star where the hydrogen is already ionized, but it will clearly fail close to the edge of the nebula. In the optically thin case, the mean intensity of radiation from the star is

$$4\pi J_\nu = \frac{L_\nu}{4\pi r^2}\,, \tag{6.1}$$

where L_ν is the luminosity of the star at frequency ν. The first 4π factor comes from integrating over all solid angles, the second from the surface of a sphere that is a radius r removed from the star. The luminosity can be calculated from the blackbody flux at the stellar surface,

$$L_\nu = 4\pi R_*^2 \times \pi J_{\rm BB}(\nu, T_*)\,, \tag{6.2}$$

where R_* is the stellar radius and T_* the surface temperature. The second factor of π is geometric in origin and accounts for the difference between the blackbody radiation field (per steradian) and outward flux from the stellar surface (e.g., Appendix 1 of O&F). We can obtain the total luminosity of the star by integrating over frequency, which gives the Stefan-Boltzmann law,

$$L_* = 4\pi R_*^2 \times \pi \int_0^\infty J_{\rm BB}(\nu, T_*){\rm d}\nu = 4\pi R_*^2 \sigma_{\rm SB} T_*^4\,. \tag{6.3}$$

The ionization rate at a distance r is then

$$G_{\rm pi}(r) \equiv \int_{\nu_0}^\infty \frac{4\pi J_\nu}{h\nu}\sigma_{\rm pi}(\nu){\rm d}\nu = \pi\frac{R_*^2}{r^2}\int_{\nu_0}^\infty \frac{J_{\rm BB}(T_*)}{h\nu}\sigma_{\rm pi}(\nu){\rm d}\nu\,. \tag{6.4}$$

Before we dive into exact calculations, let us get a sense of the numbers we can expect. Imagine we are 5 pc from an O star with $T_* \approx 40000$ K and $R_* \approx 10^{12}$ cm. The number of ionizing photons emitted from the star is

$$Q \equiv \int_{\nu_0}^\infty \frac{L_\nu}{h\nu}{\rm d}\nu \approx \frac{10^{49}}{\rm s}\,, \tag{6.5}$$

which we can either compute numerically or look up (see Table 2.3 in O&F or Table 15.1 in Draine). Taking into account the distance from the star, we estimate the intensity term in Equation 5.4,

$$\int_{\nu_0}^{\infty} \frac{4\pi J_\nu}{h\nu} \mathrm{d}\nu = \frac{Q}{4\pi r^2} \approx \frac{3.4 \times 10^9}{\mathrm{cm}^2\ \mathrm{s}}\,. \tag{6.6}$$

As a rough approximation, we take the cross-section to be that at $\nu = \nu_0$, $\sigma_{\mathrm{pi}}(\nu_0) \approx 6 \times 10^{-18}$ cm^2, and take it out of the frequency integral. We recall that this estimate is not too far from the Bohr area estimate (§5.1).[6] We now have an ionization rate per atom of

$$G_{\mathrm{pi}} \approx \frac{Q}{4\pi r^2}\sigma_{\mathrm{pi}} \approx \frac{3.4 \times 10^9}{\mathrm{cm}^2\ \mathrm{s}} \times 6 \times 10^{-18}\ \mathrm{cm}^2 \approx \frac{2 \times 10^{-8}}{\mathrm{s}}\,. \tag{6.7}$$

We can think of the ionization rate as the inverse lifetime of a neutral atom before it gets ionized,

$$t_{\mathrm{pi}} = \frac{1}{G_{\mathrm{pi}}} \approx 1.6\ \mathrm{yr}\,, \tag{6.8}$$

for our example. This time is much longer than the typical lifetime of energy levels (§3.1). We conclude that **virtually all neutral hydrogen in H II regions is in the ground state!** We used this fact when deriving the ionization cross-section. Conversely, ions can recombine to an excited level, which will quickly decay to the ground state.

Finally, we need the recombination coefficient, which we take to be $\alpha_{\mathrm{B}}(T) \approx 4 \times 10^{-13}$ cm^3/s because recombinations into the ground state will quickly lead to another ionization (§5.2). By assuming a constant $\alpha_{\mathrm{B}}(T)$, we are implicitly assuming a constant temperature of a few thousand Kelvin (Table 14.1 in Draine). We now express the densities in Equation 5.10 in terms of the neutral fraction,

$$\xi \equiv \frac{n_0}{n_{\mathrm{H}}} \quad \Longrightarrow \quad n_0 = \xi n_{\mathrm{H}} \quad \text{and} \quad n_{\mathrm{e}} = n_{\mathrm{p}} = (1-\xi) n_{\mathrm{H}}\,, \tag{6.9}$$

which gives a balance equation of

$$\xi\ n_{\mathrm{H}}\ G_{\mathrm{pi}} = (1-\xi)^2\ n_{\mathrm{H}}^2\ \alpha_{\mathrm{B}}(T)\,. \tag{6.10}$$

We try to solve for ξ and assume $n_{\mathrm{H}} = 10/\mathrm{cm}^3$, which gives the implicit equation

$$\frac{\xi}{(1-\xi)^2} = \frac{n_{\mathrm{H}} \alpha_{\mathrm{B}}(T)}{G_{\mathrm{pi}}} \approx 2 \times 10^{-4}\,. \tag{6.11}$$

The only way to solve this equation is if $\xi \ll 1$ so that the denominator is ≈ 1 and $\xi = 2 \times 10^{-4}$. Equation 6.11 makes it clear why the neutral fraction must be low: the ionization rate G_{pi} is larger than α_{B} by about five orders of magnitude. The extra factor of density "helps" the recombination side, but we would need a very high density to balance the frequent photoionizations. The bottom line is, hydrogen is almost completely ionized at a distance of 5 pc from an O star.

[6] The calculation in O&F uses a more exact number computed by integrating over the frequency-dependent cross-section. We use the simpler, approximate expression since we are mostly interested in an order-of-magnitude estimate. As a result, the following numbers differ from O&F by a factor of about 2.

6.2 Strömgren spheres

Of course, the ionized zone must be finite since the star can only emit so many photons. We can understand the nature of the transition zone to neutral gas. The width of this zone is approximately one mean free path of an ionizing photon, where as before (Appendix A.1)

$$\lambda_{\rm mfp} = \frac{1}{n\sigma} = \frac{1}{n_0\sigma_{\rm pi}} = \frac{1}{\xi_{\rm trans} n_{\rm H}\sigma_{\rm pi}} \approx 0.1 \text{ pc} \left(\frac{n}{1/\text{cm}^3}\right)^{-1} , \tag{6.12}$$

where we have assumed that the neutral fraction at the transition is $\xi_{\rm trans} = 1/2$. For the density in our example above, $n_{\rm H} = 10/\text{cm}^3$, we get a width of only 0.01 pc. Clearly, the **transition zone is extremely thin!** This fact was one of the key realizations of Strömgren (1939): we can think of the H II region as an almost totally ionized zone with a sharp boundary.

If we assume that an H II region has constant density inside the bubble and that it consists of only hydrogen, we call the resulting approximate structure a **Strömgren sphere**. Let's estimate the size of this region. We recall that Equation 6.11 is valid only at low ξ, because a significant neutral fraction would mean that the optically thin ($J_\nu \propto 1/r^2$) approximation would break down. However, we can estimate it based on the fact that all ionizing photons from the star will be absorbed within the H II region, by construction (Draine §15.1.1). Assuming that the bubble is isothermal and has a uniform density $n_{\rm H}$, we get

$$Q = \frac{4\pi}{3} R_{\rm S}^3 \alpha_{\rm B}(T) n_{\rm e} n_{\rm p} , \tag{6.13}$$

where $R_{\rm S}$ is the **Strömgren radius**, the radius where the sharp transition to neutrality occurs. By definition, all hydrogen is ionized inside $R_{\rm S}$, so that $n_{\rm e} = n_{\rm p} = n_{\rm H}$. Once again, we choose the Case B recombination coefficient because a recombination to the ground state releases another ionizing photon that will be absorbed (§5.2). We use the approximate formula from Draine Table 14.1,

$$\alpha_{\rm B}(T) \approx 2.56 \times 10^{-13} \frac{\text{cm}^3}{\text{s}} \left(\frac{T}{10^4 \text{ K}}\right)^{-0.83} . \tag{6.14}$$

We solve for the radius,

$$R_{\rm S} = \left(\frac{3Q}{4\pi n_{\rm H}^2 \alpha_{\rm B}(T)}\right)^{1/3} \approx 15 \text{ pc} \left(\frac{Q}{10^{49}/\text{s}}\right)^{1/3} \left(\frac{n_{\rm H}}{10/\text{cm}^3}\right)^{-2/3} \left(\frac{T}{10^4 \text{ K}}\right)^{0.28} . \tag{6.15}$$

The dependencies on stellar brightness and gas temperature are relatively weak, but the density does matter for the size of the Strömgren sphere. For example, choosing $n_{\rm H} = 1/\text{cm}^3$ we could get sizes of more than 100 pc (O&F Table 2.3).

6.3 The radial evolution of ionization

We now return to the (arguably) most interesting question about Strömgren spheres: what is their ionization structure $\xi(r)$? Does the neutral fraction gradually increase with radius? How small and how large is it inside and outside the Strömgren radius?

To answer these questions, we wish to solve the ionization balance equation (5.10) with the correct, radially varying radiation field. Compared to our simple estimates in §6.1, we need to take into account the absorption of ionizing photons along the way, as well as the emission of new photons following recombinations. The radiative transfer equation reads

$$\frac{{\rm d}I_\nu}{{\rm d}s} = -n_0\sigma_{\rm pi}(\nu) I_\nu + j_\nu , \tag{6.16}$$

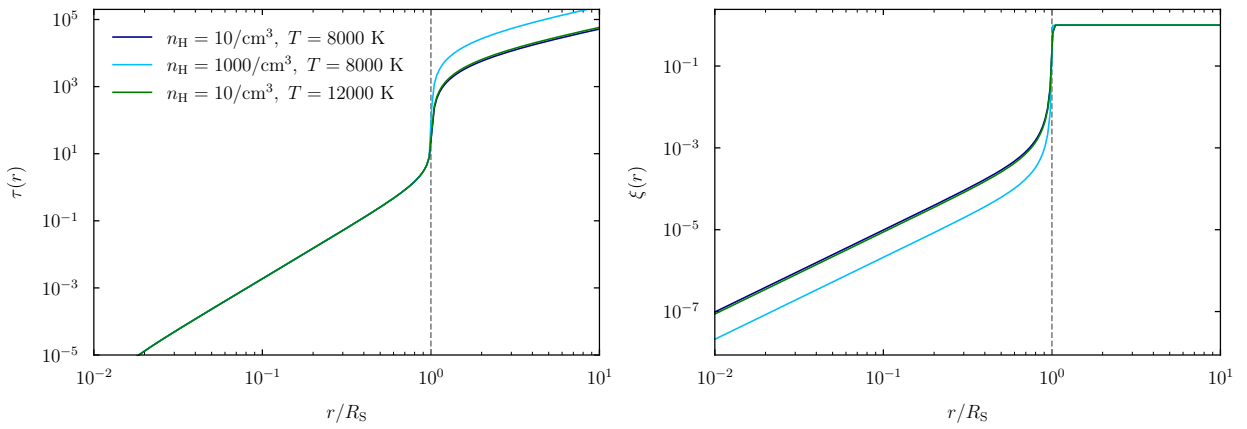


Figure 13: The radial evolution of the optical depth (left) and neutral fraction (right) in Strömgren spheres, computed numerically based on Equations 6.18 and 6.20. The radii are rescaled by the Strömgren radius. We assume a star with radius $R_* = 10^{12}$ cm and $T_* = 4 \times 10^4$ K, roughly corresponding to an O star. Solutions are shown for three different gas densities and temperatures, but those parameters have a relatively minor impact on the solution (once radii are scaled to R_S).

where I_ν is the specific intensity of radiation and j_ν is the "local emission coefficient" (O&F §2.3; J_ν is the average of I_ν over all directions). Since we care only about ionizing radiation, j_ν would depend on the local recombination rate into the ground state and on the velocity distribution of the electrons. However, as we saw in §5.2, we can ignore recombinations into the ground state as long as the emitted photons cause another ionization (Case B). Here, we are treating a spatially varying problem, which forces us to make the slightly stronger **on-the-spot approximation** that the distance between the recombination and subsequent ionization is insignificant compared to the other scales in the problem. This is an excellent approximation in H II regions, and we will thus set $j_\nu = 0$ and use the Case B recombination coefficient. With this simplification, the solution to the radiative transfer equation is clearly

$$I_\nu = I_{\nu,0} e^{-\tau_\nu} \,, \tag{6.17}$$

where τ_ν is the optical depth to ionizing photons of frequency ν that originated from the star,

$$\tau_\nu(r) = \int_0^r n_0(r') \sigma_{\rm pi}(\nu) {\rm d}r' = n_{\rm H} \sigma_{\rm pi}(\nu) \int_0^r \xi(r') {\rm d}r' \,. \tag{6.18}$$

Thus, the balance equation (5.10) becomes

$$n_0(r) \int_{\nu_0}^\infty \frac{4\pi J_\nu(r)}{h\nu} \sigma_{\rm pi}(\nu) {\rm d}\nu = n_0(r) \int_{\nu_0}^\infty \frac{L_\nu}{4\pi r^2 h\nu} e^{-\tau_\nu} \sigma_{\rm pi}(\nu) {\rm d}\nu = n_{\rm e}(r) n_{\rm p}(r) \alpha_{\rm B}(T) \,, \tag{6.19}$$

which we can more elegantly write in terms of the varying neutral fraction $\xi(r)$,

$$\boxed{\frac{\xi}{4\pi r^2} \int_{\nu_0}^\infty \frac{L_\nu e^{-\tau_\nu}}{h\nu} \sigma_{\rm pi}(\nu) {\rm d}\nu = (1-\xi)^2 \; n_{\rm H} \; \alpha_{\rm B}(T)} \tag{6.20}$$

This equation cannot be solved analytically. Draine presents an approximate calculation (§15.3), but it cannot cannot describe the shape of $\xi(r)$ near the transition to neutrality. Thus, we need to integrate the coupled differential equations 6.18 and 6.20. In this integration, we can treat $n_{\rm H}$ and T as constants or let them evolve based on more complex models for the density and temperature structure of the H II region. Figure 13 shows the result, which confirms that the transition to neutrality at the edge of the Strömgren sphere is very sharp.

6.4 Thermal balance

What is the temperature structure inside an H II region? There are a number of heating and cooling processes acting on the gas. We will consider them briefly without going into great mathematical detail, which can be found in Draine §27 and O&F §3. We start from the Lagrangian internal energy equation (Appendix A.1),

$$\frac{\mathrm{D}\varepsilon}{\mathrm{D}t} = -\frac{P}{\rho}\boldsymbol{\nabla}\cdot\boldsymbol{u} + \frac{\Gamma}{\rho} - \frac{\Lambda}{\rho}\,. \tag{6.21}$$

Since we are specifically interested in the evolution of temperature, we can cast this equation into a more suggestive form. We assume that all internal energy is in the form of thermal energy ($\gamma = 5/3$),

$$\varepsilon = \varepsilon_{\mathrm{th}} = \frac{3}{2}\frac{k_{\mathrm{B}}T}{m_{\mathrm{ptl}}}\,, \tag{6.22}$$

where m_{ptl} is the mass of the particles in question. We multiply the equation by ρ and use the Lagrangian continuity equation,

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = -\rho\boldsymbol{\nabla}\cdot\boldsymbol{u}\,, \tag{6.23}$$

to get rid of the divergence term,

$$\frac{3\rho k_{\mathrm{B}}}{2m_{\mathrm{ptl}}}\frac{\mathrm{D}T}{\mathrm{D}t} = \frac{P}{\rho}\frac{\mathrm{D}\rho}{\mathrm{D}t} + \Gamma - \Lambda = \frac{nk_{\mathrm{B}}T}{m_{\mathrm{ptl}}n}m_{\mathrm{ptl}}\frac{\mathrm{D}n}{\mathrm{D}t} + \Gamma - \Lambda\,. \tag{6.24}$$

We cancel a number of terms to get a temperature evolution equation,

$$\boxed{\frac{3}{2}nk_{\mathrm{B}}\frac{\mathrm{D}T}{\mathrm{D}t} = k_{\mathrm{B}}T\frac{\mathrm{D}n}{\mathrm{D}t} + \Gamma - \Lambda} \tag{6.25}$$

As expected, the thermal energy per unit volume is changed directly by the heating and cooling rates, and adiabatic expansion or contraction (which we do not care about here).

There are a few main contributors to heating and cooling. First, we assume that the nebula is optically thin to all non-ionizing radiation, which may excite hydrogen atoms (which then decay back to the ground state) but not necessarily contribute to the temperature of the gas. Ionizing radiation, however, gives an initial velocity to each freed electron. The resulting heating rate reads similar to the ionization rate, but adding the excess energy beyond the ionization threshold into the integral,

$$\Gamma_{\mathrm{pi}} = n_0\int_{\nu_0}^{\infty}\frac{4\pi J_\nu}{h\nu}\sigma_{\mathrm{pi}}(\nu)h(\nu-\nu_0)\mathrm{d}\nu\,. \tag{6.26}$$

We can physically interpret this expression by dividing by the ionization balance equation, $n_0 G_{\mathrm{pi}} = n_{\mathrm{e}}n_{\mathrm{p}}\alpha_{\mathrm{B}}(T)$, to get

$$\Gamma_{\mathrm{pi}} = n_{\mathrm{e}}n_{\mathrm{p}}\alpha_{\mathrm{B}}(T)\frac{\int_{\nu_0}^{\infty}\frac{4\pi J_\nu}{h\nu}\sigma_{\mathrm{pi}}(\nu)h(\nu-\nu_0)\mathrm{d}\nu}{\int_{\nu_0}^{\infty}\frac{4\pi J_\nu}{h\nu}\sigma_{\mathrm{pi}}(\nu)\mathrm{d}\nu} \equiv n_{\mathrm{e}}n_{\mathrm{p}}\alpha_{\mathrm{B}}(T)\frac{3}{2}k_{\mathrm{B}}T_{\mathrm{i}}\,, \tag{6.27}$$

where T_{i} is "ionization temperature" of the electrons freed from the atoms. It can be shown that T_{i} is approximately T_* (O&F §3.2), the temperature of the central star. If photoionization heating was totally dominant, the gas would approach a temperature of T_{i}.

There are also a number of cooling processes acting. First, when electrons recombine, their kinetic energy is lost from the fluid. We can quantify this loss with a similar argument as for

ionization heating, namely, by integrating the recombination cross-section $\sigma_{\rm nl}(u)$ over the velocity distribution function. The resulting expression is similar to the recombination coefficient in Equation 5.6,

$$\Lambda_{\rm rec} = n_{\rm e} n_{\rm p} \sum_{n=1}^{\infty} \sum_{l=0}^{n-1} \int_0^{\infty} u \sigma_{\rm nl}(u) \frac{mu^2}{2} f_{\rm MB}(u) {\rm d}u \,. \tag{6.28}$$

As in §5.2, we had to some over all levels nl (O&F §3.3 and Draine §27.3.1). Besides photoionization, gas in H II regions is cooled by the same processes as hot ionized gas: Bremsstrahlung and collisionally excited line radiation. Bremsstrahlung (or free-free emission) scales roughly as the root of temperature,

$$\Lambda_{\rm ff} \approx 1.8 \times 10^{-23} \, \frac{\rm erg}{\rm cm^3 \ s} \left(\frac{T}{10^4 \ {\rm K}} \right)^{1/2} \left(\frac{n_{\rm e}}{10/{\rm cm}^3} \right)^2 , \tag{6.29}$$

where we have assumed pure hydrogen and neglected complications due to Gaunt factors and such (see O&F §3.4 and Draine §10.3).

The final cooling process, line emission, relies on metals because the temperature is generally below 10000 K. In this regime, the collisional excitation of hydrogen (and helium) is inefficient (Draine §34.1), but ions such as O II, O III, and N II have lower-lying energy levels that can be excited. The cooling function for line emission, $\Lambda_{\rm ce}$, can more or less be worked out from first principles (O&F §3.5), at least for the lines of single-electron systems (see §2.2).

In thermal equilibrium, ${\rm D}T/{\rm D}t = 0$, and the temperature evolution Equation 6.25 simply reads $\Gamma - \Lambda = 0$ (neglecting adiabatic changes). Thus, all heating and cooling terms must balance,

$$\Gamma_{\rm pi} = \Lambda_{\rm rec} + \Lambda_{\rm ff} + \Lambda_{\rm ce} \,. \tag{6.30}$$

We can determine the equilibrium temperature of a system if we know its heating and cooling terms. In H II regions, the line emission term tends to dominate over recombination cooling and the even weaker free-free emission (Draine Figure 27.1). The higher the metallicity, the stronger line cooling becomes while the heating rate remains the same. Thus, higher metallicity shifts the equilibrium to a lower temperature (Draine Figure 27.2).

6.5 Observational diagnostics

How can we actually observe H II regions, or photoionized gas in general? The two quantities we are most interested in are density and temperature. In principle, we can use emission or absorption by any atom or ion that is sufficiently abundant to be observable. However, observing a single line cannot tell us about both n and T, and the strengths of lines from different species obviously depend on their respective abundances. Thus, we are ideally looking for systems of multiple lines in the same ion whose ratios depend on density and/or temperature.

Let us first consider the most abundant element, hydrogen. As we saw previously, recombinations into the $n = 1$ ground state produce ionizing photons that are almost immediately re-absorbed, meaning that those "Lyman-continuum photons" are not observable. For the lower Lyman-series lines, e.g., the $2 \rightarrow 1$ Lyman-α transition, the opacities are still very large (between 10 and 10^4 for typical H II regions) because those photons can easily find hydrogen in the ground state to excite. Thus, Lyman photons are scattered around until they reach the edge of the H II region. Balmer photons, those originating from transitions in the $n = 2$ shell, can escape. Since recombinations can happen into any shell, there will be cascades of excited states that eventually lead back to the ground state. It can be shown that about half of those cascades include a $3 \rightarrow 2$ (Balmer-α or Hα) transition. Thus, **Hα is a prominent tracer of the total ionizing flux.**

We can find the temperature by considering two lines of the same ion. In particular, we need two lines above the ground state that are both energetically accessible at the given temperature (i.e., whose potential is not so large that they are exceedingly rare), and whose energy level differs sufficiently so that their population is sensitive to the temperature (Equation 3.8). In such a case, it can be shown that the ratio of the line intensities depends only on the temperature and on known atomic physics such as the energy levels E_{21}, Einstein coefficients A_{21}, and collision strengths Υ_{12} (Draine §18.1). In typical H II regions, these circumstances are given for N II and O III.

A similar logic can be applied to find density diagnostics. Imagine an ion with a singlet ground state and an excited state that is split into multiple, very similar energy levels. If the density is very small, the population in the excited state is negligible and every excitation is more or less immediately followed by a decay back into the ground state. Thus, the ratio of the line intensities approaches a constant that depends on atomic physics only. In the high-density limit, the populations take on their thermal equilibrium values, but since their excitation energy is so similar, the exponential term in Equation 3.8 essentially cancels. Once again, the line ratio is determined by atomic physics, but with different coefficients that now depend on the degeneracies of the states. Since the low- and high-density limits are generally different, we can use the line ratio to infer intermediate densities, or at least lower or upper limits (Draine Figure 18.4).

Finally, we can use ratios of line intensities from different species to infer relative abundances. See Draine §18 and O&F §5 for more detail on observational diagnostics.

6.6 Diffuse ionized gas

H II regions are not the only places in the Galaxy where photoionized gas is found. In fact, up to 90% of the photoionized gas might be in a diffuse phase called the **Warm Ionized Medium (WIM)**, which has a temperature of about 8000 K and a relatively low density around $0.1/\mathrm{cm}^3$, albeit with large scatter around these numbers (Haffner et al. 2009). While the WIM makes up for a negligible fraction of the total gas in the Galaxy, its filling fraction might be as high as 0.1–0.3 because it has a larger scale height than the cooler components. This larger scale height might indicate that the WIM has slightly lower pressure than the other ISM components. In Figure 1, we assigned the WIM a density of $n_\mathrm{H} = 0.2/\mathrm{cm}^3$ to highlight pressure equilibrium, but it is possible that the WIM has lower pressure and is pushed out of the disk.

The WIM is thought to be partially ionized, mostly by leakage of ionizing radiation from the same O stars that create H II regions. In a more realistic picture with strongly inhomogeneous density structures around those stars, some photons will find channels low-density channels to escape through. Additional ionizing radiation is provided by cosmic rays and X-rays. Our knowledge about the WIM is very much evolving (see, e.g., the review by Haffner et al. 2009).

7 Atomic physics III: Molecules

Molecules are collections of atoms that are bound by electrostatic forces or covalent bonds (the sharing of electrons). These bonds are often of comparable strength to the excitation levels of atoms. However, many molecules are much easier to dissociate radiatively because they tend to have a broad spectrum of excited states that can decay in destructive ways. Thus, molecules are virtually absent from the phases of the ISM that we have studied thus far. In the most dense regions of the ISM, however, the column densities can get high enough to effectively shield gas from stellar radiation fields. It is in these dense clouds that molecules form abundantly.

In this chapter, we aim to understand the energy levels, production, and destruction of molecules. We begin by briefly introducing molecular energy levels in §7.1. As with the atoms they are composed of, the quantized energy levels of molecules largely determine their physical behavior. We then take a closer look at two of the most important molecules for astrophysics, H_2 (§7.2) and CO (§7.3).

7.1 Vibrational and rotational energy levels

For this discussion, we restrict ourselves to diatomic molecules, which are relatively simple due to their rotational symmetry around the internuclear axis. For a more detailed treatment, see Draine §5. Fundamentally, diatomic molecules have energy levels that correspond to vibration along the internuclear axis and to rotation around it. As with atoms, those energy levels are further split when we consider the relationship between the rotation of the molecule and the angular momentum of the atoms' electrons, and hyperfine-split if we additionally consider the interaction of the magnetic fields that are generated. Molecular energy levels can be described by a notation similar to that for the energy levels of atoms (§3.1, Draine §5.1.3).

To first order, we can approximate the vibrational and rotational energies in a classical sense, as a harmonic oscillator and as solid-body rotation. The electric potential energy of the two nuclei is approximately

$$V(r_\mathrm{n}) = E_0(r_\mathrm{n}) + Z_1 Z_2 \frac{q_\mathrm{e}^2}{r_\mathrm{n}}\,, \tag{7.1}$$

where r_n is the distance between the nuclei, Z_1 and Z_2 are their changes in units of the electron charge q_e, and $E_0(r_\mathrm{n})$ is an energy that depends on the quantum mechanical state of the two atoms and their electrons. In the molecular ground state (with no vibrations), the atoms are separated by some radius r_0 where the potential is minimized. We can now imagine the vibrations as a harmonic oscillator with a reduced mass

$$m_\mathrm{r} \equiv \frac{m_1 m_2}{m_1 + m_2}\,, \tag{7.2}$$

oscillating around the point $r = r_0$. We do not know the exact shape of $V(r)$ because it depends on forces other than the electrostatic repulsion between the nuclei, but we know that we can expand it around its minimum,

$$V(r) \approx V(r_0) + \frac{1}{2}k(r - r_0)^2\,, \qquad k \equiv \frac{\mathrm{d}^2 V}{\mathrm{d}r^2}\,, \tag{7.3}$$

where k is the spring constant (think Hooke's law). The stronger the molecular bond, the larger k. The fundamental angular oscillator frequency is, classically,

$$\omega_0 = \sqrt{\frac{k}{m_\mathrm{r}}}\,. \tag{7.4}$$

Larger reduced masses lead to a faster vibration frequency. The reduced mass is smaller for more uneven molecules, at fixed total mass. We recall that a QM oscillator cannot be in a state of zero energy, but instead has a ground state energy of $\hbar\omega_0/2$. Higher oscillations are described by a quantum number v, with an energy of $\hbar\omega_0(v+1/2)$.

We now turn to the rotational energy. Classically, an angular momentum of $L = J\hbar$ corresponds to a rotational energy $L^2/2I$, where I is the moment of inertia. We approximate the latter as $I \approx m_\mathrm{r} r_0^2$, and we replace J^2 by $J(J+1)$ to account for quantum-mechanical effects. In total, we thus have an energy of

$$\boxed{E(v,J) = V(r_0) + \hbar\omega_0\left(v+\frac{1}{2}\right) + \frac{\hbar^2}{2m_\mathrm{r}r_0^2}J(J+1) \equiv V(r_0) + h\nu_0\left(v+\frac{1}{2}\right) + B_\mathrm{v}J(J+1)} \quad (7.5)$$

The "fundamental vibrational energy" is now written as $h\nu_0$ (with $\nu_0 = \omega_0/2\pi$). The rotational energy depends on the vibrational state and is thus written as B_v, but purely rotational transitions $J \to J-1$ correspond to an energy of $h\nu = 2B_\mathrm{v}J$. Our treatment has been highly simplified, but it makes a basic prediction for the energy levels of molecules as a function of their nuclear properties.

Another important question is how the excited levels of molecules couple to photons, which will determine their Einstein A_{21} coefficients. For a purely rotational $J \to J-1$ transition that corresponds to a photon energy $h\nu$, this coefficient is

$$A_{J\to J-1} = \frac{\nu^3}{6h\pi^2c^3}\mu^2\frac{J}{J+1/2} \propto B_0^3\mu^2\frac{J^4}{J+1/2}\,, \quad (7.6)$$

where μ is the electric dipole moment of the molecule and B_0 is the rotation constant in the vibrational ground state (Draine §5.1.7). The main takeaway is that the transition rate strongly depends on the electric dipole: a molecule with a strong dipole finds it easier to couple to photons and to transition to a lower-J state.

7.2 Molecular hydrogen (H_2)

H_2 plays a prominent role as the by far most abundant molecule in the Universe. Its binding energy is about 4.5 eV (Draine §31.2). Molecular hydrogen comes in two states: if the proton spins are anti-aligned, we have para-H_2, and if they are aligned, we have ortho-H_2. For non-obvious quantum-mechanical symmetry reasons, the spin alignment excludes certain rotation states so that para-H_2 has $J = 0, 2, 4, ...$ whereas ortho-H_2 has $J = 1, 3, 5, ...$ (Draine §5.1.6).

The base vibrational wavelength of H_2 is $\lambda = 2.1\times10^{-4}$ cm, which corresponds to a temperature of $T = hc/\lambda k_\mathrm{B} \approx 6800$ K. Clearly, vibrations are not easy to excite thermally, given that molecules exist in cold environments! The rotational energy levels are determined by the reduced mass, $m_\mathrm{r} = m_\mathrm{p}/2$ (which is maximal for a symmetric molecule), and by the nucleon separation of about 0.75 Å in the vibrational ground state. We can compute the corresponding excitation temperature,

$$T = \frac{h\nu}{k_\mathrm{B}} = \frac{2B_0}{k_\mathrm{B}} = \frac{\hbar^2}{k_\mathrm{B}m_\mathrm{r}r_0^2} = \frac{2\hbar^2}{k_\mathrm{B}m_\mathrm{p}r_0^2} \approx 170\ \mathrm{K}\,. \quad (7.7)$$

While the rotational levels are more accessible than the vibrational ones, they are still too high to be thermally excited in cold molecular gas with $T \approx 20$ K (§8.1). Moreover, the transition probability from a rotationally excited state via a photon is, to first order, zero because the molecular has no permanent dipole, $\mu = 0$ (Equation 7.6). This is the case for all symmetric diatomic molecules, but it makes the life of the astronomer rather hard: there is no practical way to directly observe

the most common molecule in the Universe! In the next section, however, we will explore CO as a tracer of molecular hydrogen, and molecular gas in general.

As with H II and H I in the previous chapters, the abundance of H_2 is governed by a balance between creation and destruction. Molecules can form through a number of processes, including neutral-neutral and ion-neutral interactions. Naively, we might expect the dominant process to be the simplest one,

$$H + H \rightarrow H_2 + h\nu\,, \tag{7.8}$$

given that the H_2 molecule has a binding energy of about 4.5 eV (§7.2). Due to the lack of an electric dipole, however, the coupling to EM forces is weak and this process is so unlikely that it can be ignored entirely. A more likely mechanism is an ion-neutral process called "radiative association,"

$$H^- + H \rightarrow H_2 + e^- + \text{ kinetic energy}\,, \tag{7.9}$$

but the rate is still quite low. A more productive way to make molecules is provided by dust grains via so-called **grain catalysis**. Atoms settle on the surfaces of dust grains, execute a random walk, and may eventually meet another atom. The release of the binding energy propels the newly formed molecule off the grain surface. Calculating the likelihood of this process is complicated due to the diverse (and poorly understood) nature of dust grains, but the rate is roughly

$$R_{\rm gr} \approx 7.3 \times 10^{-17}\ \frac{\rm cm^3}{\rm s} \left(\frac{T}{100\ \rm K}\right)^{1/2} \left(\frac{\Sigma_{\rm gr}}{10^{-21}/\rm cm^2}\right)\,, \tag{7.10}$$

where $\Sigma_{\rm gr}$ is the total dust cross-section per H atom (integrated over dust grains of all sizes, and taking into account the relative dust abundance; Draine §31.2). This rate needs to be multiplied by $n_{\rm H} n_0$ to give the rate of molecule formations per time.

On the opposite side of the balance, we have a number of ways to destroy molecules: photoionization, photodissociation, and incorporation into other molecular species. The most important of these mechanisms is **photodissociation**, which happens when the molecule absorbs a photon that has a frequency corresponding to one of its energy levels (§7.1). The numerous vibrational and rotational levels of H_2 (Draine Figure 5.2) mean that there is an entire band in which the molecule is more or less excitable, the **Lyman-Werner band** ($912 - 1100$ Å). By symmetry, the excited state must also be able to decay via a one-photon transition, which means the decay happens after a relatively short lifetime. In 85% of cases, the H_2 molecule decays back to the ground state and emits another Lyman-Werner photon, but in 15% of the cases, the decay happens into the "vibrational continuum," meaning that a strong vibration leads to the molecule's constituent atoms flying apart. Thus, the Lyman-Werner band corresponds to energies between the lowest excited levels that can destroy the molecule upon decay and the ionization energy of hydrogen. The strength of the Lyman-Werner (or "FUV") radiation field is a key parameter in the atomic-molecular balance. It can be approximated as linear a linear radiation spectrum near 1000 Å, where the photodissociation of H_2 is most effective,

$$\nu U_\nu \equiv \chi \times 4 \times 10^{-14} \frac{\rm erg}{\rm cm^3} \left(\frac{\lambda}{1000\ \text{Å}}\right)\,, \tag{7.11}$$

where χ is inversely defined as

$$\chi \equiv \frac{(\nu U_\nu)_{1000\ \text{Å}}}{4 \times 10^{-14}\ \rm erg/cm^3}\,. \tag{7.12}$$

The measured value of χ is about unity at the solar system's position (Habing 1968; Draine 1978), but we do not know how representative this value is of the rest of the galaxy since it may strongly

depend on nearby, massive stars. Parameterizing this uncertainty with the χ parameter, we have a photodissociation rate per atom of about

$$\zeta_{\rm diss} \approx \chi \frac{4 \times 10^{-11}}{\rm s} \approx \frac{5 \times 10^{-11}}{\rm s} \,, \tag{7.13}$$

where the latter estimate refers to the local FUV field. We can now write down an equilibrium between H_2 formation and photodissociation,

$$n_2 \zeta_{\rm diss}(\chi) = n_{\rm H} n_0 R_{\rm gr} \,, \tag{7.14}$$

where we use the notation $n_2 = n(H_2)$. Using $\chi \approx 1$ and the grain catalysis rate quoted above, we get

$$\frac{n_2}{n_{\rm H}} \approx 2 \times 10^{-5} \left(\frac{n_{\rm H}}{30/{\rm cm}^3} \right) \,, \tag{7.15}$$

meaning that there would be barely any H_2 in equilibrium! Higher density helps, but not enough to counteract the destructive FUV radiation field. For gas to be molecule-dominated, it needs to be **self-shielded**, a fancy term for there being a sufficient column density in the direction of any radiation source to make a cloud optically think to Lyman-Werner radiation (Draine §31.4).

The picture that arises is one of dense, molecule-dominated clouds that are surrounded by warmer H I, which shields them from radiation. The clouds often contain significant amounts of metals and dust, which aids in the production of H_2 and larger, more complicated molecules (Draine §33). These reaction networks can get very complicated (e.g., Hollenbach & Tielens 1999). For example, H II regions typically expand into molecular clouds that formed the O stars in the first place. At the edge of their Strömgren spheres, they form a **photodissociation region (PDR)**, where molecules are progressively destroyed by UV radiation penetrating into the cloud (Draine §31.7). The structure of these regions can be solved numerically given a composition, radiation field, and so on.

In summary, the H I/H_2 balance depends mainly on three variables: the column density of gas to sources of radiation, $N_{\rm H}$, the strength of the Lyman-Werner radiation field, χ, and the abundance of dust, which is strongly correlated with the metallicity. A number of models have been put forward to describe the fraction of gas in molecules as a function of these variables (e.g., Krumholz 2013).

7.3 Carbon monoxide (CO)

Given that we cannot observe H_2, we will look at alternative observational tracers in §8.2. The most prominent such tracer is carbon monoxide (CO). This molecule's reduced mass is

$$m_{\rm r} = \frac{12 \times 16}{12 + 16} m_{\rm p} \approx 6.9 \; m_{\rm p} \,. \tag{7.16}$$

This larger mass reduces both the vibrational and rotational energies. The fundamental vibrational energy is only about a factor of 2 lower than for H_2 due to the strong covalent bond in CO (and the resulting strong spring constant in Equation 7.3). However, the fundamental rotational energy is reduced by $1/m_{\rm r}^2$, and even though the separation r_0 is slightly larger than for H_2, we get a rotational temperature of $T \approx 5.5$ K. Thus, multiple rotational energy levels of CO can be thermally excited even in the coldest clouds. Particularly, the $J = 1 \rightarrow 0$ line has a frequency of 115 GHz and a corresponding wavelength of $\lambda = 2.6$ mm, which lies in the far-IR or microwave band (Draine §19.3). The transition rate is relatively low due to a small dipole moment,

$$A_{J \rightarrow J-1} \approx \frac{1.07 \times 10^{-7}}{\rm s} \frac{J^4}{J + 1/2} \quad \Longrightarrow \quad A_{J=1 \rightarrow 0} \approx \frac{7.16 \times 10^{-8}}{\rm s} \,, \tag{7.17}$$

meaning that the $J = 1$ state of CO has a lifetime of about half a year (Draine §5.1.7). We consider the CO emission from astrophysical clouds in §8.2.

8 Molecular gas and giant molecular clouds

At the densities where molecules can form in significant quantities, the self-gravity of gas clouds starts to be important. Thus, molecular gas in the ISM is not a phase that permeates spaces at a more or less fixed pressure and density, but it instead occurs in so-called molecular clouds. In this section, we consider the general properties of molecular clouds (§8.1) and the observational tracers that allow us to infer those properties (§8.2). Finally, we consider particles even larger than molecules: dust grains, whose selective absorption of light is a key factor in observational astronomy (§8.3). A much more detailed treatment of these topics can be found in Draine §5, §19, §21–26, and §31–33. See Schinnerer & Leroy (2024) for an up-to-date review.

8.1 Properties of molecular clouds

The term "molecular cloud" is surprisingly poorly defined. Clouds are generally agglomerations of molecular and other gas, but they can be diffuse, translucent, or opaque (Draine §32.1). They are often called "dark clouds" because of their dust content (§8.3). Many clouds **form stars**, but that is not a requirement. Most clouds are thought to be **gravitationally bound**, but we will see that even that is not entirely certain in many cases. The term **Giant Molecular Cloud (GMC)** is used to describe large clouds, with masses of $10^3 - 2 \times 10^5\ M_\odot$ or so. Some clouds are collected in larger, gravitationally bound complexes (e.g., Orion with a mass of $3 \times 10^5 M_\odot$). The sizes clouds and GMCs are also diverse, ranging from a fraction of a pc to 10s of pc. The gas in clouds takes on temperatures between ≈ 8 K in the densest "cores" to ≈ 100 K in the surrounding CNM (§4). Clearly, molecular gas occurs in many shapes, densities, and temperatures (see, e.g., Bergin & Tafalla 2007, for a review).

In this section, we try to get a sense of the most important properties of molecular clouds (hereafter simply called clouds). Comparing them to the other ISM phases, we find that they are somewhat overpressurized. For example, the CNM phase that is thought to surround cold clouds has a pressure of about

$$P_{\rm CNM}/k_{\rm B} \approx 100\ {\rm K} \times 40/{\rm cm}^3 \times 1.1 = 4400\ {\rm K/cm}^3\,, \tag{8.1}$$

whereas a typical cloud might have

$$P_{\rm cloud}/k_{\rm B} \approx 10\ {\rm K} \times 1000/{\rm cm}^3 = 10000\ {\rm K/cm}^3 \tag{8.2}$$

or even more, given that the chosen values of density and temperature are relatively low. The larger pressure indicates that clouds need to be gravitationally bound, given that they do not seem to be expanding into the surrounding medium. However, turbulence may play an important role (§9). To see why, we consider the famous **size-linewidth relation** that was observationally found by Larson (1981),

$$\boxed{\sigma_{\rm u} \approx 1.1\ \frac{\rm km}{\rm s} \left(\frac{L}{1\ {\rm pc}}\right)^{\beta}} \tag{8.3}$$

where $\beta \approx 0.38$,[7] though newer investigations have found a range of different values (Draine §32.9). This relation is valid between approximately $0.1 \leq L \leq 100$ pc. This velocity dispersion is large compared to the sound speed in molecular gas,

$$c_{\rm s} = \sqrt{\frac{\gamma P}{\rho}} = \sqrt{\frac{\gamma k_{\rm B} T}{\mu m_{\rm p}}} \approx 0.24\ \frac{\rm km}{\rm s} \left(\frac{T}{10\ {\rm K}}\right)^{1/2}\,. \tag{8.4}$$

[7] The exponent is commonly denoted as γ, but we change this notation to avoid confusion with the adiabatic index of an ideal gas.

We have assumed $\mu = 2.36$ (§1.2) and $\gamma = 5/3$ because the gas is too cold for the rotational degrees of freedom of H_2 to be excited. The sound speed is slower than the velocity dispersions of all but the smallest clouds, meaning that the **turbulent motions are supersonic!**

We can infer a cloud's mass by assuming that it is in **virial equilibrium**. Broadly speaking, if a system is in virial equilibrium, we expect its kinetic energy to be half its potential energy, $2E_{\rm kin} = E_{\rm pot}$. We imagine a molecular cloud to be a sphere of uniform density ρ, mass $M_{\rm cl}$, diameter L, and velocity dispersion $\sigma_{\rm u}$. The gravitational potential energy of a self-gravitating sphere can be found by integrating the potential from each shell,

$$E_{\rm pot} = \int_0^R \frac{GM(<r)}{r} {\rm d}M = G\int_0^R \frac{4\pi\rho r^3}{3}\frac{4\pi r^2\rho}{r}{\rm d}r = 3G\left(\frac{4\pi\rho}{3}\right)^2\frac{R^5}{5} = \frac{3GM_{\rm cl}^2}{5R} = \frac{6GM_{\rm cl}^2}{5L} \tag{8.5}$$

since R is the radius and L is the diameter. The kinetic energy due to the gas motions is $M_{\rm cl}\sigma_{\rm u}^2/2$, so that we have

$$2 \times \frac{M_{\rm cl}\sigma_{\rm u}^2}{2} = \frac{6GM_{\rm cl}^2}{5L}\,. \tag{8.6}$$

We solve this equation for the **virial cloud mass**,

$$\boxed{M_{\rm cl} \approx \frac{5\sigma_{\rm u}^2 L}{6G} \approx 230\ M_\odot \left(\frac{L}{1\ {\rm pc}}\right)^{2\beta+1}} \tag{8.7}$$

where we have used Larson's relation to substitute L for $\sigma_{\rm u}$. Similarly, we can solve for the density,

$$\boxed{n_{\rm cl} \approx \frac{1.3\times10^4}{{\rm cm}^3}\left(\frac{L}{1\ {\rm pc}}\right)^{2\beta-2}} \tag{8.8}$$

or column density,

$$\boxed{N_{\rm cl} = n_{\rm cl}L \approx \frac{4\times10^{22}}{{\rm cm}^2}\left(\frac{L}{1\ {\rm pc}}\right)^{2\beta-1}} \tag{8.9}$$

These relations are approximate but roughly describe observed clouds to a factor of 2 or so. A final, fairly fundamental question we might ask about molecular clouds is how long-lived they are. We can estimate their lifetime by counting the star formation rates in clouds across a range of evolutionary stages, as well as the ages of star clusters without molecular gas. From such estimates, we find that clouds tend to be short-lived with **lifetimes of about 3 to 20 million years**. Their destruction is almost certainly determined by feedback processes to some extent, and turbulence may also create and disperse clouds (§9). Either way, the cloud lifetime is so short that some chemical processes will not have reached equilibrium yet, making molecular chemistry somewhat time-dependent in addition to all its other complexities.

8.2 Observational tracers of molecular gas

In §7.2, we already discussed why H_2 is essentially unobservable: its rotational energy levels are too high ($T > 170$ K) and it has no dipole that would efficiently couple its excited states to radiation. Thus, we must use other molecules whose abundance (hopefully) correlates well with that of H_2. Another alternative is to assume a dust-H_2 connection and measure the continuum radiation from dust (§8.3).

By far the most commonly used tracer is CO due to its low excitation temperature and relative brightness. In particular, the $J = 1 \rightarrow 0$ and $J = 3 \rightarrow 2$ lines are popular tracers, not least

because their mm wavelengths fall just to the blue side of the atmospheric radio window, making them observable from the ground. On the other hand, observing H_2 via CO introduces a number of serious complications. First, the **CO line is almost always optically thick**. The attenuation coefficient can be computed along the same lines as for H I (Equation 4.12). Multiplying with the size of a cloud, we obtain an optical depth of about 50 for $n \approx 10^3/\mathrm{cm}^3$ and a cloud size of 3 pc (with some additional assumptions; Draine §19.3.1). How can we use an optically thick line to estimate the total molecular mass? The key insight is that the velocity dispersion of clouds scales with their mass (Equation 8.3). Thus, much of the gas in a cloud does not contribute to absorption at the wavelength where it was emitted, and the total emission in the broadened line correlates well with the total CO mass. The CO-H_2 conversion factor can be derived analytically for a spherical cloud in virial equilibrium with constant density, constant CO abundance, and so on (Draine §19.4–19.6),

$$N_{\mathrm{H}_2} = X_{\mathrm{CO}} I_{\mathrm{CO}}\,, \quad X_{\mathrm{CO}} \approx 2 \times 10^{-20} \frac{1}{\mathrm{cm}^2} \frac{1}{\mathrm{K\ km/s}}\,, \tag{8.10}$$

where I_{CO} is the intensity of the $J = 1 \to 0$ (2.6 mm) line, expressed as an integral of the antenna temperature (equivalent blackbody temperature) over the line profile broadened by velocity. We can convert this factor to a surface density in astronomical units,

$$\Sigma_{\mathrm{H}_2} = \alpha_{\mathrm{CO}} I_{\mathrm{CO}}\,, \quad \alpha_{\mathrm{CO}} \approx 4.3\ \frac{M_\odot}{\mathrm{pc}^2} \frac{1}{\mathrm{K\ km/s}}\,. \tag{8.11}$$

These factors are uncertain by about ±30% (Bolatto et al. 2013). Surprisingly, the conversion factor is approximately independent of the CO abundance! This is a positive side effect of the optical thickness (Draine §19.6).

To test the conversion factor observationally, we can infer the total masses of clouds from virial equilibrium and compare to their total CO luminosity. Such investigations generally find a good correlation, albeit with large scatter (Bolatto et al. 2013). One complicating factor is metallicity: CO lives towards the center of molecular clouds, at even higher densities and lower temperatures than H_2. The relative sizes of these regions, and thus the abundances of CO and H_2, depend on the metallicity of the gas. This effect leads to a variable X_{CO} factor and even to "dark molecular gas," i.e., H_2 that lacks corresponding CO emission (Wolfire et al. 2010).

8.3 Dust

We have already alluded to dust, for example in the context of dark clouds or grain catalysis. While dust is a negligible fraction of the ISM by mass, it has an outsize impact because it is very effective at absorbing (or scattering) light over a large range of frequencies and reemitting it in the infrared. While we understand the basic principles by which dust operates, much remains mysterious about it, most notably its chemical composition. In this section, we present a cursory glance at some of these topics. For a much more detailed treatment, see Draine §21–26.

Dust is a catch-all term for tiny, solid particles with a distribution of sizes between roughly 10^{-6} cm and 10^{-4} cm, though these numbers are uncertain (Draine §21.5). The **composition** of dust is also largely a mystery. We can glean some insight from absorption spectra, which show that certain elements are suppressed compared to solar abundances in sightlines through the local ISM. These elements, most notably C, Mg, Si, and Fe, might be missing because they have been integrated into dust grains (Draine §23.1). Combined with other considerations, the abundance pattern indicate that the most promising candidates are **silicates and graphite-like** substances. The former include silicon-based molecules that we know as crystals on Earth, e.g. pyroxenes such as $MgSiO_3$ or $FeSiO_3$ and olivines such as Mg_2SiO_4 or $MgFeSiO_4$. In the latter category, we have

pure-carbon crystals such as graphite or diamond and so-called "polycyclic aromatic carbohydrates" (PAH), large molecules made of carbon rings. Additionally, oxides (e.g., SiO_2) and carbides (e.g., SiC) may contribute to dust (Draine §23.1).

The most important feature of dust grains is that they have a large cross-section for absorbing and scattering light across a wide range of frequencies. In particular, all UV, optical, and infrared light is absorbed by dust, and only microwave and radio waves with $\lambda > 1$ mm or so are unaffected. The absorption is much stronger at short wavelength (e.g., Draine Figure 21.1), which is why the absorption-wavelength relation is often called a **reddening law** (Draine §21.1). The strength of the absorption is typically expressed as an **extinction**,

$$\boxed{A_\lambda \equiv 2.5 \log_{10}\left(\frac{F_\lambda^0}{F_\lambda}\right) = 2.5 \log_{10}\left(e^{\tau_\lambda}\right) = 1.086\ \tau_\lambda} \tag{8.12}$$

where F_λ^0 is the flux from the source at wavelength λ, F_λ is the observed flux, and τ_λ is the corresponding optical depth. In other words, A_λ is the reduction of flux in magnitudes, which is almost equal to the optical depth (but not quite because $2.5 \log_{10}$ is not the same as natural log).

The electromagnetic energy absorbed by the dust goes into heating it to some equilibrium temperature, but most of it is reemitted as thermal radiation (Draine §24.3). The emission spectrum broadly covers the infrared range (Draine Figure 24.7), but spinning dust grains also emit in the microwave range seen by CMB satellites such as *Planck*. This radiation is polarized by the dust because the grains are aligned by magnetic fields in the ISM (Draine §21.3). For a much more extensive treatment of the theory of dust absorption and scattering as a function of grain size and wavelength, see Draine §22.

The temperature of dust is set by a balance of radiative heating and cooling, but the meaning of temperature refers to the internal state of the grains rather than their spatial motions. Like molecules, dust grains have internal degrees of freedom such as vibrational modes that are excited at higher temperatures (Draine §24). The result of the cooling and heating balance is a relatively low dust temperature,

$$\begin{aligned} T_{\rm dust,Si} &\approx 16\ {\rm K} \left(\frac{a}{0.1\ \mu{\rm m}}\right)^{-1/15} \left(\frac{U_*}{10^{-12}\ {\rm erg/cm^3}}\right)^{1/6} \\ T_{\rm dust,C} &\approx 22\ {\rm K} \left(\frac{a}{0.1\ \mu{\rm m}}\right)^{-1/40} \left(\frac{U_*}{10^{-12}\ {\rm erg/cm^3}}\right)^{1/6} , \end{aligned} \tag{8.13}$$

where a is the grain size and U_* is the energy density in starlight (see Draine §24.1.1 for the exact definition). The two temperatures refer to silicate-like dust grains with sizes $0.01 \lesssim a \lesssim 1\ \mu$m and graphite-like dust grains with $0.005 \lesssim a \lesssim 0.15\ \mu$m (Draine §24.1.4). We note that the dependencies on grain size and radiation intensity are weak.

Finally, the processes by which dust is created and destroyed are poorly understood at this point. Some important mechanisms are thought to be condensation from gas, accretion of additional gas onto grains, sputtering (breaking up of grains by gas), shattering of grains into smaller grains, coagulation into larger grains, and destruction by supernova shocks. However, this list is almost certainly incomplete, and much remains to be understood about the formation, destruction, and dynamics of dust (Draine §25–26).

9 Star formation in a turbulent ISM

In the previous chapter, we considered dense, mostly molecular clouds and their even denser substructure. So-called **star-forming cores** typically have a mass of 0.3–10 $M_\odot$ and form either a single star, a stellar binary, or fragment into a few stars. Clumps contain multiple cores and thus produce a population of stars. Theoretically, however, the challenge of transforming a core into a star is formidable: the material must be compressed from pc scales to stellar scales (such as $R_\odot = 7 \times 10^{10}$ cm), which represents a compression by roughly 21 orders of magnitude (Draine §41.1). Meanwhile, angular momentum and magnetic fields resist the collapse and must be removed somehow.

We begin by considering the simplest model for collapse, gravitational free-fall, in §9.1. We compare the predictions of this picture with observations of stars and star formation in the real Universe in §9.2. More recent work has established that turbulence plays a critical role in the (temporary) creation of molecular clouds and cores, and thus in the initial mass function of the formed stars. We consider its effect on star formation in §9.3.

9.1 Free-fall collapse and the Jeans mass

Assuming that gravity is the mechanism by which clouds collapse and form stars, we are faced with a basic competition between pressure supporting a cloud and gravity contracting it. The gravitational potential of a cloud grows with cloud mass, whereas the pressure corresponds to a fixed energy per unit volume. Thus, we expect there to be a mass scale where gravity "wins," called the **Jeans length** (see hydro notes),

$$\lambda_\mathrm{J} \equiv \frac{2\pi}{k_\mathrm{J}} = \sqrt{\frac{\pi c_\mathrm{s}^2}{G\rho_0}} = \sqrt{\frac{\pi\gamma k_\mathrm{B} T}{\mu m_\mathrm{p} G \rho_0}}\,. \tag{9.1}$$

In the last expression, we have inserted the ideal gas sound speed, $c_\mathrm{s}^2 = \gamma k_\mathrm{B} T/(\mu m_\mathrm{p})$. If we apply this picture to a spherical cloud, we can also define the mass within a sphere of diameter λ_J, the **Jeans mass**,

$$M_\mathrm{J} \equiv \frac{4\pi\rho_0}{3}\left(\frac{\lambda_\mathrm{J}}{2}\right)^3 = \frac{\pi^{5/2}}{6}\left(\frac{\gamma k_\mathrm{B} T}{\mu m_\mathrm{p} G}\right)^{3/2} \rho_0^{-1/2} \tag{9.2}$$

To get a sense of the numbers, we substitute $\gamma = 5/3$ and $\mu = 2.36$ to account for cold molecular gas (where the rotational degrees of hydrogen are not excited since $T \ll 170$ K). The conversion $\rho_0 = 1.4 m_\mathrm{p} n_\mathrm{H}$ holds even for molecular gas owing to the definition of n_H (§1.2). We get

$$\lambda_\mathrm{J} \approx 0.035 \text{ pc } \left(\frac{T}{10 \text{ K}}\right)^{1/2} \left(\frac{n_\mathrm{H}}{10^6/\mathrm{cm}^3}\right)^{-1/2} \tag{9.3}$$

and a resulting Jeans mass of

$$M_\mathrm{J} \approx 0.78\ M_\odot \left(\frac{T}{10 \text{ K}}\right)^{3/2} \left(\frac{n_\mathrm{H}}{10^6/\mathrm{cm}^3}\right)^{-1/2}\,. \tag{9.4}$$

To recapitulate, this mass represents the smallest mass that can collapse given the pressure (temperature) and density of star-forming cores. This result is encouraging, as it seems to explain why the masses of stars are of the order of the mass of the Sun!

The Jeans mass can also be derived in other ways. For example, by equating the gravitational (circular) velocity, $v = \sqrt{GM/R}$, with the sound speed we derive the same result within 10%. The

calculation can also be based on the idea of virial equilibrium: the internal pressure of the cloud, its gravitational energy, and the external pressure at the cloud's surface must balance. This balance results in a minimum mass at which the pressure equilibrium can be overcome by gravity called the **Bonnor-Ebert mass**, which is very similar to the Jeans mass, $M_{\rm BE} \approx 1.18\ M_{\rm J}$ (Draine §14.3).

The next pressing question is how long the collapse of a Jeans-unstable perturbation would take. An estimate is given by the **free-fall time** for a sphere of uniform density (see hydro notes),

$$t_{\rm ff} = \sqrt{\frac{3\pi}{32 G \rho_0}} = 4.4 \times 10^4\ {\rm yr}\ \left(\frac{n_{\rm H}}{10^6/{\rm cm}^3}\right)^{-1/2} . \tag{9.5}$$

Implicitly, we have made the assumption that gravity has won over pressure once the perturbation has crossed the "Jeans threshold," meaning that we neglect pressure. Even though the free-fall time depends on density, it is always quite short for the conditions found in star-forming cores.

However, the simple, spherical collapse envisioned by the free-fall time cannot actually occur in the presence of angular momentum. While different star-forming cores probably have different net rotation velocities, a relatively simple calculation shows that their collapse would be halted by the "angular momentum barrier" long before they reach the radii of stars (Draine §41.5). In reality, the gas forms a rotating accretion disk that transports angular momentum outwards and to the surrounding gas. The exact mechanisms by which this happens are still not understood in detail (Draine §41.6).

Another complication that we have yet to consider are magnetic fields: as flux lines are "frozen" in the plasma, a star would inherit the entire magnetic field strength of the collapsing cloud that created it. In reality, the magnetic fields observed in stars are orders of magnitude lower. Thus, magnetic fields must diffuse out of the cloud (Draine §41.4). Either way, magnetic fields will act as a pressure source in addition to thermal pressure and thus help to stabilize clouds against collapse, increasing the equivalent Jeans mass.

9.2 Basic observations of star formation

Observationally, we would ideally measure the star formation rate (SFR) by resolving individual stars in some population (or galaxy), determine their ages, and thus measure the rate at which new stars form. Such high-resolution observations are not generally possible though, meaning that we need to interpret the average emission from entire stellar populations. Most observational tracers of star formation rely on the signatures of the most massive stars and their H II regions because those stars have short lifetimes. For example, some O stars live for less than 10 million years, meaning that they must have formed recently if we detect their signature. Unfortunately, the brightest tracer of ionization, Hα radiation, suffers from significant dust obscuration, meaning that we can use it only in cases where we have a handle on the dust in the system. Infrared tracers can get around the dust issue. For example, the N II 205 μm line is excited in H II regions and thus a signature of recent star formation, as is free-free radio emission (Bremsstrahlung) due to the electrons (Draine §42.4). Each star formation tracer has its own systematics and tracks star formation on a different timescale. Combining different tracers helps to eliminate systematic errors.

From observations, we believe that the Milky Way has a star formation rate of about 1.3 $M_\odot$/yr. We can now compare this rate to the prediction from Jeans collapse. If all molecular gas in the Milky Way (about $10^9\ M_\odot$) was free-fall collapsing to form stars, we would expect an SFR of

$$\dot{M}_{\rm ff} = \frac{M_{\rm H_2}}{t_{\rm ff}} \approx \frac{10^9\ M_\odot}{6.2 \times 10^6\ {\rm yr}} \approx 160\ \frac{M_\odot}{\rm yr} , \tag{9.6}$$

where we have computed the free-fall time from the approximate average density of the molecular gas, $n_{\rm H} \approx 50/{\rm cm}^3$, rather than from the density of star-forming cores. Nevertheless, our estimate is orders of magnitude larger than the actual SFR, meaning that star formation must be an inefficient process! Most of the molecular gas is not actively star-forming and not all gas from star-forming clumps/cores actually makes it into the stars (for the reasons discussed in §9.1).

While we might be tempted to throw out the free-fall argument altogether at this point, it does make an interesting prediction for the scaling of SFR with density. If we assume that the free-fall time does tell us about the timescale of star formation but that this process is somehow inefficient, we can introduce a dimensionless efficiency $\epsilon_{\rm ff}$,

$$\frac{\rm SFR}{\rm unit\ volume} = \epsilon_{\rm ff} \frac{\rho}{t_{\rm ff}} \propto \frac{\rho}{\rho^{-1/2}} \propto \rho^{3/2} \,. \tag{9.7}$$

Observationally, we typically have access to surface densities rather than volume densities, but the two should be intimately related (for example, $\Sigma \propto \rho$ in a galactic disk of constant thickness). The relation between surface density of gas and SFR is known as the **Kennicutt-Schmidt relation** (Schmidt 1959; Kennicutt 1998),

$$\boxed{\Sigma_{\rm SFR} \approx (1.7 \pm 0.4) \times 10^{-4} \, \frac{M_\odot}{\rm kpc^2 \ yr} \left(\frac{\Sigma_{\rm H_2}}{M_\odot/{\rm pc}^2} \right)^{1.4 \pm 0.1}} \tag{9.8}$$

The exact normalization and exponent for this relation depend on the galaxy sample, the spatial scale over which the quantities are averaged, and other observational systematics. Nevertheless, most investigations agree relatively well on the power law with index 1.4, which is strikingly similar to the SFR $\propto \rho^{1.5}$ scaling of Equation 9.7! Similar relations exist for the surface density of neutral gas and for H I (e.g., Bigiel et al. 2008), but they are less tight than that for molecular gas (which is expected, given that stars form from dense, molecular gas).

We can think of the observed timescale of star formation as a **depletion time**, the time it would take for a system (a cloud or galaxy) to use up all its molecular gas. For example, for the Milky Way we have a molecular depletion time of

$$t_{\rm dep} = \frac{M_{\rm H_2}}{\rm SFR} \approx \frac{10^9 \ M_\odot}{1 \ M_\odot/{\rm yr}} \approx 1 \ {\rm Gyr} \,. \tag{9.9}$$

Depletion times of a Gyr are common in the local Universe. There is evidence that they were lower (about 0.5 Gyr) at $z = 2$ (e.g., Tacconi et al. 2018).

Another fundamental quantity in the study of star formation is the **initial mass function (IMF)**, the abundance of stars as a function of mass. The pioneering study of Salpeter (1955) suggested that the IMF is proportional to ${\rm d}N/{\rm d}M_* \propto M_*^{-2.35}$, and this slope has held amazingly well at high stellar masses. At low masses, the mass function must turn over and cease entirely at 0.08 $M_\odot$, the hydrogen burning limit below which gas balls turn into brown dwarfs rather than stars. The most prominent fitting functions that capture this behavior are those by Miller & Scalo (1979), Kroupa (2002), and Chabrier (2003). Figure 14 shows these functions and the cumulative fraction of stars above mass m. One important aspect of these functions is the fraction of stars above 8 $M_\odot$, since those explode as core-collapse supernovae (§2.4). An important, ongoing discussion is whether the IMF is universal or whether it depends on the properties of the gas and galaxy within which the stars are formed, such as metallicity (Bastian et al. 2010). Given that stars of different masses produce drastically different spectra, the answer may have a large impact on inferences drawn from starlight.

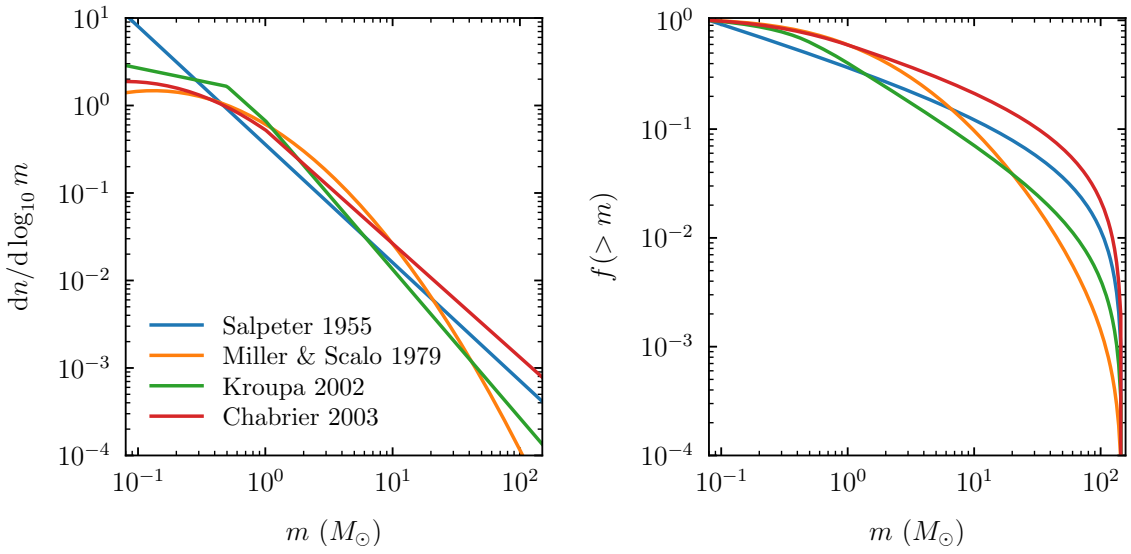


Figure 14: Fitting functions for the initial stellar mass function (Salpeter 1955; Miller & Scalo 1979; Kroupa 2002; Chabrier 2003). The panels show both the differential IMF (left) and the cumulative fraction of stars above a certain mass (right). The functions are normalized to give unity stellar mass across the range from 0.08 to 150 $M_\odot$.

9.3 The role of turbulence

One of the major riddles in star formation is what physical mechanisms give the IMF its shape. This question is far from settled, but the distribution of core masses in molecular cloud complexes can have a strikingly similar shape to the IMF, although shifted to higher mass (e.g., Alves et al. 2007). This match suggests a direct relation between the mass of the core and the star(s) it forms, with an inefficiency factor lowering the mass. Contemporary simulations show that the fragmentation of cores into multiple stars also plays an important role.

Either way, if core sizes do set stellar masses, we need to ask what distribution the core masses take on and why. This question leads us to **turbulence** because the ISM is, in general, a turbulent medium. Turbulence is continuously stirred up by the differential rotation of gas disks, by supernova blastwaves, by the expansion of H II regions, and by other feedback mechanisms such as jets (Lequeux §13.3).

There is plenty of evidence to suggest that turbulence plays an important role in shaping molecular clouds. First, their irregular shapes simply look like turbulent collections of gas rather than spherical clouds in virial equilibrium. A self-similar turbulent cascade would suggest that the average velocity on some scale l is $u_l \propto l^{1/3}$ (see hydro notes). If the velocity dispersion we see in clouds is, indeed, due to turbulent motions, we would expect a $\sigma_\mathrm{u} \propto L^{1/3}$ relation with the size of the clouds. We recall that the Larson relation (Equation 8.3) has an exponent $\gamma \approx 0.35 - 0.5$, which is close to 1/3. The difference could be due to additional motion caused by gravitational forces. We also recall that molecular clouds are over-pressurized compared to the CNM (§8.1). Turbulence could provide some of the pressure needed to maintain this disequilibrium state. Such a turbulent pressure would modify the Jeans dispersion relation, slowing gravitational collapse (Lequeux §14.1.2).

Given these arguments for the influence of turbulence, we wish to analyze the distribution of densities in a turbulent gas, imagining that the densest regions might end up forming clouds, cores, and eventually stars. Unfortunately, the problem is significantly more complicated than suggested

by the Kolmogorov-Obukhov law. First, turbulent motions within the ISM tend to be **supersonic**, and the emerging shocks violate the assumption of a constant, gradual transfer of kinetic energy between scales. As a result, the slope of the turbulent power spectrum depends on the Mach number, $\mathcal{M}$. Second, observations indicate that the ISM is magnetized to a level where MHD is important. Thus, we need to consider **MHD turbulence**, including flows that are not only supersonic but also super-Alfvénic (meaning that the velocities are larger than the Alfvén speed). Third, **self-gravity** does clearly matter, which changes the distribution of turbulent velocities and densities. And fourth, the **injection of energy happens at different scales**, unlike the constant scale L envisioned in the cascade picture. Instead, the shearing motion of a rotating galaxy induces motions on kpc scales, supernova bubbles on 100 pc scales (§2.4), and H II regions on even smaller scales.

For all of these reasons, we need simulations to adequately test the predictions of turbulence for star formation. A key result has been that the **probability distribution function (pdf) of overdensity approaches a log-normal distribution** (e.g., Vazquez-Semadeni 1994),

$$p(s) = \frac{1}{\sqrt{2\pi}\sigma_\mathrm{s}} e^{-(s-s_0)^2/2\sigma_\mathrm{s}^2}, \qquad s \equiv \ln\left(\frac{\rho}{\rho_0}\right), \tag{9.10}$$

where ρ_0 is the mean density and s_0 the mean logarithmic density. Physically, we can understand the origin of this distribution based on supersonic turbulence, where shocks compress the density field. For isothermal shocks, we learned that $\rho_2/\rho_1 \propto \mathcal{M}^2$ (see hydro notes). We expect there to be shocks with a range of Mach numbers, which means that the density is effectively multiplied by a series of random numbers. In logarithmic space, multiplications become additions, which means that we are adding random numbers the log density s. The **central limit theorem** says that such a process, it it goes on for long enough, produces a normal distribution; hence the log-normal pdf of density. Simulations have further shown that the width of this distribution, σ_s, is approximately

$$\sigma_\mathrm{s}^2 \approx \ln\left(1 + b^2\mathcal{M}_\mathrm{s}^2\right), \tag{9.11}$$

where $\mathcal{M}_\mathrm{s}$ is the typical sonic Mach number and $b \approx 0.5$ (e.g. Burkhart 2018). More recent work has shown that the picture is a little more complicated: at the high-density end, the pdf takes on a power-law shape with more gas at very high density than predicted by the rapidly declining log-normal. The reason for this tail is gravity: once a certain density threshold has been crossed, a cloud begins to collapse, which increases its density regardless of turbulent motions. This result has been confirmed observationally and implies that both turbulence and gravity play an important role.

Given a density pdf, we can now establish the connection to star formation. For example, we can posit that there is a critical density above which stars will be formed. The SFR is then the integral over the pdf above that density. Stronger turbulence (higher $\mathcal{M}_\mathrm{s}$) broadens the log-normal distribution but also makes it harder for gas to become self-gravitating. This picture is quite successful at explaining the observed SFRs in the ISM (e.g., Burkhart & Mocz 2019).

10 Global models of the ISM

After considering the various phases of the ISM and the physical processes most relevant to each, we try to tie it all together in this final chapter. We will be guided by a few classic models that have been proposed to understand the interaction of different ISM components and to explain their observed properties. All of these model invoke some sort of stable equilibrium state, which is a simplification. Most of the models are furthermore incomplete in the sense that they try to explain only certain aspects of the ISM or that they make drastic assumptions. Specifically, in §10.1 we begin with the key observation that multiple phases of neutral gas coexist at the same pressure. In §10.2, we consider the model that proposed supernovae as regulators of the ISM pressure. In §10.3, we look at a model that invokes hydrostatic balance in galactic disks. In §10.4, we present a feedback-regulated equilibrium model that predicts star formation rates.

10.1 Two-phase medium in thermal equilibrium

In §4, we observed that atomic H I exists in two phases, the WNM and CNM. They are in pressure balance because both their densities and temperatures differ, but why do they take on characteristic temperatures of $T \lesssim 300$ K and $T \gtrsim 6000$ K? Why not a temperature in between? It turns out that the answer is intimately related to thermal equilibrium and the cooling function.

We have previously considered thermal equilibrium where heating and cooling balance, $\Gamma = \Lambda$ (e.g., in H II regions, §6.4). The heating rate is approximately proportional to density, for example because higher n means that an incident radiation field can be absorbed by more atoms (in this section, we write $n = n_{\rm H}$ for simplicity). Cooling relies on processes involving two atoms, and thus scales as n^2. We use these scalings to write the heating and cooling rates as approximately density-independent rates,

$$\Gamma \approx \Gamma' n \qquad \Lambda \approx \Lambda'(T) n^2 \,. \tag{10.1}$$

The heating rate per mass is approximately independent of temperature, but the cooling rate tends to increase with temperature (though not monotonically, §2.2). For the problem at hand, we care less about cooling and heating individually and more about the **net cooling rate** per unit mass,

$$\frac{\mathcal{L}}{n} \equiv \frac{\Lambda - \Gamma}{n} \approx n\Lambda'(T) - \Gamma' \,. \tag{10.2}$$

In thermal equilibrium, we must have $\mathcal{L} = 0$, but, not every equilibrium point is stable: the evolution of temperature than depends on how the net cooling rate evolves with T. In particular, we encounter a **thermal instability** if

$$\left.\frac{\partial \mathcal{L}}{\partial T}\right|_P < 0 \tag{10.3}$$

because raising T means lowering $\mathcal{L}$, which further raises T (and vice versa). Conversely, the system is **stable** if

$$\left.\frac{\partial \mathcal{L}}{\partial T}\right|_P > 0 \,, \tag{10.4}$$

because raising T means raising $\mathcal{L}$, which pushes T back down (see Clarke & Carswell 2014 §10.3 for a more formal derivation of the stability criterion). We can further deduct how the stability criterion depends on $\Lambda'(T)$ because the ISM pressure, $P/k_{\rm B} = nT$, is approximately constant, so that the density simply adjusts to temperature,

$$\frac{{\rm d}n}{n} = -\frac{{\rm d}T}{T} \,. \tag{10.5}$$

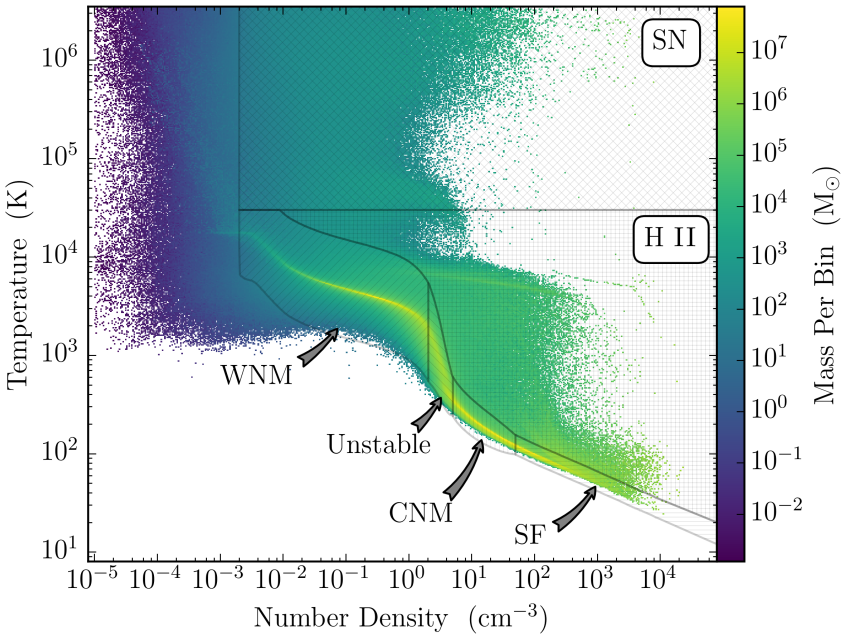


Figure 15: Density-temperature distribution of the ISM in a simulation of a Milky Way-mass disk galaxy. The gray regions highlight gas heated by supernovae and H II regions, as well as gas that is dense and cold enough to be deemed star-forming. Constant pressure would correspond to diagonal lines in n-T space, and a lot of gas occupies a narrow range in pressure (yellow ridge). A significant amount of gas does reside in the thermally unstable region between the WNM and CNM, highlighting that models that allow only two possible temperature are overly simplistic. Of course, the exact distribution depends on the feedback physics implemented in the simulation, as well as the initial conditions. Figure from Goldbaum et al. (2016).

If we raise T, the density goes down, which relatively increases the heating rate compared to the cooling rate. For stability, the **cooling rate must increase with temperature faster than the density decreases**, meaning faster than $\Lambda' \propto T$!

At temperatures of around $300 < T < 4000$ K, we encounter unstable conditions because $\Lambda'(T)$ evolves with T relatively slowly, too slowly to counteract the increase in Γ rate with decreasing n (Figure 5). Gas tends to evacuate this unstable temperature range and takes on either CNM or WNM densities and temperatures. At lower and higher temperatures, $\Lambda'(T)$ is a steeper function, leading to a net increase in $\mathcal{L}$ with T and thus stability. At low T, the behavior of the cooling function is largely driven by a sharp increase of C II and O I cooling around 100 K and by Lyman-α picking up around 10,000 K (Draine §30.4). Below the CNM and above the WNM temperatures, we need to consider other physical effects in the cooling curve, e.g., collisionally excited plasma cooling for very hot gas.

In summary, the majority of the gas in the Galaxy (neutral H I) exists as a **multi-phase medium**, where two well-defined gas phases coexist in pressure equilibrium. The temperatures and densities of these phases are largely set by the cooling curve. This concept of multi-phase pressure balance was established in Field et al. (1969), who model the ISM as a two-phase system. However, this simple picture is not entirely realistic. Figure 15 shows the n-T distribution of ISM gas in a modern simulation of a Milky Way-like galaxy (Goldbaum et al. 2016). Clearly, a significant amount of gas resides in the unstable region, presumably because the thermal instability has not yet had time to push it into either the WNM or CNM. Nonetheless, we do see that a large fraction of cool and cold gas shares an approximately constant pressure (diagonal line in n-T space).

A number of works have presented improved models of the thermal equilibrium, for example taking into account advancements in our understanding of the heating and cooling sources (e.g., Wolfire et al. 2003) and dependencies on additional variables, most notably on metallicity (e.g., Bialy & Sternberg 2019).

10.2 Three-phase medium with supernovae as pressure regulators

The idea of two-phase thermal equilibrium is elegant, and it conforms to the long-established idea that the ISM phases have to be in approximate pressure equilibrium (Spitzer 1956). But what actually sets the equilibrium pressure? After all, the CNM and WNM can co-exist at a range of pressures. However, Spitzer (1956) already predicted the need for a third, hot phase. By the 1970s, it was furthermore known that supernova (SN) remnants can fill a significant fraction of the ISM volume (e.g., Cox & Smith 1974).

Thus, McKee & Ostriker (1977) introduced a hot medium into their model. The key assumption is that the rate of supernovae essentially sets the ISM pressure: if the pressure is lower than the "SN pressure," bubbles will expand further and heat more material until the pressure has adjusted to resist the expanding bubbles. In this scenario, we would expect that a random place in the ISM is part of about one SN remnant. In §2.6, we showed that this assumption can be used to derive a characteristic pressure that roughly matches the observed one (to a factor of two or so).

McKee & Ostriker (1977) take this argument as a starting point (their Equation 2) and build a complex, **three-phase model** of the ISM. Their predicted pressure is somewhat lower than modern observations suggest (about 1600 K/cm^3; compare to §1.3), but they more or less correctly predict the densities, temperatures, and filling fractions of the CNM, WNM, and HIM. The model is explicitly not concerned with very cold gas or molecular clouds and therefore makes no predictions regarding star formation.

10.3 Hydrostatic balance

The two-phase and three-phase models in the previous sections neglected one important aspect: gravity. Effectively, we have been treating the ISM as a gas system in some volume of arbitrary shape and size. In reality, most star formation happens in relatively thin galactic disks that largely shaped by gravity and angular momentum.

To take gravity into account, Ostriker et al. (2010) assert that the pressure at the center of the disk (the **mid-plane pressure**) is set by the gravitational potential, much like the pressure of the Earth' atmosphere is set by gravity pulling the gas downwards. In their **hydrostatic equilibrium** model, Ostriker et al. (2010) balance the gravitational force (at each radius) by heating from star formation. This system forms an equilibrium because too low a gas pressure would allow gravity to push gas into the disk, which would lead to additional star formation (and vice versa). As in McKee & Ostriker (1977), the ISM gas is modeled with three components (HIM, WNM, and CNM).

The Ostriker et al. (2010) model has three important free parameters, namely, the fraction of pressure that stems from thermal energy (as opposed to turbulence, cosmic rays, etc.), the fractions of H I that reside in the CNM and WNM, and the efficiency of star formation in dense, self-gravitating gas. Once values for those parameters have been chosen, the model can be initialized with observed surface densities of gas, from which it predicts the surface density of star formation. The agreement with observations of certain disk galaxies in the local Universe is striking.

One important difference between the McKee & Ostriker (1977) and Ostriker et al. (2010) models is the heating mechanism (SN bubbles vs. stellar UV light). How can both models make valuable predictions? We note that both mechanisms are directly proportional to star formation,

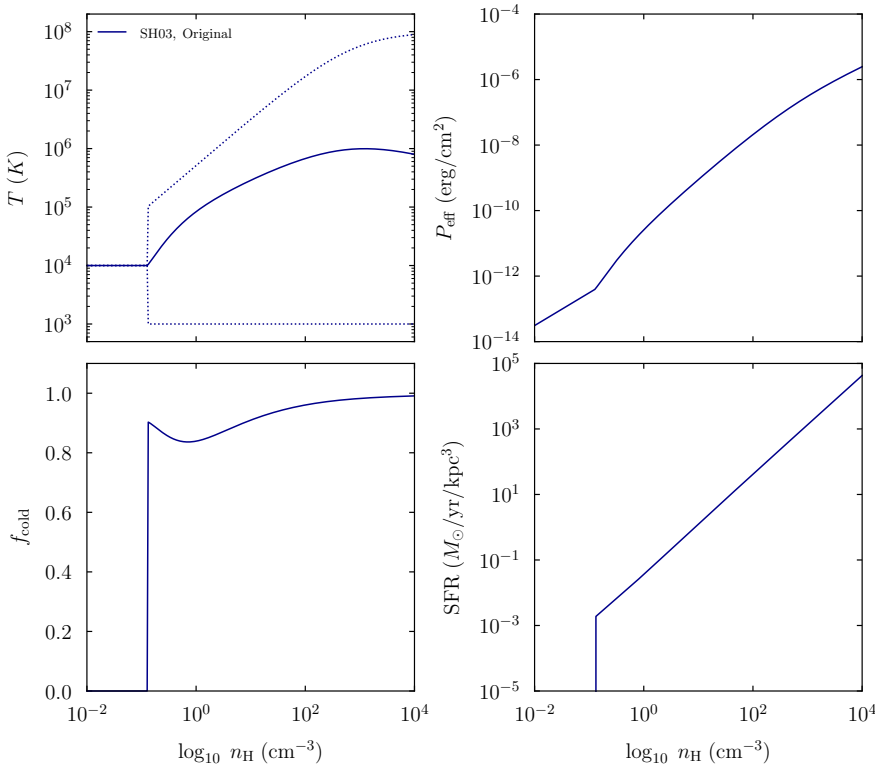


Figure 16: Predictions of the two-phase equilibrium model of Springel & Hernquist (2003). The only input to the model is density. The temperature (top left) bifurcates into an imaginary "cold" component with $T = 1000$ K and a hot component whose temperature increases with density. The pressure (top right) increases faster than suggested by an ideal gas equation of state (straight line in log space). The fraction of gas in the "cold" phase (bottom left) varies between 80% and 100%. By construction, the model predicts an SFR (bottom right) that is proportional to $\rho^{3/2}$ and thus matches the observed Kennicutt-Schmidt relation. Below the star formation threshold of about $0.13/\mathrm{cm}^3$, the model is inactive.

so that both models capture a fundamental balance between star formation and the feedback it produces.

10.4 Feedback-regulated two-phase equilibrium

The models presented thus far provide valuable insight into the interplay of the ISM phases, but they have not made any predictions regarding star formation. Modeling the SFR is crucial in simulations, where the resolution has historically been (and typically still is) too low to resolve individual molecular clouds. In this low-resolution regime, we need to add a so-called "subgrid" prescription for star formation.

Springel & Hernquist (2003) created such a model based on the idea of equilibrium. Their two-phase model predicts pressure, temperature, and SFR from only density (Figure 16). We imagine a volume of gas that contains hot ISM and "cold" clouds with a fixed $T = 1000$ K (though the exact temperature chosen for this phase is not critically important). Mathematically, Springel & Hernquist (2003) write down differential equations for star formation, cooling (using realistic cooling tables), and heating by supernovae. These equations can almost entirely be solved analytically. The SFR scales with the density of gas in the cold phase as $\rho^{3/2}$, based on the free-fall argument from §8.1. However, as the SFR goes up with density, so does the supernova rate, which evaporates more

cold gas. This interplay leads to a feedback-regulated equilibrium, where the pressure and SFR are largely determined by the supernova rate (similar to the McKee & Ostriker (1977) model).

This type of modeling can be made arbitrarily complicated by adding more phases or more processes that shift gas between the phases or change its energy, and by making those processes more realistic. For example, the Springel & Hernquist (2003) model neglects the effects of metallicity and other ISM physics such as a realistic split of the gas into a WNM, CNM, and molecular gas. To alleviate these issues, Braun & Schmidt (2012) present a model with numerous phases (including H_2) and processes (such as turbulence). However, the predictions of such models also become harder to interpret as the complexity increases.

Appendix

A Background

A.1 Hydrodynamics

Gases and fluids are described by the equations of hydrodynamics, as long as certain conditions are fulfilled. We imagine the fluid to be composed of parcels called "fluid elements." The mean free path of the constituent particles, $\lambda_{\rm mfp} = 1/n\sigma$, must be short compared to these elements (meaning that collisions are frequent within each element). Second, there must be sufficient particles per fluid element for averaged quantities such as temperature to be well-defined. And third, changes in the averaged quantities must be small across the size of the fluid elements. As long as all three conditions are met, we can apply the Euler equations of hydrodynamics to the fluid elements. We may write those equations either from the "Lagrangian" perspective of each fluid element (where the mass of the element is conserved and we imagine to be moving along with the element) or from the "Eulerian" perspective at fixed points in space. The two types of derivatives are connected by the relation

$$\frac{\mathrm{D}Q}{\mathrm{D}t} = \frac{\partial Q}{\partial t} + \boldsymbol{u}\cdot\boldsymbol{\nabla}Q\,, \tag{A.1}$$

where Q represents an arbitrary fluid quantity and $\boldsymbol{u}$ is the velocity of the fluid. Based on the microscopic properties of the particles making up the fluid, we define the density ρ, pressure P, and internal energy ε of the fluid, which then obey the Euler equations:

Eulerian fluid equations

$$\frac{\partial\rho}{\partial t} + \boldsymbol{\nabla}\cdot(\rho\boldsymbol{u}) = 0$$

$$\frac{\partial\boldsymbol{u}}{\partial t} + \boldsymbol{u}\cdot\boldsymbol{\nabla}\boldsymbol{u} = -\frac{\boldsymbol{\nabla}P}{\rho} - \boldsymbol{\nabla}\Phi$$

$$\frac{\partial\varepsilon}{\partial t} + \boldsymbol{u}\cdot\boldsymbol{\nabla}\varepsilon = -\frac{P}{\rho}\boldsymbol{\nabla}\cdot\boldsymbol{u} + \frac{\Gamma}{\rho} - \frac{\Lambda}{\rho}$$

Lagrangian fluid equations

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = -\rho\boldsymbol{\nabla}\cdot\boldsymbol{u}$$

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = -\frac{\boldsymbol{\nabla}P}{\rho} - \boldsymbol{\nabla}\Phi$$

$$\frac{\mathrm{D}\varepsilon}{\mathrm{D}t} = -\frac{P}{\rho}\boldsymbol{\nabla}\cdot\boldsymbol{u} + \frac{\Gamma}{\rho} - \frac{\Lambda}{\rho}$$

Besides the fluid quantities, we have included the effects of a gravitational potential Φ as well as heating and cooling,

$$\Gamma \equiv \rho\left(\frac{\mathrm{D}\varepsilon}{\mathrm{D}t}\right)_{\rm heating} \quad \text{and} \quad \Lambda \equiv \rho\left(\frac{\mathrm{D}\varepsilon}{\mathrm{D}t}\right)_{\rm cooling}. \tag{A.2}$$

The total energy of the fluid is the sum of the kinetic, internal, and gravitational, and magnetic energies,

$$E \equiv \rho\left(\frac{|\boldsymbol{u}|^2}{2} + \varepsilon + \Phi + \frac{\boldsymbol{B}^2}{8\pi}\right). \tag{A.3}$$

Although magnetic fields are important in the ISM, we do not solve the MHD equations in this course and have left magnetic fields out of the Euler equations. However, we do wish to calculate the magnetic energy at times (where we have assumed that $\boldsymbol{B}$ is in cgs units). The internal energy refers to all degrees of freedom (translational, rotational, or vibrational). We typically express the number of degrees of freedom, $N_{\rm dof}$, as

$$\gamma = \frac{N_{\rm dof}+2}{N_{\rm dof}}\,, \tag{A.4}$$

where almost all gases treated in this course have only translational degrees, meaning $N_{\rm dof} = 3$ and $\gamma = 5/3$. Molecular gas can additionally have rotational degrees, meaning $N_{\rm dof} = 5$ and $\gamma = 7/5$. The internal energy is in equipartition between all degrees of freedom, but only the translational degrees of freedom contribute to temperature and pressure. Thus, the translation between density, internal energy, and pressure (the equation of state) depends on γ,

$$P = nk_{\rm B}T = \frac{\rho k_{\rm B} T}{\mu m_{\rm p}} = \rho\varepsilon(\gamma - 1)\,, \tag{A.5}$$

where $m_{\rm ptl} = \mu m_{\rm p}$ is the average particle mass. This equation of state is valid only for ideal gases, but that includes all gases considered in this course. Finally, the speed of sound in an ideal gas is

$$c_{\rm s} = \sqrt{\frac{\gamma P}{\rho}}\,. \tag{A.6}$$

For more details on these concepts, please see the hydrodynamics notes.

B Derivations

B.1 Shock radius for stellar wind

Draine §38.1 proposes a slightly more accurate, albeit much more complicated, derivation for the normalization of the shock radius of a stellar wind. The idea is to not use the mass swept up during the free expansion phase as the starting point but to equate the energy injected by the shock to the energy in the bubble behind the shock. We parameterize the shock radius using a self-similar power-law scaling,

$$R_\mathrm{s} \equiv A t^\eta \quad \Longrightarrow \quad u_\mathrm{s} = \frac{\partial R_\mathrm{s}}{\partial t} = \frac{\eta R_\mathrm{s}}{t} \,. \tag{B.1}$$

Modeling shows that the shock energy is about equally distributed into kinetic and thermal energy. If we assume that the kinetic energy is roughly determined by the post-shock velocity of the gas in the shell, and that the post-shock velocity is $3/4\ u_\mathrm{s}$ (Draine §38.1), then we can write the total energy inside the shell as

$$E(t) \approx 2 \times \frac{V_\mathrm{bubble}\ \rho_0}{2} \left(\frac{3}{4} u_\mathrm{s}\right)^2 = \frac{4\pi R_\mathrm{s}^3 \rho_0}{3} \left(\frac{3}{4}\frac{\eta R_\mathrm{s}}{t}\right)^2 \approx \frac{1}{2}\dot{M} u_\mathrm{w}^2 t \,, \tag{B.2}$$

which we solve for the shock radius,

$$R_\mathrm{s}^5 = (A t^\eta)^5 = \frac{2\dot{M} u_\mathrm{w}^2}{3\pi\eta^2\rho_0} t^3 \,. \tag{B.3}$$

We see that $5\eta = 3$, or $\eta = 3/5$ as expected. We insert η,

$$R_\mathrm{s} = \left(\frac{50}{27}\frac{\dot{M} u_\mathrm{w}^2}{\pi\rho_0}\right)^{1/5} t^{3/5} \approx 0.11\ \mathrm{pc} \times t_3^{3/5} n_3^{-1/5} \dot{M}_6^{1/5} u_3^{2/5} \,. \tag{B.4}$$

This expression is identical to the one based on the free expansion phase (Equation 2.30), except that the pre-factor has changed from $(3/4)^{1/5}$ to $(50/27)^{1/5}$, a change of about 20%. The shock velocity would increase by the same factor. Given this modest difference, the more simplistic derivation purely based on self-similarity and the free expansion phase seems adequate.

References

Alves, J., Lombardi, M., & Lada, C. J. 2007, A&A, 462, L17

Bastian, N., Covey, K. R., & Meyer, M. R. 2010, ARA&A, 48, 339

Bergin, E. A., & Tafalla, M. 2007, ARA&A, 45, 339

Bialy, S., & Sternberg, A. 2019, ApJ, 881, 160

Bigiel, F., Leroy, A., Walter, F., et al. 2008, AJ, 136, 2846

Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207

Braun, H., & Schmidt, W. 2012, MNRAS, 421, 1838

Burkhart, B. 2018, ApJ, 863, 118

Burkhart, B., & Mocz, P. 2019, ApJ, 879, 129

Chabrier, G. 2003, PASP, 115, 763

Clarke, C., & Carswell, B. 2014, Principles of Astrophysical Fluid Dynamics (Cambridge University Press)

Cox, D. P., & Smith, B. W. 1974, ApJ, 189, L105

Draine, B. T. 1978, ApJS, 36, 595

—. 2011, Physics of the Interstellar and Intergalactic Medium (Princeton University Press)

Field, G. B., Goldsmith, D. W., & Habing, H. J. 1969, ApJ, 155, L149

Finkbeiner, D. P. 2003, ApJS, 146, 407

Goldbaum, N. J., Krumholz, M. R., & Forbes, J. C. 2016, ApJ, 827, 28

Habing, H. J. 1968, Bull. Astron. Inst. Netherlands, 19, 421

Haffner, L. M., Dettmar, R. J., Beckman, J. E., et al. 2009, Reviews of Modern Physics, 81, 969

Hollenbach, D. J., & Tielens, A. G. G. M. 1999, Reviews of Modern Physics, 71, 173

Kennicutt, Jr., R. C. 1998, ApJ, 498, 541

Kroupa, P. 2002, Science, 295, 82

Krumholz, M. R. 2013, MNRAS, 436, 2747

Larson, R. B. 1981, MNRAS, 194, 809

Lequeux, J. 2005, The Interstellar Medium (Springer), doi:10.1007/b137959

McKee, C. F., & Ostriker, J. P. 1977, ApJ, 218, 148

Miller, G. E., & Scalo, J. M. 1979, ApJS, 41, 513

Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of gaseous nebulae and active galactic nuclei (University Science Books)

Ostriker, E. C., McKee, C. F., & Leroy, A. K. 2010, ApJ, 721, 975

Ploeckinger, S., & Schaye, J. 2020, MNRAS, 497, 4857

Salpeter, E. E. 1955, ApJ, 121, 161

Schinnerer, E., & Leroy, A. K. 2024, ARA&A, 62, 369

Schmidt, M. 1959, ApJ, 129, 243

Spitzer, Lyman, J. 1956, ApJ, 124, 20

Springel, V., & Hernquist, L. 2003, MNRAS, 339, 289

Strömgren, B. 1939, ApJ, 89, 526

Tacconi, L. J., Genzel, R., Saintonge, A., et al. 2018, ApJ, 853, 179

Trypsteen, M. F. M., & Walker, R. 2017, Spectroscopy for amateur astronomers: recording, processing, analysis and interpretation (Cambridge University Press)

Vazquez-Semadeni, E. 1994, ApJ, 423, 681

Wolfire, M. G., Hollenbach, D., & McKee, C. F. 2010, ApJ, 716, 1191

Wolfire, M. G., McKee, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003, ApJ, 587, 278