## Our Milky Way and Other Galaxies

### Contents

#### 18 The Interstellar Medium 18.1

- 18.1 Star-gas cycle .................................................. 18.1
- 18.2 Cold-Warm-Hot phases of nearly isobaric ISM .................. 18.2
- 18.3 Molecules and dust in cold ISM: Giant Molecular Clouds .... 18.5
- 18.4 HII regions .......................................................... 18.9
- 18.5 Galactic organization of ISM and star-gas interaction along spiral arms ... 18.11

#### 19 Star Formation 19.0

- 19.1 Jeans Criterion for gravitational contraction .................. 19.0
- 19.2 Cooling by molecular emission .................................... 19.1
- 19.3 Free-fall timescale and the galactic star formation rate ...... 19.2
- 19.4 Fragmentation into cold cores and the Initial Mass Function (IMF) . . . 19.3
- 19.5 Angular momentum conservation of rotating cores and disk formation ... 19.5

#### 20 Our Milky Way Galaxy 20.0

- 20.1 Disk, halo, and bulge components of the Milky Way ............ 20.0
- 20.2 Virial mass for cluster from stellar velocity dispersion inferred from Doppler shifts ........................................................................... 20.5
- 20.3 Galactic rotation curve & dark matter ............................. 20.5
20.4 Super-massive black hole at galactic center ........................................... 20.9

21 External Galaxies 21.0

21.1 Cepheid variables as standard candle for distances to external galaxies . . . 21.0
21.2 Galactic redshift and Hubble’s law for expansion ................................. 21.0
21.3 Tully-Fisher Relation: \( L_{gal} \propto V_{rot}^4 \) ........................................ 21.3
21.4 Spiral, Elliptical, & Irregular galaxies .................................................... 21.5

22 Active Galactic Nuclei (AGNs) and Quasars 22.0

22.1 Basic properties and model ................................................................. 22.0
22.2 Lyman alpha clouds ............................................................................ 22.1
22.3 Gravitational lensing of quasar light by foreground Galaxy Clusters .......... 22.3
22.4 Gravitational redshift ......................................................................... 22.3
22.5 Apparent “super-luminal” motion of quasar jets .................................... 22.6

23 Large Scale Structure (LSS) 23.1

23.1 Galaxy clusters & super-clusters ......................................................... 23.1
23.2 Dark matter: Hot vs. Cold, WIMPs vs. MACHOs ............................... 23.3
18. The Interstellar Medium

18.1. Star-gas cycle

Compared to stars, the region between them, called the interstellar medium or “ISM”, is very low density; but it is not a completely empty vacuum. For one thing, we’ve seen above that the final remnants of stars, whether white dwarfs, neutron stars, or black holes, generally have much less mass than the initial stellar mass; this implies that a substantial fraction (30-80%) of this initial mass is recycled back into the surrounding ISM through planetary nebulae, stellar winds, or supernova explosions. Moreover, a key theme in this and the next section is that stars themselves formed out of this ISM material through gravitational contraction, making for a kind of star-gas-star cycle, as illustrated in figure 18.1.

If one assumes that, on average, a typical atom spends roughly equal fractions of time in the star vs. ISM phase of this cycle, then the average density of gas in the ISM should be roughly equal to the mass of the stars spread out over the volume between them. For example, in the region of the galaxy near the sun, the so-called “solar neighborhood”, the mean number density of stars is $n_\star \approx 0.1 \text{ pc}^{-3}$, reflecting a typical interstellar separation distance $d \approx n_\star^{-1/3} \approx 2 \text{ pc}$. (Recall that the nearest star, α Centauri, is about 1.3 pc from the sun.) If we take the average mass of each star to be roughly that of the sun, we obtain a mean mass density $\rho \approx M_\odot n_\star \approx 7 \times 10^{-24} \text{ g/cm}^3$.

This is much, much lower than the typical average density within stars, which as noted earlier for the sun is $\rho_\odot \approx 1.4 \text{ g/cm}^3$; this mostly just reflects the huge distance/size ratio, giving roughly a factor $(\text{pc}/R_\odot)^3 \sim 10^{24}$, for the volume between stars vs. that within them. Thus while most stars typically have mean densities comparable to matter (like water) here on earth (i.e., $\rho \approx 1 \text{ g/cm}^3$), this ISM density is well below (by a factor $\sim 10^4$) even the most perfect vacuum ever created in terrestrial laboratories ($\rho \sim 10^{-19} \text{ g/cm}^3$).

Indeed, for ISM densities, it is more intuitive and common to quote values in terms of the atom number density. For example, with a composition dominated by Hydrogen, the associated ISM Hydrogen-atom number density is $n \approx \rho/m_p \approx 4 \text{ cm}^{-3}$.

The assumptions and approximations behind this estimate – viz. the equal time between ISM and stars, the sun representing a typical stellar mass, the assumed star density – are all somewhat rough, and can even vary through the galaxy. Nonetheless, when averaged over the entire galaxy, the characteristic ISM number density $n \sim 1 \text{ cm}^{-3}$ is indeed comparable to this local estimate. However, within this broad average, there are wide variations, reflecting a highly complex, heterogeneous, and dynamic ISM, as discussed next.
18.2. Cold-Warm-Hot phases of nearly isobaric ISM

A key factor in the wide variations in density of the ISM is the wide variation in its temperature. Roughly speaking, gas in the ISM can be characterized in 3 distinct temperature phases, ranging from cold \((T \sim 10 - 100 \, \text{K})\), to warm \((T \sim 5000 - 10,000 \, \text{K})\), to hot \((T \sim 10^5 - 10^7 \, \text{K})\). Figure 18.2 vividly illustrates the distinct signatures of these different components in various spectral wavebands ranging from the radio to gamma rays, as mapped along the disk plane of our Milky way galaxy.

In contrast to these wide variations in temperature, the gas pressure, which is proportional to the product of density and temperature, tends to be relatively constant over the
broad ISM, with a typical value $P/k = nT \sim 10^3 \text{K/cm}^3$. This near constancy of ISM pressure stems from the fact that gravity, which declines in strength with the inverse square of the distance, is generally too weak to confine gas over the many parsec scales between stars\(^1\). In the absence of any restraining force, and ignoring any disturbances from stellar

---

\(^1\)An exception to this is in the densest cloud “cores” in star-forming regions, wherein gravity is compressing cold but very dense and thus high-pressure gas in the final contraction toward forming stars. See §\(^1\) 19.
mass ejection (e.g. from stellar winds or supernovae), the ISM gas should over time settle into a dynamical equilibrium that is roughly *isobaric*, meaning with a spatially constant gas pressure.

Within this roughly isobaric ISM, the densities of the 3 phases thus tend to scale inversely with temperature, ranging from \( n \approx 10 - 100 \text{ cm}^{-3} \) for cold clouds, to \( n \approx 0.1 - 0.2 \text{ cm}^{-3} \) for the warm gas, to \( n = 10^{-2} - 10^{-3} \text{ cm}^{-3} \) for the hot component.

Because the flux of radiation also falls with the inverse square of distance, we might expect the temperature of gas far from stars to be always very cold, for example as would be the case for the equilibrium temperature of a blackbody (see QQ 1). But the low density of interstellar gas makes it very different from a blackbody, since emitting radiation requires collisions to excite atoms or interact with free electrons, the rates of which decrease with density.

**Quick Question 1:** a. Recalling that the equilibrium blackbody temperature of the earth is \( T_e \approx (T_\odot/2)^{\sqrt{2}R_\odot/au} \approx 280 \text{ K} \), show that the corresponding temperature at a distance \( d \) from the sun is given by \( T = T_e \sqrt{au/d} \). b. Compute the temperature for \( d = 1 \text{ pc} \). c. Compute the temperature at a distance \( d = 1 \text{ pc} \) from a hot star with \( T_* = 10T_\odot \approx 60,000 \text{ K} \).

Indeed, for the hot ISM the temperature is so high that Hydrogen and Helium and are completely ionized, with only the heaviest and most complex atoms, like iron, having a few remaining bound electrons. This means that radiative emission mostly only occurs through rare collisional excitation of these few, partially ionized heavy ions, making radiative cooling very inefficient, with a characteristic cooling time of several Myr. For the warm and still mostly ionized ISM, the higher density and greater number of bound electrons in heavy ions makes cooling somewhat more effective, but cooling times are still quite long, typically of order \( 10^4 \text{ years} \).

As for the energy source that heats up the hot and warm ISM in the first place, this comes mainly from hot, massive stars. Near such a hot, luminous star, UV photoionization of the surrounding Hydrogen gas can heat it to temperatures that are a significant fraction of the stellar effective temperature, of order \( 10^4 \text{ K} \). As detailed in § 18.4, the resulting volume of ionized gas – dubbed HII regions from the standard notation for ionized hydrogen – can extend several parsecs from the star. Overall, such photo-ionization from hot stars is a significant heating source for the *warm* component of the ISM.

But an even more dramatic source of energy comes from the violent supernova (SN) explosions that end the relatively brief lifetimes of such hot, massive stars. In such SN
explosions, several solar masses of stellar material is ejected at very high speeds, approaching 10% the speed of light, implying kinetic energies of order $E_{SN} \sim M_\odot c^2/200 \sim 10^{52}$ erg. As the high-speed, expanding ejecta runs into the surrounding ISM, the resulting shock$^2$ wave heats the gas to very high temperatures, initially up to $10^8$ K. But as the ISM gas piles up, the expansion slows and cools, ending up with a temperature $T \approx 10^5 - 10^7$ K, with the total pressure comparable to the surrounding ISM. Such SN explosions are thus the primary source of the hot component of the ISM.

Reflecting the large expansion of its source SN explosions, this hot ISM component can actually occupy more than half, even up to $\sim 70-80\%$, of the volume of the galaxy. Most of the remaining volume fraction, $\sim 15 - 25\%$, makes up the warm component, with just a relatively small part, $\lesssim 5\%$, being in relatively cool clouds.

And since thermal energy density is just $E_{th} = (3/2)nkT = (3/2)P$, the near constancy of ISM pressure means that the large volume of hot gas also contains most of the ISM thermal energy.

However, in terms of overall distribution of matter, most of the ISM mass is in relatively cool, dense clouds. As discussed in §19, it is these cool clouds that provide the source material for forming new stars, so let us next examine further their nature.

### 18.3. Molecules and dust in cold ISM: Giant Molecular Clouds

The low temperature and high density of the cold ISM makes it possible for the atoms to combine into molecules, and so much of the cold ISM takes the form of Giant Molecular Clouds (GMC). Reflecting the dominant abundance of Hydrogen, the most common molecule is $\text{H}_2$. Among the heavy elements, carbon monoxide (CO) is usually the most abundant, reflecting its relative stability and the cosmic abundance of both its atomic constituents. Other common molecules include diatomic Oxygen ($\text{O}_2$) and water ($\text{H}_2\text{O}$), and in some clouds up to 100 distinct molecular species (including, e.g., alcohol, $\text{CH}_3\text{CH}_2\text{OH}$) have been detected.

The survival, and thus abundance, of GMC molecules requires both a low local gas temperature and low UV flux, both of which become problematic in the vicinity of hot stars. But the high density and low temperature of such GMC also means they tend to have quite

---

$^2$A shock wave arises whenever two gases collide with supersonic speed. It effectively converts the kinetic energy of the pre-shock gas into heat, yielding post-shock temperatures that scale with specific kinetic energy, or square of the speed, of the pre-shock gas.
high densities of ISM dust, and this can be very effective at shielding the regions from the heating and photo-dissociation by UV light from nearby stars. The dust itself is generally not formed locally, since in even the coldest clouds the density is not high enough for efficient nucleation of microscopic dust grains. Instead it is thought that most dust is formed in the outer layers of cool giant stars, and then blown away into the ISM by a strong stellar wind.

As discussed in §12, such dust can lead to very strong extinction and reddening of starlight. Figure 18.3 vividly illustrates the heavy extinction of the background starlight by the GMC Barnard 68. Note moreover how the partially extincted light from stars around the cloud edges is distinctly reddened.

Fig. 18.3.— Illustration of the heavy extinction of background starlight by the GMC Barnard 68. Note also the reddening of partially extincted light around the cloud edges.

To estimate the dust opacity\(^3\), note that a spherical dust grain of radius \(a\), mass \(m_d\), and mass density \(\rho_d = m_d/(4\pi/3)a^3\) has a physical cross section,

\[
\sigma_d \equiv \pi a^2 = \frac{3m_d}{4a\rho_d}. \tag{18.1}
\]

The overall opacity of a dust cloud is given by dividing this cross section for an individual dust particle by the mass \(m_c\) of cloud material per dust particle, i.e. \(\kappa_d = \sigma_d/m_c\). For a

\(^3\)For more extensive discussion, see http://ned.ipac.caltech.edu/level5/Sept05/Li2/Li_contents.html
mass fraction \( X_d = m_d/m_c \) of a cloud that is converted into dust, we find using (18.1) that the implied opacity is

\[
\kappa_d = \frac{3X_d}{4a\rho_d} \approx 150 \frac{\text{cm}^2}{\text{g}} \frac{X_d}{X_{d,\odot}} \frac{0.1 \mu\text{m}}{a} \frac{1\text{g/cm}^3}{\rho_d}.
\]  

(18.2)

The latter equality provides a numerical evaluation scaled by: the dust mass fraction \( X_{d,\odot} \approx 2 \times 10^{-3} \) assuming full conversion of dust-forming material at standard (solar) abundances; the dust grain internal density \( \rho_d \approx 1 \text{g/cm}^3 \) (most dust would almost float in water); and a typical dust grain size \( a \approx 0.1 \mu\text{m} \). We thus see that the corresponding dust opacity, \( \kappa_d \approx 150 \text{cm}^2/\text{g} \), is several hundred times greater than that for free electron scattering in the fully ionized gas inside a star, \( \kappa_e \approx 0.3 \text{cm}^2/\text{g} \).

As discussed previously (Appendix C), the opacity from the geometric cross section of dust grains only applies to wavelengths comparable to or smaller than the grain size, \( \lambda \lesssim a \); for \( \lambda > a \), the associated dust opacity decreases as \( \kappa_d(\lambda) \sim (\lambda/a)^{-\beta} \), where the power index \( \beta \) is sometimes referred to as the “reddening exponent”. For simple Rayleigh scattering from smooth spheres of fixed size \( a \), \( \beta = 4 \). In practice, the complex mixtures in sizes and shapes of dust typically lead to a smaller effective power index, \( \beta \approx 1 - 2 \). Still, the overall inverse dependence on wavelength means that clouds that are optically thick to dust absorption and scattering will show a substantially \textit{reddened} spectrum.

§12.3 discusses how this reddening can be quantified in terms of a formal \textit{“color excess”}, which then can be used to estimate an associated visual extinction magnitude \( A_V \equiv V_{\text{obs}} - V_{\text{intrinsic}} \). Recall that the extinction magnitude in any waveband is related to the associated optical depth, which in turns scales linearly with opacity in that waveband, \( A(\lambda) \sim \tau(\lambda) \sim \kappa(\lambda) \). For wavelengths larger than the dust size \( \lambda > a \) and a linear reddening exponent \( \beta \approx 1 \), we thus see that the optical depth, and thus the extinction magnitude declines with the inverse of the wavelength,

\[
A(\lambda) \sim \tau(\lambda) \sim \kappa(\lambda) \sim 1/\lambda.
\]  

(18.3)

For example, if a star has an extinction \( A_V \) in the visual waveband centered on \( \lambda_V \approx 500 \text{nm} \), then in the mid-infrared \textit{“M-waveband”} at roughly a factor ten higher wavelength \( \lambda_M \approx 5000 \text{nm} = 5 \mu\text{m} \), the opacity, and thus the optical depth and extinction magnitude, are all \textit{reduced} by this same factor 10, \( A_M \approx A_V/10 \). For a case with, say \( A_V = 10.8 \) magnitudes of visual extinction, the visual flux would be reduced by a factor \( e^{-\tau_V} = e^{-A_V/1.08} = e^{-10} = 4.5 \times 10^{-5} \). By contrast, in this mid-IR M-band, the factor ten lower extinction magnitude \( A_M = 1.08 \) implies a much weaker reduction, now just a factor \( e^{-\tau_M} = e^{-A_M/1.08} = e^{-1} = 0.36 \).
Stars are typically formed out of interstellar gas and dust in very dense molecular clouds, which often have 10 or 20 magnitudes of visual extinction ($A_V \approx 10 - 20$), essentially completely obscuring them at visual wavelengths. But such stars can nonetheless be readily observed with minimal extinction in mid-IR (few microns) or far-IR (millimeter) wavebands. This fact has spurred efforts to build large infra-red telescopes, both on the ground and in space. The ground-based telescopes are placed at high altitudes of very dry deserts, to minimize the effect of IR absorption by water vapor in the earth’s atmosphere. Another issue is to keep the IR detectors very cold, to reduce the thermal emission background.

Finally, the energy from dust-absorbed optical or UV light is generally reemitted in the mid-IR, at wavelengths set by the dust temperature through roughly the standard Wien’s law for peak emission of a blackbody, $\lambda_{\text{max}} \approx 30 \mu\text{m}/(T/100\text{K})$ (cf. eqn. 4.6). For GMC clouds with $T = 30 - 50\text{K}$ this gives thermal dust emission in the 60-100 $\mu\text{m}$ range, as illustrated for galactic plane dust emission in figure 18.2c.
Quick Question 2: GMC dust extinction and reddening

a. For a GMC with molecular Hydrogen density $n = 100 \text{ cm}^{-3}$, compute the associated mass density $\rho$.

b. For UV light with $\lambda = 100 \text{ nm}$ and dust with size $a = 0.1 \mu\text{m}$ and solid density $\rho_d = 1 \text{g/cm}^3$ and the solar abundance mass fraction $X_d = 2 \times 10^{-3}$, use the geometric cross section opacity derived in the text to compute the mean free path $\ell$ (in pc) for this GMC.

c. For a GMC of diameter $D = 30 \text{ pc}$, compute the optical depth $\tau$ and reduction fraction $F_{\text{obs}}/F$ for a star behind the cloud that emits such UV light.

d. Use this to compute the associated extinction magnetic for this UV light, $A_{UV}$.

e. Assuming a reddening exponent $\beta = 1$, now compute the extinction $A_V$ for visible light with $\lambda = 500 \text{ nm}$, and the extinction $A_{NIR}$ for near IR light with $\lambda = 2 \mu\text{m}$.

18.4. HII regions

Let us next consider the warm ISM that is heated by UV photo-ionization from hot, luminous OB-type stars. Specifically, consider an ISM cloud with a uniform number density $n$ of Hydrogen atoms surrounding a hot star with luminosity $L$. Stellar UV photons with energy $h\nu > h\nu_o \equiv 13.6 \text{ eV}$ can efficiently ionize neutral Hydrogen atoms, but these will then tend to quickly recombine with the free electrons.

The recombination rate scales with the electron density $n_e$ times a temperature-dependent recombination coefficient,

$$\frac{1}{t_r} = n_e \langle \sigma_r v_e \rangle_T \approx 4 \times 10^{-13} (\text{cm}^3/\text{s}) n_e \quad \text{for } T \approx 10^4 \text{ K},$$

where $t_r$ is the recombination time for an ionized Hydrogen atom, i.e. the time for a free proton to encounter and recombine with a free electron. The recombination coefficient depends on the recombination cross section $\sigma_r$ times the electron thermal speed $v_e$, with the angle brackets representing averaging over the thermal velocity distribution of electrons at a temperature $T$. As mentioned above, such UV photoionization tends to heat the gas to a temperature $T \approx 10^4 \text{ K}$, and so the latter relation evaluates this recombination coefficient for that temperature.

The total emission rate of stellar UV ionizing photons can be estimated by integrating over the Planck blackbody function $B_\nu$ from the ionization threshold frequency $\nu_o$, where
\[ h\nu_o \equiv 13.6 \text{ eV}, \]
\[ \dot{N}_{UV} \equiv L \int_{\nu_o}^{\infty} \frac{B_\nu d\nu}{B h\nu} = \frac{L}{Bh} \int_{\nu_o}^{\infty} \frac{B_\nu d\nu}{\nu}. \]  

(18.5)

Here \( B \) is the spectrally integrated Planck function, and the division by the photon energy \( h\nu \) converts the energy rate into a photon number rate.

In equilibrium, this number of ionizing photons will balance the total number of recombinations over a sphere (dubbed a Strömgren sphere after the scientist who first described it) of radius \( R_S \) centered on the star. In terms of the proton (i.e., ionized H) number density \( n_p \), each of the total number \( n_p (4/3)\pi R^3_S \) of ionized H in the sphere recombines with an electron of number density \( n_e \) over the recombination time \( t_r \). The balance with stellar ionizing photons of emission rate \( \dot{N}_{UV} \) thus requires,

\[ \dot{N}_{UV} = \frac{4\pi n_p R^3_S}{3t_r} = n_p n_e \langle \sigma_r v_e \rangle_T \frac{4\pi}{3} R^3_S. \]  

(18.6)

For full ionization of a pure H cloud of number density \( n \), we have \( n_p = n_e = n \), and so this “Strömgren radius” \( R_S \) of such an “HII region” of ionized Hydrogen (HII) is simply given by

\[ R_S = \left[ \frac{3\dot{N}_{UV}}{4\pi n^2 \langle \sigma_r v_e \rangle_T} \right]^{1/3} \approx 6.0 \text{ pc} \left[ \frac{\dot{N}_{50}}{n^2_2} \right]^{1/3}, \]  

(18.7)

where \( \dot{N}_{50} \equiv \dot{N}_{UV}/(10^{50} \text{ s}^{-1}) \) and \( n_2 \equiv n/(10^2 \text{ cm}^{-3}) \) are convenient variables scaled by typical values for this photon rate and Hydrogen number density.

The number of UV photons can be estimated from the spectral type of the exciting star, and the number density of H atoms can be inferred from the observed line emission from the HII region. Using these to compute the physical size \( R_S \), we can then use the measured angular radius \( \alpha \) to estimate the HII region’s distance, \( d = R_S/\alpha \).

In an HII region the ongoing recombination occurs through a cascade of electrons from higher to lower bound states of Hydrogen, leading to extensive emission lines for all the Hydrogen term series (Lyman, Balmer, Paschen etc. for lower final state \( n=1, 2, 3, \) etc.). But in optical images, the most prominent line emission stems from the \( n=3 \) to \( n=2 \) transition associated with the Balmer line H\( \alpha \), which is in the red part of the visible spectrum, with wavelength \( \lambda = 656.28 \text{ nm} \). Viewed in the visible, HII regions thus generally have a distinctly reddish glow, as illustrated in the left panel of figure 18.4 for the HII region known as the Rosetta nebula. But the false-color image of the same nebula in the right panel shows that, in addition to the H\( \alpha \) (now color-coded red), there is also line emission from doubly ionized Oxygen (OIII, green) and singly ionized Sulfur (SII, blue).
18.5. Galactic organization of ISM and star-gas interaction along spiral arms

In the dense regions of active star formation in the Milky Way and other galaxies, the ionization from numerous young, hot, massive stars can merge into an extended Giant HII region. Viewed from earth along the plane of the Milky Way, the projection of foreground and background stars and nebulae can make such regions appear complex and amorphous. A visually clearer view can be gleaned from external galaxies that are viewed face on, like the “Whirlpool” galaxy shown in figure 18.5. The distinctly reddish splotches seen in the optical image in the left panel are all Giant HII regions that formed in the dense clouds along this galaxy’s spiral arms.

The right panel of figure 18.5 shows a composite image in 4 distinct spectral bands, spanning the IR (red), optical (green), UV (blue), and X-rays (purple). The face-on view nicely complements the disk-embedded perspective images from multiple wavebands shown for own Milky Way galaxy in figure 18.2. Note in particular how the close link between ISM and star formation is organized by the spiral arm structure. This is discussed further in the
Fig. 18.5.— *Left:* Hubble space telescope optical image of M51, the “Whirlpool galaxy”. The reddish blotches are from Balmer series Hα line emission (which at wavelength $\lambda = 656 \text{ nm}$ is in the red part of the visible spectrum) from giant HII regions. These represent the merger of many individual HII regions that arise when dense regions of interstellar Hydrogen in otherwise cold giant molecular clouds (GMC) are photo-ionized by the UV radiation from the numerous, recently formed, hot massive stars. Note their proximity to dark bands formed from absorption of background stellar light by cold interstellar dust, which outline the galactic spiral arms. *Right:* Composite image of M51 from 4 NASA orbiting telescopes. X-rays (purple) detected by the Chandra X-ray Observatory reveal point-like sources from black holes and neutron stars in binary star systems, as well as a diffuse glow of hot ISM gas. Optical data from the Hubble Space Telescope (green) and infrared emission from the Spitzer Space Telescope (red) both highlight long lanes in the spiral arms that consist of stars and gas laced with dust. Finally, UV light (blue) from the GALEX telescope comes from hot, young stars, showing again how well these track the HII giants and star-forming GMCs along the spiral arms.

section on galaxies (§21).
19.  Star Formation

19.1.  Jeans Criterion for gravitational contraction

Stars generally form in clusters from the gravitational contraction of a dense, cold GMC. The requirements for such gravitational contraction depend on the relative magnitudes of the total internal thermal (kinetic) energy $K$ versus the gravitational binding energy $U$. For a cloud of mass $M$, uniform temperature $T$, and mean mass per particle $\mu$, the total number of particles $N = M/\mu$ have an associated total thermal energy,

$$K = \frac{3}{2} NkT = \frac{3}{2} \frac{MkT}{\mu}. \quad (19.1)$$

If the cloud is spherical with radius $R$ and uniform density $\rho = \mu n = M/(4\pi R^3/3)$, the associated gravitational binding energy (cf. eqn. 10.7) is

$$U = -\frac{3}{5} \frac{GM^2}{R}. \quad (19.2)$$

Recalling the condition $K = -U/2$ for stably bound systems in virial equilibrium, we can expect that for a cloud with $K > -U/2$, the excess internal pressure would do work to expand the cloud against gravity, leading it to be less tightly bound (or even unbound, if $K > -U$).

Conversely, for $K < -U/2$, the too-low pressure would allow the cloud to gravitationally contract, leading to a more strongly bound cloud. The critical requirement, known as the Jeans criterion, for such gravitational contraction can thus be written

$$\frac{M}{R} > \frac{5kT}{G\mu}. \quad (19.3)$$

In terms of the cloud’s atomic number density $n = \rho/\mu = N/(4\pi R^3/3)$, we can define a minimal Jeans radius for cloud contraction,

$$R_J \approx \left( \frac{15kT}{4\pi nG\mu^2} \right)^{1/2} \approx 9.6 \text{ pc} \left( \frac{T}{n} \right)^{1/2} \frac{m_p}{\mu}, \quad (19.4)$$

where the second equality assumes CGS units, with number density $n$ in cm$^{-3}$ and temperature $T$ in Kelvin.

Alternatively, one can define a minimum Jeans mass (the total mass within a Jeans radius) for a cloud to contract,

$$M_J \equiv \frac{4\pi R_J^3}{3} \mu n \approx \frac{5}{\mu^2} \left( \frac{kT}{G} \right)^{3/2} \left( \frac{15}{4\pi n} \right)^{1/2} \approx 92 M_\odot \frac{T^{3/2}}{n^{1/2}} \left( \frac{m_p}{\mu} \right)^2. \quad (19.5)$$
For typical ISM conditions, both the Jeans radius and mass are quite large, implying it can be actually quite difficult to initiate gravitational contraction. For example, for a cold cloud with $\mu = m_p$, $T = 100$ K and $n = 10$ cm$^{-3}$, we find $R_J \approx 30$ pc and $M_J \approx 30,000$ $M_\odot$, requiring then a cloud that is initially extremely large and massive. The requirements are somewhat less severe once the hydrogen atoms form into H$_2$ molecules, thus increasing the molecular weight to $\mu \approx 2m_p$, and so reducing $R_J$ by a factor 2, and $M_J$ by a factor 4.

But a general upshot of such a large Jeans mass is that stars tend typically to be formed in large clusters, resulting from an initial contraction of a GMC, with mass of order $10^4 M_\odot$ or more.

Exercise 1:

a. Assuming an isobaric ISM with the canonical pressure $P/k = nT = 10^3$ K cm$^{-3}$, derive expressions for $R_J$ (in pc) and $M_J$ (in $M_\odot$) as a function of temperature $T$ (in K).
b. Now derive analogous expressions for $R_J$ and $M_J$ as a functions of number density $n$ (in cm$^{-3}$).

19.2. Cooling by molecular emission

In contrast to the poor radiative efficiency of the ionized gas in the warm and hot phases of the ISM, in the cool ISM the formation of molecules makes such clouds much more efficient for radiative cooling. The thermal, collisional excitation of the molecules and dust leads to emission of radiation at IR wavelengths comparable to those associated with black-body emission for the given temperature. For example, for a cloud with temperature $T = 100$ K, radiation is at IR wavelengths $\lambda \approx \lambda_{max,\odot} T_\odot / T \approx 30\mu m$ (see eq. 4.6).

At low temperatures $T < 100$ K cooling by molecular radiation is dominated by carbon monoxide (CO). Both C and O are relatively abundant elements, and the molecular structure of CO provides a variety of excitation modes (rotational, vibrational, or electronic) from inelastic collision with molecular hydrogen. This converts kinetic energy of the gas to potential energy in the molecules, which de-excite radiatively to emit an IR photon that escapes the cloud, causing it to cool.

Such CO molecular cooling is a key factor in initiating and maintaining cloud contraction, by allowing the cloud to shed the increased internal energy gained from the stronger gravitational binding. In virial equilibrium only half this energy is lost, and so the interior would still heat up in proportion to the stronger gravitational binding. But in practice CO emission is often so efficient that the cloud interior can stay cool, or even become cooler, as
it contracts. The resulting dramatic reduction in interior pressure support then leads to a full gravitational collapse.

### 19.3. Free-fall timescale and the galactic star formation rate

To estimate the timescale for gravitational collapse, recall first from Kepler’s third law (see eqn. 9.5) that the period $P$ for orbit at radius $R$ around an object of mass $M$ is

$$P = \sqrt{\frac{4\pi^2 R^3}{GM}} = \sqrt{\frac{3\pi}{G\rho}},$$

(19.6)

where the second equality casts this in terms of the mean density within a sphere of this radius, $\rho = M/(4\pi R^3/3)$. Since the period from Kepler’s law does not depend on the orbital eccentricity, (19.6) also applies to a purely radial orbit (with eccentricity $\epsilon = 1$) through a central point mass from this radius. But the self-gravitational collapse of a cloud would occur at just this same rate, implying then that the free-fall time to contract to zero radius from the initial radius $R$ should be just a quarter of this orbital period,

$$t_{ff} = \frac{P}{4} = \sqrt{\frac{3\pi}{16G\rho}} = \frac{0.82 \text{ hr}}{\sqrt{\rho}} = \frac{51 \text{ Myr}}{\sqrt{n}} \sqrt{\frac{2m_p}{\mu}},$$

(19.7)

where the density evaluations assume CGS units (i.e., g/cm$^3$ for $\rho$ and cm$^{-3}$ for $n$). For a star like the sun, for which the mass density is CGS order unity, free fall would be less than an hour. But for a cold molecular cloud, with say a number density $n \approx 100 \text{ cm}^{-3}$, such free-fall would take several Myr.

**Quick Question 1:** *Free-fall time for simple constant-gravity model*

Recall the elementary physics result that an object falling under gravitational acceleration $g$ drops a distance $s = gt^2/2$ in time $t$. Fixing the gravity at a constant value $g = GM/R^2$, use this simple relation to solve for the time $t_g(R)$ to fall through a stellar radius (i.e., by setting $s = R$). Compare this with the free fall time in (19.7) by evaluating the ratio $t_g(R)/t_{ff}$.

In our galaxy, the total mass in giant molecular clouds with density $n \gtrsim 100 \text{ cm}^{-3}$ is estimated to be about $M_{GMC} \approx 10^9 M_\odot$. Since this mass should collapse to stars over a free-fall time, it suggests an overall galactic star formation rate should be given by

$$\dot{M}_{sfr} = \frac{M_{GMC}}{t_{ff}} \approx 200 \frac{M_\odot}{\text{yr}}.$$

(19.8)
But the observationally inferred star formation rate is actually much smaller, only about $1 \, M_\odot/yr$, implying an effective efficiency of only $\epsilon_{ff} \lesssim 0.01$. The reasons for this are not entirely clear, but may stem in part from inhibition of gravitational collapse by interstellar magnetic fields, and/or by interstellar turbulence. Another likely factor is the feedback from hot, massive stars, which heat up and ionize the cloud out of which they form, thus inhibiting the further gravitational contraction of the cloud into more stars. Further details are given in the somewhat advanced but still readable review of star formation by M. Krumholz:

http://ned.ipac.caltech.edu/level5/Sept10/Krumholz/Krumholz_contents.html

19.4. Fragmentation into cold cores and the Initial Mass Function (IMF)

In those portions of a GMC that do undergo gravitational collapse, the contraction soon leads to higher densities, and thus to smaller Jeans mass and Jeans radius, along with a shorter free-fall time. This tends to cause the overall cloud, with total mass $10^4 - 10^6 M_\odot$, to fragment into much smaller, stellar-mass cloud “cores” that will form into individual stars.

A key, still unsolved issue in star formation regards the physical processes and conditions that determine the mass distribution of these proto-stellar cores, leading then to what’s known as the stellar Initial Mass Function (IMF).

This IMF can be written as $dN/dm$, wherein $m = M/M_\odot$ is the stellar mass in solar units, and $dN(m) = (dN/dm)dm$ represents the fractional number of stars within a mass range $m$ to $m + dm$. Studies of the evolution of stellar clusters suggest that this can be roughly characterized by a power-law form,

$$\frac{dN}{dm} = K m^{-\alpha}, \quad (19.9)$$

where $K$ is a normalization factor that depends on the total number of stars, and the power index $\alpha$ has distinct values for high-mass vs. low-mass stars. For $m > 1$, the most commonly inferred value is $\alpha \approx 2.35$, known as the “Salpeter” IMF (dashed red line in figure 19.1), after the scientist who first quantified the concept of an IMF. The large power-index reflects the fact that higher-mass stars are much rarer than lower-mass stars.

For lower mass, there are various models, the simplest being the “flattened” IMF (from Scalo 1986, dashed blue curve in figure 19.1), with $\alpha = 0$ for $m < 1$. Another is the three-power model (from Kroupa 2001, aqua blue curved in figure 19.1), which keeps $\alpha = 2.35$ for $m > 0.5$, but then takes $\alpha = 1.3$ for Red dwarf stars in the mass range $0.08 < m < 0.5$, and $\alpha = 0.3$ for Brown dwarf stars with $m < 0.08$. Figure 19.1 compares various other IMFs.
Fig. 19.1.— Comparison of various IMFs, plotted on a log-log scale in terms of the mass fraction per logarithm interval of mass (a.k.a. “dex”), which is proportional to $m^2 dN/dm$.

**Exercise 2: Flattened Salpeter IMF**

a. For the simple flattened Salpeter IMF, with $\alpha = 2.35$ for $m > 1$ and $\alpha = 0$ for $m < 1$, integrate (19.9) over all masses to obtain an expression for the normalization $K$ in terms of the total number of stars $N_{\text{tot}}$.

b. Now use this to obtain an expression for the fraction of stars, $N(m > m_o)/N_{\text{tot}}$, with mass greater than some mass lower limit $m_o$ (assuming $m_o \geq 1$). In particular, what fraction of stars have $m > 1$?

c. For $m_o = 100$, how many total stars must a cluster have for there to be at least one star with $m \geq m_o$? How about for $m_o = 300$? What does this imply
for observational efforts to determine whether there is an upper mass cutoff to the IMF?

With a given form of the IMF for a collapsing GMC, one can model the evolution of the resulting stellar cluster, based on how each star with a given mass evolves through its various evolutionary phases, e.g. main sequence, red giant, etc.

19.5. Angular momentum conservation of rotating cores and disk formation

In general, the fragmentation of a GMC into stellar-mass cores will endow those cores with a non-zero rotation, and this can be a key factor in their final collapse toward stellar size. While material near and along the core rotation axis can still collapse to form the central star, the conservation of angular momentum for material near the rotational equator can halt the contraction and lead to formation of a protostellar disk.

For material with angular momentum per unit mass \( j \equiv vr \) in circular orbit with speed \( v \) at a radius \( r \) about a central mass \( M \), the orbital condition for balance between centrifugal and gravitational acceleration can be cast in the form,

\[
\frac{GM}{r^2} = \frac{v^2}{r} = \frac{j^2}{r^3}
\]  

(19.10)

For an initially spherical core with starting radius \( R \) and angular rotation frequency \( \Omega \), a mass parcel at the rotational equator has an angular momentum per unit mass \( j_{eq} = \Omega R^2 \). As the cloud collapses under the gravitational attraction of its own mass \( M \), conservation of angular momentum causes this parcel to rotate faster until it reaches the condition (19.10) for orbit, with associated disk radius

\[
r_d = \frac{j_{eq}^2}{GM} = \frac{\Omega^2 R^4}{GM} \equiv 2\beta_{eq}R.
\]  

(19.11)

The last equality here introduces the initial equatorial ratio of rotational to gravitational energy,

\[
\beta_{eq} \equiv \frac{\Omega^2 R^2/2}{GM/R} = \frac{3\Omega^2}{8\pi G \rho} = \frac{\Omega^2 P_{orb}^2}{8\pi^2} = \frac{1}{2} \frac{\Omega^2}{\Omega_{orb}^2}.
\]  

(19.12)

The second equality here shows that \( \beta_{eq} \) depends only on the core density \( \rho \) and its rotation frequency \( \Omega \), two quantities that can generally be readily inferred from observations, with observed cloud cores typically giving \( \beta_{eq} \approx 0.02 \). For a typical observed core size \( R \approx 0.05 \text{ pc} \), the expected disk radius \( r_d \) is a few hundred au, comparable to the inferred sizes of protostellar disks.
The last two equalities recast $\beta_{eq}$ in terms of the orbital period $P_{orb}$ or orbital frequency $\Omega_{orb} \equiv 2\pi / P_{orb}$.

Fig. 19.2.— Traces (blue lines) illustrating how conservation of angular momentum causes various locations on the surface of a rigidly rotating spherical cloud (represented by black circle) to collapse onto an orbiting disk (marked in red). The various panels are for the labeled values $\beta_{eq}$ of the equatorial rotational energy to gravitational energy. Note how material near the rotational poles contracts to the concentrated central region, while material at lower latitudes near the equator collapses onto the orbiting disk with outer radius $r_d = 2\beta_{eq}R$, as given by (19.11).

Exercise 3: Disk collapse from various latitudes
a. For a spherical cloud of radius $R$ and rotation frequency $\Omega$, consider locations away from the equator, with co-latitude $\theta$ measured from the polar axis. Derive
an expression for the associated ratio $\beta(\theta)$ of the local rotational energy to gravitational energy, writing this in terms of the equatorial ratio $\beta_{eq}$ derived in eqn. (19.12).

b. Use this to derive an expression for the associated disk radius $r(\theta)$ to which material contracts from various latitudes on the initial spherical surface of radius $R$. (You may assume that throughout the contraction, the gravitational attraction is that from a point source of mass $M$ at the cloud center.) The blue lines in figure 19.2 draw connections between this disk radius and its source location at various latitudes on the cloud surface, for various choices of the parameter $\beta_{eq}$.

**Challenge Problem:** *Disk surface density*

a. Consider a hollow thin spherical shell of radius $R$ and rotation frequency $\Omega$ that collapses under the gravitational attraction of a star of mass $M_\star$ at the shell center. Assuming the shell has a mass $M_s$ that is initially spread uniformly over its spherical surface, use the results of the previous problem to derive an expression for the disk surface density $\Sigma(r)$ as a function of disk radius $r$. Express this in terms of the shell mass $M_s$, the outer disk radius $r_d$ in (19.11), and the ratio $r/r_d$.

b. Now use this result to derive an integral expression for the total disk surface density $\Sigma(r)$ from collapse of a filled, constant-density, spherical cloud of radius $R$, mass $M$, and (rigid-body) rotation frequency $\Omega$. (You may assume that the mass $M(r)$ inside any material initially at radius $r \leq R$ remains constant throughout the contraction.) Figure 19.3 plots results for such a disk model.

Initially such disks can have a mass that is a substantial fraction of that for the central star. But in disks with Keplerian orbits, the orbital frequency increases inward with radius as $\Omega_{orb} \sim r^{-3/2}$, meaning that between two neighboring rings there is overall shear in orbital speed. Any frictional interaction – e.g., due to viscosity – between such neighboring rings will thus tend to transport angular momentum from the faster inner ring to the slower outer ring, allowing the inner mass to fall further inward, while the angular momentum receiving material moves further outward. Since the specific angular momentum increases outward as $j = vr = \Omega r^2 \sim \sqrt{r}$, this outward viscous diffusion of angular momentum allows over time for most of the mass to accrete onto the star, with just a small mass fraction retaining the original angular momentum. Eventually this remnant disk-mass can fragment into its own gravitational collapsing cores to form planets. In our own solar system the most massive
Fig. 19.3.—  Left: Disk surface density $\Sigma(r)$ (i.e., mass per unit area of the disk) vs. disk radius $r$ for the simple model of the gravitational collapse of a rotating, initially spherical cloud. The curves show results for various initial equatorial ratios of the rotational to gravitational energy, $\beta_{eq} = 0.1 - 0.5$. The disk surface density here is in units of $\rho R$, where $\rho$ and $R$ are the cloud’s initial mass density and radius. Right: Analogous plots for radial mass distribution $dm/dr = 2\pi r \Sigma$.

planet Jupiter has only 0.1% the mass of the sun, but 99% of the solar system’s angular momentum.

Of course earth too originated from the evolving proto-solar disk. You and I and everyone on earth are here today because our source material happened to stem from the equatorial regions of the proto-solar core, with too much angular momentum to fall into the sun itself, and having then been the viscous recipient of the angular momentum from other proto-solar-disk material that did diffuse inward onto the sun.
20. Our Milky Way Galaxy

![Panoramic photo of the Milky Way, taken from the European Southern Observatory’s facility in Paranal, Chile (in the dry, and isolated, Atacama desert). The left side shows the four 8m telescopes, along with a smaller 1.8m telescope in the left foreground. On the far right near the horizon the two hazy patches are dwarf galaxies that are satellites of our own Milky Way, the Large and Small Magellanic clouds. Though they appear to hang side-by-side, they are actually separated by about 15 kpc, and are removed from us by distances of 50 kpc and 60 kpc respectively.]

20.1. Disk, halo, and bulge components of the Milky Way

This tendency for conservation of angular momentum of a gravitationally collapsing cloud to form a disk is actually a quite general process that can occur on a wide range of scales: from planets, to proto-stellar cores, to even an entire proto-galaxy, with hundreds of billions of times the mass of individual stars. This indeed provides the basic rationale for the disk in our own Milky Way (MW) galaxy. We along with our sun are today still embedded within the Milky Way’s disk, orbiting about the galactic center, again because our bits of proto-galactic matter had too much angular momentum to fall further inward.

As we look up into a dark night sky, we can trace clearly the direction along this disk plane through the faint milky glow of thousands of distant, unresolved stars, from which we indeed get the name “Milky Way”. Figure 20.1 gives a vivid illustration of the Milky Way through a panoramic image of the night sky seen from the exceptionally dark and clear site of the Paranal observatory, in the Atacama desert of Chile. Indeed, toward the horizon on the right one can also see two satellite galaxies of the Milky Way, known as the Large and
Fig. 20.2.— Map of the disk plane of our Milky Way galaxy, based on IR and radio surveys. Our sun lies along the Orion spur of the Sagittarius spiral arm, between the inner Scutum-Centaurus arm, and the outer Perseus arm. The Galactic Center (GC), toward the constellation Sagittarius, is defined to be at zero galactic longitude, with the sun orbiting a distance $d \approx 8$ kpc from the GC, in the direction of longitude $90^\circ$ (i.e., to the left in the picture), toward the bright star Vega in the constellation Hercules.
Fig. 20.3.— Edge-on schematic illustration of the 3D morphology of the MW galaxy, showing the disk, halo, and bulge components, with globular clusters in halo, and the sun in the disk, offset from the galactic center. The directions from the sun away from the disk plane are dubbed the Galactic North and South Poles (GNP and GSP).

As we look along this disk plane of the MW, the background/foreground superposition of many stars and GMCs makes it very difficult to discern the overall structure, the way we readily can from the face-on view of M51 in figure 18.5. Moreover, the extinction from the extensive gas and dust means that visible images, like those in figure 20.1 or in panel e of figure 18.2, only penetrate a limited distance, typically $\sim 1$ kpc, within the disk, which itself is only about 1000 ly, or just 0.3 kpc, in thickness.

Fortunately, IR and radio images can penetrate much further, even to the other side of the galaxy, spanning the full 100,000 ly ($\sim 30$ kpc) diameter of this disk. Thus with painstaking work applying various methods for determining the distance to the myriad of stars and

---

4 "Magellanic" because they were first reported to European civilization by Ferdinand Magellan, following his first-in-history circumnavigation of the earth, with routes around southern continents showing the southern sky where these clouds are visible.
GMCs detected, it has become possible to draw a quite complete map of the overall disk structure of our MW galaxy, as given in figure 20.2. This shows that, like M51, our galaxy also has distinct spiral arms, along which are concentrations of gas, dust, HII regions, GMCs, and active star formation. The map nicely illustrates the position of our sun well away from galactic center, and also serves to define the sun-centered galactic longitude system used to chart the galactic disk.

**Quick Question 1: Galactic Year**

At the distance $d \approx 8$ kpc of the Galactic Center, the sun turns out to have an orbital speed $V_o \approx 220$ km/s. How long is one “galactic year”, i.e., the sun’s orbital period (in Myr) around the galaxy?

Figure 20.3 illustrates schematically the overall 3D morphology of the MW, which in addition to the disk, has distinct “halo” and central “bulge” components.

The halo is roughly spherical, with a diameter comparable to that of the disk, about 30 kpc. It contains very little gas or dust, and without much source for new star formation, its stars (dubbed “Population II”) are very old. This can be seen from the H-R diagrams of the globular clusters that are common in the halo, which typically have main-sequence turnoff points below the luminosity of the sun, implying ages $t > t_{ms,\odot} \approx 10$ Gyr. These old globular clusters contain of order $10^4 - 10^5$ stars, and are much more gravitationally bound and stable than the “galactic” or “open” clusters found in the disk.

Such open clusters are typically quite young, with main sequences that sometimes extend to masses of many tens of solar masses, implying ages less than their main sequence lifetimes, i.e. less than a few times 10 Myr. They are irregular in shape, and typically only contain 100 or so stars. They are so loosely bound that they tend to disperse within a few 10 Myr or less, evolving to unbound OB associations. Due to tidal effects from the galaxy, along with the shear from its differential rotation, the stars eventually disperse and mix with other stars (called “Population I”) in the disk. Figure 20.4 compares examples of a globular and an open cluster, and gives a Venn diagram showing the common and distinct properties between the two types.

The central bulge contains a mixture of traits of both the disk and halo, with both types of clusters, and both populations of stars (I and II). Because of dust absorption from within the galactic disk, it doesn’t appear in the visible to be much brighter than higher galactic longitude regions away from the galactic center; but if one corrects for this absorption, it dominates the overall galactic luminosity (as can be seen from the case of M51 shown in figure 18.5).
Fig. 20.4.— *Left:* The globular cluster M80. *Right:* The open cluster M45, a.k.a. the Pleides or Seven Sisters. *Bottom:* A ‘Venn diagram’ comparing the different and common characteristics of Globular vs. Open clusters.
20.2. Virial mass for cluster from stellar velocity dispersion inferred from Doppler shifts

By measuring the Doppler shifts of spectral lines from stars in an open cluster or a globular cluster, one can determine each star’s radial velocity \( V_r \). We can use this to define an average cluster radial velocity, \( V_c \equiv \langle V_r \rangle \), as well as an \textit{root mean square} (rms) \textit{velocity dispersion} about this mean,

\[
\sigma_v \equiv \sqrt{\langle (V_r - V_c)^2 \rangle}.
\]

The kinetic energy-per-unit-mass associated with this random component of radial velocity dispersion is \( \sigma_v^2/2 \). Assuming a similar dispersion in the two transverse directions that cannot be measured from a Doppler shift, the total associated kinetic energy from the 3 directions of motion is \( K = (3/2)M_c \sigma_v^2 \), where \( M_c \) is the total stellar mass. For a cluster of radius \( R \), the associated gravitational binding energy scales as \( U \approx -GM_c^2/R \). If the cluster is bound, then application of the usual virial condition \( K = |U|/2 \) allows one to obtain the cluster mass via

\[
M_c = \frac{3\sigma_v^2 R}{G} = 6.9 \times 10^4 M_\odot \left( \frac{\sigma_v}{10 \text{ km/s}} \right)^2 \frac{R}{\text{pc}}.
\]

In practice, application of this method requires we obtain the cluster radius through the measured angular radius \( \alpha \) and an independently known distance \( d \), through the usual relation \( R = \alpha d \).

20.3. Galactic rotation curve & dark matter

As illustrated in figure 19a, a primary diagnostic of atomic Hydrogen in the disk plane of the galaxy comes from its radio emission line\(^5\) at a wavelength of \( \lambda = 21 \text{ cm} \). As we peer into the inner disk regions of the galaxy, i.e. along galactic longitudes in the range \(-90^\circ < \ell < 90^\circ\), we find that this 21 cm line shows a distinct wavelength broadening \( \Delta \lambda(\ell) \) that varies systematically with the longitude \( \ell \). Most of this broadening arises from cumulative Doppler shift along the line of sight from the motion associated with the orbit of distinct gas clouds about the galactic center; it thus provides a key diagnostic for determining the galaxy’s “rotation curve” as a function of galactic radius \( R \).

\(^5\)This results from a “hyperfine” transition in which the spin of the electron goes from being parallel to anti-parallel to the spin of the proton. The energy difference is much smaller than for transitions between principal energy levels of the Hydrogen atom, which are a few eV, and so have wavelengths of a few hundred nm, in the visible or UV spectral bands. See QQ 2.
**Quick Question 2: Energy of Hydrogen 21-cm transition.**

What is the energy $E$, in eV, of the hyperfine, spin-flip transition that gives rise to the 21-cm emission line of neutral Hydrogen.

**Galactic Rotation**

$$\mathcal{L}(\frac{\lambda_{\text{max}}}{\Omega_o} - 1) = V_{\text{max}}$$

$$= V - V_o \sin \ell = \Omega(R) - \Omega_o R_o \sin \ell$$

$$= (\Omega - \Omega_o) R_o \sin \ell$$

Fig. 20.5.— Sketch to show how measuring the change with galactic longitude $\ell$ of the maximum Doppler-shifted wavelength $\lambda_{\text{max}}$ of the $\lambda_o \approx 21.1$ cm line from atomic Hydrogen can be used to determine the galactic rotation rate $\Omega(R)$ as a function of radius $R$, given the known rotation speed $V_o = \Omega_o R_o \approx 220$ km/s at the radius $R_o$ of the sun’s orbit. The results indicate that, inside the sun’s orbit ($R \leq R_o$), the rotation speed is nearly constant, with $V(R) \equiv R \Omega(R) \approx V_o$.

Figure 20.5 illustrates the basic geometry and associated trigonometric formulae. Focussing for convenience on longitudes in the range $0 < \ell < 90^\circ$, we find that the broadening actually takes the form of a redshift to maximum wavelength $\lambda_{\text{max}}$, from which we can readily infer the maximum line-of-sight velocity away from us, $v_{\text{rmax}} = c(\lambda_{\text{max}}/\lambda - 1)$. But for a given longitude $\ell$, a simple model with circular orbits implies that the maximum shift comes from a radius $R = R_o \sin \ell$, where $R_o \approx 8$ kpc is the radius of our own galactic orbit along with the sun. Because our line of sight along $\ell$ is tangent to an orbit at this radius, the inferred maximum velocity just depends on the difference between the orbital velocity at R and the projection of our own orbital motion along this direction,

$$V_{\text{rmax}}(\ell) = V(R) - V_o \sin \ell = \Omega(R)R - \Omega_o R_o \sin \ell = (\Omega(R) - \Omega_o)R_o \sin \ell,$$  \hspace{1cm} (20.3)
where $\Omega_o$ and $V_o = \Omega_o R_o$ are the angular and spatial velocity of the sun’s orbit at radius $R_o$. This can readily be solved to give the galactic rotation curve in terms of either the spatial or angular velocity

$$\Omega(R_o \sin \ell) = \frac{V_{rmax}(\ell)}{R_o \sin \ell} + \Omega_o ; \quad V(R_o \sin \ell) = V_{rmax}(\ell) + V_o \sin \ell.$$  \hfill (20.4)

The sun orbits the galaxy at a radial distance $R_o \approx 8\, \text{kpc}$ from the galactic center, with a speed $V_o \approx 220\, \text{km/s}$, implying then an orbital period $P_o \approx 220\, \text{Myr}$. (See QQ1.)

Applications of this approach to analyzing observations of the 21 cm line of atomic H yield the rather surprising result that, within most of the region within the sun’s galactic orbit, $R < R_o$, the orbital speed is nearly same as that of the sun, i.e.

$$V(R) \approx V_o \approx 220\, \text{km/s} ; \quad R < R_o,$$  \hfill (20.5)

which is known as a “flat” rotation curve.

Extension of this 21-cm method to longitudes $90 < \ell < 270$ that point outward to larger galactic radii $R > R_o$ is complicated by the need now to have an independent estimate of the distance to an observed Hydrogen cloud. But when this is done, the results indicate that the rotation curve remains nearly “flat”, with constant orbital speed, out to the farthest measurable radii, $R \lesssim 15\, \text{kpc}$. The left panel of figure 20.6 compares this observed flat rotation curve for our galaxy vs. what would be expected from Kepler’s law if the galaxy’s mass were as strongly centrally concentrated as its stellar luminosity.

This comparison illustrates why these flat rotation curves came as a surprise. Since most of galaxy’s luminosity comes from the central bulge within a radius $R_{\text{bulge}} \approx 1\, \text{kpc}$, it seemed reasonable to presume that most of the galaxy’s mass would be likewise contained within this central bulge. But this would then require that galactic orbital speeds should follow the same radial scaling as derived for orbits around other central concentrations of mass, like the planets around the sun. These follow the standard Keplerian scaling,

$$V_{\text{kep}}(R) = \sqrt{\frac{GM}{R}} \sim \frac{1}{\sqrt{R}},$$  \hfill (20.6)

which would thus decline with the inverse square root of the radius.

Instead, the constant orbital speed $V_o$ of a flat rotation curve implies that the amount of mass within a given radius must increase in proportion to the radius,

$$M(R) = \frac{V_o^2 R}{G} \sim R.$$  \hfill (20.7)
Fig. 20.6.— *Left:* Data for inferred galactic rotation speed (in km/s, points) vs. radius $R$ from the galactic center (in kpc), with horizontal green line showing data fit that implies a flat, or roughly constant, rotation speed for all radii $R > 5$ kpc. The red curve compares the decline of speed as $V(R) \sim 1/\sqrt{R}$ that is expected from Keplerian motion with a mass that is as centrally concentrated as the stellar light. The difference implies there is a substantial “dark matter” contribution to the mass for $R > 5$ kpc. *Right:* The top panel shows slit exposure for the negative image of an external galaxy viewed with its disk edge-on to the observer line of sight. The lower panel then shows the slit spectrum formed by plotting the wavelength spectrum of the star’s light along the vertical, against distance along the major axis of the galaxy on the horizontal axis. The flat bright emission vs. distance come from Doppler-shifted line emission lines that reflect the galactic rotation away from us on the left (longer wavelength) and toward us on the right (shorter wavelength). The flatness now shows quite directly that the rotation curve of this galaxy is also flat, as in our own Milky Way, again implying the presence of dark matter.

Since this extra gravitational mass extends to regions with very little luminosity, i.e. that are effectively very dark, it is known as *dark matter*. From studies extending up to scales well beyond our galaxy, to clusters and superclusters of external galaxies, it is now thought that there is about *five* times more dark matter in the universe than the ordinary luminous matter that makes up stars, ISM gas and dust, planets, and indeed us. The origin and exact nature of this dark matter is not known, but it is thought to interact with other matter mainly just through gravity, and not through the electromagnetic and (strong) nuclear force.
that plays such a key role in the properties of ordinary “baryonic” matter\(^6\)

Nonetheless, as discussed below, this dark matter is now thought to be crucial to the formation of large scale structure in the universe, and thus the associated galaxies that in turn provide the sites for formation of the stars, our sun, and the planets like our earth. In short, without dark matter, we wouldn’t be here today to wonder about it!

### 20.4. Super-massive black hole at galactic center

The center of our galaxy is in the direction of the constellation Sagittarius, at a distance of about 8 kpc. Over this distance the absorption by gas and dust in the disk plane contribute to some 25 magnitudes of visual extinction, completely obsurring this center in the visible parts of the spectrum. But at longer wavelengths in the infra-red and radio, for which the dust opacity is much lower, it becomes possible to see fully into the galactic center. Particularly noteworthy is Sagittarius A\(^*\) (a.k.a. Sgr A), a region of very bright radio emission.

**Quick Question 3: Angular vs. Physical Sizes at the Galactic Center**

At the distance \(d \approx 8\) kpc of the Galactic Center:

a. What is the physical size \(s\) (in AU) of an angle \(\alpha = 1\) arcsec?

b. What is the angular radius \(\alpha_c\) (in arcsec) of the central parsec cluster?

Infrared observations of the region around Sgr A shows a concentration of several hundred stars known as the “central parsec cluster”. The blurring effects of the earth’s atmosphere normally limit spatial resolution to angular sizes of order an arcsec. But using specialized techniques – known as “speckle imaging” and “adaptive optics” – it has become possible over the past couple decades to obtain IR images of individual stars within the central arcsec of the cluster, with angular resolution approaching \(\sim 0.1\) arcsec.

Monitoring of the couple dozen stars within this field since the mid-1990’s has revealed them to have small but distinctive *proper motions*, following *curved orbital* tracks that all center around a common point just slightly offset from the Sgr A radio source. Figure 20.7 illustrates the tracks of 10 stars over the 17-year period 1995-2012, with annual positions of the stars marked by dots, and individual stars identified by the color legend at the right.

\(^6\)Ordinary matter is often referred to as “baryonic” because most its mass comes from the protons and neutrons that are generally known as “baryons”. Technically though, a small fraction the mass of ordinary matter comes from electrons, which are actually classified as “leptons”, not a baryons.
Fig. 20.7.— Sub-arc resolution of central arcsec of Sgr A, showing orbital tracks of individual stars about a common central point, now identified as the location of a supermassive black hole of mass \( M_{bh} \approx 4 \times 10^6 M_\odot \). The individual dots show the annual positions of individual stars identified by the color code legend. The star S0-2 has been observed for more than its orbital period of 16.7 years.

Using the known \( d = 8 \) kpc distance, the angular sizes of the orbital tracks can be translated into physical sizes for the semi-major axes \( a \) of the orbits. Extrapolating (or following) the motion over a full cycle allows one to infer the orbit period \( P \). Application of this and the semi-axis into Kepler’s third law then gives an extremely large mass, \( M_{bh} \approx 4 \times 10^6 M_\odot \) for the central attracting object, which is inferred to be a super-massive black hole (SMBH). Exercise 1 illustrates the process for this mass determination.
Exercise 1: Using Kepler’s 3rd law to infer mass of SMBH

To compute the mass of the SMBH at the galactic center, consider the star labeled SO-02 in figure 20.7, which has recently completed a full, monitored orbit.

a. Using the arrow-key in the upper left showing the angular scale, estimate the angular extent (in arcsec) of SO-02’s projected major axis.

b. Using the known distance $d = 8$ kpc, what is the associated physical size $s$ (in AU) of the semi-major axis of SO-02’s orbit?

c. Next count the number of dots around the orbit to estimate the period $P$ (in yr) of SO-02’s orbit.

d. Assuming we have a face-on view of SO-2’s orbit, now use Kepler’s 3rd law to estimate the mass $M_{bh}$ (in $M_\odot$) of the central Black hole about which SO-2 is orbiting?

e. Suppose our view is off by a modest inclination angle $i$ from face-on. Does this increase, decrease, or have no effect on the mass estimate in part d? If it changes, by what factor?

More information on these stars in the central pc can be found at the website for the UCLA Galactic Center Group:

http://www.galacticcenter.astro.ucla.edu/
21. External Galaxies

21.1. Cepheid variables as standard candle for distances to external galaxies

What we now know as external galaxies, like our Milky Way but far outside of it, were first identified by their signature spiral form. Known merely as “spiral nebulae”, it was once thought they might be just stellar or cluster size regions like Planetary Nebulae, or the various other forms of diffuse nebulae seen in association with stars or star clusters.

The situation advanced considerably once telescopes became powerful enough to resolve individual stars within the great spiral nebula in Andromeda. In the 1920’s, using the 100-inch telescope on Mt. Wilson, Edwin Hubble was able to observe a particular kind of pulsating luminous giant star known as a Cepheid variable. Previous studies of Cepheid variables in our own Galaxy showed that they have the rather peculiar but very useful property that the period $P$ of their pulsation – which can be readily measured – is related to their intrinsic luminosity $L$. Using a Cepheid with a measured period as a luminous standard candle with a known luminosity, Hubble’s observation of the apparent brightness $F$ (actually apparent magnitude $m$) of Cepheid stars within the Andromeda nebula led him to estimate its distance $d$ as about a million light years, i.e. using the usual standard-candle formula

$$d = \sqrt{L/4\pi F}. \quad (21.1)$$

Since the Milky Way had already been inferred to have a diameter of only 100,000 light years, it was thus clear that Andromeda must lie well outside our galaxy, indeed with an angular size that implies it has a comparable physical size to the Milky Way itself.

Since this original application of Cepheid variables as standard candles, it has become clear that there are actually two distinct Cepheid classes: Types I and II, which apply respectively to Population I and II stars, with high and low metallicity. Figure 21.1 plots log $L/L_\odot$ vs. $P$ (days) for Type I and Type II Cepheids, showing that the former are about a factor four more luminous at a given period. Hubble incorrectly assumed that the Cepheids he initially used were of Type II, but they were actually of Type I. Accounting for the factor four higher luminosity within the observed apparent brightness implies that the Andromeda galaxy is actually twice as far as Hubble thought, i.e. some 2 Mly.

21.2. Galactic redshift and Hubble’s law for expansion

As Hubble applied his Cepheid method to measuring distances to other spiral nebulae, a Mt. Wilson observatory night assistant named Milton Humason, a former mule driver...
without even a high-school diploma, became especially skilled at measuring their spectra from very faint images on photographic plates. In particular, he was able to measure the Doppler shift of known spectral lines, giving then a direct measure of the galaxies’ radial velocity $V_r$.

Quite surprisingly, Humason found that, with the exception of the relatively nearby galaxies like Andromeda, all the more distant galaxies showed only redshifted spectral lines, implying from the Doppler shift formula that they are all moving away from us, with $V_r > 0$.

Even more remarkably, when combined with Hubble’s measurement of galactic distances, it lead to what is now known as the Hubble law, characterized by a linear proportionality between velocity $V_r$ and distance $d$,

$$V_r = H_0 d.$$  \hspace{1cm} (21.2)

The proportionality constant, $H_0$, is known as the Hubble constant, which has units of an inverse time. Figure 21.2 plots the original relations obtained by Hubble and Humason, with the slope of the red line fit to the data points giving $H_0$. Because of the incorrect assumption of the Cepheid type, along with a combination of other errors, the original value of nearly $H_0 \approx 500 \text{ (km/s)/Mpc}$ turns out to be a serious overestimate of the modern best value of $H_0 \approx 69 \text{ (km/s)/Mpc}$.

The implications of this Hubble law are truly profound. In particular, if we simply
assume that the velocity is constant, then dividing the distance by velocity gives the time since a distant galaxy was at zero distance from us,

\[ t = \frac{d}{V_r} = \frac{1}{H_o} \equiv t_H \approx 10 \text{ Gyr} \frac{100 \text{ (km/s)}/\text{Mpc}}{H_o}, \]  

where the second equality shows that this time, which is same for all galaxies, is given by the inverse of the Hubble constant. This is known as the Hubble time, \( t_H \equiv 1/H_o \), and as shown in the last relation of (21.3), a Hubble constant of \( H_o = 100 \text{ (km/s)}/\text{Mpc} \) gives a Hubble time of approximately \( t_H \approx 10 \text{ Gyr} \).

Modern observations of very distant galaxies show that the redshift,

\[ z \equiv \frac{\lambda_{\text{obs}}}{\lambda} - 1 = \frac{V_r}{c}, \]  

(21.4)
can become quite large, with even some cases having \( z > 1 \). If taken literally in terms of the latter velocity Doppler shift formula in (21.4), this would seem to suggest that \( V_r > c \), in apparent contradiction of special relativity.

But as discussed in the cosmology sections in part 4, a more proper interpretation of this “cosmological redshift” is that it represents the stretching of the wavelength of light by the expansion of space itself! This can readily lead to redshifts \( z > 1 \). Einstein’s limit really applies to how fast objects can travel relative to space, but that space itself can expand at a speed faster than light!
21.3. Tully-Fisher Relation: \( L_{gal} \propto V_{rot}^4 \)

For more distant galaxies, it becomes increasingly difficult to detect and resolve even giant stars like Cepheid variables as individual objects, limiting their utility in testing the Hubble law to relatively modest distances and redshifts. For much larger distances, we need another, brighter “standard candle”, like White Dwarf supernovae. But because the unpredictability of their appearance long limited the number of such WDSN\(^7\) detections, an importance alternative method has been the so-called *Tully-Fisher* relation. Empirically, it was found that the luminosity of a spiral galaxy, \( L_{gal} \), scales with maximum rotation velocity \( V_{rot} \) inferred from Doppler shift of spectral lines, with the approximate form

\[
L_{gal} \propto V_{rot}^4. \tag{21.5}
\]

The proportionality constant depends on the spectral band, but as an example, in the near-infrared “I-band” (centered at 820 nm), the relation in terms of the absolute magnitude takes the numerical form,

\[
M_I \approx -3.3 - 8.3 \log \left( \frac{V_{rot}}{\text{km/s}} \right). \tag{21.6}
\]

Since magnitude \( M \propto -2.5 \log L \), the slope of -8.3 here implies a velocity exponent \( 8.3/2.5 \approx 3.3 \), somewhat shallower than the power 4 assumed in eqn. (21.5). Figure 21.3 shows actual I-band magnitude data vs. \( \log(V_{rot}) \), compared with linear relations with slope -8.3 (blue) and -10 (red), the latter corresponding the standard Tully-Fisher law (21.5) with velocity exponent \( 4 = 10/2.5 \).

**Exercise 1:** Application of Tully-Fisher relation.

A spiral galaxy with redshift \( z = 0.23 \) and apparent I-band magnitude \( m_I = +17.6 \) has an observed total spectral line width ratio, \( \Delta \lambda/\lambda = 0.0013 \). Compute the galaxy’s:

a. Orbital velocity \( V_{rot} \);

b. Absolute I-band magnitude \( M_I \);

c. I-band luminosity \( L_I \), in units of the I-band luminosity of the sun, \( L_{I,\odot} \) (for which the I-band absolute magnitude is \( M_I \approx +4 \)).

d. Distance modulus \( m_I - M_I \);

e. Distance \( D \) (in Mpc).

f. Recession velocity \( V_r \) from redshift \( z \);

g. Associated Hubble constant \( H_o = V_r/D \).

\(^7\)In the standard, formal notation, WDSN are classified as “Type Ia”, or “SN Ia”
Fig. 21.3.— Empirical Tully-Fisher relation in the infrared, plotted as I-band absolute magnitude $M_I$ vs. logarithm of the total line-width (in km/s), set by twice the rotational velocity, $2V_{rot}$. The blue line shows best fit line with slope $-8.3$, as given in eqn. (21.6), corresponding to velocity exponent $8.3/2.5 \approx 3.3$, slightly shallower than the slope $4$ assumed in the standard log $L_{gal}$ vs. log $V_{rot}$ scaling law of eqn. (21.5). The red line shows the slope $-10 = -2.5 \times 4$ that would be implied in the I-band magnitude scaling by this standard form for the Tully-Fisher relation. The best-fit slope can differ for different wavebands.

To glean a possible physical rationale for this empirical Tully-Fisher relation, first note again that by Kepler’s law the rotational velocity $V_{rot}$ at an outer radius $R$ scales with galactic mass $M_{gal}$ as

$$V_{rot}^2 = \frac{GM_{gal}}{R}. \quad \text{(21.7)}$$

On the other hand, the galactic luminosity $L_{gal}$ scales with the galaxy surface brightness $I_o$ times the surface area $\pi R^2$ out to this outer radius,

$$L_{gal} \propto I_o \pi R^2. \quad \text{(21.8)}$$

Combining (21.7) and (21.8) gives the scaling

$$L_{gal} \propto \frac{V_{rot}^4}{I_o (M_{gal}/L_{gal})^2}, \quad \text{(21.9)}$$
which recovers the standard Tully-Fisher scaling of eqn. (21.5) if we assume a constant value for the surface brightness times the square of the mass-to-light ratio, $I_o (M_{gal}/L_{gal})^2$. Models of galaxy formation have tried to explain why this should be true, but the results are tentative and not clearly established and accepted. Nonetheless, as a strictly empirically calibrated relation, this Tully-Fisher scaling provides a luminous standard candle to infer distances beyond the range accessible to the Cepheid method, and so allows a calibration of the Hubble law to moderately large distances and redshifts.

21.4. Spiral, Elliptical, & Irregular galaxies

Fig. 21.4.— Examples of 3 types of elliptical galaxies; figure taken from http://cas.sdss.org/dr6/en/proj/basic/galaxies/ellipticals.asp.

Galaxies can be generally classified by three distinct types of morphology: Spiral, Elliptical, and Irregular.

Spiral galaxies are similar to our Milky Way, with distinct disk, halo, and bulge components. A spiral density wave in the disk forms the spiral arms that are the regions of active star formation out of the cold clouds of gas and dust. The tightness of the winding of the arms can vary, and sometimes emanate from a central “bar”. M51, a.k.a. the “Whirlpool” galaxy, shown in figure 18.5, provides a good example of a typical spiral galaxy. As illustrated
in fig. 20.2, our Milky Way galaxy is thought to be a barred spiral.

Elliptical galaxies have a spheroidal shape, with different gradations of elongation from nearly spherical (E0) to highly extended (E5), as illustrated in figure 21.4. Their stars are generally found to be Population II, and thus quite old with reduced metallicity. There appears to be a near absence of ISM gas or dust, and thus little or no new star formation. In these respects, elliptical galaxies are similar to globular clusters that orbit in the halo of our Milky way, but much bigger and more massive. Their physical sizes can span a large range, from about 0.1 to 10 times size of the 100,000 l.y. diameter of our Milky way, i.e. only $10^4 \text{ l.y.}$ for “Dwarf ellipticals” (with $M \sim 10^9 \, M_\odot$), to $10^6 \text{ l.y.}$ for giant ellipticals (with $M \sim 10^{12} \, M_\odot$). At the center of a very large cluster of galaxies, there is often a giant, “central dominant” (CD) elliptical galaxy that can have mass of $10^{12} \, M_\odot$ or more.

Irregular galaxies are just that. The overall structure is complex, though within sub-areas there can be spiral features. In many cases, it seems likely that the irregular form is because we are actually viewing two colliding galaxies, with then their mutual tidal interaction warping and disrupting whatever symmetric forms may have existed in the source galaxies. Figure 21.5 shows a mosaic of interacting galaxies (right), and a close-up the direct collision underway in the Antenna galaxy (left).

Fig. 21.5.— Left: The antenna galaxy, which is actually two galaxies undergoing a direct collision. Right: Gallery of other examples of interacting galaxies.

For any collection of objects of size $s$ separated by a mean distance $d$, the number
density is \( n \approx 1/d^3 \) while the cross section is \( \sigma \approx s^2 \). The mean free path for collision is then

\[
\ell = \frac{1}{n\sigma} = d \left( \frac{d}{s} \right)^2 .
\]

(21.10)

For individual stars, the distance/size ratio is enormous, of order \( d/s \approx \text{pc}/R_\odot \sim 4 \times 10^7 \), implying a mean-free-path \( \ell \sim 10^{15} \text{ pc}! \) Since this is more than \( 10^{10} \) larger than the \( \sim 30 \text{ kpc} \) size of a galaxy, we can conclude that individual field stars in galaxy should \textit{never} collide\(^8\).

But for galaxies, this ratio of distance/size is much smaller, about a factor 20 for us to the Andromeda galaxy, and often just a factor few for galaxy clusters. Moreover, in the early universe, the average separation among galaxies not in the same cluster was smaller, with the factor reduction just set by the redshift, \( z+1 \). As such, while somewhat rare in the current-day universe, collisions can and do occur, and they were much more common in the early universe. Indeed, some models invoke a “bottom up” scenario in which larger galaxies form from the merger of smaller galaxies.

Animations from computer simulations of two colliding galaxies can be found on web at:

\texttt{http://www.youtube.com/user/galaxydynamics}

The video dubbed “Spiral Galaxy” shows how spiral density waves can be induced by orbiting clumps of dark matter.

When galaxies do collide, their overall pattern of stars become strongly distorted by the mutual tidal interaction of the overall mass of the two galaxies; but the individual stars are too widely separate to collide, and so just pass by each other. In contrast, any gas clouds in the ISM of each galaxy do collide, with the resulting compression increasing the density of gas and dust, and thus often triggering a strong burst of new star formation. Such colliding systems are indeed often dubbed “starburst galaxies”.

While distant galaxies show a redshift that implies they are moving away from us as part of the expansion of the universe, the mutual gravitational attraction between our Milky Way and the relatively nearby Andromeda galaxy is actually pulling them toward each other. Indeed, it now seems likely that Andromeda and the Milky Way will collide in about 3 Gyr. The “Future Sky” animation in the above link shows how the sky might appear from Earth during this collision.

---

\(^8\)Some gravitational interaction can occur in the dense cores of compact globular clusters, but even there direct collision between stars is very unlikely.
22. Active Galactic Nuclei (AGNs) and Quasars

22.1. Basic properties and model

During the 1960’s sky surveys with radio telescopes discovered “QUasi-StellAr Radio” sources, now known as “Quasars”, or also Quasi-Stellar Objects (QSOs). In contrast to the extended radio emission sources seen from various regions of the galaxy, these QSOs are, like stars, point-like sources without any readily discernible angular extent. They were soon identified with similarly point-like sources in the visible and other wavebands. But quite unlike stars, their spectral energy distribution does not even roughly match that of a Black-body of any temperature; instead it has an extended power-law form over a wide range of energies from the radio through the IR, visible, UV and even extending to the X-ray and gamma-rays.

Nonetheless, this broad spectral distribution does still show patterns of absorption (and emission) lines that can be identified with known elements, but notably with a huge redshift $z$. For example, 3C273, one of the first and most famous QSOs, has $z = 0.158$, meaning that it is receding from us at a radial speed $v_r = 0.158c \approx 47,000 \text{km/s}$. Taking a Hubble constant $H_o \approx 70 \text{(km/s)/Mpc}$, this puts its distance at $d \approx v_r/H_o \approx 677 \text{Mpc} \approx 2.1 \text{Gly}$. With associated distance modulus $m - M = 5 \log(d/10 \text{pc}) \approx 39$, together with its apparent magnitude $m = +15$, this implies an absolute magnitude $M \approx -24$, or luminosity $L \approx 5 \times 10^{11} L_\odot$! This far exceeds the luminosity of any star, and indeed even outshines the luminosity of a typical galaxy of $\sim 10^{11} L_\odot$.

Modern observations, e.g. with the Hubble Space Telescope, revealed that these quasars are commonly surrounded by a comparatively faint, diffuse stellar emission from a host galaxy. It is now realized that QSOs are indeed just one example of a class of “Active Galactic Nuclei” (AGNs). In contrast to the extended galactic emission over a distance of a galactic diameter $\sim 30 \text{kpc}$, QSO/AGN emission is entirely point-like, emanating from the galactic nucleus. Indeed, since such QSO/AGNs often vary over time scales as short as a day, they must be very compact, no more than a light-day in diameter, or $\lesssim 100 \text{AU}$; this means they are roughly of order $\sim 10^8 \sim 30 \text{ kpc}/100 \text{AU}$ times smaller than their host galaxy.

This extreme luminosity from such a small volume is thought to be the result of matter accreting onto the supermassive black hole (SMBH) at the center of the QSO/AGN host galaxy. The SMBH in our Milky Way, and indeed in most galaxies in the nearby, current-day universe, are relatively inactive, with little or no ongoing accretion. But in the early universe, when the smaller inter-galactic separation meant more frequent galaxy collisions, the extreme disruption causes some stars to approach so close to the SMBH that they become tidally disrupted. The remnant stellar material typically still has too much angular
momentum to fall directly onto the SMBH, and so instead feeds an accretion disk. The viscous shear transports angular momentum outward, allowing a steady, gradual accretion in which the gravitational energy released heats the disk and powers its emitted luminosity.

**Quick Question 1:** Energy efficiency for accretion near a black hole

For accretion down to a radius that is a factor $R_{\text{acc}}/R_s$ times the Schwarzschild radius of black hole, compute the energy efficiency factor $\epsilon = E_g/mc^2$ for the gravitational energy gain $E_g$ as a fraction of the rest mass energy $mc^2$ of the accreted mass. Confirm that $\epsilon = 0.1$ for $R_{\text{acc}}/R_s = 5$.

Accretion down to the vicinity of a black hole can generate energy that is a substantial fraction of the rest mass energy of the accreting matter. (See QQ 1.) For an accretion rate $\dot{M}_{\text{acc}}$ and conversion efficiency $\epsilon$, the generated luminosity is

$$L_{\text{acc}} = \epsilon \dot{M}_{\text{acc}} c^2 = 1.4 \times 10^{12} \frac{L_\odot}{0.1} \frac{\dot{M}_{\text{acc}}}{M_\odot/\text{yr}}.$$  

(22.1)

The second equality shows the very enormous luminosity associated with accretion of $1 M_\odot/\text{yr}$ at a efficiency of $\epsilon = 0.1$ (the value for accretion to 5 Schwarzschild radii). It indeed readily equals or exceeds the luminosity inferred from the observed apparent magnitude and estimated distance of QSOs, including the example of 3C 273 mentioned above.

The SMBHs that power quasars are thought to be even more massive than those found in our and other nearby galaxies, of order a billion solar masses ($10^9 M_\odot$). But the associated Schwarzschild radii, $R_s \sim 3 \times 10^9$ km $\sim 20$ AU are still small enough to accommodate the day-timescale variation, even accounting for the fact that the emission region is likely to extend over $5 - 10 R_s$.

22.2. Lyman alpha clouds

As this enormous luminosity from distant quasars propagates through the universe, it can sometimes pass through the relatively higher density gas associated with galaxies or a galaxy cluster. Since the quasar spectral distribution extends well into the UV, the photons at wavelengths $\lambda = 121.57$ nm for the Lyman alpha (Ly-$\alpha$; $n = 1$ to $n = 2$) transition of neutral Hydrogen, for which the opacity is very high, become strongly absorbed. But along this extended Gly path length, the local Ly-$\alpha$ wavelength is Doppler shifted by the Hubble expansion, extending it to a longer wavelength that depends on the distance to the absorbing inter-galactic H-cloud. As illustrated in figure 22.1, this makes the observed quasar spectrum
have a distinct number of absorption lines. Indeed, sometimes these are so dense that they are known as the Lyman-alpha "forest", with each “tree” of the forest corresponding to a distinct inter-galactic H cloud at a distance set by the cosmological (Hubble-law) red-shift of that observed absorption feature.

In essence, the huge luminosities and huge distances of quasars provide us a set of “flashlights” to probe the inter-galactic Hydrogen gas in the universe between us and the quasars.

Quick Question 2: Lyman cloud speed and distance

Suppose a quasar shows absorption from a Lyman-alpha cloud at an observed wavelength $\lambda_{\text{obs}} = 183$ nm.

a. What is the redshift $z$ for this cloud.
b. What is its inferred recession speed $v_r$?
c. For a Hubble constant $H_0 = 67$ (km/s)/Mpc, what is its distance?

Fig. 22.1.— Illustration of Lyman-α clouds, in which Hydrogen gas in galaxies at various redshifts absorb distant quasar light in the Lyman-α line from the $n = 1$ to $n = 2$ transition of Hydrogen.
22.3. Gravitational lensing of quasar light by foreground Galaxy Clusters

As this enormous luminosity from distant quasars propagates through the universe, it can also sometimes pass so close to a galaxy cluster that the gravity from the cluster’s mass actually bends the rays of light, forming what is known as a “gravitational lense”. This basic effect of gravitational bending of light was predicted by Einstein’s General Theory of Relativity, and was famously confirmed by expeditions to measure the associated shift in the position of stars as their light passed near the sun during a solar eclipse. In the context of the passage of quasar light by a galaxy cluster, it can lead to multiple images, or even an “Einstein arc” or circle, if the quasar and galaxy’s mass-center are both closely aligned with the observer’s light of sight. Figure 22.2 illustrates the basic geometry and process.

From General Relativity, the bending angle $\theta$ depends on the mass $M$ of the lens and the “impact distance” $b$ of the light ray from the background source passing the lensing mass, 

$$
\theta = \frac{4GM}{bc^2}.
$$

(22.2)

The factor 4 is a general relativistic correction factor for the simple Newtonian calculation for bending of an object with incoming speed $V_x = c$, as illustrated in figure 22.3.

Exercise 1: Gravitational Lensing

Suppose a distant galaxy cluster with redshift $z = 0.2$ has two identical quasar images at equal angles $\theta = 10$ arcsec on each side of the cluster center.

a. Assuming the current best value for Hubble constant, $H_o = 67$ (km/s)/Mpc, what is the distance $D$ (in Mpc) to the lensing galaxy?

b. Use this distance $D$ and the angle $\theta$ to estimate the closest distance $b$ (in kpc) that the quasar’s light passes to the center of the galaxy.

c. Now use this $b$ and the angle $\theta$ in Einstein’s gravitational lensing formula to estimate the mass $M$ (in $M_\odot$) of the lensing galaxy cluster.

22.4. Gravitational redshift

Another effect related to gravitational lensing is the gravitational redshift experienced by light emitted from a radius $R_o$ near a gravitational mass $M$, which from General Relativity is given by 

$$
z = \frac{1}{\sqrt{1 - \frac{R_s}{R_o}}} - 1,
$$

(22.3)

where $R_s \equiv 2GM/c^2$ is the Schwarzschild radius for the mass $M$. (See §18.2.4 and eqn. 18.10.) As illustrated in figure 22.4, for the non-relativistic case that the initial radius is
Fig. 22.2.— Diagram to illustrate “gravitational lensing”, wherein the image of a distant galaxy or quasar is multiplied by the gravitational bending of light from passage near the very large mass of an intervening galaxy or galaxy cluster.

far above the Schwarzschild radius, $R_o \gg R_s$, straightforward Taylor expansion leads to a simple form that casts this photon redshift in terms of simple conservation of total energy of the photon plus gravity, with the loose association of an equivalent initial photon “mass” $m_o = E_o/c^2$ based on Einstein’s energy-mass equivalence principle.

**Exercise 2:** *Gravitational redshift as alternative explanation of quasar redshift*

Suppose we try to explain the redshift of 3C273 ($z=0.158$) as a gravitational redshift, rather than being from cosmological expansion.

a. Relative to Schwarzschild radius $R_s$, from what radius $R_o$ is the radiation emitted?

b. If the width of lines is 0.1% of their central wavelength, what is the range of
Fig. 22.3.— Diagram to illustrate the gravitational deflection of an object with initial speed $V_x$ impacting within a distance $b$ of a gravitational mass $M$. The same scaling applies to light with speed $V_x = c$, but with a correction factor 4 derived from General Relativity, giving then the correct scaling for gravitational bending of light.

radii (relative to $R_o$) from which the radiation can be emitted.

For a more physically reasonable assumption that any emission would come from at least radius range $\pm 10\%$ around the central radius $R_o$, what would be the relative width $\Delta \lambda/\lambda_o$ of the observed emission line?
22.5. Apparent “super-luminal” motion of quasar jets

The accretion that powers the tremendous luminosity of quasars can also drive relativistic jets from the polar axes perpendicular to the accretion disk. The jets emit in energies from the radio to gamma-rays, but can be most finely resolved (to angular resolution less than a milli-arcsecond!) spatially in the radio, using Very Long Baseline Interferometry (VLBI) from multiple radio telescopes spread across the earth. Indeed, such VLBI radio observations of such jets show they can be quite clumpy and variable on time scales of weeks to years. Quite remarkably, individual clumps in these quasar jets can sometimes show an apparent “super-luminal” motion, meaning that, for the inferred quasar distance, the propagation of individual jet clumps away from the quasar can appear to be faster than the speed of light!
As illustrated in figure 22.5, the motion is actually a fraction \( \beta = v/c \lesssim 1 \) that is near but below light speed \( c \), but with a direction toward the observer that changes the light travel time in such a way to make it appear the transverse motion is faster than light. For actual light speed fraction \( \beta < 1 \) at an angle \( \theta \) from the direction to the observers, the apparent light speed fraction is given by

\[
\beta_{\text{app}} = \frac{\beta \sin \theta}{1 - \beta \cos \theta}.
\] (22.4)

The special case with \( \beta = \cos \theta \) gives the maximum apparent speed, which in units of the speed of light is

\[
\beta_{\text{app}}^{\text{max}} = \frac{\beta}{\sqrt{1 - \beta^2}}.
\] (22.5)

From this it is clear that apparent super-luminal propagation \( \beta_{\text{app}}^{\text{max}} > 1 \) is possible whenever the propagation speed \( v = c\beta > c/\sqrt{2} = 0.707 \, c \).

**Exercise 3: Apparent super-luminal motion in a quasar**

Suppose VLBI radio monitoring shows that over 5 years a jet subcomponent of a quasar at a known distance \( D = 1 \) Gpc has moved away from the center by an angle \( \Delta \alpha = 10^{-3} \) arcsec.

a. For the quasar distance \( D \), what is the inferred apparent transverse speed of this component compared to the speed of light, \( \beta_{\text{app}} = V_{\text{app}}/c \)?

b. What is the minimum actual fraction of the speed of light \( \beta = V/c \) needed to give this apparent super-luminal speed \( \beta_{\text{app}} \)?

c. What is the associated angle \( \theta \) (in radians and degrees) between the component’s motion and our line of sight?
Apparent Super-Luminal motion of QSO jets

\[ \beta_{\text{app}} = \frac{V_{\text{app}}}{c} = \frac{\beta v \sin \theta / c}{1 - \beta \cos \theta} = \frac{\beta \sin \theta}{1 - \beta \cos \theta} = \beta_{\text{app}} \]

\[ 0 = \frac{d\beta_{\text{app}}}{d\theta} = \frac{\beta \cos \theta - \beta \sin^2 \theta}{1 - \beta \cos \theta} \]

\[ \text{max at } \cos \theta = \beta \]

\[ \beta_{\text{app}}^{\text{max}} = \frac{\beta \sqrt{1 - \beta^2}}{1 - \beta^2} = \frac{\beta}{\sqrt{1 - \beta^2}} \]

Fig. 22.5.— Derivation to show how quasar jet motion near the speed of light in a direction tilted toward the observer (on right) can lead to an apparent “super-luminal” (faster than light) propagation speed away from the quasar. The maximum apparent speed \( \beta_{\text{app}}^{\text{max}} \) (in units of the light speed \( c \)) occurs when the actual light speed fraction \( \beta = v/c < 1 \) equals \( \cos \theta \), the projection of the jet direction toward the observer.
23. Large Scale Structure (LSS)

23.1. Galaxy clusters & super-clusters

Much as stars with galaxies tend to form within stellar clusters, the galaxies in the universe also tend to collect in groups, clusters, or even in a greater hierarchy of clusters of clusters, known as “super-clusters”. Our own Milky Way is part of a small cluster known as the “Local Group”, which includes also the Andromeda galaxy, as well as up several dozen smaller, “dwarf” galaxies. Along with roughly a hundred or so other groups, this makes up the “Local supercluster”, with the highest concentration in the direction of the constellation Virgo. That concentration is also known as the Virgo (super)cluster, at distance of about 20 Mpc, but it’s outer extent could be even be defined to include the local group, making it a possible center of the local super-cluster. In any case, this is just one of millions of super-clusters in the known universe.

Over the past couple decades there have several very large surveys that aim to measure the redshift of a large number (nowadays reaching many millions!) of galaxies along selected swaths of the sky. Over these large expanses of the universe, this measured redshift \( z \) gives, for a known Hubble constant \( H_0 \), a direct measure of the distance \( D \) to the galaxy,

\[
D \approx z \frac{c}{H_0} = z D_o, \quad z \ll 1,
\]

(23.1)

where the latter equality defines the “Hubble distance” \( D_o \equiv c/H_0 \), which is just the distance that light travels over a characteristic Hubble time, \( t_H \equiv 1/H_0 \) (cf. §21.2). (This simple relation only applies for modest redshifts \( z \ll 1 \); as discussed in the next section, for \( z \geq 1 \), there is a more general relation in terms of the change in the universe’s scale factor \( R(t) \).

With the readily measured two-dimensional (galactic longitude and latitude) positions on the sky, a 2D survey along a swath on the sky can be combined with the redshift distance to form a 3D picture of the universe through that swath. The upper left panels of figure 23.1 (blue color) show a slice of this 3D picture containing one dimension of galactic position plus the distance, arranged along the radius from our own observer’s position at the origin. The result shows a remarkable “cosmic web” in the overall large-scale structure (LSS) of the universe, with a concentration of galaxies along extended, thin “walls”, with huge voids with few or no galaxies in the huge volume between the walls, but a particularly high concentrations at the intersections of the walls. Indeed, most previously identified super-clusters can be associated with one of these wall intersections.

The lower right panels of figure 23.1 (red color) show the results of very large simulations for the formation of the structure from the gravitational attraction by matter. For one such simulation, figure 23.2 shows a sequence of volume renderings at different stages of the
formation of structure, identified by the redshift $z$ associated with each epoch. As shown in the upper left panel for the earliest phase of the simulations at a redshift $z = 27.30$, one also requires a small initial seed of density fluctuations, which are then amplified by the gravitational attraction. As discussed in the next section, it is now thought that this initial seed of small-amplitude variations in density is provided by quantum fluctuations in the very early phases of the big-bang itself!

Fig. 23.1.— Comparison between observational surveys (blue, top and left) vs. computer simulations (red, bottom and right) of the large scale structure of the local universe. The observational surveys measure position and redshifts of millions of galaxies along an extended, narrow arc of the sky, using the redshift $z$ to estimate galactic distance. The simulations assume initial seed perturbations set to correspond to those inferred from Cosmic Microwave Background (CMB) fluctuations, plus cold dark matter to enhance gravitational attraction, and then compute gravitational contraction of structure starting from nearly uniform early universe at redshift $z > 25$ to the present, highly structured, local universe with redshift $z < 0.25$. 
Fig. 23.2.— Computer simulations of evolution of large-scale structure of universe, beginning with the nearly smooth, early universe at redshift \( z = 27.36 \) (upper left) to the extensive structure in the current local universe at \( z = 0 \) (lower right).

### 23.2. Dark matter: Hot vs. Cold, WIMPs vs. MACHOs

A key result of these simulations is that achieving an LSS that has the same statistical form as the observed structure requires inclusion of a significant component of *cold dark matter* (CDM), with a total mass that is factor several (∼5) times the mass of ordinary matter that makes up planets, stars, and galaxies that produce the various spectral bands of electromagnetic radiation that we can directly observe. “Cold” here means that the matter is non-relativistic, so that its gravitational contribution comes from its rest mass, and not from any relativistic enhancement in its energy. Only CDM seems able to form the deep gravitational wells from mutual attraction to frame the observed wall+void network of
galaxies observed for large-scale-structure. Hot dark matter tends to remain too distributed.

There are two candidates for CDM, dubbed by the somewhat whimsical terms “WIMPs” – for Weakly Interacting Massive Particles – and “MACHOs” – for Massive Compact Halo Objects. The latter refer to a conjectured large population of low-mass objects – perhaps roving Jupiter size bodies that are too cold to emit much radiation – thought to occupy the core and halo of galaxies. To explain the flat rotation curves of galaxies, the number density of such MACHOs would have to be so large that as they randomly pass in front of the stars they should induce a gravitational “micro-lensing” event that should be observable from monitoring of the star’s light. Extensive monitoring surveys have indeed detected such micro-lensing events, but at a rate that is well below what would be needed for MACHOs to be a significant component of dark matter mass.

There is thus now a general consensus that CDM most likely consists of some kind of WIMP. “Weakly interacting” in this context means they are not subject to either the strong nuclear force, which binds the nucleus of atoms, or electromagnetic forces, which bind electrons to atoms, and are responsible for producing light and all other forms of electromagnetic radiation. The inability to produce light is indeed what makes WIMPs a candidate for dark matter. Like neutrinos, they might be subject to the weak nuclear force, but otherwise they only interact with ordinary matter via gravity. Moreover, while neutrinos have a rest mass only only a few eV, a hypothetical WIMP could have a much larger mass, perhaps many hundreds time the GeV mass of protons and neutrons. There are several projects underway to detect WIMPs, through experiments deep underground, which shields against the flux of cosmic rays that would otherwise contaminate detections of the very few weak interactions by WIMPs.