Chapter 2
Nebulae

Abstract This chapter discusses the basic concepts and properties of ionizing nebulae. In Sect. 2.1 we describe hydrogen ionization and recombination occurring in an ionizing region highlighting the on-the-spot approximation which is frequently used by researchers. In Sect. 2.2 we give a brief overview of the categories of nebulae including an example in each case. In Sect. 2.3 we analyze the physics behind photoionization equilibrium and in Sect. 2.4 we provide the basic heating and cooling mechanisms occurring in an H II region. In Sects. 2.5 and 2.6 we discuss about the dynamical expansion of an H II region and how this leads to a strong shock front formation.

2.1 Ionization and Recombination of Hydrogen

Let us suppose there is a cosmic gas which consists of hydrogen. As we saw in the previous chapter, the ISM has various heating sources which increase its temperature. At a fixed density and as the temperature increases, molecular hydrogen dissociates producing atomic hydrogen (H I) according to the reaction

$$\mathrm{H_2} \longrightarrow 2\mathrm{HI} - 4.5\,\mathrm{eV} \quad (2.1)$$

where 4.5 eV is the energy needed in order to dissociate this hydrogen molecule. In case the gas is close to a source emitting ionizing radiation, the photons consisting this radiation interact with H I and they ionize it according to

$$\mathrm{HI} \longrightarrow \mathrm{HII} + \mathrm{e}^- - 13.6\,\mathrm{eV} \quad (2.2)$$

which creates a free electron and a free proton (ion) while absorbing energy equal to 13.6 eV.

In case a free electron and a free proton are close to each other, they may recombine to form a new H I producing either a single (but highly energetic) photon, or multiple (but lower energetic) photons. This is because during this recombination process the electron may recombine (a) directly to the ground state or (b) to any higher state but then cascading down until it reaches the ground state. In the first case, all 13.6 eV that the free electron has in the form of kinetic energy will be

T.G. Bisbas, *The Interstellar Medium, Expanding Nebulae and Triggered Star Formation*, SpringerBriefs in Astronomy, DOI 10.1007/978-3-319-26142-3_2

converted to a photon having the same amount of energy and which corresponds to the UV radiation. In the second case, various other photons may be emitted, some of them in the visible spectrum.

The *recombination coefficient*, α, describes the volumetric amount of recombinations that can be achieved in time and it has units of volume per unit time (i.e. in cgs it is $\mathrm{cm^3\,s^{-1}}$). When multiplying this coefficient with the number of electrons and the number of protons (i.e. $n_e n_p \alpha$ which is simply $n_p^2\alpha$ for H), we obtain the *recombination rate* per unit volume per unit time. The total recombination coefficient (frequently referred to as "case-A") for a hydrogen atom is given by

$$\alpha_{\mathrm{A}} = \sum_{N=1}^{\infty} \alpha_N(\mathrm{HI}, T), \tag{2.3}$$

where N is the principal quantum number. Note that the recombination coefficient depends on the temperature (the higher the temperature, the smaller the number as it is more "difficult" for a recombination event to occur). This equation takes into account *all* possible recombination states i.e. from $\infty \rightarrow 1$, or $\infty \rightarrow 2 \rightarrow 1$ etc.). Since recombinations *directly* to the ground state can produce a photon capable to ionize a nearby hydrogen atom *diffusively*, it is convenient to neglect this recombination and thus to reduce the complexity of the system of equations when examining the thermodynamics and evolution of nebulae as we will see later on in this chapter. When neglecting that specific recombination (frequently referred to as "case-B"), the recombination coefficient takes the form

$$\alpha_{\mathrm{B}} = \alpha_{\mathrm{A}} - \alpha_1(\mathrm{HI}, T) = \sum_{N=2}^{\infty} \alpha_N(\mathrm{HI}, T). \tag{2.4}$$

The above approximation is well known as the *on-the-spot* approximation (OTS; Osterbrock 1974) and is very commonly used. By invoking the OTS approximation, we simply neglect the contribution of the diffuse component of the radiation. The temperature-dependent value for the case-B recombination coefficient frequently used is $\alpha_{\mathrm{B}} \simeq 2.7 \times 10^{-10} \left(\frac{T}{[\mathrm{K}]}\right)^{-3/4}\,\mathrm{cm^3\,s^{-1}}$. For the ionizing hydrogen at a temperature of $T_{\mathrm{i}} = 10^4$ K, we obtain $\alpha_{\mathrm{B}} = 2.7 \times 10^{-13}\,\mathrm{cm^3\,s^{-1}}$.

The on-the-spot approximation tells us that *any* region that is shaded will *never* be ionized.

To understand further the effects that may be caused by diffusive radiation, Fig. 2.1 plots a schematic diagram showing such a simplified example. Here, an ionizing source (black dot at the bottom) emits ionizing radiation which interacts with the hydrogen gas around it creating free protons and free electrons (white region). As we will discuss further below (see Sect. 2.6), this ionized region expands

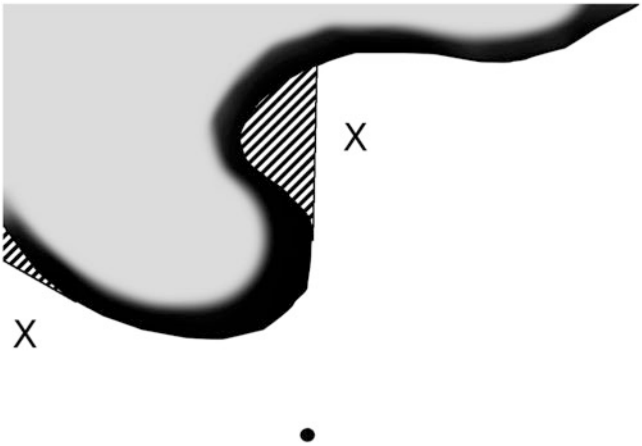

Fig. 2.1 A region interacting with an ionizing source (*black dot*). The *white area* in this diagram corresponds to the region where all hydrogen atoms are ionized. The *black region* corresponds to the shock front occurring due to the expansion of the ionized medium (see Sect. 2.6) and where the material is quite dense and absorbs all ionizing photons leaving all gas behind it undisturbed (*gray region*). The *stripped areas* correspond to the region where the diffuse component of radiation (occurred nearby such as at both places marked with "X") is the only way to ionize it. By invoking the OTS approximation, these *stripped regions* will *never* be ionized

and creates a shock front (black region). The ionizing photons cannot penetrate further and therefore behind the shock front the medium is undisturbed (gray region). The stripped region indicates ionization due to the diffuse component of radiation. Suppose that a recombination event occurs at places marked with "X". If the recombination of the free electron occurs *directly* to the ground state, then an ionizing photon will be produced—as we described earlier—which will be emitted towards a random direction. If this direction is towards the stripped region and interacts with a hydrogen atom therein, it will ionize it *diffusively* i.e. without the latter interacting directly with the ionizing source. By invoking the OTS approximation, we neglect these contributions, therefore under this assumption the stripped region will no longer be able to be ionized. It is therefore important to note also that if a liberated electron does not recombine directly to the ground state, it cannot create a photon capable to ionize a hydrogen atom and is therefore lost.

2.2 Categories of Nebulae

Nebulae[1] are regions which consist of ionized gas, particularly of hydrogen. The source of ionization is usually a single or multiple massive stars. Massive stars, as we have discussed in Chap. 1 emit copious amounts of ultraviolet radiation with

[1]Meaning "cloud" in Latin.

photons carrying more energy than the ionization potential of hydrogen (13.6 eV). Suppose that an ionizing photon hits a hydrogen atom. Upon this interaction, the photon is absorbed by the atom which in turn destroys its chemical state by forming a free electron and a free proton. The electron produced by the absorption of these UV photons are called *photoelectrons*. Any surplus in energy exceeding the ionization potential of 13.6 eV will be transferred to the photoelectron in the form of kinetic energy thus increasing the temperature of the ionized region. A characteristic feature of all nebulae is that the temperature in the ionized gas tends to be much higher than the temperature in the surrounding neutral gas. These two gas states are separated by a thin slice known as the *ionization front*. In this slice, the temperature decreases abruptly and so as the chemistry which from the phase of ionization switches to the atomic phase and deeper to the cloud to molecular phase. The difference between the temperatures of the hot ionized and cold neutral gas results in the pressure-driven expansion of nebulae.

Osterbrock (1974) divided ionizing nebulae in three main categories according to their origin and physical characteristics. These are: (1) diffuse nebulae (emission nebulae; Sect. 2.2.1), (2) planetary nebulae (Sect. 2.2.2), and (3) supernova remnants (Sect. 2.2.3). Additionally, reflection and dark nebulae can be also included in the first category (diffuse), however these do not occur due to ionization of the medium as we will see below (Sect. 2.2.4).

2.2.1 Diffuse (Emission) Nebulae

Diffuse nebulae are extended regions (of the order of pc scale) of gas and dust where there are usually embedded stars. If these stars emit UV radiation, they ionize their surroundings creating the so-called "H II regions". These H II regions are also known as *emission nebulae*. They are extended objects spanning a distance of the order of $\sim$10 pc across. They are quite rarefied with typical number densities of the order of 10–100 cm^{-3}. In regions where the youngest massive stars are located, there might be present smaller and more compact H II regions. These *ultracompact* H II regions have been identified around three decades ago (Wood and Churchwell 1989) and they may have a size of the order of a light year or less as well as their densities may even exceed 10^4 cm^{-3}.

An emission nebula is a possible site of star formation. A famous and well studied emission nebula is the Orion Nebula (M42 or NGC1976). At a distance of $\sim$414 pc (Menten et al. 2007), the Orion Nebula is a high-mass star forming region. Here, newly ionized gas found in the borders of the H II region flows into the main area of the nebula. The Orion Nebula Cluster is the closest massive star formation region to the Earth and is thus best studied. It has been claimed that most of the massive stars have already passed through the main sequence. Chandra X-ray observations (Getman et al. 2005; Preibisch et al. 2005) found $\sim$1000 sources emitting X-rays, which have masses spanning from $\sim$0.1–50 $M_{\odot}$, indicating that this area is one of the richest in X-rays ever observed. Several disks emitting small jets (microjets) and outflows have been identified in the Orion Nebula (Bally et al. 2000). Some of those disks interact from radiation and winds emitted by the nearby massive stars taking a shape of a small cocoon surrounded by a bow shock. Some of those objects create brown dwarfs if they become gravitationally unstable. Brown dwarfs have been detected in the Orion Nebula (Lucas and Roche 2000) and some of them have masses below $0.013\,M_\odot$ which is the threshold for deuterium-burning.[2]

2.2.2 Planetary Nebulae

Planetary nebulae are formed during the latter evolutionary stages of a star with mass $1 < M < 8\,M_\odot$. The dying star expands and its outer parts are ejected into the interstellar medium. The remnant core (a white dwarf) emits highly energetic radiation and ionizes this material. An expanding bubble is the result of this process. Inside the ionized region of the bubble, a plethora of cometary knots exists, which are dense clumps surviving the strong ultraviolet radiation. The life time of such a nebula is of the order of a few tens of thousand of years. One of the ongoing topics in the study of planetary nebulae is the origin of the wide range of their shapes. It has been also suggested that a star on the asymptotic giant branch is able to create a magnetic field in a dynamo located at the core-envelope interface. This magnetic field is strong enough to create bipolar or elliptical shapes, as they are widely observed. Since our Sun has mass in this range of masses, its fate is to turn up into a planetary nebula before expanding its size up to Earth's orbit!

The Helix nebula (NGC7293) is a well known and well studied planetary nebula. At its center there is a white dwarf, the remaining core of the initial star, which ionizes a spheroidal region around it. The progenitor star is estimated to had a mass of $\sim 6\,M_\odot$. Before the planetary nebula formed, it has been suggested it underwent three phases of dredge-up and hot bottom burning (Henry et al. 1999). There are two different rings (inner and outer) with the outer to be less ionized and of less temperature than the inner one. This region has a radius of 2.5–3 pc and contains thousands of photo-evaporating cometary knots (López-Martín et al. 2001). These knots consist of a dense head with an extended—relative to the size of the head—cometary tail. It has been suggested that their origin is due to instabilities in the expanding shock front, which are amplified by the ultraviolet radiation (O'dell and Handron 1996; Meaburn et al. 1998; Burkert and O'Dell 1998). Recent observations have showed also that from the emission lines of the knots examined show strong evidence for the presence of H_2. Matsuura et al. (2009) presented high resolution images of the cometary knots showing their irregular structure and pointed that they have a "tadpole" shape with an elliptical head and bright crescent inside followed by a long tail opposite to the central star.

[2]Deuterium burning occurs in stellar systems that have mass $\sim 0.013\,M_\odot$ and in which deuterium nuclei and free protons react and combine helium-3 nuclei.

2.2.3 Supernova Remnants

Supernovae remnants (SNR) are the remains of the explosion of a massive star. When a massive star with $M > 8\,M_\odot$ approaches its final stages of life, it ejects huge amounts of its initial material into the surrounding medium. This material is expanding at high speeds sweeping up the undisturbed interstellar gas and creating a strong shock front. The remaining core can either result in a fast rotating neutron star (pulsar) or even a black hole. Supernovae are extremely important for understanding the formation and evolution of our Galaxy. They heat up the interstellar medium, distribute heavy elements throughout the Galaxy, and accelerate cosmic rays. They have been also considered to play the dominant role in driving turbulence in the ISM which has significant consequences in the star formation rate of the galaxy (see Walch et al. 2014 and references therein for three-dimensional high resolution simulations showing this).

Supernovae are divided into two main categories: type I and type II. Type I supernovae are characterized by sharp maxima in their light curves which dies away gradually. This class can be subdivided in type Iα which exhibit a strong absorption line due to ionized silicon, and type Iβ which lacks of this line. On the other hand, type II supernovae are distinguished from type I by the presence of hydrogen lines in their spectra. The mass of this progenitor should be at least $\sim 8\,M_\odot$. A recent study by Mackey et al. (2014) showed that the light curve of supernovae may also result from the interaction of a static, pressure-confined shell (resulting from the movement of the massive star into the ISM in association with its emission of winds) with the strong ionizing radiation emitted after the explosion.

A famous (and rather historical) SNR is the Crab Nebula (M1, NGC 1952; see Fig. 2.2). This nebula is the remnant of a supernova explosion observed naked eye by Chinese and Arab astronomers in 1054 AD. It is also the first object ever recorded as the remnants of a catastrophic death of a star. The mass of the progenitor star is believed to have been about $M \sim 9$–$11\,M_\odot$ (Davidson and Fesen 1985; MacAlpine and Satterfield 2008). At the center of the Nebula there is a fast rotating pulsar, the remnant of the progenitor star. Using Hubble Space Telescope and Chandra observations, Hester et al. (2002) studied the inner dynamics of the nebula and Lyubarsky (2002) reported that the shock shape in this inner region is highly non-spherical because the energy flux in the pulsar wind decreases towards the axis. On the other hand, the overall structure of this nebula appears to be filamentary due to the Rayleigh-Taylor instability (Hester et al. 1996). Recently, Barlow et al. (2013) have reported the discovery of a noble gas molecular ion (namely the $^{36}ArH^+$) in the Crab Nebula. It is interesting to note that this was the first ever detection of a noble gas molecule in space and in the particular case of the Crab Nebula it is believed to have formed from the explosive nucleosynthesis in the parent massive star during its core-collapse phase. In addition, Owen and Barlow (2015) showed that the dust observed in the Crab Nebula is closely correlated with the position of the knots and filaments indicating that it is formed inside the supernova remnants and not in the ejecta.

Fig. 2.2 The Crab Nebula (M1, NGC1952). These are the remnants of a catastrophic death of a massive star (supernova explosion) observed naked eye by Chinese and Arab astronomers in 1054 AD. Filters used: Hα, O III, S II. Photo taken at the Machon Observatory Panorama Thessaloniki by Kallias Ioannidis

Both planetary nebulae and SNRs appear to be morphologically different from emission nebulae. That is primarily because of the way they are formed and which is the final stages of a dying star. Both of them, however, appear to be (at least in the early stages of their evolution) bubble shaped[3] and are in general quite small (of the order of a light year across) and while their mass is of the order of $\sim 2\,M_\odot$ or less (depending also on the mass of the parent star). Finally, the temperature of the ionized medium in the surroundings (vicinity) of the dead star may get even twice as that observed in an emission nebula.

2.2.4 Reflection and Dark Nebulae

Reflection and dark nebulae have one common characteristic: their presence indicate the presence of dust. As we have discussed in Sect. 1.2, the ISM consists also of dust particles which have size similar to the wavelength of the blue light. Therefore, if a nearby bright star shines them, they illuminate reflecting only the blue light (since all other longer wavelengths will not be scattered). On the other hand, if a region containing high amounts of dust is in-between a bright region and the observer (i.e. the Earth), then they are able to fully absorb the visible light and they appear as dark

[3]Not to be confused with the bubble shaped emission nebulae (H II regions).

areas in the sky. If we place a dark nebula further and further away, stars will start to overlap in the foreground and along the line of sight, so there will be a point which the nebula will not appear as dark anymore. Therefore, in general a dark nebula is a "local" object in this regard. Figure 2.3 shows the ρ-Ophiuchus region which is an example showing simultaneously both a reflection and a dark nebula.

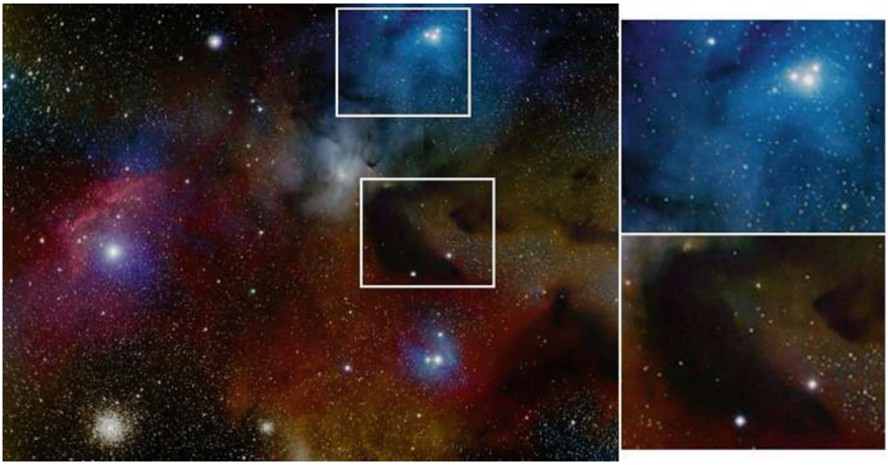

Fig. 2.3 This spectacular image shows the ρ-Ophiuchus region in which one can observe a *bluish* reflection and nearby a dark nebula simultaneously, indicating the presence of high amounts of dust. Filters used: RGB. Photo taken at Thassos island by Kallias Ioannidis

At a distance of $\sim$150 pc, ρ-Ophiuchus is a star forming region in which many young stellar object have been identified (i.e. Nutter et al. 2006). The system is known to contain a brown dwarf with a temperature of $\sim$3000 K which was formed by the gravitational collapse of the surrounding gas. Three-dimensional hydrodynamical simulations by Stamatellos et al. (2007) including radiative transfer techniques showed that the gas temperature of the dark cloud (lower framed image of Fig. 2.3) is colder than $T \sim 10$–11 K.

2.3 Photoionization Equilibrium

In this book we will focus on the evolution and the characteristics of emission nebulae, since in planetary nebulae and supernova remnants the corresponding physics is different from what we intent to present here and also their chemistry is much more complicated. Consider an arbitrary density distribution $\rho(\mathbf{r})$ consisting of neutral hydrogen. The corresponding number density of hydrogen nuclei in all forms (atomic and molecular) is $n(\mathbf{r}) = \frac{X}{m_p}\rho(\mathbf{r})$, where X is the cosmic fraction of hydrogen when account is taken of the contribution of other elements i.e. helium.

Suppose now that we switch on a massive star capable of emitting copious amounts of Lyman continuum photons (having energy $h\nu \geq 13.6\,\text{eV}$) and that this position defines the center of a Cartesian co-ordinate system. The photons interact with the neutral medium and produce free electrons (with an assumed number density n_e) and free protons (n_p). Since in our approximation we neglect ionization of helium, we have $n_e = n_p$ (because a hydrogen atom consists of one electron and one proton only).

By invoking the OTS approximation described in Sect. 2.1, the recombination rate per unit volume is $\alpha_B n_e n_p$ and assuming that inside the H II region the hydrogen is fully ionized, $\alpha_B n_e n_p = \frac{\alpha_B \rho^2(\mathbf{r})}{m^2}$, where $m = m_p/X$ and by assuming that $\rho_e(\mathbf{r}) = \rho_p(\mathbf{r}) m_e/m_p$.

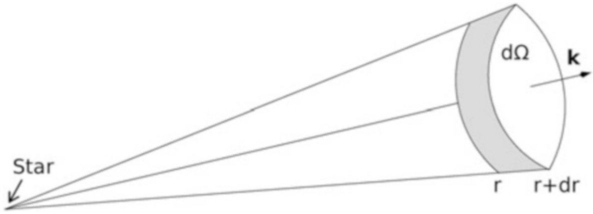

Fig. 2.4 The solid angle $d\Omega$ and the unit vector $\hat{\mathbf{k}}$ emanated from the star (ionizing source)

Consider the element of volume defined by the infinitesimal solid angle $d^2\Omega$ about the unit vector $\hat{\mathbf{k}}$ (as seen from the ionizing source) and the infinitesimal range of radii $(r, r + dr)$, as measured from the star. Figure 2.4 shows this setup. If the number flux of ionizing photons in the direction $\hat{\mathbf{k}}$ is $\dot{N}(r)$, then the equation of ionization balance gives

$$\dot{N}(r) r^2 d^2\Omega = \frac{\alpha_B \rho^2(\mathbf{r})}{m^2} r^2 d^2\Omega dr + \dot{N}(r + dr) \cdot (r + dr)^2 d^2\Omega, \tag{2.5}$$

or if we cancel $d^2\Omega$,

$$\dot{N}(r) r^2 = \frac{\alpha_B \rho^2(\mathbf{r})}{m^2} r^2 dr + \dot{N}(r + dr) \cdot (r + dr)^2. \tag{2.6}$$

From the above equation we obtain

$$\dot{N}(r + dr) \cdot (r + dr)^2 - \dot{N}(r) r^2 = -\frac{\alpha_B \rho^2(\mathbf{r})}{m^2} r^2 dr. \tag{2.7}$$

The lefthand side of the latter is

$$\dot{N}(r + dr) \cdot (r + dr)^2 - \dot{N}(r) r^2 = \frac{d}{dr} \left(\dot{N}(r) r^2 \right) dr, \tag{2.8}$$

so if we combine this relation with Eq. 2.7 we obtain

$$\frac{d}{dr}\left(\dot{N}(r)r^2\right)dr = -\frac{\alpha_B \rho^2(\mathbf{r})}{m^2} r^2 dr. \tag{2.9}$$

By canceling dr from both sides and integrating, we obtain

$$\dot{N}(r)r^2 = \frac{\dot{\mathcal{N}}_{\mathrm{LyC}}}{4\pi} - \frac{\alpha_B}{m^2}\int_{r'=0}^{r'=r} \rho^2(\mathbf{r}')r'^2 dr', \tag{2.10}$$

where $\frac{\dot{\mathcal{N}}_{\mathrm{LyC}}}{4\pi}$ is the constant of integration and $\dot{\mathcal{N}}_{\mathrm{LyC}}$ is the total emission rate of the star. Equation 2.10 gives the photoionization balance condition inside an H II region.

Let us consider R_{IF} to be the position of the ionization front (IF), where the flux of the arriving ionizing photons is zero ($\dot{N}_{\mathrm{tot}}(R_{\mathrm{IF}}) = 0$). We define here that the ionization front is the moving edge of the H II region. Assuming that the material inside the H II region is fully ionized, Eq. 2.10 is then

$$\int_{r=0}^{r=R_{\mathrm{IF}}} \rho^2(\mathbf{r})r^2 dr = \frac{\dot{\mathcal{N}}_{\mathrm{LyC}} m^2}{4\pi\alpha_B}. \tag{2.11}$$

If the density distribution is uniform ($\rho(\mathbf{r}) = \rho_o$), then Eq. 2.11 becomes

$$\frac{\rho_o^2 R_{\mathrm{IF}}^3}{3} = \frac{\dot{\mathcal{N}}_{\mathrm{LyC}} m^2}{4\pi\alpha_B}, \tag{2.12}$$

and so the resultant H II region is spherical with radius

$$R_{\mathrm{St}} \equiv R_{\mathrm{IF}} = \left(\frac{3\dot{\mathcal{N}}_{\mathrm{LyC}} m^2}{4\pi\alpha_B \rho_o^2}\right)^{1/3}, \tag{2.13}$$

where R_{St} is the Strömgren radius (Strömgren 1939).

Strömgren (1939) was the first to show that the transition from a state of almost completely ionized material to a state of almost completely neutral material occurs in a very short distance compared to the dimensions of the H II region. Thus we may treat the front as a sharp discontinuity. The distance over which the degree of ionization changes from $\simeq$90 % to $\simeq$10 % is given by

$$\Delta R_{\mathrm{St}} \simeq \frac{20m}{\rho_o\bar{\sigma}} \simeq 2\times 10^{-4}\,\mathrm{pc}\left(\frac{\rho_o}{10^{-20}\,\mathrm{g\,cm^{-3}}}\right)^{-1}. \tag{2.14}$$

Here $\bar{\sigma} = 7\times 10^{-18}\,\mathrm{cm}^2$ is the photoionization cross section presented by a hydrogen atom to Lyman continuum photons from an OB star.

2.4 Heating and Cooling Mechanisms

As we have previously mentioned, the source of ionizing radiation in H II regions is a single or multiple massive stars, in the spectral type of O or B (hence known as OB stars). An OB star has a typical effective temperature[4] of $T_{\rm eff} \sim 3\text{–}6 \times 10^4$ K. Therefore one would expect that such a star would be capable of heating up the surrounding medium in temperatures up to these values. However, typical H II regions have $T_{\rm i} \simeq 10^4$ K which is much lower than $T_{\rm eff}$. This difference in temperatures result from various cooling processes that occur simultaneously and that do not allow the ionizing temperature to go higher than $T_{\rm i} \simeq 10^4$ K.[5] Below we briefly describe the most important heating and cooling sources occurring in an H II region and which result in its thermal balance.

The main heating process in an H II region comes, clearly, from the UV ionizing radiation emitted by the central source. Consider a photon carrying energy more than the 13.6 eV ionization potential of hydrogen, i.e. $h\nu = 15$ eV. When this photon interacts with a hydrogen atom it will ionize it producing a photoelectron. The excess of $15 - 13.6 = 1.4$ eV will be converted to the form of kinetic energy in the photoelectron. In a medium, the faster the particles move, the more hot the medium is, therefore the above kinetic energy simply increases the temperature of the H II region. Other heating processes may apply such as cosmic rays, X-rays, etc. but due to their small contribution in the total heating of the ionized medium we will not discuss them further. If no cooling process was present, the H II region temperature would gradually increase to reach the temperature of the star.

There are several cooling processes that act to decrease the temperature of an H II region down to its observed value. As long as an electron and a proton recombine to form a hydrogen atom, the electron kinetic energy is converted into radiation which can escape the nebula while carrying energy, therefore cooling it. This radiation is able to cool the medium if the recombination does not occur directly to the ground state, otherwise as we have discussed earlier it can ionize a nearby hydrogen atom diffusively. Other cooling processes include the bremsstrahlung radiation which is emitted by the thermal plasma and which is observed at radio frequencies, as well as the collisional excitations of fast moving electrons with neutral hydrogen. These collisions can convert the kinetic energy of an electron to radiation which can also escape the nebula.

However, the most important cooling process is the forbidden line cooling. Species responsible for this effect are metals, such as O and Ne. In case of low abundance in these metals (i.e. low metallicity), forbidden line cooling is not strong enough which act to allow the temperature of the emission nebula to increase.

[4]The effective temperature corresponds to the temperature of the object as if it was black body and emitting the equivalent amount of electromagnetic radiation.

[5]In emission nebulae, $T_{\rm i}$ can reach higher values if we decrease the metallicity, since in this case we change the chemical processes. For the same reason and due to the more powerful heating sources, planetary nebulae and supernova remnants have also in general $T_{\rm i} > 10^4$ K.

Forbidden lines arise when an electron of a metal is excited by a collision (i.e. with a free fast moving electron) into a metastable state (also known as fine structure levels). These metastable states act as an "energy trap" meaning that when an electron occupies this state, it gets significantly longer time to spontaneously de-excite back to its lower energy value in comparison with the rest of states. For example, a spontaneous de-excitation from a metastable state may take from $t \sim 10\,\mathrm{s}$ to $t > 10^4\,\mathrm{s}$ compared to $t \sim 10^{-8}\,\mathrm{s}$ in the "normal" case. When the medium is dense, during the above period of i.e. 10 s, it is very highly probable for the metal to collide again with another particle of lower energy, hence for the excited electron to transfer its energy excess to the other particle via a collision. Then, the excited electron of the metastable state will lose energy by converting it to kinetic energy. However, if the medium is rarefied enough to avoid these collisions, lower than a value known as "critical density" ($n_{\rm crit}$), then the electron will spontaneously de-excite to a lower state by releasing a photon. This photon has a frequency which cannot ionize a hydrogen atom, therefore it escapes the nebula contributing in its cooling. These are the known forbidden lines (or fine structure lines) and are very frequently observed in nebulae, particularly in planetary nebulae.[6]

All above heating and cooling mechanisms result in the thermal balance of an H II region which reaches a final temperature of $T_{\rm i} \simeq 10^4\,\mathrm{K}$. Forbidden lines can act as nebular diagnostics. They can tell much about the properties of the nebula such as the abundances of species therein, the density of the ionized and atomic medium, the strength and the profile of the ionizing profile of the radiation emitted by the central star and others. These lines are mostly observed in radio wavelengths and (far-)infrared wavelengths, the observation of which requires ground-based arrays such as ALMA and JVLA, or space-borne telescopes such as *Spitzer* and *Herschel* respectively.

2.5 R-Type Expansion

When an ionizing source is switched on in an initially neutral medium (which for simplicity we assume that it is spherically symmetric of uniform number density, n), photons start to rapidly ionize it. The initial phase of ionization is extremely quick with the ionization front reaching velocities significantly higher than the sound

[6]Under the same effect the famous 21-cm forbidden line emission of H I is produced however there are two exceptions: (1) the 21-cm line is not emitted from an H II region and does not contribute in its cooling (2) the 21-cm line occurs when there is a spin flip of the electron when compared to the spin of the proton. The more stable case is when the two spins are anti-parallel (para-hydrogen) contrary to the parallel case (ortho-hydrogen). For a spontaneous flip of the ortho to para state, it is required a time of 1–10 million years(!) in order not for the ortho-hydrogen atom to undergo a single collision. It is astonishing to consider that a 21-cm map of a galaxy has been produced by photons that only Astrophysicists can see and that will almost certainly never be reproduced in any lab on Earth.

speed of the ISM. This velocity is so high that the medium cannot hydrodynamically response and it therefore remains approximately undisturbed. This initial phase is known as '*R*-type expansion' (R: Rarefied).

If we consider that $\dot{\mathscr{N}}_{\rm LyC}$ is the number of photons emitted by an ionizing source and assuming that the medium remains undisturbed once the photons start to propagate in it causing ionizations, we can determine the position of the expanding IF as follows. During the R-type expansion phase, photoionization equilibrium is not met hence there will be always a surplus of photons remaining after a given volume has been ionized. This is expressed as

$$\dot{\mathscr{N}}_{\rm surplus} = \dot{\mathscr{N}}_{\rm LyC} - \frac{4}{3}\pi r_{\rm IF}^3 \alpha_{\rm B} n^2. \tag{2.15}$$

The above equation tells us that the remaining photons are equal to the total number emitted by the source minus those that have been lost by recombinations of electrons with protons to other than the ground state (i.e. OTS approximation). The surplus amount of photons are causing ionizations to an expanding volume equal to $4\pi r_{\rm IF}^2 dr_{\rm IF}/dt$ (that is the surface of the sphere at $r_{\rm IF}$ times the velocity of expansion). This expanding volume has n number density therefore

$$\dot{\mathscr{N}}_{\rm surplus} = n 4\pi r_{\rm IF}^2 \frac{dr_{\rm IF}}{dt}. \tag{2.16}$$

Combining Eqs. 2.15 and 2.16 we obtain the differential equation

$$n 4\pi r_{\rm IF}^2 \frac{dr_{\rm IF}}{dt} = \dot{\mathscr{N}}_{\rm LyC} - \frac{4}{3}\pi r_{\rm IF}^3 \alpha_{\rm B} n^2. \tag{2.17}$$

Using the Strømgren radius relation (Eq. 2.13), we solve the above to find the position of the IF in the *R*-type phase

$$r_{\rm IF}(t) = R_{\rm St}\left(1 - e^{-n\alpha_{\rm B}t}\right)^{1/3} \tag{2.18}$$

where $\alpha_{\rm B}$ is the case-B recombination coefficient (see Sect. 2.1) and t is the time. The *R*-type expansion will be terminated as soon as $e^{-n\alpha_{\rm B}t} \sim 0$ which occurs when $t_{\rm rec} \sim 3/n\alpha_{\rm B}$. The quantity $t_{\rm rec} = 1/n\alpha_{\rm B}$ is commonly referred to as "recombination time". When the *R*-type phase terminates, the Strømgren sphere has been created and the system has obtain photoionization balance as we have explained above. Thereafter, the dynamical expansion begins which is described in the section below.

2.6 D-Type Expansion

Once the Strömgren sphere has been formed and photoionization equilibrium has been met, the H II region continues to expand due to the high pressure difference observed between the ionized and the neutral medium. This expansion phase, called

'D-type expansion' (D: Dense), continues until pressure equilibrium is met. Once the system obtains this equilibrium, no further dynamical evolution of the H II occurs. In the D-type expansion, the H II region expands at a high-supersonic speed relative to the sound speed of the neutral medium. This drives the formation of a strong shock front ahead of the expanding ionized region (i.e. the ionization front) which sweeps material in time.

> To understand why a shock front is formed, it is important that we should describe what the sound speed does represent. The sound speed is the speed at which a disturbance is propagating in a given medium. The sound itself is actually a disturbance. Consider the atmospheric air at the sea level. The speed of sound at that level is $\sim$0.3 km/s. If a moving object (i.e. a supersonic airplane) exceeds that speed, it means that the disturbance it creates (from its motion) cannot move faster than the object itself. Hence material of the atmospheric air is constantly being packed up in the area just ahead of the edge of the object towards its moving direction.

For the standards of the ISM, the sound speed, c_s is connected with the gas temperature according to the relation:

$$c_s = \sqrt{\frac{k_B T_{gas}}{\mu m_p}} \tag{2.19}$$

where k_B is Boltzmann's constant and μ is the mean molecular weight of the gas. A medium consisted purely by atomic hydrogen (so $\mu = 1$) at $T_{gas} = 10$ K has a sound speed of $c_o \sim 0.2$ km/s. If this medium is heated up to $T_{gas} = 10^4$ K it is ionized (so $\mu = 0.5$) and its sound speed is much higher $c_i \sim 12.8$ km/s. As we have seen, these temperatures are the standard (average) temperatures found in H II regions. The speed at which an H II region expands exceeds the sound speed of the neutral medium by approximately an order of magnitude, hence a shock front is formed.

The motion of an expanding H II regions associated with the formation of a shock front has been discussed initially by Kahn (1954). However, Spitzer (1978) was the first to attempt an analytical model of the D-type expansion and further discussions followed by Dyson and Williams (1980) have explored this model in more depth. In this approximation, the time-dependent thermal pressure $P_i(t)$ of the ionized gas is $P_i(t) = \rho_i(t)c_i^2$, where c_i is the sound speed of the ionized medium. However, the first to fully explore the D-type expansion to the point of pressure equilibrium was Raga et al. (2012a) while in Raga et al. (2012b) further additions concerning the inertia of the moving shocked material have been considered.

Raga et al. (2012a) considered a model in which the two pressures of (1) the neutral gas in the shell confined between the ionization front and the shock front and of (2) the ram pressure of the undisturbed neutral material (i.e. the one that the shock front has not yet overrun) are equal. They then approximated the D-type expansion of the modelled H II region by using the *piston relation* in which the velocity of the shock front, V_{SF}, is equal to the velocity of the ionization front, $\dot{R}$, plus the post-shock velocity, V_{ps}. However, V_{ps} can be replaced with $V_{SF}/\mathscr{M}_o^2$, where $\mathscr{M}_o = V_{SF}/c_o$ is the Mach number, therefore

$$V_{SF} = \dot{R} + \frac{c_o}{V_{SF}}. \tag{2.20}$$

Under the above assumptions, the velocity of the shock front can be also expressed in terms of the Strømgren radius, the position of the ionization front and the sound speed of the ionized medium as

$$V_{SF} = \left(\frac{R_{St}}{R}\right)^{3/4} c_i, \tag{2.21}$$

and by replacing Eq. 2.21 to 2.20 we obtain the differential equation

$$\frac{1}{c_i}\dot{R} = \left(\frac{R_{St}}{R}\right)^{3/4} - \left(\frac{c_o}{c_i}\right)^2 \left(\frac{R}{R_{St}}\right)^{3/4}. \tag{2.22}$$

Its solution gives the position of the ionization front in unit time. In the above differential equation, the second term on the righthand side can be neglected at early times, since both ratios c_o/c_i and R/R_{St} are in general small. In this case, Eq. 2.22 has the simple solution

$$R(t) = R_{St}\left(1 + \frac{7}{4}\frac{c_i t}{R_{St}}\right)^{4/7}, \tag{2.23}$$

which is known as "Spitzer equation" (Spitzer 1978). However, as the H II expands, the ratio $(R/R_{St})^{3/4}$ becomes much larger than $(c_o/c_i)^2$ and the negative term in the righthand side of Eq. 2.22 starts to dominate. In this case, the H II region stagnates and no further expansion is allowed. This stagnation radius is given by the relation

$$R_{STAG,I} = \left(\frac{c_i}{c_o}\right)^{4/3} R_{St}. \tag{2.24}$$

The expansion of an H II region however, can be also studied using the equation of motion of the expanding shell (note that in the method described above, the differential equation resulted by equating the pressures acting on the shell). This has been studied by Hosokawa and Inutsuka (2006) and independently by Raga et al. (2012b), the latter of which considered additionally the external ram pressure of the undisturbed medium on the H II region. Raga et al. (2012b) described in their work that the differential equation of motion of the shell i.e.

$$\frac{d}{dt}\left(M\dot{R}\right) = 4\pi R^2 \left(P_{\rm i} - P_{\rm o}\right) \tag{2.25}$$

includes the inertia of the shocked gas. In particular, if we expand Eq. 2.25, we then end up to a second order differential equation of the form

$$\ddot{R} + \left(\frac{3}{R}\right)\dot{R}^2 = \frac{3R_{\rm St}^{3/2}c_{\rm i}^2}{R^{5/2}} - \frac{3c_{\rm o}^2}{R}. \tag{2.26}$$

If we were to omit the $\ddot{R}$ term, Eqs. 2.22 and 2.26 are then the same. Similarly to what we have discussed previously, the term $c_{\rm o}^2/R$ is in general small at early times when the H II region expands, thus we may neglect it. In this case, Eq. 2.26 has as analytical solution

$$R(t) = R_{\rm St}\left(1 + \frac{7}{4}\sqrt{\frac{4}{3}}\frac{c_{\rm i}t}{R_{\rm St}}\right)^{4/7}. \tag{2.27}$$

This solution was presented by Hosokawa and Inutsuka (2006). It differs by Eq. 2.23 by the $\sqrt{4/3}$ factor which, according to Raga et al. (2012b), results from the inclusion of the inertia. As the ionized medium expands further, however, the negative term of Eq. 2.26 starts to dominate and, as before, the H II region stagnates. The stagnation radius obtained here is given by the relation

$$R_{\rm STAG,II} = R_{\rm St}\left(\frac{8}{3}\right)^{2/3}\left(\frac{c_{\rm i}}{c_{\rm o}}\right)^{4/3} \tag{2.28}$$

2.6.1 Further Discussion

Figure 2.5 plots the solutions of Eq. 2.22 (Raga-I), Eq. 2.26 (Raga-II), Eq. 2.23 (Spitzer) and Eq. 2.27 (Hosokawa-Inutsuka). During two different workshops,[7] many groups focusing on computational star formation emphasizing on expanding H II regions gathered to validate their codes against the above set of equations (Bisbas et al. 2015). The results of their simulations agreed on a different expansion law, marked as "Simulations" in Fig. 2.5. This law follows in the early times of the D-type expansion the Raga-II and Hosokawa-Inutsuka equations, but at later times it stagnates to the value of Raga-I equation, which is the generalized case of Spitzer's approximation. During intermediate times, it is between the Raga-I and Raga-II equations.

[7] The STARBENCH workshops in 2013 (Exeter, U.K.) and in 2014 (Bonn, Germany).

Currently, there is no universal equation that can explain this behavior. However, we may provide here some remarks that can help us towards this. Once the H II region starts expanding, a strong shock front forms. At early times, this shock front is quite thin and very dense. As time progresses, the shock front continues to sweep up material from the undisturbed medium, however it tends to expand somehow faster than the ionization front as long as the latter follows a $\propto t^{4/7}$ law. This means that at later times (i.e. when the H II region stagnates) the shock front will be completely detached. Hence, the stagnated H II region will not have a shock front at all immediately after the ionization front.

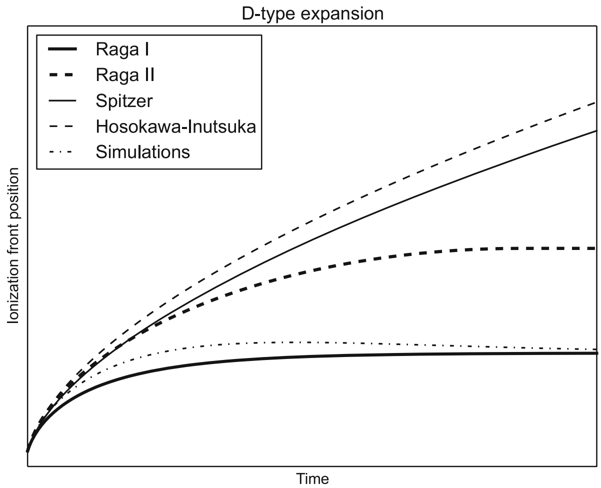

Fig. 2.5 Solutions of the D-type expansion of an H II region. The *dot-dashed line* shows the result of the 12 different codes participating in the StarBench project

This points to the assumption that Spitzer's well known equation and its generalized form by Raga et al. (2012a) (Raga-I) is an, in fact, *no-shell* approximation i.e. at all times no shock front is developed. On the contrary, the Hosokawa-Inutsuka equation as well as its generalized form by Raga et al. (2012b) (Raga-II) is produced by solving the equation of motion of the expanding shell. In this way, it *is* considered the development of a shock front which is infinitesimally thin (thin shell approximation), meaning that it has an infinite density (δ-function).

Raga-I and Raga-II are then acting as boundary conditions in which a family of different expansion laws exist. Due to the conditions assumed in the simulations of the StarBench project, the codes followed the particular dash-dotted law of Fig. 2.5. However, work in this direction has to been done by the community which will give a much better understanding of how an H II region expands.

References

Osterbrock, D. E. 1974, Research supported by the Research Corp., Wisconsin Alumni Research Foundation, John Simon Guggenheim Memorial Foundation, Institute for Advanced Studies, and National Science Foundation. San Francisco, W. H. Freeman and Co., 1974. 263 p.
Wood, D. O. S., & Churchwell, E. 1989, ApJS, 69, 831
Menten, K. M., Reid, M. J., Forbrich, J., & Brunthaler, A. 2007, A&A, 474, 515
Getman, K. V., Flaccomio, E., Broos, P. S., et al. 2005, ApJS, 160, 319
Preibisch, T., Kim, Y.-C., Favata, F., et al. 2005, ApJS, 160, 401
Bally, J., O'Dell, C. R., & McCaughrean, M. J. 2000, AJ, 119, 2919
Lucas, P. W., & Roche, P. F. 2000, MNRAS, 314, 858
Henry, R. B. C., Kwitter, K. B., & Dufour, R. J. 1999, ApJ, 517, 782
López-Martín, L., Raga, A. C., Mellema, G., Henney, W. J., & Cantó, J. 2001, ApJ, 548, 288
O'dell, C. R., & Handron, K. D. 1996, AJ, 111, 1630
Meaburn, J., Clayton, C. A., Bryce, M., et al. 1998, MNRAS, 294, 201
Burkert, A., & O'Dell, C. R. 1998, ApJ, 503, 792
Matsuura, M., Speck, A. K., McHunu, B. M., et al. 2009, ApJ, 700, 1067
Walch, S. K., Girichidis, P., Naab, T., et al. 2014, arXiv:1412.2749
Mackey, J., Mohamed, S., Gvaramadze, V. V., et al. 2014, Nature, 512, 282
Davidson, K., & Fesen, R. A. 1985, ARA&A, 23, 119
MacAlpine, G. M., & Satterfield, T. J. 2008, AJ, 136, 2152
Hester, J. J., Mori, K., Burrows, D., et al. 2002, ApJ, 577, L49
Lyubarsky, Y. E. 2002, MNRAS, 329, L34
Hester, J. J., Stone, J. M., Scowen, P. A., et al. 1996, ApJ, 456, 225
Barlow, M. J., Swinyard, B. M., Owen, P. J., et al. 2013, Science, 342, 1343
Owen, P. J., & Barlow, M. J. 2015, ApJ, 801, 141
Nutter, D., Ward-Thompson, D., & André, P. 2006, MNRAS, 368, 1833
Stamatellos, D., Whitworth, A. P., & Ward-Thompson, D. 2007, MNRAS, 379, 1390
Strömgren, B. 1939, ApJ, 89, 526
Kahn, F. D. 1954, Bull. Astron. Inst. Netherlands, 12, 187
Spitzer, L. 1978, New York Wiley-Interscience, 1978. 333 p.,
Dyson, J. E., & Williams, D. A. 1980, New York, Halsted Press, 1980. 204 p.,
Raga, A. C., Cantó, J., & Rodríguez, L. F. 2012a, Rev. Mexicana Astron. Astrofis., 48, 149
Raga, A. C., Cantó, J., Rodríguez, L. F., & Velázquez, P. F. 2012b, MNRAS, 424, 2522
Hosokawa, T., & Inutsuka, S.-i. 2006, ApJ, 646, 240
Bisbas, T. G., et al. 2015, MNRAS, 453, 1324

http://www.springer.com/978-3-319-26140-9

The Interstellar Medium, Expanding Nebulae and
Triggered Star Formation
Theory and Simulations
Bisbas, Th.G.
2016, XI, 71 p. 22 illus., 15 illus. in color., Softcover
ISBN: 978-3-319-26140-9